Energy Efficiency Maximization in Large-Scale Cell-Free Massive MIMO: A Projected Gradient Approach

Trang C. Mai, *Member, IEEE*, Hien Quoc Ngo, *Senior Member, IEEE*, and Le-Nam Tran, *Senior Member, IEEE*

Abstract—This paper considers the fundamental power allocation problem in cell-free massive mutiple-input and multiple-output (MIMO) systems which aims at maximizing the total energy efficiency (EE) under a sum power constraint at each access point (AP) and a quality-of-service (QoS) constraint at each user. Existing solutions for this optimization problem are based on solving a sequence of second-order cone programs (SOCPs), whose computational complexity scales dramatically with the network size. Therefore, they are not implementable for practical large-scale cell-free massive MIMO systems. To tackle this issue, we propose an iterative power control algorithm based on the frame work of an accelerated projected gradient (APG) method. In particular, each iteration of the proposed method is done by simple closed-form expressions, where a penalty method is applied to bring constraints into the objective in the form of penalty functions. Finally, the convergence of the proposed algorithm is analytically proved and numerically compared to the known solution based on SOCP. Simulations results demonstrate that our proposed power control algorithm can achieve the same EE as the existing SOCPs-based method, but more importantly, its run time is much lower (one to two orders of magnitude reduction in run time, compared to the SOCPs-based approaches).

Index Terms—Cell-free massive MIMO, massive MIMO, spectral efficiency, MMSE-SIC, power control.

I. INTRODUCTION

CELL-FREE massive mutiple-input and multiple-output (MIMO) has attracted a lot of research interest recently, for its ability to overcome the inherent intercell-interference of cellular networks [2]–[7]. Basically, cell-free massive MIMO

uses a massive number of distributed access points (APs) together with simple linear processing to coherently serve many users using the same time and frequency resources. Each APs can be equipped with several antennas. It relies on the favorable propagation and channel hardening property of massive MIMO technology [8], and the macro-diversity of network MIMO technique [9]. Therefore, it can provide universally good service to all users in the network regardless their locations.

In cell-free massive MIMO, because the APs and users are distributed over a large area, power controls are very important to control the near-far effect, and hence, can significantly improve the system performance as well as to save the radiated powers from the APs in the downlink and the users in the uplink. Thus, many research works on power allocations in cell-free massive MIMO have been studied [2], [3], [10]–[13]. In [2], [3], power control coefficients at the APs and users were optimally chosen to maximize the minimum spectral efficiency of all users. In [10], the max-min power control under limited backhaul was investigated. The downlink transmission power optimization of cell-free massive MIMO with spatially correlated Rayleigh fading channels for noncoherent joint and coherent joint transmission was exploited in [11]. A joint downlink transmit powers and the number of active APs optimization was proposed and solved in [12]. In [13], the total energy efficiency maximization taken into account the hardware and backhaul power consumption was proposed and exploited. In the context of cell-free massive MIMO, the solutions to power control problems in most, if not all, of previous works are based on successive convex approximation principle, which approximates a non-convex problem by a sequence of convex second order cone programs (SCOPs) [14]. As a result, these methods have very high computational complexity, as they rely on interior point methods (through the use of off-the-shelf convex solvers) to solve these convex problems. Therefore, they are not implementable for large-scale cell-free massive MIMO with many APs and users (e.g. in stadium or shopping malls where we may have thousands of APs and active users).

To deal with the large-scale problem in cell-free massive MIMO, [15] proposed a scalable framework, which uses AP selection to cope with computational complexity and backhaul requirements. However, the power allocation is quite simple and heuristic, and thus, it may underestimate the capacity of

the system. Another approach to deal with the large-scale problem is presented in [16]–[18]. The main idea of this approach is to decompose a large optimization problem into smaller optimization subproblems to reduce computational complexity. However, this method cannot be applied to solve optimization problems in which the variables are coupled, such as those with quality of service (QoS) constraints.

In this paper, we consider the energy efficiency maximization problem with QoS constraints with an emphasis on large-scale settings. For such scenarios, the number of power control coefficients (i.e., the product of the number of APs and the number of users) which need to be optimized can be extremely large. Thus, our goal is to propose a novel power control algorithm for the energy efficiency maximization problem, which has much lower computational complexity compared to the traditional SOCP-based method. It is apparent from the above discussions that an efficient numerical method for this particular problem is still demanding. To this end, we combine the penalty method and the accelerated projected gradient (APG) method. More specifically, the penalty method is used to handle the QoS constraints in our problem, resulting in more tractable subproblems. Note that the penalty method is widely used to deal with constrained optimization [19]. In principle, the penalty method penalizes a set of constraints by proper terms and adds the penalty terms into the objective, creating the so-called penalized objective. In this way, an optimal problem with sophisticated constraints can be converted into a regularized optimization problem with simple constraints for which efficient solutions are easier to derive. By increasing the penalty parameter, the solutions to these regularized problems converge to a solution of the original problem. In this paper, to solve the regularized problems obtained from the penalty method, we then apply the APG method which is a variant of the accelerated proximal gradient method proposed in [20] for nonconvex programming. As shall be numerically shown in Section V, compared to the sequential SOCP-based method, the proposed method achieves the same total EE but with much lower run time and computational complexity since it is entirely based on first order oracle (i.e the value of the objective and its gradient). Thus, the proposed method can be readily modified to tackle the high complexity of other resource allocation problems for large-scale cell-free massive MIMO. The main contributions of this paper are as follows.

- We provide the mathematical background of the APG method with a detailed proof as an alternative solution for sequential SOCP-based method to deal with many resource allocation problems in large-scale cell-free massive MIMO.
- In our proposed APG method, no external optimization solver is needed as the projection, which is the main operation of the proposed method, is done by closed form expression. As the result, it is much faster to output a solution, compared to the known sequential SOCP-based method.
- We customize the presented APG method to solve the total EE maximization problem in cell-free massive MIMO.

- For our specific problem, we first transform the problem of total EE maximization, subject to transmit power constraints at APs and the individual quality-of-service (QoS) constraints at each user, into a form amenable to the application of the APG method.
- We then combined the penalty method and the APG method to achieve a good and low-complexity power control algorithm. In particular, the QoS constraints are penalized by a proper smooth penalty term which in controlled by a penalty parameter. The penalty term is then added to the original objective, giving rise to the penalized problem. The APG is applied to solve the penalized problem, whereby each iteration admits closed-form expressions. The computational complexity of the proposed algorithm is provided.
- We provide numerical results to show that the proposed algorithm can achieve the same performance as an SCOPs-based method but with much reduced run time.
- We also verify that our proposed algorithm converges to a feasible solution regardless the choice of starting point.

The rest of this paper is organized as follows. Section II provide the preliminaries of the APG method. Next, Section III recalls total EE optimization problem in cell-free massive MIMO. Then, Section IV proposes to use APG method for total EE optimization problem. Section V evaluates the system performance by using numerical results. Finally, the conclusion is drawn in Section VI.

Notation: Standard notations are used in this paper. The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and the Hermitian, respectively. Notation $[\mathbf{x}]_+$ denotes the projector onto the positive orthant. Notation ∇, and ∂ denote the gradient, and sub-gradient, respectively. Notation $\mathrm{dom} f$ denotes a domain of function f, and $||.||$ denotes the l_2-norm. Notation $\equiv$ is used to define an equivalent quantity. Finally, we use $\odot$, $\otimes$, and $\langle .,. \rangle$ to denote the Hadamard, the Kronecker products and an inner product, respectively.

II. MATHEMATICAL PRELIMINARIES: ACCELERATED PROJECTED GRADIENT METHOD

In this section, we first provide the general framework of an accelerated proximal gradient method and then present a variant of the accelerated proximal gradient method, which is termed the accelerated *projected* gradient (APG) method, to deal with the EE maximization problem to be considered in Section III. First, we recall some definitions. A function f is said to be proper if $\mathrm{dom} f \neq 0$. A function f is said to have an L-Lipschitz continuous gradient if there exists some $L > 0$ such that $||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||, \forall \mathbf{x}, \mathbf{y}$. If $f(x) \geq x_0, \ \forall x \in X$, then the function is said to be bounded from below by x_0. A function f is lower semicontinuous at point x_0 if $\liminf_{x \to \mathbf{x}0} f(x) \geq f(\mathbf{x}_0)$. $f(x)$ is called coercive if f is bounded from below and $f(\mathbf{x}) \to \infty$ when $||\mathbf{x}|| \to \infty$.

We now present a general mathematical framework, called the accelerated proximal gradient method for non-convex problems presented in [20], which concerns the following

optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{T(\mathbf{x}) \equiv f(\mathbf{x}) + g(\mathbf{x})\}, \tag{1}$$

where $f(\mathbf{x})$ is L-Lipschitz continuous gradient, $g(\mathbf{x})$ is proper and lower semicontinuous, and $T(\mathbf{x})$ is coercive. Then the accelerated proximal gradient method for solving (1), consists of the following iterations:

$$
\begin{aligned}
\mathbf{y}_k &= \mathbf{x}_k + \frac{t_{k-1}}{t_k}(\mathbf{z}_k - \mathbf{x}_k) + \frac{t_{k-1} - 1}{t_k}(\mathbf{x}_k - \mathbf{x}_{k-1}), \\
\mathbf{z}_{k+1} &= \mathrm{prox}_{\alpha_y g}(\mathbf{y}_k - \alpha_y \nabla f(\mathbf{y}_k)), \\
\mathbf{v}_{k+1} &= \mathrm{prox}_{\alpha_x g}(\mathbf{x}_k - \alpha_x \nabla f(\mathbf{x}_k)), \\
\mathbf{x}_{k+1} &= \begin{cases} \mathbf{z}_{k+1}, & T(\mathbf{z}_{k+1}) \le T(\mathbf{v}_{k+1}) \\ \mathbf{v}_{k+1}, & \text{otherwise}, \end{cases} \\
t_{k+1} &= \frac{\sqrt{4t_k^2 + 1} + 1}{2},
\end{aligned} \tag{2}
$$

where α_x, α_y are step sizes, and $\mathrm{prox}_{\alpha g}$ is the proximal operator defined as

$$\mathrm{prox}_{\alpha g}(\mathbf{x}) \triangleq \arg\min_{\mathbf{u}} \, g(\mathbf{u}) + \frac{1}{2\alpha}||\mathbf{x} - \mathbf{u}||^2. \tag{3}$$

In general, accelerated proximal gradient method in (2) is designed to cope with the *unconstrained optimization* problem (1). However, most of resource allocation problems in cell-free massive MIMO are constrained optimization problems. Therefore, in this paper, we present a special case of the accelerated proximal gradient method, which is called the APG method, to deal with those problems. Specifically, we consider the following optimization problem

$$\min_{\mathbf{x} \in \mathcal{C}} \, f(\mathbf{x}), \tag{4}$$

where $\mathcal{C}$ is the feasible convex set of the considered problem, which is often defined by a set of constraints. Again, assume that $f(\mathbf{x})$ is a proper function with Lipschitz continuous gradient, and bounded from below. We remark that $g(\mathbf{x})$ in (1) is not necessarily smooth. Thus, to apply the iterations in (2) to solve (4), we can let $g(\mathbf{x})$ in (1) be the indicator function of the feasible set $\mathcal{C}$. In this way, the proximal operator in (3) reduces to the Euclidean projection onto $\mathcal{C}$ [20]. As a result, the APG method for solving (4), consists of the following iterations:

$$
\begin{aligned}
\mathbf{y}_k &= \mathbf{x}_k + \frac{t_{k-1}}{t_k}(\mathbf{z}_k - \mathbf{x}_k) + \frac{t_{k-1} - 1}{t_k}(\mathbf{x}_k - \mathbf{x}_{k-1}) & (5\mathrm{a}) \\
\mathbf{z}_{k+1} &= P_{\mathcal{C}}(\mathbf{y}_k - \alpha_y \nabla f(\mathbf{y}_k)) & (5\mathrm{b}) \\
\mathbf{v}_{k+1} &= P_{\mathcal{C}}(\mathbf{x}_k - \alpha_x \nabla f(\mathbf{x}_k)) & (5\mathrm{c}) \\
\mathbf{x}_{k+1} &= \begin{cases} \mathbf{z}_{k+1}, & f(\mathbf{z}_{k+1}) \le f(\mathbf{v}_{k+1}), \\ \mathbf{v}_{k+1}, & \text{otherwise} \end{cases} & (5\mathrm{d}) \\
t_{k+1} &= \frac{\sqrt{4t_k^2 + 1} + 1}{2}, & (5\mathrm{e})
\end{aligned}
$$

where $P_{\mathcal{C}}(\mathbf{x})$ denotes the Euclidean projection of $\mathbf{x}$ onto $\mathcal{C}$, which is defined as

$$P_{\mathcal{C}}(\mathbf{x}) \triangleq \arg\min_{\mathbf{u} \in \mathcal{C}} ||\mathbf{x} - \mathbf{u}||^2. \tag{6}$$
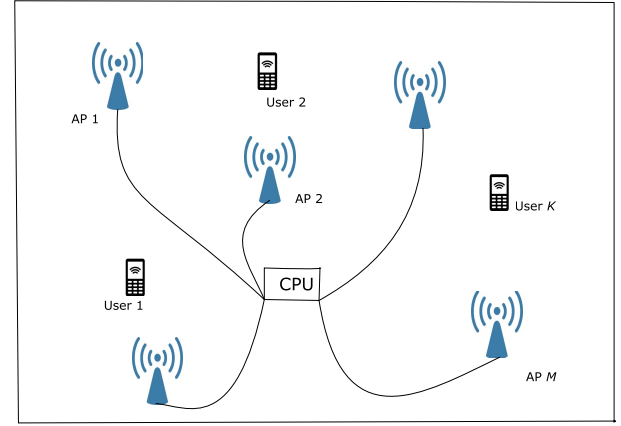

Fig. 1. System model.

III. TOTAL ENERGY EFFICIENCY OPTIMIZATION IN CELL-FREE MASSIVE MIMO

In this section, we first briefly introduce the system model of cell-free massive MIMO and formulate the total EE optimization problem with conjugate beamforming at the APs, taking into account arbitrary pilot sequence assignments, and imperfect channel estimation. Then, by using penalty functions (PFs), we reformulate the total EE optimization problem into the form so that the APG method can be applied.

A. System Model

1) Spectral Eficiency: We consider a cell-free massive MIMO downlink, which is shown in Fig. 1, where M APs coherently serve K users. To take the advantage of channel reciprocity, we consider time division duplex (TDD) operation, where channel information only needs to be estimated in the uplink training phase, and is used in both uplink and downlink data transmission phases. All M APs connect to a central processing unit (CPU) through a backhaul network. Each user has a single antenna, while each AP is equipped with N antennas. The propagation channel between AP m and user k is modeled as

$$\mathbf{g}_{mk} = \beta_{mk}^{1/2} \mathbf{h}_{mk}, \tag{7}$$

where β_{mk} is the large-scale fading, and $\mathbf{h}_{mk}$ is the small-scale fading, whose elements are i.i.d. $\mathcal{CN}(0,1)$ RVs. The downlink transmission needs two phases: uplink training and downlink payload data transmission phases. In the training phase, all users send their pilot sequences, $\sqrt{\tau_{\mathrm{p}}} \boldsymbol{\varphi}_k \in \mathbb{C}^{\tau_{\mathrm{p}} \times 1}$, $\forall k$, where $||\boldsymbol{\varphi}_k||^2 = 1$, to all APs in the system. Then, pilot signal received at AP m is

$$\mathbf{Y}_{\mathrm{p},m} = \sqrt{\tau_{\mathrm{p}} \rho_{\mathrm{p}}} \sum_{k=1}^{K} \mathbf{g}_{mk} \boldsymbol{\varphi}_k^H + \mathbf{W}_{\mathrm{p},m}, \tag{8}$$

where ρ_{p} is the normalized transmit signal-to-noise ratio (SNR) of each pilot symbol, $\mathbf{W}_{\mathrm{p},m}$ is the noise matrix whose elements are i.i.d. $\mathcal{CN}(0,1)$ RVs. After that, each AP uses its received pilot signals from all K users to estimate its

local channels using the minimum mean-square error (MMSE) technique [21]. The channel estimate of $\mathbf{g}_{mk}$ is

$$\hat{\mathbf{g}}_{mk} = \frac{\sqrt{\tau_{\mathrm{p}}\rho_{\mathrm{p}}}\beta_{mk}}{\tau_{\mathrm{p}}\rho_{\mathrm{p}}\sum_{k'=1}^{K}\beta_{mk'}\left|\varphi_{k'}^{H}\varphi_{k}\right|^{2}+1}\check{\mathbf{y}}_{\mathrm{p},m\ k}, \qquad (9)$$

where

$$\check{\mathbf{y}}_{\mathbf{p},m\ k} \triangleq \sqrt{\tau_{\mathrm{p}}\rho_{\mathrm{p}}}\mathbf{g}_{mk} + \sqrt{\tau_{\mathrm{p}}\rho_{\mathrm{p}}}\sum_{k'\neq k}^{K}\mathbf{g}_{mk'}\varphi_{k'}^{H}\varphi_{k} + \mathbf{W}_{\mathrm{p},m}\varphi_{k}. \qquad (10)$$

In the downlink data transmission phase, APs use conjugate beamforming technique and the channels estimated in the training phase to precode the symbols intended for all users. The vector of transmitted signals from AP m is

$$\mathbf{x}_{m} = \sqrt{\rho_{\mathrm{d}}}\sum_{k=1}^{K}\sqrt{\eta_{mk}}\hat{\mathbf{g}}_{mk}^{*}q_{k}, \qquad (11)$$

where q_{k} is the symbol intended for user k which satisfies $\mathbb{E}\left\{|q_{k}|^{2}\right\} = 1$, $\forall k$, ρ_{d} is the normalized transmit power at each AP, and η_{mk} is power coefficient between the AP m and user k, satisfying

$$\sum_{k=1}^{K}\eta_{mk}\gamma_{mk} \leq \frac{1}{N}, \quad \forall m, \qquad (12)$$

where

$$\gamma_{mk} = \frac{\tau_{\mathrm{p}}\rho_{\mathrm{p}}\beta_{mk}^{2}}{\tau_{\mathrm{p}}\rho_{\mathrm{p}}\sum_{k'=1}^{K}\beta_{mk'}\left|\varphi_{k'}^{H}\varphi_{k}\right|^{2}+1}. \qquad (13)$$

Then the signal received at user k is given by

$$r_{k} = \sum_{m=1}^{M}\mathbf{g}_{mk}^{T}\mathbf{x}_{m} + w_{k}. \qquad (14)$$

By applying the same technique as the one in [13], the downlink achievable spectral efficiency (SE) of the k-th user can be expressed as

$$\mathsf{S}_{\mathrm{e}k}\left(\{\eta_{mk}\}\right)$$
$$= \frac{\tau_{\mathrm{c}} - \tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\log_{2}\Bigg(1$$
$$+ \frac{\rho_{\mathrm{d}}N^{2}|\bar{\boldsymbol{\gamma}}_{kk}^{T}\bar{\boldsymbol{\eta}}_{k}|^{2}}{\rho_{\mathrm{d}}N^{2}\sum_{k'\neq k}^{K}|\bar{\boldsymbol{\gamma}}_{k'k}^{T}\bar{\boldsymbol{\eta}}_{k'}|^{2} + \rho_{\mathrm{d}}N\sum_{k'=1}^{K}||\boldsymbol{\kappa}_{k'k}\odot\bar{\boldsymbol{\eta}}_{k'}||_{2}^{2}+1}\Bigg), \qquad (15)$$

where τ_{c} is the length of each coherence interval, τ_{p} is the length of the training phase for each coherence interval, $\boldsymbol{\kappa}_{k'k} \triangleq \left[\sqrt{\gamma_{1k'}\beta_{1k}};\sqrt{\gamma_{2k'}\beta_{2k}};\dots;\sqrt{\gamma_{Mk'}\beta_{Mk}}\right] \in \mathbb{R}_{+}^{M}$, $\bar{\boldsymbol{\eta}}_{k} \triangleq \left[\sqrt{\eta_{1k}},\dots,\sqrt{\eta_{Mk}}\right]^{T} \in \mathbb{R}_{+}^{M}$, and $\bar{\boldsymbol{\gamma}}_{k'k} \triangleq |\varphi_{k'}^{H}\varphi_{k}|\left[\gamma_{1k'}\frac{\beta_{1k}}{\beta_{1k'}},\gamma_{2k'}\frac{\beta_{2k}}{\beta_{2k'}},\dots,\gamma_{Mk'}\frac{\beta_{Mk}}{\beta_{Mk'}}\right]^{T}$.

2) Power Consumption Model: In this paper, the total power consumption is modeled as [13]

$$P_{\mathrm{total}} = \sum_{m=1}^{M}P_{m} + \sum_{m=1}^{M}P_{\mathrm{bh},m}, \qquad (16)$$

where P_{m} is the power consumption at the m-th AP, and $P_{\mathrm{bh},m}$ is the power consumed by the backhaul link connecting the CPU and the m-th AP. Specifically, P_{m} is modeled as

$$P_{m} = \frac{1}{\alpha_{m}}\rho_{\mathrm{d}}N_{0}\left(N\sum_{k=1}^{K}\eta_{mk}\gamma_{mk}\right) + NP_{\mathrm{tc},m}, \qquad (17)$$

where $0 \leq \alpha_{m} \leq 1$ is the power amplifier efficiency, N_{0} is the noise power, and $P_{\mathrm{tc},m}$ is the internal power required to run the circuit components at each antenna of the m-th AP. Next, $P_{\mathrm{bh},m}$ is modeled as

$$P_{\mathrm{bh},m} = P_{0,m} + B \cdot \mathsf{S}_{\mathrm{e}}\left(\{\eta_{mk}\}\right) \cdot P_{\mathrm{bt},m}, \qquad (18)$$

where $P_{0,m}$ is a fixed power consumption of each backhaul, $P_{\mathrm{bt},m}$ is the traffic-dependent power, and B is the system bandwidth.

3) Total Energy Efficiency: While spectral efficiency has been a common performance measure for wireless communication design, in this paper we aim to maximize the total energy efficiency of the system, which is defined as how many bits can be transmitted by one Joule. Specifically, the total EE (bit/Joule) can be calculated as

$$\mathsf{E}_{\mathrm{e}}\left(\{\eta_{mk}\}\right) = \frac{B\sum_{k=1}^{K}\mathsf{S}_{\mathrm{e}k}(\{\eta_{mk}\})}{P_{\mathrm{total}}}, \qquad (19)$$

B. Optimization Problem Formulation

Our problem is to maximize the total EE (19) by allocating the power coefficients $\{\eta_{mk}\}$, under a sum power constraint at each AP and a QoS constraint, i.e., SE constraint $\mathsf{S}_{\mathrm{o}k}$ at each user. The optimization problem is stated as

$$(\mathcal{P}): \begin{cases} \displaystyle\max_{\{\eta_{mk}\}}, & \mathsf{E}_{\mathrm{e}}(\{\eta_{mk}\}) \\ \mathrm{s.t.} & \mathsf{S}_{\mathrm{e}k}(\{\eta_{mk}\}) \geq \mathsf{S}_{\mathrm{o}k}, \quad \forall k, \\ & \sum_{k=1}^{K}\eta_{mk}\gamma_{mk} \leq 1/N, \quad \forall m, \\ & \eta_{mk} \geq 0, \quad \forall k, \forall m, \end{cases} \qquad (20)$$

In this paper $(\mathcal{P})$ is assumed to be feasible. An equivalent form of problem $(\mathcal{P})$ can be rewritten as

$$(\mathcal{P}_{1}): \begin{cases} \displaystyle\max_{\{\eta_{mk}\}}, & \dfrac{B\sum_{k=1}^{K}\mathsf{S}_{\mathrm{e}k}(\{\eta_{mk}\})}{\bar{P}_{\mathrm{fix}} + \rho_{\mathrm{d}}N_{0}N\sum_{m=1}^{M}\frac{1}{\alpha_{m}}\sum_{k=1}^{K}\eta_{mk}\gamma_{mk}} \\ \mathrm{s.t.} & \mathsf{S}_{\mathrm{e}k}(\{\eta_{mk}\}) \geq \mathsf{S}_{\mathrm{o}k}, \quad \forall k, \\ & \sum_{k=1}^{K}\eta_{mk}\gamma_{mk} \leq 1/N, \quad \forall m, \\ & \eta_{mk} \geq 0, \quad \forall k, \forall m, \end{cases} \qquad (21)$$

where $\bar{P}_{\mathrm{fix}} \triangleq \sum_{m=1}^{M}(NP_{\mathrm{tc},m}+P_{0,m})$. Note that the objective function of the problem $(\mathcal{P}_{1})$ is nonconvex. The common method to tackle such a nonconvex problem is to iteratively approximate a nonconvex function by a series of convex

functions under the framework of successive convex approximation. In fact, this is the method presented in [13], in which the EE maximization problem is solved by a sequence of SOCPs. However, as the complexity dramatically increases when the system scales up (i.e. the numbers of APs and users increase), such method cannot provide a solution for large-scale optimization problems in cell-free massive MIMO systems with thousands of APs and users. In the next section, we will propose a new algorithm based on the APG method to solve problem $(\mathcal{P}_1)$. Our proposed algorithm has very low complexity, and hence, and can efficiently deal with the systems with many APs and users.

IV. PROPOSED APG ALGORITHM FOR THE TOTAL ENERGY EFFICIENCY OPTIMIZATION

In this section, we apply the APG method to efficiently solve the total energy efficiency optimization problem in Section III. We first reformulate the optimization problem $(\mathcal{P}_1)$ by change of variables such that the gradient of the function and the resulting projection can be computed more efficiently. We also apply a penalty method to convert the total EE maximization problem into the form which is amenable to applying the APG method. Note that our algorithm is done over large-scale fading time scale and is performed at the CPU. The details of these steps are described in the following.

A. Problem Reformulation

In order to apply the APG method for solving problem $(\mathcal{P}_1)$, we need to reformulate $(\mathcal{P}_1)$ into the form of (4) and to make sure that the gradient of the objective is Lipschitz continuous. To this end, we first introduce a new variable $\theta_{mk} = \sqrt{\eta_{mk}\gamma_{mk}}$ and define new notations as follows:

- $\boldsymbol{\theta} \triangleq (\boldsymbol{\theta}_1; \boldsymbol{\theta}_2; \ldots; \boldsymbol{\theta}_M) \in \mathbb{R}_+^{MK}$, where $\boldsymbol{\theta}_m \triangleq [\theta_{m1}; \ldots; \theta_{mK}] \in \mathbb{R}_+^K$, is the vector of all power control coefficients associated with AP m.
- $\mathbf{A}_k \triangleq \mathbf{I}_M \otimes \mathbf{e}_k^T$, where $\mathbf{e}_k \in \mathbb{R}^K$ denotes the k-th unit vector, i.e., the vector such that $e_k = 1$ and $e_j = 0$, $\forall j \neq k$.
- $\tilde{\boldsymbol{\gamma}}_{k'k} \triangleq |\boldsymbol{\varphi}_{k'}^H \boldsymbol{\varphi}_k| \left[\sqrt{\gamma_{1k'}} \frac{\beta_{1k}}{\beta_{1k'}}; \sqrt{\gamma_{2k'}} \frac{\beta_{2k}}{\beta_{2k'}}; \ldots; \sqrt{\gamma_{Mk'}} \frac{\beta_{Mk}}{\beta_{Mk'}} \right]$, and $\tilde{\boldsymbol{\kappa}}_k \triangleq [\sqrt{\beta_{1k}}; \sqrt{\beta_{2k}}; \ldots; \sqrt{\beta_{Mk}}] \in \mathbb{R}_+^M$.

Then, $(\mathcal{P}_1)$ can be rewritten as

$$(\mathcal{P}_2) : \begin{cases} \max_{\boldsymbol{\theta} \in \mathcal{C}}, & B \dfrac{u(\boldsymbol{\theta})}{v(\boldsymbol{\theta})} \triangleq f(\boldsymbol{\theta}) \\ \text{s.t.} & u_k(\boldsymbol{\theta}) \geq \mathsf{S}_{\mathrm{o}k}, \quad \forall k, \end{cases} \tag{22}$$

where

$$\mathcal{C} = \{\boldsymbol{\theta} \mid ||\boldsymbol{\theta}_m||^2 \leq \frac{1}{N}, \quad m = 1, 2, \ldots, M; \boldsymbol{\theta} \geq 0\}, \tag{23}$$

$$v(\boldsymbol{\theta}) = \bar{P}_{\mathrm{fix}} + \rho_{\mathrm{d}} N_0 N \sum_{m=1}^{M} \frac{1}{\alpha_m} ||\boldsymbol{\theta}_m||^2, \tag{24}$$

$$u(\boldsymbol{\theta}) \triangleq \sum_{k=1}^{K} u_k(\boldsymbol{\theta}), \tag{25}$$

and, (26), shown at the bottom of the page. As shall be seen shortly, the projection onto $\mathcal{C}$ can be done by closed form expressions. Thus, the main obstacle in deriving an efficient algorithm for solving $(\mathcal{P}_2)$ is the QoS constraints. To overcome this issue we propose to combine the penalty method and the APG method as described in the next subsection.

B. Proposed Algorithm

The overall structure of the proposed method is as follows:

- The penalty method is invoked to bring the QoS constraints into the objective by some form of a loss function through a penalty parameter, leading to the penalized problem.
- The APG method is then applied to solve the penalized problem. This process is repeated until a stopping criterion is achieved.

1) Penalty Method: The constraint $u_k(\boldsymbol{\theta}) \geq \mathsf{S}_{\mathrm{o}k}$ can be written as, (27), shown at the bottom of the page, where $a_k \triangleq \sqrt{\dfrac{2^{\mathsf{S}_{\mathrm{o}k} \frac{\tau_{\mathrm{c}}}{\tau_{\mathrm{c}} - \tau_{\mathrm{p}}}} - 1}{\rho_{\mathrm{d}} N^2}}$. Then, for each QoS constraint we introduce the following quadratic loss function [19]

$$\Psi_k(\boldsymbol{\theta}) \triangleq [\max(0, g_k(\boldsymbol{\theta}))]^2, \tag{28}$$

where $g_k(\boldsymbol{\theta})$ is defined by (29), shown at the bottom of the page. Note that $g_k(\boldsymbol{\theta})$ is convex and $\Psi_k(\boldsymbol{\theta})$ is smooth. Then, for a given penalty coefficient ξ, the penalized objective function of $(\mathcal{P}_2)$, denoted by $f_\xi(\boldsymbol{\theta})$, is given by

$$f_\xi(\boldsymbol{\theta}) \triangleq B \frac{u(\boldsymbol{\theta})}{v(\boldsymbol{\theta})} - \xi \sum_{k}^{K} \Psi_k(\boldsymbol{\theta}). \tag{30}$$

We remark that the above regularized objective is formed in the context of maximization. Also note that the value of the penalty coefficient ξ should be selected appropriately. If this parameter is large, the feasibility is guaranteed but the resulting optimization problem is numerically ill-conditioned.

$$u_k(\boldsymbol{\theta}) = \frac{\tau_{\mathrm{c}} - \tau_{\mathrm{p}}}{\tau_{\mathrm{c}}} \log_2 \left(1 + \frac{\rho_{\mathrm{d}} N^2 (\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta})^2}{\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} (\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta})^2 + \rho_{\mathrm{d}} N \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + 1} \right). \tag{26}$$

$$\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta} \geq a_k \sqrt{\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} (\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta})^2 + \rho_{\mathrm{d}} N \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + 1}, \tag{27}$$

$$g_k(\boldsymbol{\theta}) \triangleq a_k \sqrt{\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} (\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta})^2 + \rho_{\mathrm{d}} N \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + 1} - \tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta}. \tag{29}$$

On the other hand, if it is too small, it may produce a suboptimal solution or even converges to an infeasible point, i.e. the constraints are violated.

In practice, to avoid the above issues, we can first solve the penalized optimization problem for a small value of ξ and check if the stopping criterion is met. If not, we can increase ξ by $\rho > 1$ times and repeat this process until the stopping criterion is met. In this iterative process, it is critical to use the solution of the previous iteration as the starting point of the next. In essence, the key to the penalty method is to solve the following regularized optimization problem for a given ξ

$$\max_{\boldsymbol{\theta} \in \mathcal{C}} \quad f_\xi(\boldsymbol{\theta}), \tag{31}$$

which has the same form as (4). We are now in a position to apply the APG method to solve (31) which is detailed next section.

2) APG Method: We first show that $f_\xi(\boldsymbol{\theta})$ is a proper function with Lipschitz continuous gradient and bounded from above, and thus, the APG method is applicable to solve (31). Towards this end, it is easy to see that the function $f(\boldsymbol{\theta})$ is proper and bounded from above,[1] which is shown by the following inequalities

$$f_\xi(\boldsymbol{\theta}) = B \frac{\tau_{\mathrm{c}} - \tau_{\mathrm{p}}}{\tau_{\mathrm{c}}} \frac{u(\boldsymbol{\theta})}{v(\boldsymbol{\theta})} - \xi \sum_{k}^{K} \left[\max\left(0, g_k(\boldsymbol{\theta})\right)\right]^2$$

$$\leq B \frac{\tau_{\mathrm{c}} - \tau_{\mathrm{p}}}{\tau_{\mathrm{c}}} \frac{u(\boldsymbol{\theta})}{v(\boldsymbol{\theta})} < \infty.$$

The above inequality holds since the total EE is bounded from above as its numerator, i.e. total SE, is limited by inter-user interference and total power consumption. The Lipschitz continuity of the gradient of $f_\xi(\boldsymbol{\theta})$ is stated in Proposition 1.

Proposition 1: The function $f_\xi(\boldsymbol{\theta})$ shown in (30) is L_f-Lipschitz continuous gradient with a Lipschitz contant L_f given by (84) in Appendix B.

Proof: See Appendix B. ∎

It is now obvious that we can apply the APG method in (5) to solve (31). Our proposed method that combines the penalty method and the APG method is summarized in Algorithm 1, where $\boldsymbol{\theta}_\xi$ denotes an optimal solution to (31). Regarding the APG procedure in Algorithm 1, we note that we have modified (5), accounting for the maximization context, where we move along the gradient to increase the objective of the current point. Note also that, for a practical purpose we stop the APG procedure when the relative increase in the objective during the last 10 iterations is less than ς.

It is clear that the key operations in the implementation of Algorithm 1 are the computation of the gradient $\nabla f_\xi(\boldsymbol{\theta})$ and the projections in (32) and (33). In particular, these two operations can be done in closed-form as shown in Proposition 2 and Proposition 3, respectively.

Proposition 2: $\nabla f_\xi(\boldsymbol{\theta})$ can be calculated as

$$\nabla f_\xi(\boldsymbol{\theta}) = B \frac{v(\boldsymbol{\theta}) \nabla u(\boldsymbol{\theta}) - u(\boldsymbol{\theta}) \nabla v(\boldsymbol{\theta})}{v(\boldsymbol{\theta})^2} - \xi \sum_{k=1}^{K} \nabla \Psi_k(\boldsymbol{\theta}), \tag{34}$$

[1]Note that for a minimization problem the objective should be bounded from below.

Algorithm 1 The Proposed Algorithm for Solving (22)

Input: $\boldsymbol{\theta}^{(0)} \in \mathbb{R}_+^{MK}$, $0 < \alpha_\theta, \alpha_y < 1/L_f$, $\rho > 1$, $\delta > 0$, $\varsigma > 0$, ξ

Initialization: $\boldsymbol{\theta}^{(1)} = \mathbf{z}^{(1)} = \boldsymbol{\theta}^{(0)}$

repeat   /* outer loop: penalty method */
  Set $t^{(1)} = t^{(0)} = 1$; $n \leftarrow 1$; $m \leftarrow 1$
  repeat   /* inner loop: APG method */

$$\mathbf{y}^{(n)} = \boldsymbol{\theta}^{(n)} + \frac{t^{(n-1)}}{t^{(n)}}(\mathbf{z}^{(n)} - \boldsymbol{\theta}^{(n)})$$
$$+ \frac{t^{(n-1)} - 1}{t^{(n)}}(\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta}^{(n-1)})$$
$$\mathbf{z}^{(n+1)} = P_\mathcal{C}(\mathbf{y}^{(n)} + \alpha_y \nabla f_{\xi_m}(\mathbf{y}^{(n)})) \tag{32}$$
$$\mathbf{v}^{(n+1)} = P_\mathcal{C}(\boldsymbol{\theta}^{(n)} + \alpha_\theta \nabla f_{\xi_m}(\boldsymbol{\theta}^{(n)})) \tag{33}$$
$$\boldsymbol{\theta}^{(n+1)} = \begin{cases} \mathbf{z}^{(n+1)}, & \text{if } f_{\xi_m}(\mathbf{z}^{(n+1)}) \\ & \geq f_{\xi_m}(\mathbf{v}^{(n+1)}) \\ \mathbf{v}^{(n+1)}, & \text{otherwise,} \end{cases}$$
$$t^{(n+1)} = \frac{\sqrt{4(t^{(n)})^2 + 1} + 1}{2}$$
$$n \leftarrow n + 1$$

  until $\left| \frac{f_{\xi_m}(\boldsymbol{\theta}^{(n)}) - f_{\xi_m}(\boldsymbol{\theta}^{(n-10)})}{f_{\xi_m}(\boldsymbol{\theta}^{(n)})} \right| \leq \varsigma$;
  Update the starting point for the next iteration:
  $\boldsymbol{\theta}^{(1)} = \mathbf{z}^{(1)} = \boldsymbol{\theta}^{(n)}$
  Set $\boldsymbol{\theta}_{\xi_m} = \boldsymbol{\theta}^{(n)}$
  Increase the penalty parameter: $\xi_{m+1} = \xi_m \times \rho$
  $m \leftarrow m + 1$
until *convergence*;

---

where

$$\nabla u(\boldsymbol{\theta}) = \sum_{k=1}^{K} \nabla u_k(\boldsymbol{\theta}), \tag{35}$$

$$\nabla v(\boldsymbol{\theta}) = \rho_{\mathrm{d}} N_0 N \left[\frac{2}{\alpha_1}\boldsymbol{\theta}_1; \frac{2}{\alpha_2}\boldsymbol{\theta}_2; \dots, \frac{2}{\alpha_M}\boldsymbol{\theta}_M\right], \tag{36}$$

$$\nabla \Psi_k(\boldsymbol{\theta}) = 2 \left[\max\left(0, g_k(\boldsymbol{\theta})\right)\right] \nabla g_k(\boldsymbol{\theta}), \tag{37}$$

and $\nabla g_k(\boldsymbol{\theta})$, $\nabla u_k(\boldsymbol{\theta})$, which are shown at the top of the next page, are given by (38), and (39), shown at the bottom of the next page, respectively.

Proof: See Appendix C. ∎

Proposition 3: The projection $P_\mathcal{C}(\mathbf{u})$ admits the following analytical solution

$$\boldsymbol{\theta}_m = \frac{\sqrt{1/N}}{\max(\|\left[\mathbf{u}_m\right]_+\|, \sqrt{1/N})} \left[\mathbf{u}_m\right]_+, \quad \forall m = 1, 2, \dots, M.$$
$$\tag{40}$$

Proof: See Appendix D. ∎

C. Proposed Algorithm With Line Search

In Algorithm 1 is guaranteed to converge for any fixed step sizes smaller than L_f. However, it is possible that L_f given in (84) is significantly larger than the best Lipschitz constant of the gradient of $f_\xi(\boldsymbol{\theta})$ which is practically difficult to find.

Algorithm 2 Backtracking Line Search for Finding a Step Size for (32) and (33)

Input: $\nu < 1$, $\delta > 0$

$\mathbf{s}^{(n)} = \mathbf{z}^{(n)} - \mathbf{y}^{(n-1)}$; $\mathbf{r}^{(n)} = \nabla f_\xi(\mathbf{z}^{(n)}) - \nabla f_\xi(\mathbf{y}^{(n-1)})$

Set $\alpha_y = \frac{(\mathbf{s}^{(n)})^T \mathbf{s}^{(n)}}{(\mathbf{s}^{(n)})^T \mathbf{r}^{(n)}}$ or $\alpha_y = \frac{(\mathbf{s}^{(n)})^T \mathbf{r}^{(n)}}{(\mathbf{r}^{(n)})^T \mathbf{r}^{(n)}}$,

$\mathbf{s}^{(n)} = \mathbf{v}^{(n)} - \boldsymbol{\theta}^{(n-1)}$; $\mathbf{r}^{(n)} = \nabla f_\xi(\mathbf{v}^{(n)}) - \nabla f_\xi(\boldsymbol{\theta}^{(n-1)})$,

Set $\alpha_\theta = \frac{(\mathbf{s}^{(n)})^T \mathbf{s}^{(n)}}{(\mathbf{s}^{(n)})^T \mathbf{r}^{(n)}}$ or $\alpha_\theta = \frac{(\mathbf{s}^{(n)})^T \mathbf{r}^{(n)}}{(\mathbf{r}^{(n)})^T \mathbf{r}^{(n)}}$,

repeat   /* step size for (32) */

  $\mathbf{z}^{(n+1)} = P_\mathcal{C}(\mathbf{y}^{(n)} + \alpha_y \nabla f_\xi(\mathbf{y}^{(n)}))$,

  $\alpha_y = \alpha_y \nu$,

until $f_\xi(\mathbf{z}^{(n+1)}) \geq f_\xi(\mathbf{y}^{(n)}) + \delta||\mathbf{z}^{(n+1)} - \mathbf{y}^{(n)}||^2$;

repeat   /* step size for (33) */

  $\mathbf{v}^{(n+1)} = P_\mathcal{C}(\boldsymbol{\theta}^{(n)} + \alpha_\theta \nabla f_\xi(\boldsymbol{\theta}^{(n)}))$,

  $\alpha_\theta = \alpha_\theta \nu$,

until $f_\xi(\mathbf{v}^{(n+1)}) \geq f_\xi(\boldsymbol{\theta}^{(n)}) + \delta||\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}||^2$;

In order to find a larger step size, and thus faster convergence, we can carry out a line search to tune the step size in (32) and (33). In this paper, we can perform a line search as described in Algorithm 2, inspired from [20], which works as follows. In each iteration, the backtracking line search starts with a large step size, and then decreases it until a better feasible solution is found. As we can see from Algorithm 2, the algorithm will always terminate with a better point, in the sense of maximizing the objective function $f_\xi(\boldsymbol{\theta})$. Note that the backtracking line search in Algorithm 2 follows the Barzilai-Borwein (BB) rule [22]. As $\nabla f_\xi(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant L_f given in (84), the line search procedure is guaranteed to terminate after finite steps.

D. Convergence Analysis of Proposed Method

The convergence of Algorithm 1 is guaranteed due to that of the APG method and the penalty method. Specifically, for a given ξ, similar to Theorem 1 of [20], we can show that the objective sequence $\{f_\xi(\boldsymbol{\theta}^{(n)})\}$ is monotonically increasing. Also, the sequence $\{\boldsymbol{\theta}^{(n)}\}$ is bounded and thus has accumulation points. Each accumulation point is also a stationary solution to (31). Furthermore, following the arguments in [19, Chap. 10] we can show that the iterate sequence $\{\boldsymbol{\theta}_{\xi_m}\}$ converges (in the subsequence sense) to a feasible point of $(\mathcal{P}_2)$ when $\xi_m \to \infty$. Thus the obtained solution of Algorithm 1 is also a stationary point of $(\mathcal{P}_2)$. The proof of these claims is

given in Appendix E. We note however that since Algorithm 1 will terminate for some finite ξ_m when a pre-determined error tolerance is met, it can only produce an approximate stationary solution of $(\mathcal{P}_2)$.

E. Computational Complexity Analysis

It is obvious that the complexity of Algorithm 1 in each iteration is dominated by that of (32), and (33) and the computation of the objective. Here we use the big-$\mathcal{O}$ notation to analyse the complexity of (32) and (33). From Proposition 2, it is easy to see that the complexity to calculate the gradient of $\nabla f_\xi(\boldsymbol{\theta})$ is $\mathcal{O}(KM^2)$. The projection $P_\mathcal{C}(\mathbf{u})$ requires the complexity of $\mathcal{O}(KM)$ which is obvious from Proposition 3. Similarly, the complexity of computing $f_\xi(\boldsymbol{\theta})$ is $\mathcal{O}(KM^2)$. As a result, the overall complexity of Algorithm 1 is $\mathcal{O}(I_P I_{APG} KM^2)$ where I_P and I_{APG} are the number of iterations of the outer loop (i.e. the penalty method) and the inner loop (i.e. the APG method) in Algorithm 1, respectively. Note that the line search procedure contributes negligible complexity since the gradient can be reused and the projection requires much less complexity. In [13], a SCA method based on solving a sequence of SOCPs was presented. We remark that the complexity of solving an SOCP in each SCA iteration is $\mathcal{O}(\sqrt{K+M}M^3 K^4)$ [23]. Thus the complexity of the SCA method in [13] is $\mathcal{O}(I_{SCA}\sqrt{K+M}M^3 K^4)$ where I_{SCA} is the number of SCA iterations. It is apparent that the computational complexity of our proposed method is much lower than the SCA method in [13]. This point is numerically demonstrated in the next section.

V. NUMERICAL RESULTS

In this section, numerical results will be provided to evaluate as well as show the benefits of our proposed algorithm.

A. System Setup

We consider cell-free massive MIMO systems, where locations of M APs and K users randomly uniformly generated within an area of 1×1 km^2. The wrapped around technique is used. The large-scale fading coefficient is modeled as:

$$\beta_{mk} = \text{PL}_{mk} \cdot z_{mk}, \tag{41}$$

where z_{mk} is the log-normal shadowing with the standard derivation $\sigma_{sh} = 8$ dB, and PL_{mk} is the three-slope-based

$$\nabla g_k(\boldsymbol{\theta}) = Na_k\rho_{\text{d}}\sqrt{\bar{\mu}_k}\Big(\sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T(\tilde{\boldsymbol{\gamma}}_{k'k}\tilde{\boldsymbol{\gamma}}_{k'k}^T + \frac{1}{N}\mathbf{B}_k)\mathbf{A}_{k'} + \frac{1}{N}\mathbf{A}_k^T\mathbf{B}_k\mathbf{A}_k\Big)\boldsymbol{\theta} - \mathbf{A}_k^T\tilde{\boldsymbol{\gamma}}_{kk}, \tag{38}$$

$$\nabla u_k(\boldsymbol{\theta}) = \frac{\rho_{\text{d}}\mu_k(\tau_{\text{c}} - \tau_{\text{p}})}{\tau_{\text{c}}\ln 2}\Big(\mathbf{A}_k^T\tilde{\boldsymbol{\gamma}}_{kk}\tilde{\boldsymbol{\gamma}}_{kk}^T\mathbf{A}_k - \bar{\mu}_k\rho_{\text{d}}(\tilde{\boldsymbol{\gamma}}_{kk}^T\mathbf{A}_k\boldsymbol{\theta})^2\Big(\sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T(\tilde{\boldsymbol{\gamma}}_{k'k}\tilde{\boldsymbol{\gamma}}_{k'k}^T + \frac{1}{N}\mathbf{B}_k)\mathbf{A}_{k'} + \frac{1}{N}\mathbf{A}_k^T\mathbf{B}_k\mathbf{A}_k\Big)\Big)\boldsymbol{\theta},$$

$$\mu_k \triangleq \frac{2}{\rho_{\text{d}}\sum_{k'=1}^{K}(\tilde{\boldsymbol{\gamma}}_{k'k}^T\mathbf{A}_{k'}\boldsymbol{\theta})^2 + \frac{\rho_{\text{d}}}{N}\sum_{k'=1}^{K}||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'}\boldsymbol{\theta})||^2 + \frac{1}{N^2}};$$

$$\bar{\mu}_k \triangleq \frac{1}{\rho_{\text{d}}\sum_{k'=1, k' \neq k}^{K}(\tilde{\boldsymbol{\gamma}}_{k'k}^T\mathbf{A}_{k'}\boldsymbol{\theta})^2 + \frac{\rho_{\text{d}}}{N}\sum_{k'=1}^{K}||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'}\boldsymbol{\theta})||^2 + \frac{1}{N^2}}. \tag{39}$$
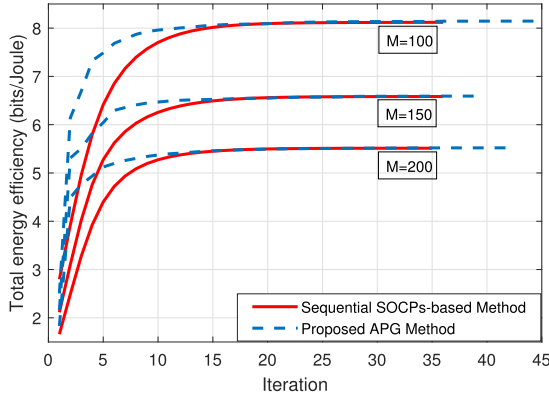
Fig. 2.   Convergence rate of the proposed APG algorithm with $K = 40$, $N = 1$.

path loss, which is modeled (in dB) as

$$
\mathrm{PL}_{mk} = \begin{cases} -L - 35\log_{10}(d_{mk}), & \text{if } d_{mk} > d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_{mk}), & \text{if } d_0 < d_{mk} \le d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_0), & \text{if } d_{mk} \le d_0, \end{cases}
$$

where we choose $d_0 = 10$ m, $d_1 = 50$ m, and $L = 140.7$ dB. The power consumption is summarized as follows: power amplifier coefficient $a_m = 0.4$, $\forall m$; internal power consumption per antenna $P_{\mathrm{tc},m} = 0.2$ $\forall m$; fixed power consumption per each backhaul. In addition, we choose $\mathsf{S}_{\mathrm{o}k} = 1$ bit/s/Hz, $B = 20$ MHz, $\rho_{\mathrm{d}} = 1$ W, $\rho_{\mathrm{p}} = 0.2$ W, and noise figure is 9 dB. In simulation, we implement Algorithm 1 with the line search described in Algorithm 2 where $\rho = 0.5$.

B. Convergence of Proposed Algorithm

In the first experiment, we show the performance of the proposed APG method in comparison with the sequential SOCPs-based method in [13], with $D = 1$ km, $\tau_c = 200$, and $\tau_p = K$. As can be seen in Fig. 2, the proposed method achieves the same EE performance as the SOCPs-based method. In terms of the number of iterations required to output a solution, although our proposed APG method requires more iterations to converge, compared to the SOCPs-based method. However, as we mentioned previously, the proposed method requires very cheap iteration cost, and thus is far more efficient in terms of the actual run time. This point is clearly illustrated in Fig. 3, where we plot the run time of the proposed algorithm and the SOCPs-based method as a function of M. The simulations are built using MATLAB and the results are obtained on a Dell laptop with Intel Core™ i7-9750H and RAM of 16 GB. The stopping criterion is $\varsigma = 10^{-3}$. Compared to the sequential SOCPs-based method, our proposed scheme reduces the run time significantly, i.e, about 62 times and 53 times when $M = 100$ and $M = 400$, respectively.

Next, we need to verify that the proposed APG algorithm will not violate any PFs, or equivalently, the total loss, $\sum_k^K \Psi_k(\boldsymbol{\theta})$, will converge to 0, regardless the starting point. The numerical results are shown in Fig. 4 using two different
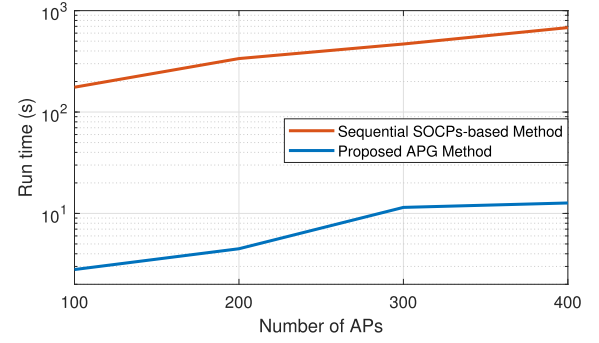

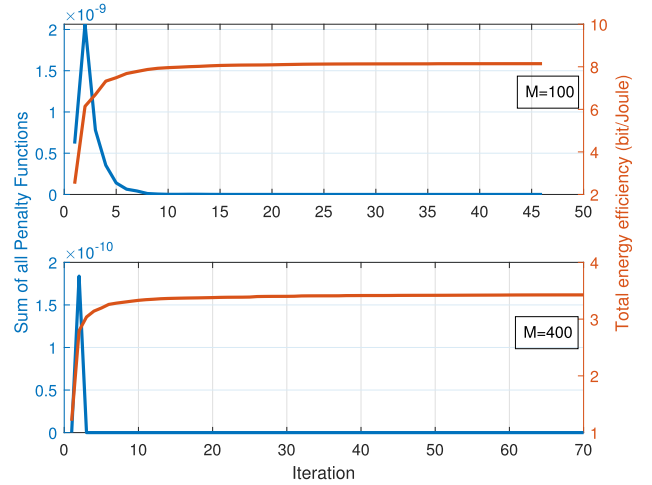Fig. 3.   Run time versus the number of APs M. The number of users is $K = 40$.

Fig. 4.   Convergence rate vs sum of all penalty functions with $K = 40$, $N = 1$.

scenarios with the number of APs, $M = 100$ and $M = 400$, respectively. In this figure, blue curves and orange curves represent total PFs (total loss) and total EE, respectively. In the both scenarios, total PFs starts in infeasible domain, and gradually converges to 0 when the algorithm terminate, as the result of increasing the penalty parameter ξ.

C. Multi-Antenna APs

Fig. 5 examines the effect of multiple antennas at the APs. We consider two scenarios: ($N = 1$ and M changes) and (N changes, $M = 100$). For a fair comparison, both scenarios have the same total number of antennas of all APs, i.e. MN is fixed. The numerical results show that, with fixed number of APs $M = 100$, run time just changes a small amount, or even faster when changing the number of antennas per AP from 1 to 10. However, run time increase proportionally with the number of APs using single antenna. This is indeed an expected result since the complexity of our proposed algorithm only depends on the number of APs M. Regarding to the total EE, Fig. 5 (b) shows that multi-antenna APs always outperform single-antenna APs, on the condition that they have the same total number of antennas at APs. The main reason is that more energy is consumed when single-antena's AP is used. Therefore, based on our numerical results, instead of
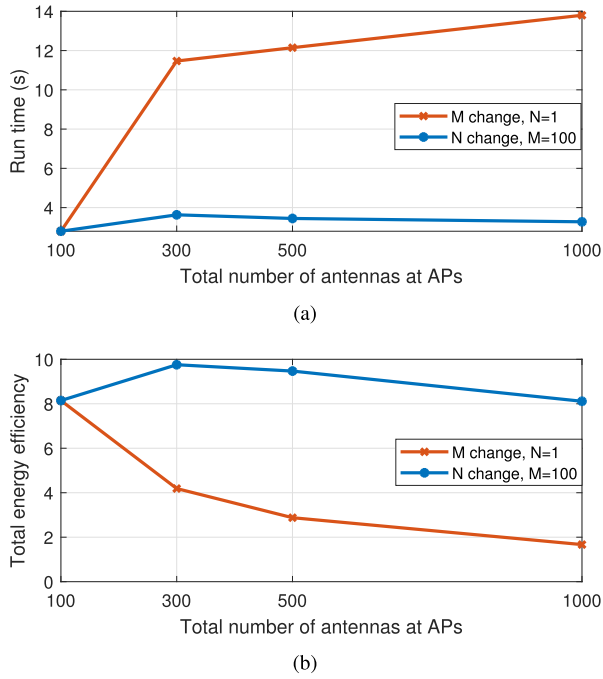
(a)

(b)

Fig. 5. Run time and total energy efficiency vs total number of antennas at APs with $K = 40$.
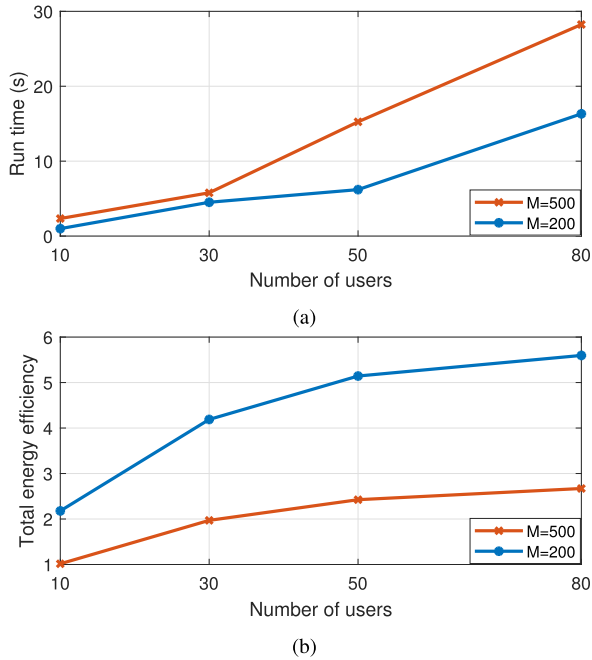

(a)

(b)

Fig. 6. Run time and total energy efficiency vs number of users with $N = 10$.

increasing the number of APs, we should increase the number of anntennas per APs to take the advantages of both run time and total EE of the proposed algorithm.

D. System Setup Based on Number of Users

E. Multi-Antenna APs

Fig. 6 compares two setups of the system with different number of APs using multi-antennas. It clear that both
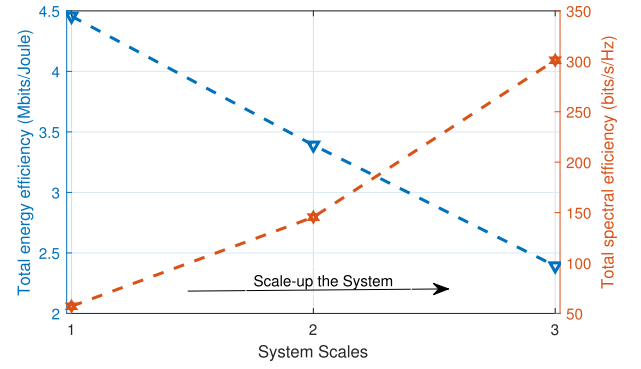

Fig. 7. Total energy efficiency and total spectral efficiency vs system scale with $N = 1$.

system's performance, i.e. total EE, and run time heavily depend on number of APs in the system regardless number of users. In this case, we can see that system with $M = 200$ always outperforms the one with $M = 500$ in term of run time and total EE, when the number of users changes from 10 to 80. Therefore, based on number of users in the system, we can setup the system by just activating the suitable number of APs to achieve higher performance and faster running time.

F. System Scale v.s. Total Energy Efficiency

Fig. 7 considers three different system scale setups, where the first, the second, and the third system scales are corresponding to different number of APs and users $M \times K$, which are 200×40, 500×100, and 1000×200, respectively. We choose $\tau_c = 500$, and $\tau_p = 200$ for all scenarios. The result shows the trade-off between the spectral efficiency and energy efficiency. While the total SE proportionally increases when extending the system scale (because the number of users increases), the total EE shows the opposite trend (because the energy consumption increases). Depending on practical applications and requirements, suitable numbers of APs and users can be designed.

VI. CONCLUSION

We applied a APG method to deal with a large-scale EE optimization problem in cell-free massive MIMO systems, taking into account of power contraints at the APs and QoS contraints at each user. Compared to the sequential SOCPs-based approaches, our proposed method achieves the same performance, while its run time is much faster, i.e., one to two orders of magnitude reduction. Therefore, this method has a great potential to deal with large-scale optimization problems in cell-free massive MIMO, and hence, can be applied to practical scenarios. In addition, our optimization problem is done over large-scale fading time scale. It means the optimal power control coefficients can be updated only once for each large-scale fading realization. Since large-scale fading coefficients change very slowly with time, the system can fully operate in a bursty communication scenario with some random activation of the users.

APPENDIX

A. Useful Properties and Lemmas

In this section, we provide some useful properties and lemmas related to Lipschitz continuity that shall be used to analyse the Lipschitz continuity of the gradient of the objective in (31). Some of the following properties and lemmas are in fact extension of the results in [24, Section 1.5] for scalar-valued functions. We also provide proofs to these for the sake of completeness.

1) Linear Combinations: Let $f_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous with a Lipschitz constants L_1 and L_2, respectively. Then $f_1 \pm f_2$ is Lipschitz continuous with a Lipschitz constant $L_1 + L_2$.

Proof:

$$
\begin{aligned}
&||(f_1 + f_2)(\mathbf{x}) - (f_1 + f_2)(\mathbf{y})|| \\
&\overset{(a)}{\leq} ||f_1(\mathbf{x}) - f_1(\mathbf{y})|| + ||f_2(\mathbf{x}) - f_2(\mathbf{y})|| \\
&\leq (L_1 + L_2) ||\mathbf{x} - \mathbf{y}||,
\end{aligned}
$$

where (a) is the triangle inequality. ∎

2) Product of Functions: Let $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^m$ be bounded functions over a bounded set $\mathcal{S}$, such that $|f_1(\mathbf{x})| \leq M_1$ and $||f_2(\mathbf{x})|| \leq M_2$ for all $\mathbf{x} \in \mathcal{S}$. Further suppose that f_1 and f_2 are Lipschitz continuous with a Lipschitz constants L_1 and L_2, respectively. Then the product of f_1 and f_2 is Lipschitz continuous with a Lipschitz constant $(M_1 L_1 + M_2 L_2)$. *Proof:*

$$
\begin{aligned}
&||f_1 f_2(\mathbf{x}) - f_1 f_2(\mathbf{y})|| \\
&\leq ||f_1(\mathbf{x})(f_2(\mathbf{x}) - f_2(\mathbf{y})) + (f_1(\mathbf{x}) - f_1(\mathbf{y})) f_2(\mathbf{y})|| \\
&\leq ||f_1(\mathbf{x})(f_2(\mathbf{x}) - f_2(\mathbf{y}))|| + ||(f_1(\mathbf{x}) - f_1(\mathbf{y})) f_2(\mathbf{y})|| \\
&\leq M_1 ||f_2(\mathbf{x}) - f_2(\mathbf{y})|| + M_2 ||f_1(\mathbf{x}) - f_1(\mathbf{y})|| \\
&\leq (M_1 L_1 + M_2 L_2) ||\mathbf{x} - \mathbf{y}||.
\end{aligned}
$$

∎

3) The Composition of Functions: Let $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constants L_1 and L_2, respectively. Then the composite function $f_1 \circ f_2$ is Lipschitz continuous with a Lipschitz constant $L_1 L_2$. *Proof:*

$$
\begin{aligned}
||f_1(f_2(\mathbf{x})) - f_1(f_2(\mathbf{y}))|| &\leq L_1 ||f_2(\mathbf{x}) - f_2(\mathbf{y})|| \\
&\leq L_1 L_2 ||\mathbf{x} - \mathbf{y}||.
\end{aligned}
$$

∎

4) The Quotient of Functions: Let $f_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constants L_1 and L_2, respectively on a bounded set $\mathcal{S}$, such that $||f_1(\mathbf{x})|| \leq M$ for all $\mathbf{x} \in \mathcal{S}$, and further assume that there is a constant $c > 0$ such that $|f_2(\mathbf{x})| \geq c$ for all $\mathbf{x} \in \mathcal{S}$. Then f_1/f_2 is Lipschitz continuous with a Lipschitz constant $(ML_1 + L_2/c^3)$.

Proof: First, since f_2 is Lipschitz continuous with a Lipschitz constants L_2, and $|f_2(\mathbf{x})| \geq c$, we have

$$
|1/f_2(x) - 1/f_2(y)| = \frac{|f_2(x) - f_2(y)|}{|f_2(x) f_2(y)|} \leq \frac{L_2}{c^2} ||\mathbf{x} - \mathbf{y}||.
$$

Thus $1/f_2$ is Lipschitz continuous with a Lipschitz constant $\frac{L_2}{c^2}$. Next, applying the product property in A.2 for f_1 and $1/f_2$, we have

$$
||(f_1/f_2)(\mathbf{x}) - (f_1/f_2)(\mathbf{y})|| \leq (ML_1 + L_2/c^3) ||\mathbf{x} - \mathbf{y}||.
$$

∎

Lemma 1: Let $f : \mathbb{R}^n \to \mathbb{R}$, and assume that $||\mathbf{x} + \mathbf{y}|| \leq c$, $\forall \mathbf{x}, \mathbf{y} \in \mathcal{S}$. Then $f(\mathbf{x}) = (\mathbf{a}^T \mathbf{x})^2$ is Lipschitz continuous with a Lipschitz constant $c||\mathbf{a}||^2$.

Proof:

$$
\begin{aligned}
&|f(\mathbf{x}) - f(\mathbf{y})| \\
&= |(\mathbf{a}^T \mathbf{x})^2 - (\mathbf{a}^T \mathbf{y})^2| = |(\mathbf{a}^T \mathbf{x} - \mathbf{a}^T \mathbf{y})(\mathbf{a}^T \mathbf{x} + \mathbf{a}^T \mathbf{y})| \\
&\overset{(b1)}{\leq} |\mathbf{a}^T \mathbf{x} + \mathbf{a}^T \mathbf{y}| \, ||\mathbf{a}|| ||\mathbf{x} - \mathbf{y}|| \overset{(b2)}{\leq} c||\mathbf{a}||^2 ||\mathbf{x} - \mathbf{y}||,
\end{aligned} \tag{42}
$$

where $(b1)$ and $(b2)$ base on the Cauchy-Schwarz inequality. ∎

Lemma 2: Let $f : \mathbb{R}^n \to \mathbb{R}$, and assume that $(||\mathbf{a} \odot \mathbf{x}|| + ||\mathbf{a} \odot \mathbf{y}||) \leq d$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then a Lipschitz constant of $f(\mathbf{x}) = ||\mathbf{a} \odot \mathbf{x}||^2$ is Lipschitz continuous with a Lipschitz constant $d||\mathbf{a}||$.

Proof:

$$
\begin{aligned}
&||f(\mathbf{x}) - f(\mathbf{y})|| \\
&= |\,||\mathbf{a} \odot \mathbf{x}||^2 - ||\mathbf{a} \odot \mathbf{y}||^2| \\
&= |(||\mathbf{a} \odot \mathbf{x}|| - ||\mathbf{a} \odot \mathbf{y}||)(||\mathbf{a} \odot \mathbf{x}|| + ||\mathbf{a} \odot \mathbf{y}||)| \\
&\overset{(c1)}{\leq} ||\mathbf{a} \odot \mathbf{x} - \mathbf{a} \odot \mathbf{y}||(||\mathbf{a} \odot \mathbf{x}|| + ||\mathbf{a} \odot \mathbf{y}||) \\
&\overset{(c2)}{\leq} d||\mathbf{a}|| ||\mathbf{x} - \mathbf{y}||,
\end{aligned} \tag{43}
$$

where $(c1)$ and $(c2)$ are due to the triangle inequality and the Cauchy-Schwarz inequality, respectively. ∎

B. Proof of Proposition 1

First, note that $f_\xi(\boldsymbol{\theta})$ is proper as $\mathrm{dom} f_\xi \neq 0$. Next, we need to prove that $f_\xi(\boldsymbol{\theta})$ is Lipschitz continuous gradient. $\nabla f_\xi(\boldsymbol{\theta})$ can be calculated as

$$
\begin{aligned}
\nabla f_\xi(\boldsymbol{\theta}) = B \frac{v(\boldsymbol{\theta}) \nabla u(\boldsymbol{\theta}) - u(\boldsymbol{\theta}) \nabla v(\boldsymbol{\theta})}{v(\boldsymbol{\theta})^2} \\
- 2\xi \sum_{k=1}^{K} [\max(0, g_k(\boldsymbol{\theta}))] \nabla g_k(\boldsymbol{\theta}).
\end{aligned} \tag{44}
$$

As $\nabla f_\xi(\boldsymbol{\theta})$ is computed from $u_k(\boldsymbol{\theta})$, $v(\boldsymbol{\theta})$, $g_k(\boldsymbol{\theta})$, $\nabla u_k(\boldsymbol{\theta})$, $\nabla v(\boldsymbol{\theta})$, and $\nabla g_k(\boldsymbol{\theta})$ we now need to find the Lipschitz constants of these terms, and then apply properties in Appendix A to conclude the Lipschitz constant of $\nabla f_\xi(\boldsymbol{\theta})$. To this end the following results are in order

1) $\nabla u_k(\boldsymbol{\theta})$ Is Lipschitz Continuous, and $||\nabla u_k(\boldsymbol{\theta})||$ Is Bounded From Above: Recall that

$$
\nabla u_k(\boldsymbol{\theta}) = \frac{\nabla n_k(\boldsymbol{\theta})}{\ln 2 (n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta}))} - \frac{n_k(\boldsymbol{\theta}) \nabla d_k(\boldsymbol{\theta})}{\ln 2 (n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta})) d_k(\boldsymbol{\theta})}, \tag{45}
$$

where

$$n_k(\boldsymbol{\theta}) \triangleq \rho_{\mathrm{d}} N^2 \big(\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta} \big)^2, \tag{46}$$

$$d_k(\boldsymbol{\theta}) \triangleq \rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} \big(\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta} \big)^2$$

$$+ \rho_{\mathrm{d}} N \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + 1, \tag{47}$$

$$\nabla n_k(\boldsymbol{\theta}) = 2\rho_{\mathrm{d}} N^2 \mathbf{A}_k^T \tilde{\boldsymbol{\gamma}}_{kk} \tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta}, \tag{48}$$

and

$$\nabla d_k(\boldsymbol{\theta}) = 2\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta}$$

$$+ 2\rho_{\mathrm{d}} N \sum_{k'=1}^{K} \mathbf{A}_{k'}^T \mathbf{B}_k \mathbf{A}_{k'} \boldsymbol{\theta}. \tag{49}$$

In order to show that $\nabla u_k(\theta)$ is Lipschitz continuous, we will prove that the first term and the second term of the right hand side in (45) are Lipschitz continuous, respectively. First, note that by applying Lemma 1 and Lemma 2, it is easy to see that $n_k(\boldsymbol{\theta})$, $d_k(\boldsymbol{\theta})$ and $n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta})$ are Lipschitz continuous and their Lipschitz constants are

$$L_{n_k} = 2\rho_{\mathrm{d}} N^{3/2} ||\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k||^2, \tag{50}$$

$$L_{d_k} = 2\rho_{\mathrm{d}} N^{3/2} \sum_{k' \neq k}^{K} ||\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'}||^2 + 2\sqrt{K/N} \tilde{\boldsymbol{\kappa}}_k^T \mathbf{1}_M ||\tilde{\boldsymbol{\kappa}}_k||, \tag{51}$$

and

$$L_{n_k, d_k} = 2\rho_{\mathrm{d}} N^{3/2} \sum_{k=1}^{K} ||\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k||^2 + 2\sqrt{K/N} \tilde{\boldsymbol{\kappa}}_k^T \mathbf{1}_M ||\tilde{\boldsymbol{\kappa}}_k||, \tag{52}$$

respectively. Next, we have $\nabla n_k(\boldsymbol{\theta}) = 2\rho_{\mathrm{d}} N^2 \mathbf{A}_k^T \tilde{\boldsymbol{\gamma}}_{kk} \tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta} \triangleq \mathbf{C}_{n_k} \boldsymbol{\theta}$ is Lipschitz continuous, as

$$||\nabla n_k(\mathbf{x}) - \nabla n_k(\mathbf{y})|| = ||\mathbf{C}_{n_k}(\mathbf{x} - \mathbf{y})||$$
$$\leq \lambda_{\max}(\mathbf{C}_{n_k}) ||\mathbf{x} - \mathbf{y}||, \tag{53}$$

where $\lambda_{\max}(\mathbf{C}_{n_k})$ is the largest eigenvalue of $\mathbf{C}_{n_k}$. Moreover,

$$||\nabla n_k(\boldsymbol{\theta})|| = ||\mathbf{C}_{n_k} \boldsymbol{\theta}|| \leq \frac{\lambda_{\max}(\mathbf{C}_{n_k}) \sqrt{M}}{\sqrt{N}}, \tag{54}$$

and

$$\ln 2 \left(n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta}) \right) \geq \ln 2. \tag{55}$$

Then, from (52), (53), (54), and (55), by applying the quotient of functions in Section A.4, we have that $\frac{\nabla n_k(\boldsymbol{\theta})}{\ln 2(n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta}))}$ is Lipschitz continuous with a Lipschitz constant as follow

$$L_{t1} = \frac{(\lambda_{\max}(\mathbf{C}_n))^2 \sqrt{M}}{\sqrt{N}} + \frac{L_{n,d}}{(\ln 2)^2}. \tag{56}$$

By applying the similar method to $\nabla n_k(\boldsymbol{\theta})$, we can prove that $\nabla d_k(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant written as

$$L_{\nabla d_k} = \lambda_{\max}(\mathbf{C}_{d_k}), \tag{57}$$

where $\mathbf{C}_{d_k} \triangleq 2\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} + 2\rho_{\mathrm{d}} N \sum_{k'=1}^{K} \mathbf{A}_{k'}^T \mathbf{B}_k \mathbf{A}_{k'}$. Moreover, we have

$$||\nabla d_k(\boldsymbol{\theta})|| \leq \frac{\lambda_{\max}(\mathbf{C}_{d_k}) \sqrt{M}}{\sqrt{N}}, \tag{58}$$

and

$$|n_k(\boldsymbol{\theta})| \overset{(d1)}{<} n_{\max} < \infty, \tag{59}$$

where $(d1)$ is based on the fact that $|n_k(\boldsymbol{\theta})|$ is bounded by the power constraints. Then, from (50), (57), (58), and (59), by applying the product of functions in Section A.2, it is easy to see that $n_k(\boldsymbol{\theta}) \nabla d_k(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant given by

$$L_{n_k, \nabla d_k} = n_{\max} L_{n_k} + \frac{(\lambda_{\max}(\mathbf{C}_{d_k}))^2 \sqrt{M}}{\sqrt{N}}, \tag{60}$$

and

$$||n_k(\boldsymbol{\theta}) \nabla d_k(\boldsymbol{\theta})|| \leq \frac{n_{\max} \lambda_{\max}(\mathbf{C}_{d_k}) \sqrt{M}}{\sqrt{N}}. \tag{61}$$

Similar to (59), we have

$$|d_k(\boldsymbol{\theta})| < d_{\max} < \infty, \tag{62}$$

for some d_{max}, and thus

$$|n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta})| < d_{\max} + n_{\max} < \infty. \tag{63}$$

Then, from (51), (52), (62), (63) and by applying the product of functions in Section A.2, we can see that $\left(n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta}) \right) d_k(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant given by

$$L_{n_k + d_k, d_k} = d_{\max} L_{d_k} + (n_{\max} + d_{\max}) L_{n_k, d_k}. \tag{64}$$

Next, we have

$$\ln 2 \left(n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta}) \right) d_k(\boldsymbol{\theta}) \geq \ln 2. \tag{65}$$

We now can conclude, from (60), (61), (64), and (65) and by applying the quotient of functions in the Section A.4, that $\frac{n_k(\boldsymbol{\theta}) \nabla d_k(\boldsymbol{\theta})}{\ln 2(n_k(\boldsymbol{\theta}) + d_k(\boldsymbol{\theta})) d_k(\boldsymbol{\theta})}$ is Lipschitz continuous with the following Lipschitz constant

$$L_{t2} = \frac{n_{\max} \lambda_{\max}(\mathbf{C}_{d_k}) \sqrt{M} L_{n_k, \nabla d_k}}{\sqrt{N}} + \frac{L_{n_k + d_k, d_k}}{(\ln 2)^2}. \tag{66}$$

Finally, from (56), and (66), by applying the linear combination of the functions in Section A.1, it follows that $\nabla u_k(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant given by

$$L_{\nabla u_k} = L_{t1} + L_{t2}, \tag{67}$$

and $||\nabla u_k(\boldsymbol{\theta})||$ is bounded as

$$||\nabla u_k(\boldsymbol{\theta})|| \leq \frac{(\lambda_{\max}(\mathbf{C}_{n_k}) + n_{\max} \lambda_{\max}(\mathbf{C}_{d_k})) \sqrt{M}}{\ln 2 \sqrt{N}} \triangleq \zeta_{\nabla u_k}. \tag{68}$$

2) $u_k(\boldsymbol{\theta})$ Is Bounded and Lipschitz Continuous: First, we have

$$u_k(\boldsymbol{\theta}) = \frac{\tau_c - \tau_p}{\tau_c} \log_2\left(1 + \frac{n_k(\boldsymbol{\theta})}{d_k(\boldsymbol{\theta})}\right)$$
$$\leq \frac{\tau_c - \tau_p}{\tau_c} \log_2(1 + n_{\max}) \triangleq \zeta_{u_k}. \tag{69}$$

Next, to prove $u_k(\boldsymbol{\theta})$ is Lipschitz continuous, we first consider the function $h_1(x) = \log_2(x)$ over the domain $1 \leq x \leq x_{\max}$. Note that $h_1(x)$ is continuously differentiable and thus we have

$$\sup_{t \in (1, x_{\max})} |h_1'(t)| = \sup_{t \in (1, x_{\max})} \log_2 e\left|\frac{1}{t}\right| \leq \log_2 e. \tag{70}$$

By the mean value theorem, there exists some ξ between y and z, $y, z \in (1, x_{\max})$, such that

$$|h_1(z) - h_1(y)| = |h_1'(\xi)(z - y)|$$
$$\leq \sup_{t \in (1, x_{\max})} |h_1'(t)| |z - y| \leq \log_2 e |z - y|. \tag{71}$$

In other words, $h_1(x) = \log_2(x)$ is Lipschitz continuous with a constant $\log_2 e$. Next we consider the function $h_2(x) = 1 + \frac{n_k(\boldsymbol{\theta})}{d_k(\boldsymbol{\theta})}$, for which we have

$$1 \leq |d_k(\boldsymbol{\theta})| \overset{(d3)}{<} d_{\max} < \infty \tag{72}$$

where $(d3)$ is based on the fact that $|d_k(\boldsymbol{\theta})|$ is bounded by the power constraints. Then, from (50), (51), (59), and (72), the quotient property in Section A.4 implies that $h_2(x)$ is Lipschitz continuous with a Lipschitz constant found as

$$L_{h_2} = (n_{\max} + d_{\max})(L_n + L_d) + L_d. \tag{73}$$

Finally, from (71), and (73) by applying composition property in Section A.3, we can prove that $u_k(\boldsymbol{\theta})$ is Lipschitz continuous with the following Lipschitz constant

$$L_{u_k} = L_{h_2} \log_2 e. \tag{74}$$

3) $v(\boldsymbol{\theta})$ and $\nabla v(\boldsymbol{\theta})$ Are Bounded and Lipschitz Continuous: It is easy to see that $v(\boldsymbol{\theta})$ is bounded by

$$v(\boldsymbol{\theta}) \leq \bar{P}_{\text{fix}} + \rho_d N_0 N \sum_{m=1}^{M} \frac{1}{N\alpha_m} \triangleq \zeta_v. \tag{75}$$

Next, let us rewrite $v(\boldsymbol{\theta})$ as

$$v(\boldsymbol{\theta}) = \bar{P}_{\text{fix}} + \rho_d N_0 N \sum_{m=1}^{M} \frac{1}{\alpha_m} ||\boldsymbol{\theta}_m||^2 \triangleq \bar{P}_{\text{fix}} + \mathbf{C}_v ||\boldsymbol{\theta}||^2. \tag{76}$$

Then, following the same steps in Section B.1, we can show that $v(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant L_v expressed as

$$L_v = \frac{2\sqrt{M}\lambda_{\max}(\mathbf{C}_v)}{\sqrt{N}}. \tag{77}$$

Next, we have

$$\nabla v(\boldsymbol{\theta}) = \rho_d N_0 N \left[\frac{2}{\alpha_1}\boldsymbol{\theta}_1; \frac{2}{\alpha_2}\boldsymbol{\theta}_2; \ldots, \frac{2}{\alpha_M}\boldsymbol{\theta}_M\right] \triangleq \mathbf{C}_{\nabla v}\boldsymbol{\theta}, \tag{78}$$

and thus $||\nabla v(\boldsymbol{\theta})||$ is bounded by

$$||\nabla v(\boldsymbol{\theta})|| \leq \frac{\sqrt{M}\lambda_{\max}(\mathbf{C}_{\nabla v})}{\sqrt{N}} \triangleq \zeta_{\nabla v}, \tag{79}$$

and $\nabla v(\boldsymbol{\theta})$ is Lipschitz continuous with the following Lipschitz constant

$$L_{\nabla v} = \lambda_{\max}(\mathbf{C}_{\nabla v}). \tag{80}$$

4) $g_k(\boldsymbol{\theta})$ and $\nabla g_k(\boldsymbol{\theta})$ Are Bounded and Lipschitz Continuous: By following the same method in the Section B.1 and using the fact that $d_k(\boldsymbol{\theta}) \geq 1$, we can show that $g_k(\boldsymbol{\theta})$ is Lipschitz continuous with $L_{g_k} = a_k L_{d_k} + ||\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k||$ and bounded as

$$g_k(\boldsymbol{\theta}) \leq a_k d_k(\boldsymbol{\theta}) + \tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta}$$
$$\leq a_k |d_k(\boldsymbol{\theta})| + \frac{n_{\max}}{\rho_d N^2} < a_k d_{\max} + \frac{n_{\max}}{\rho_d N^2} \triangleq \zeta_{g_k}. \tag{81}$$

$\nabla g_k(\boldsymbol{\theta})$ is also Lipschitz continuous with $L_{\nabla g_k} = a_k L_{\nabla d_k}$ and bound as

$$||\nabla g_k(\boldsymbol{\theta})|| < \frac{a_k \rho_d \lambda_{\max}(\mathbf{C}_{d_k})\sqrt{M}}{\sqrt{N}} + \lambda_{\max}(\mathbf{A}_k^T \tilde{\boldsymbol{\gamma}}_{kk}) \triangleq \zeta_{\nabla g_k}. \tag{82}$$

5) $\nabla f_\xi(\boldsymbol{\theta})$ Is Lipschitz Continuous: From (24) we have

$$v(\boldsymbol{\theta})^2 \geq \bar{P}_{\text{fix}}^2, \tag{83}$$

which is due to the fact that the second term in (24) is always greater than or equal to zero. Then, following the same method to find the Lipschitz constant of $\nabla u_k(\theta)$ in Section B.1, we can prove that $\nabla f_\xi(\boldsymbol{\theta})$ is Lipschitz continuous with a Lipschitz constant given by

$$L_{f_\xi} = L_1 + L_2 + L_3, \tag{84}$$

where $L_1 = \frac{B\sum_{k=1}^{K} L_{\nabla u_k}}{\bar{P}_{\text{fix}}^2}$; $L_2 = \frac{B\left((\sum_{k=1}^{K} L_{\nabla u_k})(\sum_{k=1}^{K} \zeta_{\nabla u_k}) + L_v \zeta_v\right)}{\bar{P}_{\text{fix}}^4}$; and $L_3 = 2\xi \sum_{k=1}^{K}(\zeta_{g_k} L_{g_k} + \zeta_{\nabla g_k} L_{\nabla g_k})$.

C. Proof of Proposition 2

Recall that the PF is

$$\Psi_k(\boldsymbol{\theta}) = [\max(0, g_k(\boldsymbol{\theta}))]^2, \tag{85}$$

then

$$\nabla\Psi_k(\boldsymbol{\theta}) = \begin{cases} 0, & g_k(\boldsymbol{\theta}) \leq 0 \\ 2g_k(\boldsymbol{\theta})\nabla g_k(\boldsymbol{\theta}), & g_k(\boldsymbol{\theta}) > 0, \end{cases} \tag{86}$$

which can be rewritten as

$$\nabla\Psi_k(\boldsymbol{\theta}) = 2[\max(0, g_k(\boldsymbol{\theta}))]\nabla g_k(\boldsymbol{\theta}). \tag{87}$$

Using the quotient rule we can write $\nabla f(\boldsymbol{\theta})$ as

$$\nabla f_\xi(\boldsymbol{\theta}) = B\frac{v(\boldsymbol{\theta})\nabla u(\boldsymbol{\theta}) - u(\boldsymbol{\theta})\nabla v(\boldsymbol{\theta})}{v(\boldsymbol{\theta})^2} - \xi\sum_{k=1}^{K}\nabla\Psi_k(\boldsymbol{\theta}), \tag{88}$$

where

$$\nabla v(\boldsymbol{\theta}) = \rho_d N_0 N\left[\frac{2}{\alpha_1}\boldsymbol{\theta}_1; \frac{2}{\alpha_2}\boldsymbol{\theta}_2; \ldots, \frac{2}{\alpha_M}\boldsymbol{\theta}_M\right], \tag{89}$$

and

$$\nabla u(\boldsymbol{\theta}) = \sum_{k=1}^{K} \nabla u_k(\boldsymbol{\theta}). \qquad (90)$$

To find the gradient of $\nabla g_k(\boldsymbol{\theta})$ we recall the following equalities

$$\nabla \left(\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta}\right)^2 = 2\mathbf{A}_{k'}^T \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta}, \qquad (91)$$

$$\nabla \left(||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2\right) = 2\mathbf{A}_{k'}^T \mathbf{B}_k \mathbf{A}_{k'} \boldsymbol{\theta}, \qquad (92)$$

where $\mathbf{B}_k \in R_+^{M \times M}$ is a diagonal matrix whose m-th element is $[\mathbf{B}_k]_m = \beta_{mk}$. Then, by applying the chain rule, we can easily compute the gradient of $\nabla g_k(\boldsymbol{\theta})$ as shown in (38). To find the gradient of $\nabla u_k(\boldsymbol{\theta})$, we first apply the chain rule together with the quotient rule, we have

$$\nabla u_k(\boldsymbol{\theta}) = \frac{d(\boldsymbol{\theta})}{\ln 2 \left(n(\boldsymbol{\theta}) + d(\boldsymbol{\theta})\right)} \frac{d(\boldsymbol{\theta})\nabla n(\boldsymbol{\theta}) - n(\boldsymbol{\theta})\nabla d(\boldsymbol{\theta})}{d(\boldsymbol{\theta})^2}$$
$$= \frac{1}{\ln 2 \left(n(\boldsymbol{\theta}) + d(\boldsymbol{\theta})\right)} \left(\nabla n(\boldsymbol{\theta}) - \frac{n(\boldsymbol{\theta})\nabla d(\boldsymbol{\theta})}{d(\boldsymbol{\theta})}\right), \qquad (93)$$

where

$$n(\boldsymbol{\theta}) \triangleq \rho_{\mathrm{d}} N^2 \left(\tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta}\right)^2, \qquad (94)$$

$$d(\boldsymbol{\theta}) \triangleq \rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} \left(\tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta}\right)^2$$
$$+ \rho_{\mathrm{d}} N \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + 1, \qquad (95)$$

$$\nabla n(\boldsymbol{\theta}) = 2\rho_{\mathrm{d}} N^2 \mathbf{A}_k^T \tilde{\boldsymbol{\gamma}}_{kk} \tilde{\boldsymbol{\gamma}}_{kk}^T \mathbf{A}_k \boldsymbol{\theta}, \qquad (96)$$

and

$$\nabla d(\boldsymbol{\theta}) = 2\rho_{\mathrm{d}} N^2 \sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta}$$
$$+ 2\rho_{\mathrm{d}} N \sum_{k'=1}^{K} \mathbf{A}_{k'}^T \mathbf{B}_k \mathbf{A}_{k'} \boldsymbol{\theta}. \qquad (97)$$

The substitution of (96) and (97) into (93) yields (39).

D. Proof of Proposition 3

We now show that the projection onto $\mathcal{C}$ admits an *analytical solution* and is *parallelizable*. Recall that $P_{\mathcal{C}}(\mathbf{u})$ is explicitly written as

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{MK}} \quad ||\boldsymbol{\theta} - \mathbf{u}||^2 \qquad (98\mathrm{a})$$

$$\text{s.t. } ||\boldsymbol{\theta}_m||^2 \leq \frac{1}{N}, \quad m = 1, 2, \dots, M \qquad (98\mathrm{b})$$

$$\boldsymbol{\theta} \geq 0. \qquad (98\mathrm{c})$$

Note that the objective in (98) is separable with $\boldsymbol{\theta}_m$. Thus (98) boils down to solving the following subproblem for each m

$$\min_{\boldsymbol{\theta}_m \in \mathbb{R}^K} \quad ||\boldsymbol{\theta}_m - \mathbf{u}_m||^2 \qquad (99\mathrm{a})$$

$$\text{s.t. } ||\boldsymbol{\theta}_m||^2 \leq \frac{1}{N} \qquad (99\mathrm{b})$$

$$\boldsymbol{\theta}_m \geq 0. \qquad (99\mathrm{c})$$

Problem (99) is actually the projection onto the intersection of an Euclidean ball and the positive orthant. Finally, the result (40) is a direct application of [25, Theorem 7.1].

E. Convergence Proof of Algorithm 1

The convergence proof is divided into two parts. In the first part of the proof we show that, for a given ξ_m, the APG iterations converge to a stationary solution to the penalized problem (31). In the second part we show that $\{\boldsymbol{\theta}_\xi\}$ converges to a feasible point of $(\mathcal{P}_2)$. Thus the convergent point of Algorithm 1 is indeed a stationary point of $(\mathcal{P}_2)$.

We begin with the first part of the proof by recalling an important inequality of a L_f-Lipschitz continuous gradient function. Specifically, for a function $f(\mathbf{x})$ has a Lipschitz continuous gradient with a constant L_f, the following inequality holds

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla_{\mathbf{x}} f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle - \frac{L_f}{2} ||\mathbf{y} - \mathbf{x}||^2. \quad (100)$$

The projection in (33) is equivalent to

$$\mathbf{v}^{(n+1)} = \operatorname*{argmin}_{\boldsymbol{\theta} \in \mathcal{C}} ||\boldsymbol{\theta} - \boldsymbol{\theta}^{(n)} - \alpha_\theta \nabla f_\xi(\boldsymbol{\theta}^{(n)})||^2 \qquad (101\mathrm{a})$$

$$= \operatorname*{argmax}_{\boldsymbol{\theta} \in \mathcal{C}} \langle \nabla f_\xi(\boldsymbol{\theta}^{(n)}), \boldsymbol{\theta} - \boldsymbol{\theta}^{(n)} \rangle - \frac{1}{2\alpha_\theta} ||\boldsymbol{\theta} - \boldsymbol{\theta}^{(n)}||^2 \qquad (101\mathrm{b})$$

where $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$ and we have used the fact that $||\mathbf{a} - \mathbf{b}||^2 = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 - 2\langle \mathbf{a}, \mathbf{b} \rangle$. Note that when $\boldsymbol{\theta} = \boldsymbol{\theta}^{(n)}$, the objective in (101b) is 0 and $\mathbf{v}^{(n+1)}$ is the optimal solution to (101b). Thus the following inequality is obvious

$$\langle \nabla f_\xi(\boldsymbol{\theta}^{(n)}), \mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)} \rangle - \frac{1}{2\alpha_\theta} ||\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}||^2 \geq 0 \qquad (102)$$

Combining (100) and (102) we obtain

$$f_\xi(\mathbf{v}^{(n+1)})$$
$$\geq f_\xi(\boldsymbol{\theta}^{(n)}) + \langle \nabla f_\xi(\boldsymbol{\theta}^{(n)}), \mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)} \rangle - \frac{L}{2} ||\mathbf{v}^{(n+1)}$$
$$- \boldsymbol{\theta}^{(n)}||^2$$
$$\geq f_\xi(\boldsymbol{\theta}^{(n)}) + \left(\frac{1}{2\alpha_\theta} - \frac{L}{2}\right) ||\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}||^2. \qquad (103)$$

It is easy to see that if $\alpha_\theta < \frac{1}{L_f}$, then $f_\xi(\mathbf{v}^{(n+1)}) \geq f_\xi(\boldsymbol{\theta}^{(n)})$. Next, if $f_\xi(\mathbf{z}^{(n+1)}) \geq f_\xi(\mathbf{v}^{(n+1)})$, then $\boldsymbol{\theta}^{(n+1)} = \mathbf{z}^{(n+1)}$, and

$$f_\xi(\boldsymbol{\theta}^{(n+1)}) = f_\xi(\mathbf{z}^{(n+1)}) \geq f_\xi(\mathbf{v}_{k+1}). \qquad (104)$$

If $f_\xi(\mathbf{z}^{(n+1)}) < f_\xi(\mathbf{v}^{(n+1)})$, then $\boldsymbol{\theta}^{(n+1)} = \mathbf{v}^{(n+1)}$, and

$$f_\xi(\boldsymbol{\theta}^{(n+1)}) = f_\xi(\mathbf{v}^{(n+1)}) \geq f_\xi(\boldsymbol{\theta}^{(n)}). \qquad (105)$$

In summary we have shown that

$$f_\xi(\boldsymbol{\theta}^{(n+1)}) \geq f_\xi(\mathbf{v}^{(n+1)}) \geq f_\xi(\boldsymbol{\theta}^{(n)}). \qquad (106)$$

Since $\mathcal{C}$ is compact convex, $\{\boldsymbol{\theta}^{(n)}\}$ and $\{\mathbf{v}^{(n)}\}$ are bounded. Thus $\{\boldsymbol{\theta}^{(n)}\}$ has accumulation points.

In the second part of the proof, we now assert that any accumulation point is a stationary solution of (31). As $f_\xi(\boldsymbol{\theta}^{(n)})$

is non-decreasing, the objective at all the accumulation points is the same which is denoted by f_ξ^*. Then, from (103) we have

$$
\left(\frac{1}{2\alpha_\theta} - \frac{L}{2}\right)\left\|\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}\right\|^2 \le f_\xi\left(\mathbf{v}^{(n+1)}\right) - f_\xi\left(\boldsymbol{\theta}^{(n)}\right)
$$
$$
\le f_\xi\left(\boldsymbol{\theta}^{(n+1)}\right) - f_\xi\left(\boldsymbol{\theta}^{(n)}\right).
$$
(107)

Summing over $n = 1, 2, \cdots, \infty$, we have

$$
\left(\frac{1}{2\alpha_\theta} - \frac{L}{2}\right)\sum_{n=1}^{\infty}\left\|\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}\right\|^2 \le f_\xi^* - f_\xi\left(\boldsymbol{\theta}^{(1)}\right) < \infty.
$$
(108)

Since $\alpha_\mathbf{x} < \frac{1}{L}$, we can conclude that

$$
\mathbf{v}^{(n+1)} \to \boldsymbol{\theta}^{(n)} \quad \text{as} \quad n \to \infty.
$$
(109)

The optimality condition of (101b) results in

$$
\left\langle \frac{1}{\alpha_\theta}\left(\mathbf{v}^{(n+1)} - \boldsymbol{\theta}^{(n)}\right) - \nabla_{\boldsymbol{\theta}}f_\xi\left(\boldsymbol{\theta}^{(n)}\right), \boldsymbol{\theta} - \boldsymbol{\theta}^{(n)}\right\rangle \le 0, \quad \forall \boldsymbol{\theta} \in \mathcal{C}.
$$
(110)

Let $\boldsymbol{\theta}^*$ be any accumulation point of $\{\boldsymbol{\theta}^{(n)}\}$, i.e. $\{\boldsymbol{\theta}^{(n_j)}\} \to \boldsymbol{\theta}^*$ as $j \to \infty$. From (109) we immediately have that $\{\boldsymbol{\theta}^{(n_j+1)}\} \to \boldsymbol{\theta}^*$. We also note that $\nabla_{\boldsymbol{\theta}}f_\xi(\boldsymbol{\theta})$ is continuous, and thus $\nabla_{\boldsymbol{\theta}}f_\xi(\boldsymbol{\theta}^{(n_j)}) \to \nabla_{\boldsymbol{\theta}}f_\xi(\boldsymbol{\theta}^*)$. By letting $j \to \infty$ in (110), we have $\left\langle \nabla_{\boldsymbol{\theta}}f_\xi(\boldsymbol{\theta}^*), \boldsymbol{\theta} - \boldsymbol{\theta}^*\right\rangle \ge 0$, $\forall \boldsymbol{\theta} \in \mathcal{C}$. This inequality simply means that $\boldsymbol{\theta}^*$ is a stationary solution to (31) which completes the first part of the proof.

Now we show that $\{\boldsymbol{\theta}_{\xi_m}\}$ indeed converges to a feasible point of $(\mathcal{P}_2)$. Note that for small ξ_m, $\boldsymbol{\theta}_{\xi_m}$ may not be feasible to $(\mathcal{P}_2)$, and that the following inequalities always hold for lager ξ_m

$$
f(\boldsymbol{\theta}_{\xi_m}) \ge f_{\xi_m}(\boldsymbol{\theta}_{\xi_m}) = \max_{\boldsymbol{\theta} \in \mathcal{C}} f_{\xi_m}(\boldsymbol{\theta}) \ge f^*,
$$
(111)

where f^* is the optimal objective of $(\mathcal{P}_2)$. In the above, the first equality is due to the negativity of the penalty term and the second inequality is true for two reasons. First, the problem $\max_{\boldsymbol{\theta} \in \mathcal{C}} f_{\xi_m}(\boldsymbol{\theta})$ becomes a convex problem for large ξ_m since $f_{\xi_m}(\boldsymbol{\theta})$ becomes concave. Second, the APG method can find the optimal solution since the problem is now convex. Thus the second inequality in (111) holds because the optimal objective is no less than the objective at any feasible solution.

Let us consider a sequence $\xi_m \to \infty$. Since the sequence $\boldsymbol{\theta}_{\xi_m}$ belongs to compact set, it has a convergent subsequence (i.e. the Bolzano-Weierstrass theorem). Thus we can assume without loss of optimality that $\boldsymbol{\theta}_{\xi_m}$ converges to a certain point $\boldsymbol{\theta}^*$ by abuse of notation. We will show that $\boldsymbol{\theta}^*$ is indeed feasible to $(\mathcal{P}_2)$. First note that since $\boldsymbol{\theta}_{\xi_m} \to \boldsymbol{\theta}^*$ and thus $f(\boldsymbol{\theta}_{\xi_m}) \to f(\boldsymbol{\theta}^*)$ due to the continuity of f. Then from (111) we have $f(\boldsymbol{\theta}^*) \ge f^*$. Suppose to the contrary that $\boldsymbol{\theta}^*$ is infeasible. Since $\boldsymbol{\theta}_{\xi_m} \to \boldsymbol{\theta}^*$ and $\Psi_k(\cdot)$ is continuous, for sufficiently large m we have

$$
\Psi_k(\boldsymbol{\theta}_{\xi_m}) \ge \Psi_k(\boldsymbol{\theta}^*) > 0, \quad k = 1, 2, \dots, K.
$$
(112)

Thus for these k we would have

$$
f_{\xi_m}(\boldsymbol{\theta}_{\xi_m}) = f(\boldsymbol{\theta}_{\xi_m}) - \xi_m\Psi_k(\boldsymbol{\theta}_{\xi_m}) \le f(\boldsymbol{\theta}_{\xi_m}) - \xi_m\Psi_k(\boldsymbol{\theta}^*).
$$
(113)

Thus it is easy to see that $f_{\xi_m}(\boldsymbol{\theta}_{\xi_m}) \to -\infty$ when $\xi_m \to \infty$, which contradicts (111) and thus completes the proof.

REFERENCES

[1] L. N. Tran, H. Quoc Ngo, and T. Q. Duong, "First-order methods for energy-efficient power control in cell-free massive MIMO," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 848–852.

[2] 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.

[3] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[4] G. Interdonato, E. Björnson, H. Quoc Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 197, Dec. 2019.

[5] 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.

[6] C. D'Andrea, A. Garcia-Rodriguez, G. Geraci, L. G. Giordano, and S. Buzzi, "Analysis of UAV communications in cell-free massive MIMO systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 133–147, 2020.

[7] M. Attarifar, A. Abbasfar, and A. Lozano, "Modified conjugate beamforming for cell-free massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 616–619, Apr. 2019.

[8] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[9] S. Shamai (Shitz) and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE VTS 53rd Veh. Technol. Conf., Spring*, vol. 3, May 2001, pp. 1745–1749.

[10] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max–min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6796–6815, Oct. 2019.

[11] J. Qiu, K. Xu, X. Xia, Z. Shen, and W. Xie, "Downlink power optimization for cell-free massive MIMO over spatially correlated Rayleigh fading channels," *IEEE Access*, vol. 8, pp. 56214–56227, 2020.

[12] T. Van Chien, E. Bjornson, and E. G. Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6798–6812, Oct. 2020.

[13] 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.

[14] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program.*, vol. 95, no. 1, pp. 3–51, 2003.

[15] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[16] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[17] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[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] A. Nemirovski, *Optimization II. Numerical Methods for Nonlinear Continuous Optimization*. Haifa, Israel: Technion–Israel Institute of Technology, 1999.

[20] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 379–387.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[22] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, Jan. 1988.

[23] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: SIAM, 2001.

[24] N. Weaver, *Lipschitz Algebras*, 2nd ed. Singapore: World Scientific, 2018.

[25] 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.

Trang C. Mai (Member, IEEE) received the B.E. degree (Hons.) in electrical and electronic engineering from Le Quy Don Technical University, Vietnam, in 2008, the M.E. degree in communication systems from The University of Electro-Communications, Japan, in 2013, and the Ph.D. degree in electrical and electronic engineering from Queen's University Belfast, U.K., in 2019.

He is currently a Research Fellow at Queen's University Belfast. His research interests include cell-free massive MIMO, massive MIMO, applications of convex optimization, machine learning, and deep learning on wireless communications. He serves as a Reviewer for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, and IEEE WIRELESS COMMUNICATIONS LETTERS.

Hien Quoc Ngo (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2007, the M.S. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2010, and the Ph.D. degree in communication systems from Linköping University (LiU), Sweden, in 2015.

In 2014, he visited the Nokia Bell Labs, Murray Hill, NJ, USA. From January 2016 to April 2017, he was a VR Researcher at the Department of Electrical Engineering (ISY), LiU. He was also a Visiting Research Fellow at the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, U.K., funded by the Swedish Research Council. He is currently a Reader (Associate Professor) at Queen's University Belfast. He has coauthored many research papers in wireless communications and coauthored the Cambridge University Press textbook *Fundamentals of Massive MIMO* (2016). His main research interests include massive (large-scale) MIMO systems, cell-free massive MIMO, physical layer security, and cooperative communications.

Dr. Ngo has been a member of Technical Program Committees for several IEEE conferences, such as ICC, GLOBECOM, WCNC, and VTC. He 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 an IEEE COMMUNICATIONS LETTERS Exemplary Reviewer for 2014, an IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer for 2015, and an IEEE WIRELESS COMMUNICATIONS LETTERS Exemplary Reviewer for 2016. He was awarded the UKRI Future Leaders Fellowship in 2019. He currently serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, *Digital Signal Processing*, *Physical Communication* (PHYCOM) (Elsevier), and *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*. He was a Guest Editor of *IET Communications* Special Issue on "Recent Advances on 5G Communications" and IEEE ACCESS Special Issue on "Modelling, Analysis, and Design of 5G Ultra-Dense Networks", in 2017.

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 had held post-doctoral positions at 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 interest is 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 articles 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 was a co-recipient of an IEEE GLOBECOM 2021 Best Paper Award. 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. He is an Associate Editor of *EURASIP Journal on Wireless Communications and Networking*.