| | |
|---|---|
| **Title** | First-Order Methods for Energy-Efficient Power Control in Cell-Free Massive MIMO |
| **Authors(s)** | Tran, Le-Nam, Ngo, Hien Quoc |
| **Publication date** | 2019-11-06 |
| **Publication information** | Tran, Le-Nam, and Hien Quoc Ngo. “First-Order Methods for Energy-Efficient Power Control in Cell-Free Massive MIMO.” IEEE, November 6, 2019. https://doi.org/10.1109/IEEECONF44664.2019.9048924. |
| **Conference details** | The 2019 53rd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, United States of America, 3-6 November 2019 |
| **Publisher** | IEEE |
| **Item record/more information** | http://hdl.handle.net/10197/11911 |
| **Publisher's statement** | © 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. |
| **Publisher's version (DOI)** | 10.1109/IEEECONF44664.2019.9048924 |

# First-Order Methods for Energy-Efficient Power Control in Cell-Free Massive MIMO

*Invited Paper*

Le-Nam Tran* and Hien Quoc Ngo†
*University College Dublin, Ireland. Email: nam.tran@ucd.ie
†Queen's University Belfast, U.K. Email: hien.ngo@qub.ac.uk

***Abstract*—This paper considers a cell-free massive MIMO system with multiple-antenna access points and single-antenna users. The APs use conjugate beamforming to beamform the data to all users in the network. Total energy efficiency maximization is investigated. This optimization problem is nonconvex and thus difficult to solve. Existing solutions are based on second-order optimization methods in connection with convex approximations. These methods have been shown to perform very well but their complexity does not scale favorably with the network size. To tackle this issue, in this paper we propose to use a first-order method for nonconvex programming to our energy efficiency problem. Compared to the second-order methods, the proposed method achieves the same performance, while its run time is much faster. Thus, it is considered as a feasible solution for resource allocation in cell-free massive MIMO systems.**

## I. Introduction

Conventional mobile networks (called cellular networks) are based on cellular topology. The inherent limitation of these systems is the inter-cell interference. This interference becomes more serious for users at the cell boundaries since these users suffer very high interference from neighbouring cells. Therefore, there have been many technologies dealing with the inter-cell interference in cellular networks over the past 40 years. One of a promising technologies is network mutiple-input and multiple-output (MIMO) (a.k.a. coordinated multi-point with joint transmission) which was introduced in the 2000s [1]. The basic idea of network MIMO is that several base stations cooperate to form clusters and jointly serve users in the networks. There are still cells, so inter-cell interference persists. Furthermore, to obtain good performance, network MIMO requires complicated signal co-processing with high computational complexity, and high backhaul overhead to exchange channel state information among the base stations.

Cell-free massive MIMO was introduced in [2] as a scalable and practical version of network MIMO. In cell-free massive MIMO, many access points (APs) distributed in a large area coherently serve many users. Massive MIMO technologies are used to simplify the signal processing and to reduce the backhaul overhead. It is shown in [2] that cell-free massive MIMO can offer uniformly good service for all users in the networks with simple and distributed conjugate beamforming/matched filtering schemes. Therefore, there has been a great deal of interest in cell-free massive MIMO recently [3]–[5].

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 H. Q. Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1.

In cell-free massive MIMO, resource allocation is crucial since it can improve the system performance significantly. Thus, there are many papers working on this aspect. For example, in [2], max-min power allocation and pilot assignment algorithm were proposed and exploited. In [3], the authors investigated a total energy efficiency maximization problem which aims at choosing power control coefficients to maximize the total energy efficiency under the per-AP power and per-user spectral efficiency constraints. The solutions to the resource allocations in most of previous work are based on second-order optimization methods. These methods have very high computational complexity, and hence, does not scalable in large-scale cell-free massive MIMO where we have thousands of APs and users in a very large area.

Motivated by the aforementioned discussion, in this paper, we propose to use a first-order method for nonconvex programming in resource allocation of cell-free massive MIMO systems. We focus on the total energy maximization problem of cell-free massive MIMO as in [3]. The proposed method converges very quickly and achieves the same energy efficiency as the second-order method proposed in [3] does. More importantly, it requires very low computational complexity, and hence, its run time is much quicker than that of the second-order method in [3].

*Notation:* The superscripts $(\cdot)^T$ stands for the transpose. Notation $\odot$ stands for the Hadamard product. Finally, we use $\nabla f(\mathbf{x})$ and $[\mathbf{x}]_+$ to denote the gradient of $f(\mathbf{x})$, and the projector onto the positive orthant, respectively.

## II. Total Energy Efficiency Maximization Problem

This section summarizes the system model, spectral efficiency, and total energy efficiency maximization problem of cell-free massive MIMO given in [3].

We consider the downlink of a cell-free massive MIMO which includes $M$ APs and $K$ users as in [3]. Each AP has $N$ antennas, while each user has a single antenna. All APs coherently serve all users using TDD operation. The downlink transmission includes two main phases: uplink training and downlink data payload transmission phases. During the uplink training phase, all users send their pilot sequences to the APs. Then the APs will estimate the channels to all users based on the received pilot signals. These channel estimates will be used at the APs to precode the symbols intended for all users during the downlink payload data transmission phase.

Since the system model is the same as that in [3], we skip the details. In addition, as in [3], notation is adopted as follows:

- $\beta_{mk}$ is the large-scale fading coefficient of the channel between the $m$-th AP and the $k$-th user.
- $\tau_{\rm c}$ and $\tau_{\rm p}$ are the lengths of each coherence interval and uplink training duration (in samples), respectively.
- The $n$-th component of the estimate of the channel between AP $m$ and user $k$ has variance $\gamma_{mk}$:

$$\gamma_{mk} = \frac{\tau_{\rm p}\rho_{\rm p}\beta_{mk}^2}{\tau_{\rm p}\rho_{\rm p}\sum_{k'=1}^{K}\beta_{mk'}\left|\boldsymbol{\varphi}_{k'}^H\boldsymbol{\varphi}_k\right|^2+1}, \quad (1)$$

  where $\rho_{\rm p}$ is the normalized power of each pilot symbol (normalized by the noise power $N_0$), and $\boldsymbol{\varphi}_k \in \mathbb{C}^{\tau_{\rm p}\times 1}$, where $\|\boldsymbol{\varphi}_k\|^2 = 1$, is the pilot sequence transmitted from the $k$-th user.
- $\rho_{\rm d}$ is the maximum normalized power at each AP.
- The power control coefficients at AP $m$ associated with user $k$ is denoted by $\eta_{mk}$, chosen to satisfy the transmitted power at AP $m$ is constrained by $\rho_{\rm d}$, and hence,

$$\sum_{k=1}^{K}\eta_{mk}\gamma_{mk} \le \frac{1}{N}, \quad \text{for all } m = 1,\dots,M. \quad (2)$$

- $B$ is again the system bandwidth.

As a consequence, we can obtain an achievable spectral efficiency (bit/s/Hz) of the $k$-th user as (3) shown at the top of the next page [3], where $\bar{\boldsymbol{\eta}}_k \triangleq [\sqrt{\eta_{1k}},\dots,\sqrt{\eta_{Mk}}]^T \in \mathbb{R}_+^M$, consists of all power control coefficients associated with user $k$, $\boldsymbol{\kappa}_{k'k} = [\sqrt{\gamma_{1k'}\beta_{1k}};\sqrt{\gamma_{2k'}\beta_{2k}};\dots;\sqrt{\gamma_{Mk'}\beta_{Mk}}] \in \mathbb{R}_+^M$ , and

$$\bar{\boldsymbol{\gamma}}_{k'k} \triangleq |\boldsymbol{\varphi}_{k'}^H\boldsymbol{\varphi}_k|\left[\gamma_{1k'}\frac{\beta_{1k}}{\beta_{1k'}},\gamma_{2k'}\frac{\beta_{2k}}{\beta_{2k'}},\dots,\gamma_{Mk'}\frac{\beta_{Mk}}{\beta_{Mk'}}\right]^T.$$

The corresponding total energy efficiency (bit/Joule) is

$$\mathsf{E}_{\rm e}\left(\{\eta_{mk}\}\right) = \frac{B\cdot\mathsf{S}_{\rm e}\left(\{\eta_{mk}\}\right)}{P_{\rm total}}. \quad (4)$$

In (4), $\mathsf{S}_{\rm e}\left(\{\eta_{mk}\}\right)$ is the sum spectral efficiency given by

$$\mathsf{S}_{\rm e}\left(\{\eta_{mk}\}\right) = \sum_{k=1}^{K}\mathsf{S}_{{\rm e}k}(\{\eta_{mk}\}), \quad (5)$$

and $P_{\rm total}$ is the total power consumption given by [3, Eq. (20)]

$$P_{\rm total} = \rho_{\rm d}N_0\sum_{m=1}^{M}\frac{1}{\alpha_m}\left(N\sum_{k=1}^{K}\eta_{mk}\gamma_{mk}\right) + \sum_{m=1}^{M}(NP_{{\rm tc},m}+P_{0,m}) + B\left(\sum_{m=1}^{M}P_{{\rm bt},m}\right)\mathsf{S}_{\rm e}\left(\{\eta_{mk}\}\right), \quad (6)$$

where $0 < \alpha_m \le 1$ is the power amplifier efficiency, $P_{{\rm tc},m}$ is the internal power required to run the circuit components corresponding to each antenna of the $m$-th AP, $P_{0,m}$ is the traffic-independent power of each backhaul, and $P_{{\rm bt},m}$ is the traffic-dependent power (in Watt per bit/s).

Our goal is allocating the power coefficients $\{\eta_{mk}\}$ to maximize the total energy efficiency (4), under a transmit power constraint at each AP. The optimization problem is formulated as follows:

$$(\mathcal{P}): \begin{cases} \max\limits_{\{\eta_{mk}\}} & \mathsf{E}_{\rm e}(\{\eta_{mk}\}) \\ \text{s.t.} & \sum_{k=1}^{K}\eta_{mk}\gamma_{mk} \le 1/N,\ \forall m, \\ & \eta_{mk} \ge 0,\ \forall k,\ \forall m, \end{cases} \quad (7)$$

which is equivalent to

$$(\mathcal{P}_1): \begin{cases} \max\limits_{\{\eta_{mk}\}} & \dfrac{B\cdot\mathsf{S}_{\rm e}(\{\eta_{mk}\})}{\bar{P}_{\rm fix}+\rho_{\rm d}N_0N\sum\limits_{m=1}^{M}\frac{1}{\alpha_m}\sum\limits_{k=1}^{K}\eta_{mk}\gamma_{mk}} & \text{(8a)} \\ \text{s.t.} & \sum\limits_{k=1}^{K}\eta_{mk}\gamma_{mk} \le 1/N,\ \forall m, & \text{(8b)} \\ & \eta_{mk} \ge 0,\ \forall k,\ \forall m. & \text{(8c)} \end{cases}$$

The problem above is intractable since the objective function is nonconvex. To deal with such nonconvex problems, sequential convex approximation (SCA) has gradually become a standard mathematical tool. The idea of SCA is to approximate a nonconvex program by a series of convex subproblems. In all known solutions for the considered problem, interior point methods (through the use of off-the-shelf convex solvers) are invoked to solve these convex problems. However, it is well known that interior point methods do not scale with the problem size. Thus the existing solutions are unable to explore the full potential of cell-free massive MIMO systems where the number of APs can be in order of thousands. In the following section we present a mathematical framework on which our proposed solution is based to tackle this scalability problem.

## III. Preliminaries

One method of particular interest is the first-order method presented in [6] which concerns the following problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{F(\mathbf{x}) \equiv f(\mathbf{x}) + g(\mathbf{x})\}, \quad (9)$$

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

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \le L||\mathbf{x}-\mathbf{y}||, \forall \mathbf{x},\mathbf{y}. \quad (10)$$

- A2: $g(\mathbf{x})$ is proper and lower semicontinuous.
- A3: $F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ is coercive, i.e., $F(\mathbf{x})$ is bounded from below and $F(x) \to \infty$ when $||\mathbf{x}|| \to \infty$.

Problem (9) includes constrained optimization as a special case. Specifically, let $\mathcal{C}$ be a closed convex set and let $\delta_{\mathcal{C}}(\mathbf{x})$ be its indicator function defined as

$$\delta_{\mathcal{C}}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x}\in\mathcal{C} \\ +\infty & \mathbf{x}\notin\mathcal{C}. \end{cases} \quad (11)$$

Then the constrained minimization problem $\min\ \{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{C}\}$ can be equivalently rewritten as the form of (9) by letting

$$\mathsf{S}_{ek}\left(\{\eta_{mk}\}\right)=\frac{\tau_c-\tau_p}{\tau_c}\log_2\left(1+\frac{\rho_d N^2|\bar{\boldsymbol{\gamma}}_{kk}^T\bar{\boldsymbol{\eta}}_k|^2}{\rho_d N^2\sum\limits_{k'\neq k}^{K}|\bar{\boldsymbol{\gamma}}_{k'k}^T\bar{\boldsymbol{\eta}}_{k'}|^2+\rho_d N\sum\limits_{k'=1}^{K}||\boldsymbol{\kappa}_{k'k}\odot\bar{\boldsymbol{\eta}}_{k'}||_2^2+1}\right) \quad (3)$$

---

$g(\mathbf{x})\equiv\delta_{\mathcal{C}}(\mathbf{x})$. We will focus on this special case of (9) for the rest of the paper.

The authors in [6] propose an accelerated proximal gradient (APG) method for solving (9), consisting of the following iterations:[1]

$$\mathbf{y}^k=\mathbf{x}^k+\frac{t_{k-1}}{t_k}(\mathbf{z}^k-\mathbf{x}^k)+\frac{t_{k-1}-1}{t_k}(\mathbf{x}^k-\mathbf{x}^{k-1}) \quad (12a)$$
$$\mathbf{z}^{k+1}=\mathrm{prox}_{\alpha_y g}(\mathbf{y}^k-\alpha_y\nabla f(\mathbf{y}^k)) \quad (12b)$$
$$\mathbf{v}^{k+1}=\mathrm{prox}_{\alpha_x g}(\mathbf{x}^k-\alpha_x\nabla f(\mathbf{x}^k)) \quad (12c)$$
$$\mathbf{x}^{k+1}=\begin{cases}\mathbf{z}^{k+1} & F(\mathbf{z}^{k+1})\leq F(\mathbf{v}^{k+1})\\ \mathbf{v}^{k+1} & \text{otherwise}\end{cases} \quad (12d)$$
$$t_{k+1}=\frac{\sqrt{4t_k^2+1}+1}{2}, \quad (12e)$$

where $\mathrm{prox}_{\alpha g}(\mathbf{x})$, called the proximal operator, is defined as

$$\mathrm{prox}_{\alpha g}(\mathbf{x}):=\operatorname*{argmin}_{\mathbf{u}}\ g(\mathbf{u})+\frac{1}{2\alpha}||\mathbf{x}-\mathbf{u}||^2. \quad (13)$$

The above algorithm is an extension of Beck and Teboulle's APG method in [7] to solve general nonconvex and nonsmooth programs. For nonconvex programming, $\mathbf{y}^k$ can be a bad extrapolation, and thus $\mathbf{v}^{k+1}$ is computed to address this issue.

In the special case where $g(\mathbf{x})=\delta_{\mathcal{C}}(\mathbf{x})$, where $\mathcal{C}$ is nonempty set, the proximal operator $\mathrm{prox}_{\alpha g}(\mathbf{x})$ reduces to projection onto $\mathcal{C}$:

$$\begin{aligned}\mathrm{prox}_{\alpha g}(\mathbf{x})&:=\operatorname*{argmin}_{\mathbf{u}}\ \delta_{\mathcal{C}}(\mathbf{x})+\frac{1}{2\alpha}||\mathbf{x}-\mathbf{u}||^2\\&=\operatorname*{argmin}_{\mathbf{u}\in\mathcal{C}}\ ||\mathbf{x}-\mathbf{u}||^2\triangleq P_{\mathcal{C}}(\mathbf{x}).\end{aligned} \quad (14)$$

In this case, the resulting APG is more commonly known as accelerated projected gradient method.

**Theorem 1** ( [6, Theorem 1])**.** *Under assumptions A1-A3, the iterates* $\{\mathbf{x}^k\}$ *generated by* (12) *are bounded. Let* $\mathbf{x}^*$ *be any accumulation point of* $\{\mathbf{x}^k\}$*, we have* $0\in\partial F(\mathbf{x}^*)$*, i.e.,* $\mathbf{x}^*$ *is a critical point.*

## IV. Proposed Method

### A. Problem Reformulation

We are in a position to describe our proposed algorithm for solving (8). To proceed we first perform a change of variable $\theta_{mk}=\sqrt{\eta_{mk}\gamma_{mk}}$ and rewrite the considered problem with respect to the new variable. Let $\boldsymbol{\theta}_m\triangleq[\theta_{m1};\dots;\theta_{mK}]\in\mathbb{R}_+^K$ be the vector of all power control coefficients associated with AP $m$ and let $\boldsymbol{\theta}\triangleq(\boldsymbol{\theta}_1;\boldsymbol{\theta}_2;\dots.;\boldsymbol{\theta}_M)\in\mathbb{R}_+^{MK}$. It is easy to see that the feasible set of the problem is simply expressed as

$$\mathcal{C}=\{\boldsymbol{\theta}\mid||\boldsymbol{\theta}_m||^2\leq\frac{1}{N},m=1,2,\dots,M;\boldsymbol{\theta}\geq 0. \quad (15)$$

To rewrite the objective function in $(\mathcal{P}_1)$ as a function of $\boldsymbol{\theta}$, we define $\mathbf{A}_k=\mathbf{I}_M\otimes\mathbf{e}_k^T$, where $\mathbf{e}_k\in\mathbb{R}^K$ denotes the $k$-th unit vector, i.e., the vector such that $e_k=1$ and $e_j=0,\forall j\neq k$. Then we can express $(\mathcal{P}_1)$ in a compact form as

$$\min\quad f(\boldsymbol{\theta}) \quad (16a)$$
$$\text{s.t.}\quad \boldsymbol{\theta}\in\mathcal{C}, \quad (16b)$$

where $f(\boldsymbol{\theta})$ is given by

$$f(\boldsymbol{\theta})\triangleq -B\frac{\tau_c-\tau_p}{\tau_c}\frac{u(\boldsymbol{\theta})}{v(\boldsymbol{\theta})}, \quad (17)$$

where $u(\boldsymbol{\theta})$ and $v(\boldsymbol{\theta})$ are given by (18) and (19), shown at the top of the following page, where

$$\tilde{\boldsymbol{\gamma}}_{k'k}\triangleq|\boldsymbol{\varphi}_{k'}^H\boldsymbol{\varphi}_k|\left[\sqrt{\gamma_{1k'}}\frac{\beta_{1k}}{\beta_{1k'}};\sqrt{\gamma_{2k'}}\frac{\beta_{2k}}{\beta_{2k'}};\dots;\sqrt{\gamma_{Mk'}}\frac{\beta_{Mk}}{\beta_{Mk'}}\right], \quad (20)$$

and $\tilde{\boldsymbol{\kappa}}_k=[\sqrt{\beta_{1k}};\sqrt{\beta_{2k}};\dots;\sqrt{\beta_{Mk}}]\in\mathbb{R}_+^M$. Note that we have rewritten maximization in (8a) as minimization in (16a). Note also that the natural logarithm is used in (18) for mathematical convenience.

Further let $\delta_{\mathcal{C}}(\mathbf{x})$ be the indicator function of the feasible set $\mathcal{C}$. Then we can simply rewrite (16) as

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{MK}}\quad f(\boldsymbol{\theta})+\delta_{\mathcal{C}}(\boldsymbol{\theta}). \quad (21)$$

### B. Algorithm Description

Based on the APG presented above the proposed method is outlined in Algorithm 1. Note that the step sizes $\alpha_x$ and $\alpha_y$ should be less than $1/L_f$ where $L_f$ is a Lipschitz constant of $\nabla f(\boldsymbol{\theta})$. The implementation of the proposed algorithm heavily depends on the computation of $\nabla f(\boldsymbol{\theta})$ and the projection onto the feasible set, which are detailed in what follows.

*1) The gradient of $f(\boldsymbol{\theta})$:* Using the quotient rule we can write $\nabla f(\boldsymbol{\theta})$ as

$$\nabla f(\boldsymbol{\theta})=-B\frac{\tau_c-\tau_p}{\tau_c}\frac{v(\boldsymbol{\theta})\nabla u(\boldsymbol{\theta})-u(\boldsymbol{\theta})\nabla v(\boldsymbol{\theta})}{v(\boldsymbol{\theta})^2}, \quad (22)$$

where $\nabla v(\boldsymbol{\theta})$ is found as

$$\nabla v(\boldsymbol{\theta})=\rho_d N_0 N\big[\frac{2}{\alpha_1}\boldsymbol{\theta}_1;\frac{2}{\alpha_2}\boldsymbol{\theta}_2;\dots,\frac{2}{\alpha_M}\boldsymbol{\theta}_M\big]. \quad (23)$$

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

$$\nabla\big(\tilde{\boldsymbol{\gamma}}_{k'k}^T\mathbf{A}_{k'}\boldsymbol{\theta}\big)^2=2\mathbf{A}_{k'}^T\tilde{\boldsymbol{\gamma}}_{k'k}\tilde{\boldsymbol{\gamma}}_{k'k}^T\mathbf{A}_{k'}\boldsymbol{\theta}, \quad (24)$$

[1]For the sake of brevity, we only present herein the monotone APG. The nonmonotone version is also proposed in [6] which has better convergence rate.

$$u(\boldsymbol{\theta})=\sum_{k=1}^{K}\log\left(1+\frac{\rho_{\mathrm{d}}N^2\big(\tilde{\boldsymbol{\gamma}}_{kk}^T\mathbf{A}_k\boldsymbol{\theta}\big)^2}{\rho_{\mathrm{d}}N^2\sum\limits_{k'\neq k}^{K}\big(\tilde{\boldsymbol{\gamma}}_{k'k}^T\mathbf{A}_{k'}\boldsymbol{\theta}\big)^2+\rho_{\mathrm{d}}N\sum\limits_{k'=1}^{K}||\tilde{\boldsymbol{\kappa}}_k\odot(\mathbf{A}_{k'}\boldsymbol{\theta})||^2+1}\right) \tag{18}$$

$$v(\boldsymbol{\theta})=\bar{P}_{\mathrm{fix}}+\rho_{\mathrm{d}}N_0N\sum_{m=1}^{M}\frac{1}{\alpha_m}||\boldsymbol{\theta}_m||^2 \tag{19}$$

---

**Input:** $\boldsymbol{\theta}^0$, $t_1=t_0=1$, $0<\alpha_\theta<\frac{1}{L_f}$, $0<\alpha_y<\frac{1}{L_f}$

1 Set $\boldsymbol{\theta}^1=\mathbf{z}^1=\boldsymbol{\theta}^0$

2 **for** $n=1,2,\ldots$ **do**

3 $\quad\mathbf{y}^n=\boldsymbol{\theta}^n+\frac{t_{n-1}}{t_n}(\mathbf{z}^n-\boldsymbol{\theta}^n)+\frac{t_{n-1}-1}{t_n}(\boldsymbol{\theta}^n-\boldsymbol{\theta}^{n-1})$

4 $\quad\mathbf{z}^{n+1}=P_{\mathcal{C}}(\mathbf{y}^n-\alpha_y\nabla f(\mathbf{y}^n))$

5 $\quad\mathbf{v}^{n+1}=P_{\mathcal{C}}(\boldsymbol{\theta}^n-\alpha_\theta\nabla f(\boldsymbol{\theta}^n))$

6 $\quad\boldsymbol{\theta}^{n+1}=\begin{cases}\mathbf{z}^{n+1} & F(\mathbf{z}^{n+1})<F(\mathbf{v}^{n+1})\\ \mathbf{v}^{n+1} & \text{otherwise}\end{cases}$

7 $\quad t_{n+1}=\frac{\sqrt{4t_n^2+1}+1}{2}$

8 **end**

9

**Algorithm 1:** Proposed Algorithm.

$$\nabla\big(||\tilde{\boldsymbol{\kappa}}_k\odot(\mathbf{A}_{k'}\boldsymbol{\theta})||^2\big)=2\mathbf{A}_{k'}^T\mathbf{B}_k\mathbf{A}_{k'}\boldsymbol{\theta}, \tag{25}$$

where $\mathbf{B}_k\in R_+^{M\times M}$ is a diagonal matrix whose $m$-th element is $[\mathbf{B}_k]_m=\beta_{mk}$. Using the composition rule for gradient we can write $\nabla u(\boldsymbol{\theta})$ as in (26), shown at the top of the next page.

*2) Projection onto $\mathcal{C}$:* 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

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{MK}}\quad ||\boldsymbol{\theta}-\mathbf{u}||^2 \tag{27a}$$

$$\text{s.t.}\quad ||\boldsymbol{\theta}_m||^2\leq\frac{1}{N}, m=1,2,\ldots,M \tag{27b}$$

$$\boldsymbol{\theta}\geq 0. \tag{27c}$$

Note that the objective in (27) is separable with $\boldsymbol{\theta}_m$. Thus (27) boils down to solving the following subproblem for each $m$

$$\min_{\boldsymbol{\theta}_m\in\mathbb{R}^{K}}\quad ||\boldsymbol{\theta}_m-\mathbf{u}_m||^2 \tag{28a}$$

$$\text{s.t.}\quad ||\boldsymbol{\theta}_m||^2\leq\frac{1}{N} \tag{28b}$$

$$\boldsymbol{\theta}_m\geq 0. \tag{28c}$$

Problem (28) is actually the projection onto the intersection of an Euclidean ball and the positive orthant.

**Lemma 2.** *Problem* (28) *admits the following analytical solution*

$$\boldsymbol{\theta}_m=\frac{\sqrt{1/N}}{\max(||[\mathbf{u}_m]_+||,\sqrt{1/N})}[\mathbf{u}_m]_+. \tag{29}$$

*Proof:* The result is a direct application of [8, Theorem 7.1]. □

---

**Input:** $\rho<1$, $\delta>0$

$\mathbf{s}_n=\mathbf{z}_n-\mathbf{y}_n$, $\mathbf{r}_n=\nabla f(\mathbf{z}_n)-\nabla f(\mathbf{y}_n)$

Set $\alpha_y=\frac{\mathbf{s}_n^T\mathbf{s}_n}{\mathbf{s}_n^T\mathbf{r}_n}$ or $\alpha_y=\frac{\mathbf{s}_n^T\mathbf{r}_n}{\mathbf{r}_n^T\mathbf{r}_n}$

**repeat**

$\quad\mathbf{z}^{n+1}=P_{\mathcal{C}}(\mathbf{y}^n-\alpha_y\nabla f(\mathbf{y}^n))$

$\quad\alpha_y=\alpha_y\rho$

**until** $f(\mathbf{z}^{n+1})\leq f(\mathbf{y}^n)-\delta||\mathbf{z}^{n+1}-\mathbf{y}^{n+1}||^2$;

**Algorithm 2:** Backtracking linear search

### *C. Convergence Analysis*

From (22) and (26), we can show that $\nabla f(\boldsymbol{\theta})$ is Lipschitz continuous, using the results in [9, Section 1.5]. For the sake of brevity, we skip the details here. In reality we do not need to find a Lipschitz of $\nabla f(\boldsymbol{\theta})$ to implement the proposed algorithm. Instead we can carry out a line search to tune the step size. For example, instead of using a fixed step size as in Line 4 of Algorithm 1, we can perform a line search as described in Algorithm 2.

Note that the backtracking line search in Algorithm 2 follows the Barzilai-Borwein (BB) rule. A line search procedure can also be used to replace Line 5 of Algorithm 1. The idea of a line search is to start with a large step size and decrease it until a better feasible solution is found. As $\nabla f(\mathbf{z}_n)$ is Lipschitz continuous for some Lipschitz constant $L_f$, the line search procedure is guaranteed to terminate after finite steps. In the worst case, the line search procedure stops when the step size is smaller than $L_f$.

## V. Numerical results

This section provides numerical results to highlight the benefits of our proposed algorithm. $M$ APs and $K$ users are uniformly located at random within a square of $D\times D$ km$^2$. The wrapped around technique and random pilot assignment are used. The large-scale fading is modeled using [3, Eqs. (45)-(46)]. All parameters related to large-scale fading and power consumption models are chosen the same as those in [3]. Furthermore, we choose $\tau_{\mathrm{c}}=200$, $\tau_{\mathrm{p}}=20$, $B=20$ MHz, $\rho_{\mathrm{d}}=1$ W, $\rho_{\mathrm{p}}=0.2$ W, and noise figure is 9 dB. In all figures, the results are obtained by implementing Algorithm 1 with a line search where $\rho=0.5$.

In the first experiment we show the convergence rate of the proposed first order method in comparison with the SCA-based method in [3]. Here $D=1$ km. As can be seen, the proposed method converges very quickly and achieves the

$$\nabla u(\boldsymbol{\theta}) = \rho_{\rm d} \sum_{k=1}^{K} \Big( \mu_k \sum_{k'=1}^{K} \mathbf{A}_{k'}^T \big( \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T + \frac{1}{N} \mathbf{B}_k \big) \mathbf{A}_{k'} - \bar{\mu}_k \big( \sum_{k' \neq k}^{K} \mathbf{A}_{k'}^T \big( \tilde{\boldsymbol{\gamma}}_{k'k} \tilde{\boldsymbol{\gamma}}_{k'k}^T + \frac{1}{N} \mathbf{B}_k \big) \mathbf{A}_{k'} + \frac{1}{N} \mathbf{A}_k^T \mathbf{B}_k \mathbf{A}_k \big) \Big) \boldsymbol{\theta}, \tag{26a}$$

$$\mu_k = \frac{2}{\rho_{\rm d} \sum_{k'=1}^{K} \big( \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta} \big)^2 + \frac{\rho_{\rm d}}{N} \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + \frac{1}{N^2}}, \tag{26b}$$

$$\bar{\mu}_k = \frac{2}{\rho_{\rm d} \sum_{k'=1, k' \neq k}^{K} \big( \tilde{\boldsymbol{\gamma}}_{k'k}^T \mathbf{A}_{k'} \boldsymbol{\theta} \big)^2 + \frac{\rho_{\rm d}}{N} \sum_{k'=1}^{K} ||\tilde{\boldsymbol{\kappa}}_k \odot (\mathbf{A}_{k'} \boldsymbol{\theta})||^2 + \frac{1}{N^2}}, \tag{26c}$$

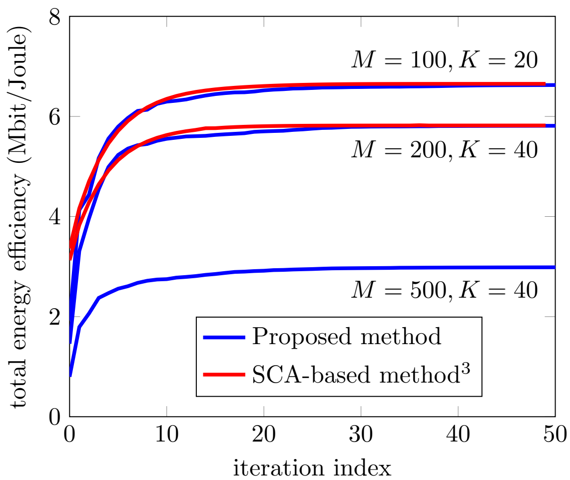


Fig. 1. Convergence rate of the proposed algorithm.

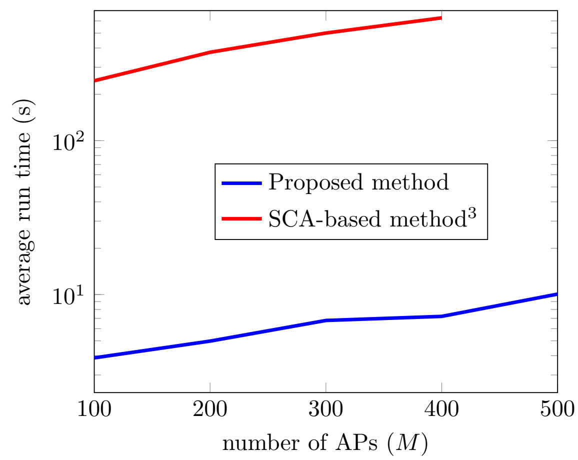


Fig. 2. Run time of proposed algorithm versus the number of APs $M$. The number of users is $K = 40$.

same EE performance as the SCA-based method. In terms of the number of iterations required to output a solution, both methods perform similarly. However, as we mentioned previously, the proposed method requires very cheap iteration cost, and thus is far more efficient in terms of actual run time. This point is clearly illustrated in Fig. 2, where we plot the run time of the proposed algorithm and the SCA-based method as a function of $M$. The results are obtained on a PC with Intel® Core™ i7-8650U and RAM of 16 GB. The stopping criterion is $\left|\frac{f(\boldsymbol{\theta}^n)-f(\boldsymbol{\theta}^{n-10})}{f(\boldsymbol{\theta}^n)}\right| \leq 10^{-4}$. Compared to the SCA-based method, our proposed scheme reduces the run time significantly, i.e, about 60 times and 75 times when $M = 100$ and $M = 400$, respectively. The gain is more significant when the network size is large (i.e. thousands of APs and users).

## VI. Conclusion

In this paper we proposed to apply a first-order method to solve an energy-efficiency power control problem in very large-scale cell-free massive MIMO systems. The proposed method worked very well and achieved the same energy efficiency as the SCA approaches in previous work did. More importantly, our proposed method could offer about one to two orders of magnitude reduction in run time, compared to the SCA approaches. This enables us to fully characterize the potential of these systems as well as design them efficiently.

## References


[1] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitter end," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2001.

[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] 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.

[4] 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.

[5] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive mimo communications," *EURASIP J. Wireless Commun. Netw.*, Aug. 2019.

[6] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Proceedings of NIPS'15 - Volume 1*. MIT Press, pp. 379–387. [Online]. Available: http://dl.acm.org/citation.cfm?id=2969239.2969282

[7] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[8] H. Bauschke, M. Bui, and X. Wang, "Projecting onto the intersection of a cone and a sphere," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2158–2188, 2018. [Online]. Available: https://doi.org/10.1137/17M1141849

[9] N. Weaver, *Lipschitz Algebras*, 2nd ed. WORLD SCIENTIFIC, 2018. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9911