Fast First-Order Algorithm for Large-Scale Max-Min Fair Multi-Group Multicast Beamforming

Chong Zhang, *Student Member, IEEE,* Min Dong, *Senior Member, IEEE,* and Ben Liang, *Fellow, IEEE*

Abstract—We propose a first-order fast algorithm for the weighted max-min fair (MMF) multi-group multicast beamforming problem in large-scale systems. Utilizing the optimal multicast beamforming structure obtained recently, we convert the nonconvex MMF problem into a min-max weight minimization problem and show that it is a weakly convex problem. We propose using the projected subgradient algorithm (PSA) to solve the problem directly, instead of the conventional method that requires iteratively solving its inverse problem. We show that PSA for our problem has closed-form updates and thus is computationally cheap. Furthermore, PSA converges to a near-stationary point of our problem within finite time. Simulation results show that our PSA-based algorithm offers near-optimal performance with considerably lower computational complexity than existing methods for large-scale systems.

Index Terms—Multicast beamforming, optimal beamforming structure, large scale, projected subgradient, weakly convex optimization.

I. Introduction

Content distribution through wireless multicasting has become increasingly popular among wireless applications. Efficient transmission techniques via multicast beamforming have become crucial to support high-speed content distribution. With massive multiple-input multiple-output (MIMO) becoming the essential technology for future networks, it is critical to develop effective and computationally efficient multicast beamforming solutions suitable for large-scale systems.

Early works studied the multicast beamforming design for traditional multi-antenna systems in various scenarios, including a single user group or multiple user groups [1], [2], multi-cell networks [3], and relay networks [4]. Since the family of multicast beamforming problems are nonconvex and NP-hard, the existing works have focused on developing numerical algorithms or signal processing techniques for good suboptimal solutions. Semi-definite relaxation (SDR) has been a widely adopted common approach [1]–[3]. However, as wireless systems are becoming large-scale, the successive convex approximation (SCA) method [5] becomes more popular for its computational and performance advantages over SDR as the size of the problem grows [6]–[8]. Despite the improvement, SCA relies on second-order interior-point methods (IPMs) to solve each convex approximation, where the computational complexity is still too high for massive MIMO systems. Several algorithms were proposed to improve the computational

efficiency at each SCA iteration, such as zero-forcing pre-processing [9] and alternating direction method of multipliers (ADMM) [10] for multi-group scenarios, and first-order methods [11] for single-group scenarios. The optimal multicast beamforming structure has been obtained recently in [12], which is shown to be a weighted minimum mean square error (MMSE) filter with an inherent low-dimensional structure. This structure helps convert the beamforming problem into a weight optimization problem of a much lower dimension [12], allowing design opportunities for efficient algorithms for massive MIMO systems.

The multi-group multicast beamforming design can be cast into either a quality-of-service (QoS) problem for power minimization with signal-to-interference-and-noise (SINR) guarantees, or a max-min fair (MMF) problem for maximizing the minimum SINR subject to some transmit power budget. They are inverse problems [2]. Although both problems are nonconvex, the MMF problem is a max-min problem that is much more complicated to solve than the QoS problem [1], [2]. Typically, the solution to the MMF problem is obtained via iteratively solving its inverse QoS problem along with a bi-section search [2], [8], [10], [12]. The QoS problem at each iteration can then be solved by either SDR or SCA. This additional layer of iteration leads to high computational complexity, especially for large-scale systems.

To address the above issue, in this letter, we propose a fast first-order algorithm for the weighted MMF multi-group multicasting problem. We focus on the min-max weight optimization problem, which is transformed from the original MMF problem by using the optimal beamforming structure [12]. We show that this converted problem is weakly convex, and the projected subgradient algorithm (PSA) [13] can be efficiently used to solve it *directly*. In particular, we show that for our problem, PSA provides closed-form subgradient update and projection and thus, is computationally cheap. Furthermore, based on the recent convergence result for weakly convex problems, we show that PSA converges to a near-stationary point of our problem within finite time. We further propose an initialization method for faster convergence. Our simulation results show that our PSA-based algorithm offers near-optimal performance with substantially lower computational complexity than the existing state-of-the-art algorithms for large-scale systems.

II. Problem Formulation

We consider a downlink multi-group multicast beamforming scenario, where the base station (BS) equipped with N antennas transmits messages to G multicast groups. Each

group receives a common message that is independent of the messages to other groups. Denote the set of group indices by $\mathcal{G} \triangleq \{1, \cdots, G\}$. Assume that there are K_i single-antenna users in group i, with the set of user indices denoted by $\mathcal{K}_i \triangleq \{1, \cdots, K_i\}, i \in \mathcal{G}$. The total number of users in all groups is denoted by $K_{\text{tot}} \triangleq \sum_{i=1}^{G} K_i$. Let $\mathbf{w}_i \in \mathbb{C}^N$ be the multicast beamforming vector for group i, and let $\mathbf{h}_{ik} \in \mathbb{C}^N$ be the channel vector from the BS to user k in group i, for $k \in \mathcal{K}_i$, $i \in \mathcal{G}$. The received signal at user k in group i is given by

$$y_{ik} = \mathbf{w}_i^H \mathbf{h}_{ik} s_i + \sum_{j \neq i} \mathbf{w}_j^H \mathbf{h}_{ik} s_j + n_{ik}$$

where s_i is the symbol transmitted to group i with $E[|s_i|^2] = 1$, and n_{ik} is the receiver additive white Gaussian noise with zero mean and variance σ^2. The received SINR at this user is

$$\text{SINR}_{ik} = \frac{|\mathbf{w}_i^H \mathbf{h}_{ik}|^2}{\sum_{j \neq i} |\mathbf{w}_j^H \mathbf{h}_{ik}|^2 + \sigma^2}. \quad (1)$$

The BS transmit power is given by $\sum_{i=1}^{G} \|\mathbf{w}_i\|^2$.

This letter focuses on the weighted MMF multicast beamforming problem, *i.e.,* maximizing the minimum weighted SINR among all users, subject to the BS transmit power constraint. We assume that all $\mathbf{h}_{ik}$'s are perfectly known at the BS. Define $\mathbf{w} \triangleq [\mathbf{w}_1^H, \cdots, \mathbf{w}_G^H]^H$. The weighted MMF problem is given by

$$\mathcal{S}_o : \max_{\mathbf{w}} \min_{i,k} \frac{\text{SINR}_{ik}}{\gamma_{ik}} \quad \text{s.t.} \quad \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \leq P$$

where P is the maximum power budget at the BS, and $\{\gamma_{ik}\}$ are the weights to control the grade of service or fairness among users.

Problem $\mathcal{S}_o$ is a nonconvex max-min optimization problem and is known to be NP-hard. Existing methods in the literature are through iteratively solving the dual problem of $\mathcal{S}_o$ – the QoS problem, *i.e.,* minimizing the transmit power subject to minimum SINR targets [2], [8]. Specifically, consider the following equivalent problem to $\mathcal{S}_o$:

$$\mathcal{S}_1 : \max_{\mathbf{w},t} t$$
$$\text{s.t.} \ \text{SINR}_{ik} \geq t\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \quad \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \leq P.$$

The dual QoS problem to $\mathcal{S}_1$ is given as follows:

$$\mathcal{P}_o : \min_{\mathbf{w}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \quad \text{s.t.} \quad \text{SINR}_{ik} \geq t\gamma_{ik}, \ k \in \mathcal{K}_i, i \in \mathcal{G}.$$

The solution to $\mathcal{S}_o$ is computed by solving $\mathcal{P}_o$ along with a bi-section search over t until the transmit power is equal to P. The popular methods in the literature to solve the nonconvex problem $\mathcal{P}_o$ are SDR and, recently, SCA. SCA has an advantage in both performance and computational efficiency for large-scale problems. It convexifies the problem first and relies on the second-order IPM to solve the corresponding convex approximation problem [2], [8], [12]. However, the computational complexity of the IPM is still high for large-scale problems. As a result, the iterative method to solve $\mathcal{S}_o$ via $\mathcal{P}_o$ incurs high computational complexity for wireless systems with large-scale antenna arrays or a large number of users.

In this letter, we propose a fast first-order algorithm to solve $\mathcal{S}_o$ directly with low computational complexity.

III. PRELIMINARY: OPTIMAL MULTICAST BEAMFORMING STRUCTURE

The structure of the optimal multicast beamforming solution to $\mathcal{S}_o$ has recently been obtained in [12], which is shown to be a weighted MMSE filter. Under this solution structure, $\mathcal{S}_o$ is transformed into an equivalent weight optimization problem of a much smaller size that is independent of the number of antennas N. This structure can be utilized for substantial computational saving for a solution in large-scale systems. The optimal multicast beamforming solution is given by [12, Theorem 2]

$$\mathbf{w}_i^o = \mathbf{R}^{-1} \mathbf{H}_i \mathbf{a}_i^o, \quad i \in \mathcal{G} \quad (2)$$

where $\mathbf{H}_i \triangleq [\mathbf{h}_{i1}, \cdots, \mathbf{h}_{iK_i}]$ is the channel matrix for group i, $\mathbf{a}_i^o \in \mathbb{C}^{K_i}$ is the optimal weight vector for group i, and $\mathbf{R}$ is the (normalized) noise plus weighted channel covariance (of all users) matrix given in a semi-closed form. To further simplify the required computation, the approximate expression of $\mathbf{R}$ is obtained in [12]. Express each channel as $\mathbf{h}_{ik} = \sqrt{\beta_{ik}} \mathbf{g}_{ik}$, where β_{ik} is the channel variance, and $\mathbf{g}_{ik}$ is the normalized channel vector representing the small-scale fading whose elements are i.i.d. zero mean. The approximate expression of $\mathbf{R}$, for large N, is given by

$$\widetilde{\mathbf{R}} = \mathbf{I} + \frac{P}{\sigma^2} \sum_{i=1}^{G} \sum_{k=1}^{K_i} \frac{\eta_{ik}}{\sum_{i'=1}^{G} \sum_{k'=1}^{K_{i'}} \frac{\eta_{i'k'}}{\beta_{i'k'}}} \mathbf{g}_{ik} \mathbf{g}_{ik}^H \quad (3)$$

where $\eta_{ik} \triangleq \frac{\gamma_{ik}}{(N - \sum_{jl \neq ik} \gamma_{jl})}$. In particular, for $\gamma_{ik} = \gamma, \forall i, k$, $\widetilde{\mathbf{R}}$ in (3) is further simplified to

$$\widetilde{\mathbf{R}} = \mathbf{I} + \frac{P\bar{\beta}}{\sigma^2 K_{\text{tot}}} \sum_{i=1}^{G} \sum_{k=1}^{K_i} \mathbf{g}_{ik} \mathbf{g}_{ik}^H \quad (4)$$

where $\bar{\beta} \triangleq \frac{K_{\text{tot}}}{\sum_{i=1}^{G} \sum_{k=1}^{K_i} \frac{1}{\beta_{ik}}}$ is the harmonic mean of the channel variances of all users. With the solution $\mathbf{w}_i^o$ in (2) and $\widetilde{\mathbf{R}}$ to approximate $\mathbf{R}$, the original problem $\mathcal{S}_o$ can be transformed into the following weight optimization problem

$$\mathcal{S}_2 : \max_{\mathbf{a}} \min_{i,k} \frac{1}{\gamma_{ik}} \frac{\mathbf{a}_i^H \widetilde{\mathbf{A}}_{iik} \mathbf{a}_i}{\sum_{j \neq i} \mathbf{a}_j^H \widetilde{\mathbf{A}}_{jik} \mathbf{a}_j + \sigma^2}$$

$$\text{s.t.} \quad \sum_{i=1}^{G} \|\widetilde{\mathbf{C}}_i \mathbf{a}_i\|^2 \leq P$$

where $\mathbf{a} \triangleq [\mathbf{a}_1^H, \cdots, \mathbf{a}_G^H]^H$, $\widetilde{\mathbf{C}}_i \triangleq \widetilde{\mathbf{R}}^{-1} \mathbf{H}_i$, and $\widetilde{\mathbf{A}}_{jik} \triangleq \widetilde{\mathbf{C}}_j^H \mathbf{h}_{ik} \mathbf{h}_{ik}^H \widetilde{\mathbf{C}}_j$, $k \in \mathcal{K}_i$, $i, j \in \mathcal{G}$. Note that the dimension of weight vector $\mathbf{a}$ (K_{tot}) is much lower than that of the beamforming vector $\mathbf{w}$ (GN) in massive MIMO systems with $K_i \ll N$. Hence, the complexity in computing the beamforming solution is substantially reduced by optimizing $\mathbf{a}$ in $\mathcal{S}_2$, instead of $\mathbf{w}$ in $\mathcal{S}_o$.

Note that [12] focuses on obtaining the optimal beamforming structure, while it still adopts the commonly used numerical algorithms, such as SDR or SCA, when solving the optimization problems. Different from [12], in this letter, we utilize the optimal structure and focus on proposing a

fast numerical algorithm for solving the max-min optimization problem, which provides a more efficient computational method to obtain a solution to the MMF problem.

IV. FIRST-ORDER FAST ALGORITHM

Using the optimal beamforming structure, in this section, we propose a fast first-order algorithm to solve $\mathcal{S}_2$. Problem $\mathcal{S}_2$ is a nonconvex max-min problem. Based on the structure of $\mathcal{S}_2$, we show that PSA can be applied to compute a near-stationary solution to $\mathcal{S}_2$ efficiently.

A. Problem Reformulation

For the purpose of computation, we express all complex quantities in $\mathcal{S}_2$ using their real and imaginary parts. Define
$$\mathbf{x}_i \triangleq [\Re\mathfrak{e}\{\mathbf{a}_i\}^T, \Im\mathfrak{m}\{\mathbf{a}_i\}^T]^T, \ \mathbf{C}_i \triangleq \begin{bmatrix} \Re\mathfrak{e}\{\widetilde{\mathbf{C}}_i\} & -\Im\mathfrak{m}\{\widetilde{\mathbf{C}}_i\} \\ \Im\mathfrak{m}\{\widetilde{\mathbf{C}}_i\} & \Re\mathfrak{e}\{\widetilde{\mathbf{C}}_i\} \end{bmatrix},$$
$$\mathbf{A}_{jik} \triangleq \begin{bmatrix} \Re\mathfrak{e}\{\widetilde{\mathbf{A}}_{jik}\} & -\Im\mathfrak{m}\{\widetilde{\mathbf{A}}_{jik}\} \\ \Im\mathfrak{m}\{\widetilde{\mathbf{A}}_{jik}\} & \Re\mathfrak{e}\{\widetilde{\mathbf{A}}_{jik}\} \end{bmatrix}, \text{ for } k \in \mathcal{K}_i, \ i,j \in \mathcal{G}.$$
It follows that $\|\widetilde{\mathbf{C}}_i\mathbf{a}_i\|^2 = \|\mathbf{C}_i\mathbf{x}_i\|^2$ and $\mathbf{a}_j^T\widetilde{\mathbf{A}}_{jik}\mathbf{a}_j = \mathbf{x}_j^T\mathbf{A}_{jik}\mathbf{x}_j$. Using the above, we can express problem $\mathcal{S}_2$ equivalently in the real domain as

$$\mathcal{S}_3 : \ \max_{\mathbf{x}\in\mathcal{X}} \min_{i,k} \frac{1}{\gamma_{ik}} \frac{\mathbf{x}_i^T\mathbf{A}_{iik}\mathbf{x}_i}{\sum_{j\neq i}\mathbf{x}_j^T\mathbf{A}_{jik}\mathbf{x}_j + \sigma^2}$$

where $\mathbf{x} \triangleq [\mathbf{x}_1^T,\ldots,\mathbf{x}_G^T]^T$, and $\mathcal{X} \triangleq \{\mathbf{x} : \sum_{i=1}^G \|\mathbf{C}_i\mathbf{x}_i\|^2 \leq P\}$ is the compact convex feasible set of $\mathcal{S}_3$. Define $\phi_{ik}(\mathbf{x}) \triangleq \frac{-\mathbf{x}_i^T\mathbf{A}_{iik}\mathbf{x}_i}{\gamma_{ik}(\sum_{j\neq i}\mathbf{x}_j^T\mathbf{A}_{jik}\mathbf{x}_j+\sigma^2)}, \ k \in \mathcal{K}_i, i \in \mathcal{G}$. Then, we can rewrite $\mathcal{S}_3$ in an equivalent min-max form as $\min_{\mathbf{x}\in\mathcal{X}} \max_{i,k} \phi_{ik}(\mathbf{x})$, which is further equivalent to

$$\mathcal{S}_4 : \ \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \ f(\mathbf{x},\mathbf{y})$$

where $f(\mathbf{x},\mathbf{y}) \triangleq \boldsymbol{\phi}^T(\mathbf{x})\mathbf{y}$, with $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^{K_{\text{tot}}}$ containing all $\phi_{ik}(\mathbf{x})$'s, and $\mathcal{Y} \triangleq \{\mathbf{y} : \mathbf{y} \succcurlyeq \mathbf{0}, \mathbf{1}^T\mathbf{y} = 1\}$ is a probability simplex, which is a compact convex set.[1] An optimal solution to the inner maximization of $\mathcal{S}_4$ is $\mathbf{y} = [0,\ldots,0,1,0,\ldots,0]^T$, with 1 at some ith position. Note that $f(\mathbf{x},\mathbf{y})$ is concave in $\mathbf{y}$ and nonconvex in $\mathbf{x}$. Thus, $\mathcal{S}_4$ is a nonconvex-concave min-max problem and is NP-hard. Let $g(\mathbf{x}) \triangleq \max_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x},\mathbf{y})$. Then, we express $\mathcal{S}_4$ as

$$\mathcal{S}_5 : \ \min_{\mathbf{x}\in\mathcal{X}} \ g(\mathbf{x}).$$

Note that $g(\mathbf{x})$ is nonconvex. If $g(\mathbf{x})$ is differentiable, one can use the projected gradient descent [14] to solve $\mathcal{S}_5$. However, in our problem, $g(\mathbf{x})$ may not be differentiable, and its gradient $\nabla g(\mathbf{x})$ may not exist. In what follows, by examining the structure of the problem, we propose to use PSA [13] to find a solution at the vicinity of a stationary point for $\mathcal{S}_5$.

B. The Projected Subgradient Algorithm

We first show the structure of our problem. We assume the channel gain is finite for each user: $\|\mathbf{h}_{ik}\| < \infty, \ \forall k,i$. Thus, all elements in $\mathbf{A}_{jik}$ are finite, $\forall k \in \mathcal{K}_i, \ i,j \in \mathcal{G}$. Also, note that $\gamma_{ik} > 0, \ \sigma^2 > 0$. It follows that, since $\mathcal{X}$

---

[1]Note that by introducing $\mathbf{y}$, we transform $\mathcal{S}_3$ into $\mathcal{S}_4$. The structure in $\mathcal{S}_4$ will benefit our exposition in Section IV-B for the proposed algorithm and convergence analysis.

---

and $\mathcal{Y}$ are compact, the gradient $\nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y})$ is finite for any $\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}$. Thus, there exists a constant $L > 0$, such that $\|\nabla_{\mathbf{x}}f(\mathbf{x}_1,\mathbf{y}) - \nabla_{\mathbf{x}}f(\mathbf{x}_2,\mathbf{y})\| \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|$, for any $\mathbf{x}_1,\mathbf{x}_2 \in \mathcal{X}, \ \mathbf{y} \in \mathcal{Y}$. This means that, $f(\mathbf{x},\mathbf{y})$ is an L-smooth function of $\mathbf{x} \in \mathcal{X}$, which satisfies the following [15]

$$f(\mathbf{x}_2,\mathbf{y}) \geq f(\mathbf{x}_1,\mathbf{y}) + \nabla_{\mathbf{x}}f(\mathbf{x}_1,\mathbf{y})^T(\mathbf{x}_2 - \mathbf{x}_1)$$
$$- \frac{L}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|^2, \ \forall \mathbf{x}_1,\mathbf{x}_2 \in \mathcal{X}, \forall \mathbf{y} \in \mathcal{Y}. \quad (5)$$

Next, we show that $\nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y})$ is a subgradient of $g(\mathbf{x})$. The Fréchet subdifferential of $g(\mathbf{x})$ is the set of subgradients of $g(\mathbf{x})$ defined by $\partial g(\mathbf{x}) \triangleq \{\mathbf{v} \mid \liminf_{\mathbf{x}'\to\mathbf{x}} \frac{g(\mathbf{x}')-g(\mathbf{x})-\mathbf{v}^T(\mathbf{x}'-\mathbf{x})}{\|\mathbf{x}'-\mathbf{x}\|} \geq 0\}$ [16]. By the definition of $g(\mathbf{x})$ and from (5), for any $\mathbf{x}' \in \mathcal{X}$, we have

$$g(\mathbf{x}') \geq f(\mathbf{x},\mathbf{y}) + \nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y})^T(\mathbf{x}' - \mathbf{x}) - \frac{L}{2}\|\mathbf{x}' - \mathbf{x}\|^2$$
$$= g(\mathbf{x}) + \nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y})^T(\mathbf{x}' - \mathbf{x}) - \frac{L}{2}\|\mathbf{x}' - \mathbf{x}\|^2. \quad (6)$$

After rearranging the terms at both sides of the inequality in (6) and taking $\liminf$ for $\mathbf{x}' \to \mathbf{x}$, we conclude that $\nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y}) \in \partial g(\mathbf{x})$.

Following the above result, we propose to solve $\mathcal{S}_5$ by PSA with the following updating procedure:

At iteration j:
$$\mathbf{y}^{(j)} \in \arg\max_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x}^{(j)},\mathbf{y}), \quad (7)$$
$$\mathbf{x}^{(j+1)} = \Pi_{\mathcal{X}}\big(\mathbf{x}^{(j)} - \alpha\nabla_{\mathbf{x}}f(\mathbf{x}^{(j)},\mathbf{y}^{(j)})\big) \quad (8)$$

where $\alpha > 0$ is the step size, and $\Pi_{\mathcal{X}}(\mathbf{x})$ denotes the projection of point $\mathbf{x}$ onto set $\mathcal{X}$, given by

$$\Pi_{\mathcal{X}}(\mathbf{x}) = \begin{cases} \sqrt{\frac{P}{P_{\mathbf{x}}}}\mathbf{x} & \mathbf{x} \notin \mathcal{X} \\ \mathbf{x} & \mathbf{x} \in \mathcal{X} \end{cases} \quad (9)$$

where $P_{\mathbf{x}} \triangleq \sum_{i=1}^G \|\mathbf{C}_i\mathbf{x}_i\|^2$.

Note that the inherent structure of our problem makes PSA particularly suitable for solving $\mathcal{S}_5$. First, $\mathbf{y}^{(j)}$ in (7) can be directly obtained by taking the maximum among $\phi_{ik}(\mathbf{x}^{(j)})$'s. Second, the projection $\Pi_{\mathcal{X}}(\mathbf{x})$ is a simple closed-form function in (9), and $\nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{y})$ has a closed-form expression. Thus, the computation of $\mathbf{x}^{(j+1)}$ in (8) is inexpensive. Below, we discuss the convergence result for the proposed PSA.

1) Convergence Analysis: Based on the recent results on weakly convex problems [17]–[19], we show that PSA converges within finite time to a near-stationary point of $\mathcal{S}_5$. Recall that $f(\mathbf{x},\mathbf{y})$ is L-smooth over $\mathcal{X}$, and $\mathcal{Y}$ is compact. It follows that $g(\mathbf{x}) = \max_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x},\mathbf{y})$ is L-weakly convex over $\mathcal{X}$, i.e., $g(\mathbf{x}) + \frac{L}{2}\|\mathbf{x}\|^2$ is convex for $\mathbf{x} \in \mathcal{X}$ [20, Lemma 1]. Consider the extension of $g(\mathbf{x})$ to $\mathbb{R}^{2K_{\text{tot}}}$: $\tilde{g}(\mathbf{x}) = g(\mathbf{x}) + \mathbb{I}_{\mathcal{X}}(\mathbf{x})$, where $\mathbb{I}_{\mathcal{X}}(\mathbf{x})$ is an indicator function, taking 0 if $\mathbf{x} \in \mathcal{X}$ and ∞ otherwise. Define the Moreau envelope [21] of $\tilde{g}(\mathbf{x})$ as

$$\tilde{g}_\lambda(\mathbf{x}) \triangleq \min_{\mathbf{x}'\in\mathbb{R}^{2K_{\text{tot}}}} \{\tilde{g}(\mathbf{x}') + \frac{1}{2\lambda}\|\mathbf{x}' - \mathbf{x}\|^2\} \quad (10)$$

where $\lambda < 1/L$. The Moreau envelope $\tilde{g}_\lambda(\mathbf{x})$ is a smooth approximation to the non-smooth but L-weakly convex function $g(\mathbf{x})$ over $\mathcal{X}$. Note from the earlier discussion that the objective function of the minimization problem in (10) is

strictly convex. Let $\widehat{\mathbf{x}} \triangleq \arg\min_{\mathbf{x}'}\{\tilde{g}(\mathbf{x}') + \frac{1}{2\lambda}\|\mathbf{x}' - \mathbf{x}\|^2\}$. Then, we have $\widehat{\mathbf{x}} \in \mathcal{X}$ and

$$\|\widehat{\mathbf{x}} - \mathbf{x}\| = \lambda\|\nabla\tilde{g}_\lambda(\mathbf{x})\|. \tag{11}$$

Thus, $\|\nabla\tilde{g}_\lambda(\mathbf{x})\| \leq \epsilon$ implies that [21]

$$\|\widehat{\mathbf{x}} - \mathbf{x}\| \leq \lambda\epsilon, \quad \text{and} \quad \min_{\mathbf{u}\in\partial\tilde{g}(\widehat{\mathbf{x}})}\|\mathbf{u}\| \leq \epsilon. \tag{12}$$

The above means that a small gradient $\|\nabla\tilde{g}_\lambda(\mathbf{x})\| \leq \epsilon$ implies that $\mathbf{x}$ is close to a point $\widehat{\mathbf{x}}$ that is a near-stationary (*i.e.*, ϵ-stationary) point of $\mathcal{S}_5$. Hence, $\|\nabla\tilde{g}_\lambda(\mathbf{x})\|$ provides a near-stationarity measure of $\mathbf{x}$ to a stationary point of $\mathcal{S}_5$. Based on this, we have the convergence result of PSA for $\mathcal{S}_5$ below.

Theorem 1. Assume the continuous function $f(\mathbf{x}, \mathbf{y})$ is C-Lipschitz over $\mathcal{X} \times \mathcal{Y}$. Define $D \triangleq \max_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|$, $M \triangleq \max_{\mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y}} \|\nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{y})\|$, and $\Delta \triangleq \min\{LD^2, CD\}$. Starting from $\mathbf{x}^{(0)} \in \mathcal{X}$, let J be the total number of iterations in PSA. Let step size $\alpha = \sqrt{\frac{\Delta}{LM^2(J+1)}}$. Let the output of PSA be $\bar{\mathbf{x}} = \mathbf{x}^{(j)}$, where $j \sim \text{Uniform}[0, J]$. Then, $\bar{\mathbf{x}}$ satisfies

$$E\|\nabla\tilde{g}_{\frac{1}{2L}}(\bar{\mathbf{x}})\|^2 \leq \frac{4\sqrt{\Delta LM^2}}{\sqrt{J+1}}. \tag{13}$$

Proof: See Appendix A. ∎

Theorem 1 indicates that if we take a random sample in $\{\mathbf{x}^{(j)}\}_{j=0}^J$ as the output of PSA $\bar{\mathbf{x}}$, then $E\|\nabla\tilde{g}_{\frac{1}{2L}}(\bar{\mathbf{x}})\|^2$ decreases in the order of (at most) $O(\frac{1}{\sqrt{J}})$. A more direct way to interpret this result is that, to obtain the output $\bar{\mathbf{x}}$ satisfying $E\|\nabla\tilde{g}_{\frac{1}{2L}}(\bar{\mathbf{x}})\| \leq \epsilon$, the required number of iterations J for PSA is *at most* $O(\epsilon^{-4})$. Thus, for weakly convex problem $\mathcal{S}_5$, Theorem 1 shows that PSA converges within finite time, upper bounded by $O(\epsilon^{-4})$, to an ϵ-accuracy point of $\mathcal{S}_5$.

By Theorem 1, we can also set the stopping criterion for PSA. The convergence analysis in Theorem 1 is based on a random sample in $\{\mathbf{x}^{(j)}\}_{j=0}^J$. Thus, to implement PSA, we can set a random stopping point $J^o < J$ for PSA and use $\mathbf{x}^{(J^o)}$ as the algorithm output.

2) Initialization: An easy-to-compute good initial point is essential to accelerate the convergence of PSA. Using (2) to transfer $\mathcal{P}_o$ into optimizing $\mathbf{a}$ (denoted by $\mathcal{P}'_o$) instead of $\mathbf{w}$, we propose to use SDR with Gaussian randomization (GR) [12] to solve $\mathcal{P}'_o$ along with one bi-section search over t to generate the initial point. The one-step bi-section is inexpensive and is intended to find $\mathbf{x}^{(0)}$ that is closer to the optimal solution. Note that this initial point may not be feasible in $\mathcal{X}$. Nonetheless, after one iteration via the projection step in (8), the subsequent points are feasible. This initialization method has low computational complexity and generates an initial point very close to a stationary point when $\mathcal{S}_5$ is of a small to moderate size.

V. SIMULATION RESULTS

We set $G = 3$, $K_i = K$, and SINR target $\gamma_{ik} = \gamma = 10$ dB, $\forall k, i$. The channels are generated i.i.d. as $\mathbf{h}_{ik} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and the receiver noise variance is $\sigma^2 = 1$. We set $P/\sigma^2 = 10$ dB. The approximate expression $\widetilde{\mathbf{R}}$ in (4) is used in the simulation. For PSA, we set the step size $\alpha = 0.01$ and set the stopping criterion as $|g(\mathbf{x}^{(j+1)}) - g(\mathbf{x}^{(j)})| \leq 10^{-5}$. Based on various simulation experiments, PSA generally converges
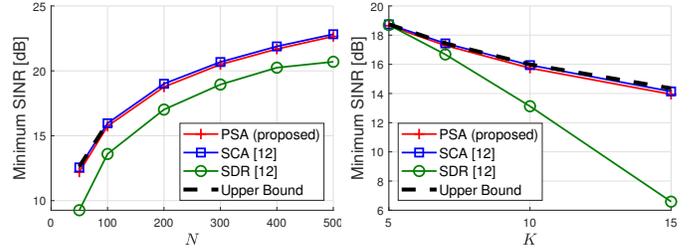

Fig. 1. Left: Average minimum SINR vs. N. Right: Average minimum SINR vs. K ($G = 3, N = 100$).

TABLE I
AVERAGE COMPUTATION TIME OVER N (SEC.) ($G = 3, K = 10$)

N	100	200	300	400	500
PSA (proposed)	1.830	2.031	2.524	2.753	2.821
SCA [12]	87.12	81.17	81.71	86.63	90.54
SDR [12]	11.47	12.36	13.07	14.41	14.90

TABLE II
AVERAGE COMPUTATION TIME OVER K (SEC.) ($G = 3, N = 100$)

K	5	7	10	15
PSA (proposed)	0.814	1.128	1.851	3.676
SCA [12]	24.54	47.20	87.53	174.1
SDR [12]	6.063	8.342	11.33	21.60

within $30 \sim 5000$ iterations.[2] Besides our proposed PSA with the initialization method, we consider the following methods for comparison: 1) The upper bound for $\mathcal{S}_o$: It is obtained by solving $\mathcal{P}_o$ using SDR along with the bi-section search over t; 2) SDR: It uses the optimal structure in (2) with $\widetilde{\mathbf{R}}$ in (4) and solves $\mathcal{S}_o$ by solving $\mathcal{P}_o$ using SDR plus GR along with the bi-section search over t [12]; 3) SCA: It uses the optimal structure in (2) with $\widetilde{\mathbf{R}}$ in (4) and solves $\mathcal{S}_o$ by solving $\mathcal{P}_o$ via SCA and the bi-section search over t. CVX is used in each SCA iteration [12]. The proposed initialization method is used in all methods.

Fig. 1-Left shows the average minimum SINR vs. N for $K = 10$. Both PSA and SCA nearly attain the upper bound for all values of N. SDR is about 2dB worse, as the approximation by SDR deteriorates when the number of constraints (GK) becomes large. Compared with other methods, our proposed PSA is much more computationally efficient in obtaining the solution. Table I shows the corresponding computation time, which includes computing the initial point. The average computation time of PSA is only about 3% of that of SCA and about 20% of that of SDR.

Fig. 1-Right shows the average minimum SINR vs. K for $N = 100$. Again, both PSA and SCA nearly attain the upper bound for all values of K, while SDR deteriorates substantially as K becomes large. The corresponding average computation times (including the initial point) are shown in Table II, which again show that our proposed PSA is a fast algorithm with much lower computational complexity than SCA and SDR.

VI. CONCLUSION

In this letter, we have proposed a fast algorithm for multi-group multicast MMF beamforming using the optimal

[2] We have studied different values of α and found $\alpha = 0.01$ generally provides suitable trade-off between performance and convergence speed.

beamforming structure. We have shown that the nonconvex MMF problem can be transformed into an L-weakly convex optimization problem, which we have proposed using PSA to solve directly. Under our problem structure, PSA yields a closed-form updating procedure that is highly computationally inexpensive. We provide the convergence result to the proposed PSA. Simulation results demonstrate that PSA provides a near-optimal performance with a substantially lower computational complexity than the existing algorithms for large-scale systems.

APPENDIX A
PROOF OF THEOREM 1

Proof: Our proof follows the proof techniques of Theorem 3.1 in [18].[3] Let $\widehat{\mathbf{x}}^{(j)} \triangleq \arg\min_{\mathbf{x}}\{\tilde{g}(\mathbf{x}) + L\|\mathbf{x} - \mathbf{x}^{(j)}\|^2\}$. Based on $\tilde{g}_\lambda(\mathbf{x})$ in (10), with $\lambda = \frac{1}{2L}$, we have

$$\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j+1)}) \leq \tilde{g}(\widehat{\mathbf{x}}^{(j)}) + L\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j+1)}\|^2$$
$$= \tilde{g}(\widehat{\mathbf{x}}^{(j)}) + L\|\Pi_{\mathcal{X}}(\widehat{\mathbf{x}}^{(j)}) - \Pi_{\mathcal{X}}(\mathbf{x}^{(j)} - \alpha\nabla_{\mathbf{x}}f(\mathbf{x}^{(j)}, \mathbf{y}^{(j)}))\|^2$$
$$\overset{(a)}{\leq} \tilde{g}(\widehat{\mathbf{x}}^{(j)}) + L\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)} + \alpha\nabla_{\mathbf{x}}f(\mathbf{x}^{(j)}, \mathbf{y}^{(j)})\|^2$$
$$\leq \tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)}) + 2\alpha L\nabla_{\mathbf{x}}f(\mathbf{x}^{(j)}, \mathbf{y}^{(j)})^T(\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}) + \alpha^2 LM^2 \quad (14)$$

where (a) is due to $\|\Pi_{\mathcal{X}}(\mathbf{x}_1) - \Pi_{\mathcal{X}}(\mathbf{x}_2)\| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|$, $\forall \mathbf{x}_1, \mathbf{x}_2$. From (6), the second term in (14) is given by

$$\nabla_{\mathbf{x}}f(\mathbf{x}^{(j)}, \mathbf{y}^{(j)})^T(\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)})$$
$$\leq \tilde{g}(\widehat{\mathbf{x}}^{(j)}) - \tilde{g}(\mathbf{x}^{(j)}) + \frac{L}{2}\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2$$
$$= \left(\tilde{g}(\widehat{\mathbf{x}}^{(j)}) + L\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2\right) - \left(\tilde{g}(\mathbf{x}^{(j)}) + L\|\mathbf{x}^{(j)} - \mathbf{x}^{(j)}\|^2\right)$$
$$\quad - \frac{L}{2}\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2$$
$$\overset{(a)}{\leq} -L\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2 \overset{(b)}{=} -\frac{1}{4L}\|\nabla\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)})\|^2 \quad (15)$$

where (a) is because $\varphi(\mathbf{x}) \triangleq \tilde{g}(\mathbf{x}) + L\|\mathbf{x} - \mathbf{x}^{(j)}\|$ is an L-strongly convex function, which leads to

$$\varphi(\mathbf{x}^{(j)}) - \varphi(\widehat{\mathbf{x}}^{(j)}) = \varphi(\mathbf{x}^{(j)}) - \min_{\mathbf{x}}\varphi(\mathbf{x})$$
$$\geq \nabla\varphi(\widehat{\mathbf{x}}^{(j)})^T(\mathbf{x}^{(j)} - \widehat{\mathbf{x}}^{(j)}) + \frac{L}{2}\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2 = \frac{L}{2}\|\widehat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}\|^2$$

where the last equality is because $\nabla\varphi(\widehat{\mathbf{x}}^{(j)}) = 0$ as $\widehat{\mathbf{x}}^{(j)} = \arg\min_{\mathbf{x}}\varphi(\mathbf{x})$. Also, (b) in (15) is by (11). Applying (15) to (14) yields

$$\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j+1)}) \leq \tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)}) - \frac{\alpha}{2}\|\nabla\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)})\|^2 + \alpha^2 LM^2. \quad (16)$$

Summing both sides of (16) over j, rearranging the terms, and noting from (10) that $\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(J+1)}) \geq \min_{\mathbf{x}}\tilde{g}(\mathbf{x})$, we have

$$\frac{1}{J+1}\sum_{j=0}^{J}\|\nabla\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)})\|^2 \leq \frac{2\left(\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(0)}) - \min_{\mathbf{x}}\tilde{g}(\mathbf{x})\right)}{\alpha(J+1)}$$
$$+ 2\alpha LM^2. \quad (17)$$

Note from (10) that $\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(0)}) \leq \min_{\mathbf{x}}\tilde{g}(\mathbf{x}) + L\|\mathbf{x}^o - \mathbf{x}^{(0)}\|^2$, where $\mathbf{x}^o \triangleq \arg\min_{\mathbf{x}}\tilde{g}(\mathbf{x})$. Thus, $\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(0)}) - \min_{\mathbf{x}}\tilde{g}(\mathbf{x}) \leq L\|\mathbf{x}^o - \mathbf{x}^{(0)}\|^2 \leq LD^2$. Also, since $g(\mathbf{x}) = \max_{\mathbf{y}\in\mathcal{Y}}f(\mathbf{x}, \mathbf{y})$, $g(\mathbf{x})$ is also C-Lipschitz over $\mathcal{X}$, *i.e.*, $|g(\mathbf{x}_1) - g(\mathbf{x}_2)| \leq$

$C\|\mathbf{x}_1 - \mathbf{x}_2\|$, $\forall\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$. It follows that $\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(0)}) - \min_{\mathbf{x}}\tilde{g}(\mathbf{x}) \leq \tilde{g}(\mathbf{x}^{(0)}) - \min_{\mathbf{x}}\tilde{g}(\mathbf{x}) \leq CD$. Combining the above, let $\Delta \triangleq \min\{LD^2, CD\}$. Then, (17) becomes

$$\frac{1}{J+1}\sum_{j=0}^{J}\|\nabla\tilde{g}_{\frac{1}{2L}}(\mathbf{x}^{(j)})\|^2 \leq \frac{2\Delta}{\alpha(J+1)} + 2\alpha LM^2. \quad (18)$$

Minimizing RHS of (18) over α yields the optimal step size $\alpha = \sqrt{\frac{\Delta}{LM^2(J+1)}}$. Substituting this optimal α into (18) and noting that LHS of (18) is $E\|\nabla\tilde{g}_{\frac{1}{2L}}(\bar{\mathbf{x}})\|^2$, we have (13). ∎

REFERENCES

[1] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, pp. 2239–2251, Jun. 2006.

[2] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, pp. 1268–1279, Mar. 2008.

[3] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 12–21, Jan. 2013.

[4] M. Dong and B. Liang, "Multicast relay beamforming through dual approach," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, Dec. 2013, pp. 492–495.

[5] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Aug. 1978.

[6] L. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.

[7] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, pp. 804–808, Jul. 2015.

[8] D. Christopoulos, S. Chatzinatas, and B. Ottersten, "Multicast multigroup beamforming for per-antenna power constrained large-scale arrays," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun.*, Jun. 2015, pp. 271–275.

[9] M. Sadeghi, L. Sanguinetti, R. Couillet, and C. Yuen, "Reducing the computational complexity of multicasting in large-scale antenna systems," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 2963–2975, May 2017.

[10] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Trans. Commun.*, vol. 65, pp. 2685–2698, Jun. 2017.

[11] A. Konar and N. D. Sidiropoulos, "Fast approximation algorithms for a class of non-convex QCQP problems using first-order methods," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3494–3509, Jul. 2017.

[12] M. Dong and Q. Wang, "Multi-group multicast beamforming: Optimal structure and efficient algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 3738–3753, May 2020.

[13] B. Polyak, *Introduction to Optimization*. Optimization Software, Inc., New York, 1987.

[14] E. S. Levitin and B. T. Polyak, "Constrained minimization methods," *USSR Comput. Math. Math. Phys.*, vol. 6, no. 5, pp. 787–823, 1966.

[15] A. Beck, *First-Order Methods in Optimization*. SIAM, 2017.

[16] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer Science & Business Media, 2009.

[17] C. Jin, P. Netrapalli, and M. Jordan, "What is local optimality in nonconvex-nonconcave minimax optimization?" in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4880–4889.

[18] D. Davis and D. Drusvyatskiy, "Stochastic model-based minimization of weakly convex functions," *SIAM J. Optim.*, vol. 29, no. 1, pp. 207–239, 2019.

[19] S. Chen, A. Garcia, and S. Shahrampour, "On distributed nonconvex optimization: Projected subgradient method for weakly convex problems in networks," *IEEE Trans. Autom. Control*, vol. 67, no. 2, pp. 662–675, 2022.

[20] K. K. Thekumparampil, P. Jain, P. Netrapalli, and S. Oh, "Efficient algorithms for smooth minimax optimization," in *Proc. Advances Neural Inf. Process. Syst.*, vol. 32, 2019.

[21] R. T. Rockafellar, *Convex Analysis*. Princeton Univ. Press, 2015.

---

[3]The convergence analysis in [18] is for a proximal stochastic subgradient method for stochastic optimization of weakly convex functions. Since PSA is different from the stochastic method, that convergence result cannot be directly used.