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 preprocessing [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
The solution to $S_o$ 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 non-convex 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 $S_o$ 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.
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 $S_2$. Problem $S_2$ is a nonconvex max-min problem. Based on the structure of $S_2$, we show that PSA can be applied to compute a near-stationary solution to $S_2$ efficiently.

A. Problem Reformulation

For the purpose of computation, we express all complex quantities in $S_2$ using their real and imaginary parts. Define

$$x_i \triangleq [\text{Re}(a_i)^T, \text{Im}(a_i)^T]^T, \quad C_i \triangleq \begin{bmatrix} \text{Re}(C_i) & -\text{Im}(C_i) \\ \text{Im}(C_i) & \text{Re}(C_i) \end{bmatrix},$$

and

$$A_{ijk} \triangleq \begin{bmatrix} \text{Re}(A_{ijk}) & -\text{Im}(A_{ijk}) \\ \text{Im}(A_{ijk}) & \text{Re}(A_{ijk}) \end{bmatrix},$$

for $k \in K_i, i, j \in G$.

It follows that $\|C_i a_i\|^2 = \|C_i x_i\|^2$ and $a_j^T A_{ijk} a_i = x_j^T A_{ijk} x_j$. Using the above, we can express problem $S_2$ equivalently in the real domain as

$$S_3: \max_{x \in \mathcal{X}} \min_{i, k} \frac{1}{\gamma_{ik}} \sum_{j \neq i} x_j^T A_{ijk} x_j + \sigma^2$$

where $x \triangleq [x_1^T, \ldots, x_G^T]^T$, and $\mathcal{X} \triangleq \left\{ x : \sum_{i=1}^G \|C_i x_i\|^2 \leq P \right\}$ is the compact convex feasible set of $S_3$. Define $\phi_{ik}(x) \triangleq -x_i^T A_{ijk} x_i / \gamma_{ik} \sum_{j \neq i} x_j^T A_{ijk} x_j + \sigma^2$, $k \in K_i, i \in G$. Then, we can rewrite $S_3$ in an equivalent min-max form as $\min_{x \in \mathcal{X}} \max_{i, k} \phi_{ik}(x)$, which is further equivalent to

$$S_4: \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$$

where $f(x, y) \triangleq \phi^T(x) y$, with $\phi(x) \in \mathbb{R}^{K_u}$ containing all $\phi_{ik}(x)$'s, and $\mathcal{Y} \triangleq \{ y : y \geq 0, 1^T y = 1 \}$ is a probability simplex, which is a compact convex set. An optimal solution to the inner maximization of $S_4$ is $y = [0, 0, \ldots, 0, 1, \ldots, 0]^T$, with 1 at some $i$th position. Note that $f(x, y)$ is concave in $x$ and nonconvex in $y$. Thus, $S_4$ is a nonconvex-concave min-max problem and is NP-hard. Let $g(x) \triangleq \max_{y \in \mathcal{Y}} f(x, y)$. Then, we express $S_4$ as

$$S_5: \min_{x \in \mathcal{X}} g(x).$$

Note that $g(x)$ is nonconvex. If $g(x)$ is differentiable, one can use the projected gradient descent [14] to solve $S_5$. However, in our problem, $g(x)$ may not be differentiable, and its gradient $\nabla g(x)$ 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 $S_5$.

B. The Projected Subgradient Algorithm

We first show the structure of our problem. We assume the channel gain is finite for each user: $\|h_{ik}\| < \infty, \forall k, i$. Thus, all elements in $A_{ijk}$ are finite, $\forall k \in K_i, i, j \in G$. Also, note that $\gamma_{ik} > 0, \sigma^2 > 0$. It follows that, since $\mathcal{X}$ and $\mathcal{Y}$ are compact, the gradient $\nabla_x f(x, y)$ is finite for any $x \in \mathcal{X}, y \in \mathcal{Y}$. Thus, there exists a constant $L > 0$, such that $\|\nabla_x f(x, y) - \nabla_x f(x, y')\| \leq L \|x - x\|$, for any $x, x' \in \mathcal{X}, y, y' \in \mathcal{Y}$. This means that, $f(x, y)$ is an $L$-smooth function of $x \in \mathcal{X}$, which satisfies the following [15]

$$f(x, y) \geq f(x, y) + \nabla_x f(x, y)^T (x' - x) - \frac{L}{2} \|x' - x\|^2,$$

for any $x' \in \mathcal{X}$, we have

$$g(x') \geq g(x) + \nabla_x f(x, y)^T (x' - x) - \frac{L}{2} \|x' - x\|^2. \tag{6}$$

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

$$g(x') \geq g(x) + \nabla_x f(x, y)^T (x' - x) - \frac{L}{2} \|x' - x\|^2. \tag{6}$$

After rearranging the terms at both sides of the inequality in (6) and taking lim inf for $x' \rightarrow x$, we conclude that $\nabla_x f(x, y) \in \partial g(x)$.

Following the above result, we propose to solve $S_5$ by PSA with the following updating procedure:

$$\text{At iteration } j: \begin{align*}
    y^{(j)} & \in \arg \max_{y \in \mathcal{Y}} f(x^{(j)}, y), \\
    x^{(j+1)} & = x^{(j)} - \alpha \nabla_x f(x^{(j)}, y^{(j)}) \tag{8}
\end{align*}$$

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

$$\Pi_{\mathcal{X}}(x) = \begin{cases} P_x x & x \notin \mathcal{X} \\
    x & x \in \mathcal{X} \end{cases} \tag{9}$$

where $P_x \triangleq \sum_{i=1}^G \|C_i x_i\|^2$.

Note that the inherent structure of our problem makes PSA particularly suitable for solving $S_5$. First, $y^{(j)}$ in (7) can be directly obtained by taking the maximum among $\phi_{ik}(x^{(j)})$'s. Second, the projection $\Pi_{\mathcal{X}}(x)$ is a simple closed-form function in (9), and $\nabla_x f(x, y)$ has a closed-form expression. Thus, the computation of $x^{(j+1)}$ 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 $S_5$. Recall that $f(x, y)$ is $L$-smooth over $\mathcal{X}$, and $\mathcal{Y}$ is compact. It follows that $g(x) = \max_{y \in \mathcal{Y}} f(x, y)$ is $L$-weakly convex over $\mathcal{X}$, i.e., $g(x) + \frac{L}{2} \|x\|^2$ is convex for $x \in \mathcal{X}$ [20, Lemma 1]. Consider the extension of $g(x)$ to $\mathbb{R}^{2K_u}$: \( \tilde{g}(x) = g(x) + \lambda \|x\| \), where $I_{\mathcal{X}}(x)$ is an indicator function, taking 0 if $x \in \mathcal{X}$ and $\infty$ otherwise. Define the Moreau envelope [21] of $\tilde{g}(x)$ as

$$\tilde{g}_\lambda(x) \triangleq \min_{x' \in \mathbb{R}^{2K_u}} \{ \tilde{g}(x') + \frac{1}{2\lambda} \|x' - x\|^2 \} \tag{10}$$

where $\lambda < 1/L$. The Moreau envelope $\tilde{g}_\lambda(x)$ is a smooth approximation to the non-smooth but $L$-weakly convex function $g(x)$ over $\mathcal{X}$. Note from the earlier discussion that the objective function of the minimization problem in (10) is
strictly convex. Let $\tilde{x} \triangleq \arg \min_{x} \{ \hat{g}(x') + \frac{1}{2\lambda} \|x' - x\|^2 \}$. Then, we have $\tilde{x} \in X$ and

$$\|\tilde{x} - x\| = \lambda \|\nabla \hat{g}_\lambda(x)\|.$$  \hspace{1cm} (11)

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

$$\|\tilde{x} - x\| \leq \lambda \epsilon,$$

and $\min_{u \in \nabla \hat{g}_\lambda(x)} \|u\| \leq \epsilon$. \hspace{1cm} (12)

The above means that a small gradient $\|\nabla \hat{g}_\lambda(x)\| \leq \epsilon$ implies that $x$ is close to a point $\tilde{x}$ that is a near-stationary (i.e., $\epsilon$-stationary) point of $S_5$. Hence, $\|\nabla \hat{g}_\lambda(x)\|$ provides a near-stationarity measure of $x$ to a stationary point of $S_5$. Based on this, we have the convergence result of PSA for $S_5$ below.

**Theorem 1.** Assume the continuous function $f(x,y)$ is $C$-Lipschitz over $X \times Y$. Define $D \triangleq \max_{x_1,x_2 \in X} \|x_1 - x_2\|$, $M \triangleq \max_{x \in X,y \in Y} \|\nabla_x f(x,y)\|$, and $\Delta \triangleq \min \{LD^2, CD\}$. Starting from $x(0) \in X$, 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 $\bar{x} = \bar{x}^{(j)}$, where $j \sim \text{Uniform}[0,J]$. Then, $\bar{x}$ satisfies

$$E[\|\nabla \hat{g}_\lambda(x)\|] \leq \frac{4\sqrt{\Delta LM^2}}{\sqrt{J+1}}.$$  \hspace{1cm} (13)

**Proof:** See Appendix A.

Theorem 1 indicates that if we take a random sample in $\{x^{(j)}\}_{j=0}^J$ as the output of PSA $\bar{x}$, then $E[\|\nabla \hat{g}_\lambda(x)\|]$ 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 $\bar{x}$ satisfying $E[\|\nabla \hat{g}_\lambda(x)\|] \leq \epsilon$, the required number of iterations $J$ for PSA is at most $O(\epsilon^{-4})$. Thus, for weakly convex problem $S_5$, Theorem 1 shows that PSA converges within finite time, upper bounded by $O(\epsilon^{-4})$, to an $\epsilon$-accuracy point of $S_5$.

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 $\{x^{(j)}\}_{j=0}^J$. Thus, to implement PSA, we can set a random stopping point $J^o < J$ for PSA and use $x^{(J^o)}$ 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 $P_o$ into optimizing a (denoted by $P'_o$) instead of $w$, we propose to use SDR with Gaussian randomization (GR) [12] to solve $P'_o$ 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 $x^{(0)}$ that is closer to the optimal solution. Note that this initial point may not be feasible in $X$. 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 $S_5$ is of a small to moderate size.

V. Simulation Results

We set $G = 3$, $K_t = K_r$, and SINR target $\gamma_{ik} = \gamma = 10 \text{ dB}$, $\forall k, i$. The channels are generated i.i.d. as $h_{ik} \sim \mathcal{CN}(0,1)$, and the receiver noise variance is $\sigma^2 = 1$. We set $P/\sigma^2 = 10 \text{ dB}$. The approximate expression $R$ in (4) is used in the simulation. For PSA, we set the step size $\alpha = 0.01$ and set the stopping criterion as $|g(x^{(j+1)}) - g(x^{(j)})| \leq 10^{-5}$. Based on various simulation experiments, PSA generally converges 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 $S_5$: It is obtained by solving $P_o$ using SDR along with the bi-section search over $t$; 2) SDR: It uses the optimal structure in (2) with $R$ in (4) and solves $S_5$ by solving $P_o$ using SDR with $P_o$ along with the bi-section search over $t$ [12]; 3) SCA: It uses the optimal structure in (2) with $R$ in (4) and solves $S_5$ by solving $P_o$ 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: Average minimum SINR vs. $N$. Right: Average minimum SINR vs. $K$ ($G = 3, N = 100$).](image)

**TABLE I** \hspace{1cm}

| $N$ | 100 | 200 | 300 | 400 | 500 |
|-----|-----|-----|-----|-----|-----|
| PSA (proposed) | 1.130 | 2.051 | 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** \hspace{1cm}

| $K$ | 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 |

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 $\alpha$ 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]. Let ̂x(j) ≜ arg minx{ĝ(x) + L∥x − x(j)∥2}. Based on ̂g(x) in (10), with λ = 1/2, we have

\[ \frac{g_{\lambda}}{g_{\lambda}}(x(j + 1)) \leq \frac{g_{\lambda}}{g_{\lambda}}(x(j)) + L∥x(j) − x(j + 1)∥^2 \]

where (a) is due to Πx(x(j)) − Πx(x(j)) ≤ ∥x(j) − x(j)∥, ∀ x(j), x(j). From (6), the second term in (14) is given by

\[ \nabla g_{\lambda}(x(j)) = \frac{1}{2} \nabla \nabla g_{\lambda}(x(j)) \]

where the last equality is because \nabla g(x) = 0 as ̂x(j) = arg minx g(x). Also, (b) in (15) is by (11). Applying (15) to (14) yields

\[ \frac{g_{\lambda}}{g_{\lambda}}(x(j + 1)) \leq \frac{g_{\lambda}}{g_{\lambda}}(x(j)) - \frac{1}{2} ∥\nabla g_{\lambda}(x(j))∥^2 + αL∥x(j) − x(j)∥^2 \]

where α = 2/\nabla g(x). It follows that \nabla g_{\lambda}(x(j)) = \nabla g_{\lambda}(x(j)) = \nabla g_{\lambda}(x(j)) = \nabla g_{\lambda}(x(j)) = \nabla g_{\lambda}(x(j)) = \nabla g_{\lambda}(x(j)).

Summing both sides of (16) over j, rearranging the terms, and noting from (10) that ̂g_{\lambda}(x(j + 1)) ≥ minx ̂g(x), we have

\[ \frac{1}{J + 1} \sum_{j=0}^{J} (∥\nabla g_{\lambda}(x(j))∥^2 ≤ 2∥\nabla g_{\lambda}(x(j)) − \nabla g_{\lambda}(x(j))∥^2 + 2αL∥x(j) − x(j)∥^2 \]

Note from (10) that ̂g_{\lambda}(x(j)) ≤ minx ̂g(x), we have

\[ \frac{1}{J + 1} \sum_{j=0}^{J} (∥\nabla g_{\lambda}(x(j))∥^2 ≤ 2∥\nabla g_{\lambda}(x(j)) − \nabla g_{\lambda}(x(j))∥^2 + 2αL∥x(j) − x(j)∥^2 \]

where x(j) = arg minx ̂g(x). Thus, ̂g_{\lambda}(x(j)) = ̂g(x(j)) ≤ L∥x(j) − x(j)∥^2 ≤ LD^2. Also, since sgn(x) = maxx∈(x,y) f(x,y), g(x) is also C-Lipschitz over X, i.e., g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x) = g(x).

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.

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 multiconnected 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. Advantages Multi-Sensor Adaptive Process., Dec. 2013, pp. 492–495.
[5] R. Marks and G. P. Wright, “A general interior 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 parametric approach to physical layer multicasting for large-scale antenna arrays,” IEEE Signal Process. Lett., vol. 21, no. 1, pp. 114–117, Jan. 2014.
[7] J. 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] X. Chang, S. Chatzinotas, and B. Ottersten, “Multicast multi-group 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] D. 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] D. Poljak, 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.