Low-Complexity Joint Antenna Selection and Robust Multi-Group Multicast Beamforming for Massive MIMO

Niloofar Mohamadi, Min Dong, *Fellow, IEEE*, and Shahram ShahbazPanahi, *Senior Member, IEEE*

***Abstract*—We consider low-complexity design for joint antenna selection and robust multi-group multicast beamforming in massive multiple-input multiple-output (MIMO) systems. Relying on the estimated channel covariance and assuming a limited number of antennas for transmission at the base station, we aim to minimize the transmit power subject to the worst-case signal-to-interference-plus-noise-ratio (SINR) guarantee and per selected antenna power budget. Converting the worst-case SINR constraints to a set of non-convex constraints, we propose a two-phase approach to solve the problem efficiently: the antenna selection phase, followed by the robust multicast beamforming generation phase. We propose an SINR-based approach for antenna selection, where the challenging mixed-integer problem is converted into an approximate joint optimization problem via a sequence of transformation, relaxation, and SINR approximation. We develop a fast two-layered alternating direction method of multipliers (ADMM)-based algorithm to compute an approximate solution. In particular, with our ADMM construction, we obtain semi-closed-form solutions for antenna selection and beamforming subproblems at each ADMM iteration for fast updates. To further reduce the computational complexity, we propose a signal-to-leakage-ratio (SLR)-based approach using the SLR constraints in the joint optimization problem. This allows us to develop a two-layered ADMM-based algorithm, which can compute a solution more efficiently due to the SLR structure. The robust multicast beamforming solution for the selected antennas is computed using the fast algorithm we developed recently. Simulation results show the effectiveness of our two proposed approximated approaches for antenna selection and the overall two-phase approach in both overall performance and substantially low computational complexity in a massive MIMO setting.**

***Index Terms*—Antenna selection, robust multicast beamforming, large-scale systems, ADMM, low complexity.**

Manuscript received 26 March 2023; revised 15 August 2023 and 9 December 2023; accepted 27 December 2023. Date of publication 12 January 2024; date of current version 7 February 2024. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Discovery Grants. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Renato L. G. Cavalcante. *(Corresponding author: Min Dong.)*

The authors are with the Department of Electrical, Computer and Software Engineering, Ontario Tech University, Oshawa, ON L1G 0C5, Canada (e-mail: niloofar.mohamadi@ontariotechu.net; min.dong@ontariotechu.ca; shahram.shahbazpanahi@ontariotechu.ca).

Digital Object Identifier 10.1109/TSP.2024.3351478

I. Introduction

MULTICAST applications for content distribution have become increasingly popular in wireless services. Efficient multicast transmission is critical to support these services in the evolved wireless networks and beyond. Multicast beamforming enables common data to be delivered to multiple users simultaneously with improved power and spectrum efficiency and is the enabling physical-layer technique for multicast transmission [1]. However, the multicast nature renders beamforming design optimization challenging [1], [2], [3], [4], [5]. Moreover, designing multicast beamforming for massive multiple-input multiple-output (MIMO) wireless systems must overcome several new challenges to ensure its effectiveness in supporting multicast transmissions.

For massive MIMO systems with a large number of transmit antennas at the base station (BS), accurate acquisition of channel state information (CSI) is challenging, especially in a fading environment where channels vary over time. This uncertainty imposes design challenges in providing quality-of-service (QoS) guarantees. Furthermore, to limit the hardware complexity and cost, the BS may have a limited number of radio frequency (RF) chains that are fewer than the antenna elements. In this case, the antenna selection technique can be adopted at the BS, thereby choosing a subset of antennas for transmission based on the channel condition over each antenna element. In the presence of CSI uncertainty, this additional decision of antenna selection further complicates the multicast beamforming design process. Finally, for the large-scale multicast beamforming problem in massive MIMO systems, it is vital to devise a low-complexity algorithm to solve this challenging problem effectively. The literature on multicast beamforming has mainly focused on the algorithm design assuming perfect CSI. The approaches proposed in the earlier literature are mainly suitable for traditional multi-antenna wireless systems [1], [2], [3], [4], [5], and the more recent works developed computationally efficient methods or algorithms for massive MIMO systems [6], [7], [8], [9], [10], [11], [12], [13], [14]. The antenna selection for multicast beamforming under perfect CSI has been studied in [15], [16], [17], [18]. Only a few studies considered imperfect CSI and proposed different robust design methods while using all available antennas at the BS [19], [20], [21], [22].

This paper considers the issues mentioned earlier and proposes an effective design of antenna selection and multi-group multicast beamforming for massive MIMO wireless systems. In a fading environment, channel covariance evolves much slower than the instantaneous CSI. In time-division duplexing (TDD) massive MIMO systems, channel covariance acquisition requires much less uplink pilot overhead [23], [24], [25]. In frequency-division duplexing (FDD) systems, due to overwhelming downlink pilot overhead and computational complexity, channel covariance estimation is resorted by using uplink channel estimates and array processing techniques [26], [27], [28]. Thus, the channel covariance estimate is adopted in our multicast beamforming design. In the presence of estimation error, we develop a robust design approach to account for channel uncertainty and guarantee QoS requirements. In particular, with a given number of RF chains, we aim to jointly optimize the antenna selection and robust multicast beamforming to minimize the transmit power while meeting QoS targets. In tackling this challenging large-scale joint optimization problem, we mainly focus on devising low-complexity algorithms to compute a fast solution with good performance.

A. Related Works

Multicast beamforming design has been investigated in traditional wireless systems for downlink transmissions in a single-group or multi-group setting [1], [2], [3], the multi-cell environment [4], and the relay network scenario [5]. The design formulation is typically either the QoS problem for transmit power minimization while meeting the QoS targets, or the max-min fair (MMF) problem for maximizing the minimum signal-to-interference-and-noise-ratio (SINR) among users under the transmit power budget. The multicast beamforming problems are non-convex quadratically constrained quadratic programming (QCQP) problems, whose general form is known to be NP-hard [29]. Semi-definite relaxation (SDR) was the popular method to find a good suboptimal solution [1], [2], [3], [4]. However, SDR has a high computational complexity unsuitable for large systems. Successive convex approximation (SCA) was then considered to improve both performance and computational efficiency in computing a solution [6], [8]. For massive MIMO systems, a fast algorithm based on the alternating direction method of multipliers (ADMM) was proposed [7], and low-complexity suboptimal multicast beamforming schemes were devised [8], [9], [10]. The optimal multi-group multicast beamforming structure was established in [11], showing that the high-dimensional beamforming problem can be converted into a much smaller problem. Based on this structure, fast first-order algorithms have been developed for multicast beamforming [12], [13], [14]. All of the above works assume perfect CSI knowledge.

Existing works on multicast beamforming design under imperfect CSI are mostly limited to traditional multi-antenna systems, where robust multicast beamforming methods were proposed for downlink transmission scenarios [20], [21], [30], cognitive radio networks [19], and cloud radio access networks [31], [32]. In these works, either the worst-case SINR guarantee [19], [20], [32] or SINR outage probability target [20], [30] is considered for robust beamforming formulation. These works generally reformulated the robust SINR constraints and proposed SDR-based methods to obtain a feasible solution. To reduce the computational complexity, [21], [31] proposed using a lower bound on the worst-case SINR to simplify the problem, for which SCA was adopted to compute a solution. However, these lower bounds are often overly conservative, causing substantial performance degradation as the channel estimation error increases. Downlink beamforming using channel covariance is another common approach in the literature [33], since channel covariance matrices evolve much slower than instantaneous CSI and are easier to acquire. In particular, for massive MIMO systems, [23], [24], [25] studied channel covariance estimation and showed that uplink pilot overhead could be reduced substantially. Robust multicast beamforming using channel covariance matrices in traditional multi-antenna systems has been considered in [34], [35], where SDR-based methods were proposed. However, they are not scalable for massive MIMO systems.

The existing study on computationally efficient robust multicast beamforming for massive MIMO systems is scarce. Although several ADMM-based algorithms have been proposed under perfect CSI [7], [36], the non-convex robust design formulation renders the problem much more challenging to apply the ADMM technique. In the recent work in [22], a fast ADMM-based robust multicast beamforming algorithm using channel covariance has been developed, which constructed ADMM to directly solve the non-convex robust design problem with a convergence guarantee.

The problem of joint antenna selection and multicast beamforming has been studied in several existing works [15], [16], [17], all assuming perfect CSI. In [15], antenna selection was formulated as the l_0-norm constraint on the beamforming vectors, and the convex approximation with an SDR-based algorithm was proposed for the QoS problem. In [16], the energy efficiency maximization subject to QoS and power constraints was considered in a multi-cell scenario. In this work, antenna selection was formulated either as the constraint on the L_0-norm of beamforming vectors or as the binary selection variables to form a mixed-integer program, and SCA-based algorithms were proposed. Besides the above, antenna selection with hybrid beamforming for simultaneous wireless information and power transfer in a multicast beamforming scenario was also considered [17]. As joint antenna selection and multicast beamforming is a very challenging problem, the existing work for a low-complexity design suitable for massive MIMO is scarce. To the best of our knowledge, only [18] considered a computationally efficient design for this joint optimization problem in a single group setting for the MMF problem. An SCA-based approach was proposed, followed by first-order fast algorithms to solve the SCA subproblem efficiently at each iteration. No available low-complexity solution exists in the literature for joint antenna selection and multi-group multicast beamforming, especially for a robust design under imperfect CSI, for massive MIMO systems.

B. Contributions

This paper considers a massive MIMO system scenario, where the BS selects a subset of antennas for downlink multicast beamforming based on imperfect CSI. In particular, we provide a low-complexity design of joint antenna selection and robust multi-group multicast beamforming suitable for massive MIMO systems. Assuming the estimated channel covariance matrices are available at the BS with a bounded error, we formulate the joint antenna selection and robust multicast beamforming problem to minimize the total transmit power subject to the worst-case SINR guarantee and the transmit power budget for each selected antenna. Our main contributions are summarized below:

- The formulated joint antenna selection and robust multicast beamforming problem is a challenging mixed-integer program with binary antenna selection variables and worst-case SINR requirements. First, we transform the worst-case SINR constraints into a set of non-convex constraints with a more tractable form. Then, we propose a two-phase approach to solve this joint optimization problem efficiently: the antenna selection phase, followed by the robust multicast beamforming generation phase. In Phase 1, we devise efficient algorithms to determine antenna selection. In Phase 2, we compute the robust multicast beamforming solution with the selected antenna using the fast ADMM-based algorithm proposed in our recent work [22].
- We propose an SINR-based approach for antenna selection in Phase 1. We first transform the reformulated mixed-integer problem into an equivalent problem with all continuous variables. Then, we apply a penalty-term-based relaxation technique and SINR constraint simplification to arrive at an approximated problem. This new form allows us to develop a fast two-layer ADMM-based algorithm to compute a solution efficiently. In particular, at each layer, our particular ADMM construction enables us to exploit the structure of the subproblem at each ADMM block, and we are able to obtain semi-closed-form solutions for both antenna selection and beamforming vectors[1]. These solutions lead to fast ADMM updates at each iteration in the inner and outer layers with low computational complexity.
- To further reduce the computational complexity in Phase 1, we propose a signal-to-leakage-ratio (SLR)-based approach, where the SINR constraints are replaced by the SLR constraints for QoS guarantee. A two-layer ADMM-based algorithm is also developed for this problem. Due to the SLR metric structure, the proposed algorithm significantly reduces computational complexity from the SINR-based approach.
- The simulation results show that our proposed ADMM-based algorithms for SINR-based and SLR-based approaches have fast convergence behaviors. Also, their performances are nearly identical to that of the interior point method (IPM) used in the standard nonlinear solvers but with significantly lower computational complexity. Furthermore, the SINR-based and SLR-based approximated approaches in Phase 1 outperform other alternative methods for antenna selection, and the resulting performance is nearly identical to that of the original problem without approximation, demonstrating the overall effectiveness of our proposed low-complexity approaches.

[1]The semi-closed-form solution provides an explicit expression for the solution but some parameters involved in the expression are unknown and need to be determined numerically.

C. Organization and Notations

The rest of this paper is organized as follows. Section II provides the system model and problem formulation. In Section III, we propose our two-phase approach for joint antenna selection and robust multicast beamforming. In Section IV, we propose an SINR-based approach for antenna selection and develop a fast ADMM-based algorithm to obtain the solution. In Section V, we further propose an SLR-based approach with a fast ADMM-based algorithm for antenna selection. Section VI describes the robust multicast beamforming design given antenna selection. The simulation results are provided in Section VII, followed by the conclusion in Section VIII.

Notations: Hermitian, transpose, and conjugate are denoted by $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$, respectively. The Frobenius norm and the Euclidean norm are denoted by $\|\cdot\|_F$ and $\|\cdot\|$, respectively. The Hadamard matrix product is denoted by $\odot$. The inequality $\mathbf{A} \succcurlyeq 0$ means that matrix $\mathbf{A}$ is positive semi-definite, and $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. The real part of a complex variable a is denoted by $\mathfrak{Re}\{a\}$. A complex Gaussian random vector $\mathbf{a}$ with zero mean and covariance matrix $\mathbf{C}$ is denoted by $\mathbf{a} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$.

II. System Model and Problem Formulation

A. System Model

We consider a downlink multi-group multicasting scenario in a massive MIMO system, where a multi-antenna BS serves G groups of users. Users in each group receive a common message, which is independent of the messages sent to other groups. Let $\mathcal{G} \triangleq \{1, ..., G\}$ denote the index set of the multicast groups and $\mathcal{K}_i \triangleq \{1, \ldots, K_i\}$ the index set of the single-antenna users in group i, for $i \in \mathcal{G}$. Each user is associated with only one group.

The BS is equipped with M antennas and L transmit RF chains, where $M \gg 1$ and $L \leq M$. With a limited number of RF chains, the BS selects L antennas, one for each RF chain, for downlink transmission. Let $\mathbf{h}_{ik} \in \mathbb{C}^{M\times 1}$ denote the channel vector from the BS to user k in group i. Also, let $\mathbf{w}_i = [w_{i1}, \cdots, w_{iM}]^T \in \mathbb{C}^{M\times 1}$ denote the multicast beamforming vector for group i, and $\mathbf{a} \triangleq [a_1, \ldots, a_M]^T$ the antenna selection vector, where $a_m \in \{0, 1\}$ is the selection indicator for antenna $m \in Mc \triangleq \{1, \ldots, M\}$, with 1 meaning being selected and 0 otherwise. With the selected antennas in $\mathbf{a}$, the signal received at user k in group i is given by

$$y_{ik} = \mathbf{w}_i^H \mathbf{A}\mathbf{h}_{ik} s_i + \sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H \mathbf{A}\mathbf{h}_{ik} s_l + n_{ik},\ k \in \mathcal{K}_i,\ i \in \mathcal{G}.$$

where $\mathbf{A} = \text{diag}(\mathbf{a})$, s_i is the symbol intended for group i, n_{ik} is the receiver additive white Gaussian noise at user k in group i with zero mean and variance σ^2, and $\mathcal{G}_{-i} \triangleq \mathcal{G}\backslash\{i\}$.

In a massive MIMO system, acquiring the instantaneous downlink CSI at the BS can be challenging. One approach in the literature is to estimate the channel covariance matrix over time, since it evolves much slower than the instantaneous CSI [23], [24], [25]. Thus, we consider that the BS performs antenna selection and beamforming design based on the knowledge of channel covariance matrices. Let $\bar{\mathbf{R}}_{ik} \triangleq \mathbb{E}\{\mathbf{h}_{ik}\mathbf{h}_{ik}^H\}$ denote the *true* full channel covariance matrix between the BS and user k in group i. The BS computes SINR of user k in group i based on $\bar{\mathbf{R}}_{ik}$. Given the antenna selection vector $\mathbf{a}$, the SINR is expressed as

$$\text{SINR}_{ik} = \frac{\mathbf{w}_i^H \mathbf{A}\bar{\mathbf{R}}_{ik}\mathbf{A}\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H \mathbf{A}\bar{\mathbf{R}}_{ik}\mathbf{A}\mathbf{w}_l + \sigma^2}, k \in \mathcal{K}_i,\ i \in \mathcal{G}. \quad (1)$$

Note that the transmit power constraint can be imposed by involving the antenna selection vector $\mathbf{a}$. Specifically, let the maximum transmit power at each selected antenna be $P_{\max}$ and 0 at unselected antennas. Then, the per-antenna power constraint can be written as

$$\sum_{i=1}^{G} |w_{im}|^2 \le a_m P_{\max}, \quad m \in \mathcal{M}. \quad (2)$$

Under constraint (2), since for $a_m = 0$, $w_{im} = 0, \forall i \in \mathcal{G}$, we can remove selection matrix $\mathbf{A}$ from (1) and re-write SINR as

$$\text{SINR}_{ik} = \frac{\mathbf{w}_i^H \bar{\mathbf{R}}_{ik}\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H \bar{\mathbf{R}}_{ik}\mathbf{w}_l + \sigma^2}, \quad k \in \mathcal{K}_i,\ i \in \mathcal{G}.$$

Our goal is to jointly determine antenna selection and multicast beamforming vectors to minimize the transmit power, subject to per-antenna transmit power budget and the minimum SINR requirements. This problem is formulated as follows:

$$\mathcal{P}_{\text{perf}}: \min_{\mathbf{a},\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2$$
$$\text{s.t.}\ \ \text{SINR}_{ik} \ge \gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \quad (3a)$$
$$\sum_{i=1}^{G} |w_{im}|^2 \le a_m P_{\max}, \quad a_m \in \{0,1\},\ m \in \mathcal{M} \quad (3b)$$
$$\sum_{m=1}^{M} a_m = L \quad (3c)$$

where γ_{ik} is the SINR target for user k in group i, and constraint (3c) limits the number of selected antennas to be L.

B. Robust Formulation Based on Channel Covariance

In practical systems, the BS only has the estimated channel covariance matrices for the transmission design. Thus, we consider a robust design of antenna selection and multicast beamforming, which accounts for the uncertainty due to the channel covariance estimation error to achieve a robust performance. Let $\mathbf{R}_{ik}$ denote the $M \times M$ *estimated* channel covariance matrix for user k in group i. Its relation to the true channel covariance is modeled as $\bar{\mathbf{R}}_{ik} = \mathbf{R}_{ik} + \mathbf{E}_{ik}$, where $\mathbf{E}_{ik}$ is the corresponding error matrix. We follow a spherical error model for channel uncertainty commonly considered in robust designs [19], [20], [31], [32], [33]. In particular, we assume $\mathbf{E}_{ik}$ is bounded within a hyper-spherical region as

$$\|\mathbf{E}_{ik}\|_F \le \epsilon_{ik}, \quad k \in \mathcal{K}_i, i \in \mathcal{G} \quad (4)$$

where ϵ_{ik} is the error bound. Let $\mathcal{B}(\epsilon_{ik})$ denote the set of all error matrices satisfying (4), given by

$$\mathcal{B}(\epsilon_{ik}) \triangleq \{\mathbf{E}_{ik} :\ \|\mathbf{E}_{ik}\|_F \le \epsilon_{ik}\}. \quad (5)$$

With the above error model, we formulate the joint optimization of antenna selection and robust multi-group multicast beamforming, where the SINR constraints in (3a) of $\mathcal{P}_{\text{perf}}$ are replaced by the SINR guarantee for any $\mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik})$. The problem is given as follows:

$$\mathcal{P}_{\text{rob}}: \min_{\mathbf{a},\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2$$
$$\text{s.t.}\ \min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l + \sigma^2} \ge \gamma_{ik},$$
$$k \in \mathcal{K}_i, i \in \mathcal{G}, \quad (6)$$
$$\text{(3b) and (3c)}.$$

Note that both $\mathcal{P}_{\text{perf}}$ and $\mathcal{P}_{\text{rob}}$ are mixed-integer programs due to binary antenna selection. In $\mathcal{P}_{\text{perf}}$, for given $\mathbf{a}$, the problem becomes a non-convex QCQP problem, which is known to be NP-hard in general [29]. Compared with $\mathcal{P}_{\text{perf}}$, the robust formation in SINR constraint (6) makes $\mathcal{P}_{\text{rob}}$ even more challenging to solve. Furthermore, the size of the problem is large for $M \gg 1$. Therefore, it is critical to design an effective and scalable solution with low computational complexity. With this goal, in the next section, we reformulate problem $\mathcal{P}_{\text{rob}}$, and derive a low-complexity solution for this problem.

III. Joint Antenna Selection and Robust Multicast Beamforming

A. Problem Reformulation

For problem $\mathcal{P}_{\text{rob}}$, we first reformulate the worst-case SINR constraints in (6) for a given antenna selection vector $\mathbf{a}$ by employing the approach proposed in [22] for a pure robust multicast beamforming problem. In particular, the constraint in (6), for $k \in \mathcal{K}_i$, $i \in \mathcal{G}$, can be equivalently written as

$$\frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \ge \gamma_{ik},\ \forall \mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik}),$$

which is equivalent to

$$\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l \ge \sigma^2\gamma_{ik},$$
$$\forall \mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik}). \quad (7)$$

The above constraint can be further expressed as the following equivalent constraint:

$$\min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i$$

$$- \gamma_{ik} \sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l \geq \sigma^2\gamma_{ik}.$$

The left-hand side (LHS) of this constraint is a convex optimization problem over $\mathbf{E}_{ik}$, with a closed-form solution given by [22] $\mathbf{E}_{ik}^\star = -\epsilon_{ik}\frac{\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H}{\|\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H\|_F}$. Defining $\mathbf{W}\triangleq[\mathbf{w}_1,\ldots,\mathbf{w}_G]\in\mathbb{C}^{M\times G}$ and diagonal matrix $\mathbf{D}_{ik}\in\mathbb{C}^{G\times G}$ with the ith diagonal entry being 1 and the rest being $-\gamma_{ik}$, we then can rewrite $\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H = \mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H$. Substituting the expression of $\mathbf{E}_{ik}^\star$ into the LHS of the above constraint, we then have the following constraint equivalent to the worst-case SINR constraint in (6) for each k and i:

$$\text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}. \quad (8)$$

Note that in the above constraint function, the first term is related to the estimated SINR based on $\mathbf{R}_{ik}$, while the second term reflects the worst-case error (from true SINR) under $\mathbf{W}$.

For the per-antenna power constraint in (3b), let $\mathbf{e}_m$ denote an $M\times 1$ selection vector with the mth entry being 1 and the rest 0's. Then, $\mathbf{e}_m^T\mathbf{W}$ selects the mth row of $\mathbf{W}$, and we have $\sum_{i=1}^G |w_{im}|^2 = \|\mathbf{e}_m^T\mathbf{W}\|^2$. Based on the above, we transform $\mathcal{P}_{\text{rob}}$ into the following equivalent problem:

$$\begin{aligned}\mathcal{P}_1: \min_{\mathbf{a},\mathbf{W}}\ & \text{tr}(\mathbf{W}\mathbf{W}^H)\\ \text{s.t.}\ & \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik},\\ & \qquad k\in\mathcal{K}_i, i\in\mathcal{G}, \quad (9a)\\ & \|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\max},\ m\in\mathcal{M}, \quad (9b)\\ & \sum_{m=1}^M a_m = L. \quad (9c)\\ & a_m\in\{0,1\},\ m\in\mathcal{M}. \quad (9d)\end{aligned}$$

Compared with $\mathcal{P}_{\text{rob}}$, $\mathcal{P}_1$ has a more tractable form where each constraint in (9a) is now an explicit function of $\mathbf{W}$. In the rest of the paper, we will focus on solving $\mathcal{P}_1$.

B. Proposed Two-Phase Approach for Solving $\mathcal{P}_1$

Even though the constraints in (9a) are more tractable, $\mathcal{P}_1$ is still a challenging large-scale mixed-integer program. Our goal is to develop a method that is both efficient and effective to solve $\mathcal{P}_1$. Aiming at this goal, we propose a low-complexity two-phase approach to efficiently compute a solution for $\mathcal{P}_1$:[2]

- **Phase 1 – Antenna selection**: This phase is to determine the binary antenna selection vector $\mathbf{a}$ efficiently with a good performance. To achieve this, we need to further reformulate and relax $\mathcal{P}_1$ into an amenable form that allows us to develop a fast algorithm to solve the joint optimization problem to obtain $\mathbf{a}$. Towards this, we propose two approaches:
  - We first devise an SINR-based approach and develop a fast algorithm to obtain $\mathbf{a}$.
  - To further reduce the computational complexity, we then propose an SLR-based approach with a fast algorithm to determine $\mathbf{a}$.
- **Phase 2 – Robust multicast beamforming solution**: After obtaining $\mathbf{a}$ in Phase 1, we solve the robust multicast beamforming problem in $\mathcal{P}_1$ for $\mathbf{W}$ with the selected antennas.

Phase 1 is crucial since the overall performance depends on the quality of the solution for antenna selection. In addition, the computational efficiency in determining antenna selection affects the overall computational complexity of the solution to this problem. Next, we present our proposed SINR-based approach for antenna selection in Phase 1.

IV. Antenna Selection: SINR-Based Approach

In Phase 1, we determine antenna selection $\mathbf{a}$ in $\mathcal{P}_1$. Since $\mathcal{P}_1$ is a mixed-integer program, we consider replacing each binary variable a_m with a continuous variable in $[0,1]$. In particular, it is shown in [17, Lemma 1] that the following three constraints together

$$\sum_{m=1}^M a_m^2 = L,\ \sum_{m=1}^M a_m = L,\ \ 0\leq a_m\leq 1, m\in M \quad (10a)$$

are equivalent to the binary constraints (9c) and (9d) for a_m's. As such, we can equivalently replace constraints (9c) and (9d) by the set of three constraints in (10a) and convert the mixed-integer programming problem into the following equivalent non-convex optimization problem with continuous variables:

$$\begin{aligned}\mathcal{P}_2: \ & \min_{\mathbf{a},\mathbf{W}} \text{tr}(\mathbf{W}\mathbf{W}^H)\\ & \text{s.t. (9a), (9b), and (9c)}\\ & \sum_{m=1}^M a_m^2 = L, \quad 0\leq a_m\leq 1,\ m\in M. \quad (10b)\end{aligned}$$

Although $\mathbf{a}$ and $\mathbf{W}$ are now all continuous variables, $\mathcal{P}_2$ is still a non-convex QCQP problem, which is NP-hard in the general form. To further make the problem more tractable, we relax $\mathcal{P}_2$ by transferring the equality constraint in (10b) to the objective function as a penalty term with penalty parameter $\zeta>0$, given by

$$\begin{aligned}\mathcal{P}_3: \ \min_{\mathbf{a},\mathbf{W}}\ & \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L|\\ \text{s.t.}\ & \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik},\\ & \qquad k\in\mathcal{K}_i, i\in\mathcal{G}, \quad (11a)\\ & \|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\max},\ m\in\mathcal{M}, \quad (11b)\\ & \mathbf{1}^T\mathbf{a} = L, \quad \mathbf{0}\preccurlyeq\mathbf{a}\preccurlyeq\mathbf{1}, \quad (11c)\end{aligned}$$

where the constraints on $\mathbf{a}$ are now expressed in vector forms.

Note that $\mathcal{P}_3$ is a relaxed version of $\mathcal{P}_2$ (or $\mathcal{P}_1$). The main difficulty in solving $\mathcal{P}_3$ is in the non-convex constraint (11a), which is for the worst-case SINR guarantee. In particular, the Frobenius norm $\|\cdot\|_F$ w.r.t. $\mathbf{W}$ reflects the impact of maximum covariance matrix error ϵ_{ik} from (4) on the SINR. This term

[2] Note that one may consider using the alternating optimization approach to solve $\mathbf{W}$ and $\mathbf{a}$ iteratively. However, $\mathcal{P}_3$ is non-convex w.r.t. $\mathbf{W}$, and the alternating procedure does not guarantee to converge. Also, we need to make a choice to map solution $\mathbf{a}$ to binary variables of 0 and 1, without violating any constraints, which is challenging to do.

complicates the algorithm design and makes it difficult to solve $\mathcal{P}_3$ efficiently.

Since our main goal in this phase is to determine antenna selection $\mathbf{a}$, while computing $\mathbf{W}$ in this phase is to facilitate the determination of $\mathbf{a}$, we propose to approximate this SINR constraint in $\mathcal{P}_3$. In particular, note that the first term at the LHS of (11a) is a function of the estimated channel covariance $\mathbf{R}_{ik}$ and is related to the estimated SINR. The second term shows the worst-case error under $\mathbf{W}$. It is some power function of $\mathbf{W}$ but does not depend on $\mathbf{R}_{ik}$. The choice of $\mathbf{a}$ only affects the first term: it determines the choice of sub-matrix in $\mathbf{R}_{ik}$, and thus the maximum SINR can be achieved (once $\mathbf{W}$ is optimized). The second term only reflects the power of $\mathbf{W}$ and, therefore, is not very sensitive to the choice of $\mathbf{a}$. Also, this error term is relatively small compared to the first term. Thus, the effect of this second term is only secondary for antenna selection. Following this, for the purpose of determining $\mathbf{a}$, and with consideration of the complexity of solving the problem, we ignore the second term in (11a) and only use the first term.[3] In other words, we ignore the worst-case error and express SINR under the estimated channel covariance matrices $\{\mathbf{R}_{ik}\}$ to find antenna selection vector $\mathbf{a}$. This leads to the following approximated problem from $\mathcal{P}_3$:

$$\begin{aligned}\mathcal{P}_3^{\text{SINR}}:\ &\min_{\mathbf{a},\mathbf{W}}\ \mathrm{tr}(\mathbf{W}\mathbf{W}^H)+\zeta|\mathbf{a}^T\mathbf{a}-L|\\ \text{s.t.}\ &\mathrm{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H)\geq\sigma^2\gamma_{ik}, k\in\mathcal{K}_i, i\in\mathcal{G},\\ &(11b)\text{ and }(11c).\end{aligned}$$

Although $\mathcal{P}_3^{\text{SINR}}$ is still non-convex, both the objective and constraint functions are in a quadratic form that is more amenable to efficient algorithm design. In particular, ADMM [37] is a robust and fast numerical method that can efficiently solve many large-scale structured problems. Through construction, it can decompose a complicated problem into smaller subproblems, which often can be solved with much lower computational complexity and in parallel. Following this, we design an ADMM-based fast algorithm to directly solve $\mathcal{P}_3^{\text{SINR}}$. The convergence of ADMM for a wide range of non-convex problems is recently established in [38]. Following this, we develop an ADMM-based algorithm to solve $\mathcal{P}_3^{\text{SINR}}$ that is of low computational complexity, which is shown below.

Note that the elements in the solution $\mathbf{a}^\star$ to $\mathcal{P}_3^{\text{SINR}}$ are not guaranteed to be binary. After solving $\mathcal{P}_3^{\text{SINR}}$, we determine the binary selection vector $\mathbf{a}$ by setting the L largest elements in $\mathbf{a}^\star$ to 1 and the rest to 0.

A. The ADMM-Based Algorithm for $\mathcal{P}_3^{SINR}$

To construct ADMM blocks, we introduce auxiliary variable $\mathbf{Y}\in\mathbb{C}^{M\times G}$ and transform $\mathcal{P}_3^{\text{SINR}}$ into the following equivalent problem:

$$\begin{aligned}\mathcal{P}_{\text{ADMM}}^{\text{SINR}}:\ &\min_{\mathbf{a},\mathbf{W},\mathbf{Y}}\ \mathrm{tr}(\mathbf{W}\mathbf{W}^H)+\zeta|\mathbf{a}^T\mathbf{a}-L|\\ \text{s.t.}\ &\mathrm{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H)\geq\sigma^2\gamma_{ik}, k\in\mathcal{K}_i, i\in\mathcal{G}, &(12a)\\ &\|\mathbf{e}_m^T\mathbf{Y}\|^2\leq a_m P_{\max},\quad m\in\mathcal{M}, &(12b)\\ &\mathbf{1}^T\mathbf{a}=L,\quad \mathbf{0}\preccurlyeq\mathbf{a}\preccurlyeq\mathbf{1}, &(12c)\\ &\mathbf{Y}=\mathbf{W}. &(12d)\end{aligned}$$

We use $\mathcal{C}$ to denote the feasible set of $(\mathbf{a},\mathbf{W},\mathbf{Y})$ that satisfies constraints (12a)–(12c) in $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$ and define the indicator function for $\mathcal{C}$ as $\mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y})\triangleq\{0: \text{if } (\mathbf{a},\mathbf{W},\mathbf{Y})\in\mathcal{C};\ \infty: \text{ otherwise}\}$. Then, by moving the constraints (12a)–(12c) into the objective function in $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$, we arrive at the following equivalent problem

$$\begin{aligned}\min_{\mathbf{a},\mathbf{W},\mathbf{Y}}\ &\mathrm{tr}(\mathbf{W}\mathbf{W}^H)+\zeta|\mathbf{a}^T\mathbf{a}-L|+\mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y}) &(13)\\ \text{s.t.}\ &(12d).\end{aligned}$$

The augmented Lagrangian for this problem is given by

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})=\ &\mathrm{tr}(\mathbf{W}\mathbf{W}^H)+\zeta|\mathbf{a}^T\mathbf{a}-L|+\mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y})\\ &+\rho\|\mathbf{W}-\mathbf{Y}+\mathbf{Z}\|_F^2. &(14)\end{aligned}$$

where $\mathbf{Z}\in\mathbb{C}^{M\times G}$ is the dual variable associated with constraint (12d), and ρ is the penalty parameter.

With the auxiliary variable $\mathbf{Y}$ and constraint (12d), we now can minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ w.r.t. $\mathbf{a}$, $\mathbf{W}$, and $\mathbf{Y}$ separately in an iterative fashion. Our proposed ADMM-based algorithm is summarized in Algorithm 1. The algorithm consists of two main updating blocks. In the first two blocks, we minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ w.r.t. $(\mathbf{a},\mathbf{W})$ and $\mathbf{Y}$, respectively, to obtain the updates. We will show that the solutions to these two optimization subproblems can be computed efficiently using semi-closed-form expressions. Finally, we round each a_m to the nearest 0 or 1 to obtain antenna selection.

In the first block of Algorithm 1, given $(\mathbf{Y}^l,\mathbf{Z}^l)$ in iteration l, the joint optimization subproblem (15) for $(\mathbf{a},\mathbf{W})$ can be further decoupled into two subproblems for $\mathbf{W}$ and $\mathbf{a}$ separately, since $\mathbf{W}$ and $\mathbf{a}$ are in separate terms of $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ in (14). Thus, in the following subsections, we will first describe the

---

Algorithm 1 The ADMM-Based Algorithm for $\mathcal{P}_3^{\text{SINR}}$ to Obtain Antenna Selection $\mathbf{a}$.

Initialization: Set ρ; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\mathbf{Z}^0=\mathbf{0}$; Set $l=0$.
repeat
1) Update optimization variables $\mathbf{W}^{l+1}$, $\mathbf{a}^{l+1}$:
$$(\mathbf{W}^{l+1},\mathbf{a}^{l+1})=\arg\min_{\mathbf{a},\mathbf{W}}\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y}^l,\mathbf{Z}^l). \quad (15)$$
2) Update the auxiliary variable $\mathbf{Y}^{l+1}$:
$$\mathbf{Y}^{l+1}=\arg\min_{\mathbf{Y}}\mathcal{L}_\rho(\mathbf{a}^{l+1},\mathbf{W}^{l+1},\mathbf{Y},\mathbf{Z}^l). \quad (16)$$
3) Update dual variable $\mathbf{Z}^{l+1}$:
$$\mathbf{Z}^{l+1}=\mathbf{Z}^l+\mathbf{Y}^{l+1}-\mathbf{W}^{l+1}. \quad (17)$$
4) Set $l\leftarrow l+1$.
until convergence
Obtain $\mathbf{a}$: Set the L largest elements in $\mathbf{a}^l$ to 1 and the rest to 0.

---

[3]We also note that the second term and the objective are both some power functions of $\mathbf{W}$. Both are desired to be minimized by optimizing $\mathbf{W}$. Thus, ignoring the second term in (11a) is not expected to have an adversarial effect on the resulting choice of $\mathbf{a}$. We will study the impact of this approximation in our simulation result.

solution for updating $\mathbf{W}$ in Section IV-B, and then the solution for updating $\mathbf{a}$ in Section IV-C. The solution to subproblem (16) for updating $\mathbf{Y}$ is described in Section IV-D.

B. Updating $\mathbf{W}$

For subproblem (15), given $(\mathbf{Y}^l, \mathbf{Z}^l)$ in iteration l, we first minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ w.r.t. $\mathbf{W}$. After removing the terms in (14) that are irrelevant to $\mathbf{W}$, the problem is equivalent to

$$\begin{aligned}&\min_{\mathbf{W}} \operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) + \rho\|\mathbf{W} - \mathbf{Y}^l + \mathbf{Z}^l\|_F^2 \\ &\text{s.t. } \operatorname{tr}\left(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\right) \geq \sigma^2\gamma_{ik},\ k \in \mathcal{K}_i, i \in \mathcal{G}.\end{aligned} \tag{18}$$

Subproblem (18) is still a challenging non-convex QCQP problem.[4] However, compared with $\mathcal{P}^{\text{SINR}}_{\text{ADMM}}$, it is a much smaller problem w.r.t. $\mathbf{W}$ only and has fewer constraints. More importantly, since the objective and constraint functions are in the quadratic form, we can exploit the consensus-ADMM technique to further break down the problem into quadratic convex subproblems to solve. Thus, to obtain a good approximate solution that performs well and is also computationally efficient, we propose an inner-layer consensus-ADMM-based algorithm. As we will see below, under our ADMM construction, each subproblem is a quadratic convex problem, either unconstrained or with a single linear constraint, for which a closed-form solution can be derived. This leads to fast computation of $\mathbf{W}$ to subproblem (18).

The approach for ADMM construction is similar to Algorithm 1 for $\mathcal{P}^{\text{SINR}}_{\text{ADMM}}$. First, after expanding the second term of the objective function in (18) and removing the constant terms, problem (18) is equivalent to the following problem

$$\begin{aligned}\mathcal{P}_{\mathbf{W}}: &\min_{\mathbf{W}} (\rho+1)\operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) + 2\rho\Re\left\{\operatorname{tr}\left(\mathbf{B}^l\mathbf{W}^H\right)\right\} \\ &\text{s.t.}\quad \operatorname{tr}\left(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\right) \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}.\end{aligned}$$

where $\mathbf{B}^l \triangleq -\mathbf{Y}^l + \mathbf{Z}^l$.

Introducing auxiliary variables $\mathbf{V}_{ik} \in \mathbb{C}^{M\times G}$ and $\mathbf{W}_{ik} \in \mathbb{C}^{M\times G}$, for $k \in \mathcal{K}_i$, $i \in \mathcal{G}$, we arrive at the following problem equivalent to $\mathcal{P}_{\mathbf{W}}$:

$$\mathcal{P}^{\text{ADMM}}_{\mathbf{W}}: \min_{\mathbf{W},\{\mathbf{V}_{ik},\mathbf{W}_{ik}\}} (\rho+1)\operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) + 2\rho\Re\left\{\operatorname{tr}\left(\mathbf{B}^l\mathbf{W}^H\right)\right\}$$

$$\text{s.t.}\quad \Re\left\{\operatorname{tr}\left(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H\right)\right\} \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{19a}$$

$$\mathbf{W}_{ik} = \mathbf{W}, k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{19b}$$

$$\mathbf{V}_{ik} = \mathbf{W}\mathbf{D}_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{19c}$$

Denote the feasibility set of $\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}$ satisfying the constraints in (19a) as $\mathcal{F}$, and define the indicator function $\mathbb{I}_{\mathcal{F}}(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}) \triangleq \{0: \text{if } \{\mathbf{V}_{ik}, \mathbf{W}_{ik}\} \in \mathcal{F}; \infty: \text{ otherwise}\}$. We transfer $\mathcal{P}^{\text{ADMM}}_{\mathbf{W}}$ to the following equivalent problem with only linear equality constraints:

$$\begin{aligned}\min_{\mathbf{W},\{\mathbf{V}_{ik},\mathbf{W}_{ik}\}}\quad &(\rho+1)\operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) + 2\rho\Re\left\{\operatorname{tr}\left(\mathbf{B}^l\mathbf{W}^H\right)\right\} \\ &+ \mathbb{I}_{\mathcal{F}}\left(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}\right) \\ \text{s.t.}\quad &(19b) \text{ and } (19c).\end{aligned} \tag{20}$$

[4]The constraint in (18) is similar to those in the multicast beamforming QoS problems.

Algorithm 2 Consensus-ADMM-Based Algorithm for $\mathcal{P}_{\mathbf{W}}$.

Initialization: Set $\tilde{\rho}$; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\mathbf{S}^0_{ik} = \mathbf{0}$, $\mathbf{T}^0_{ik} = \mathbf{0}$, $\forall k, i$; Set $j = 0$.

repeat

1) Update the auxiliary variables $\{\mathbf{W}^{j+1}_{ik}\}$:

$$\{\mathbf{W}^{j+1}_{ik}\} = \arg\min_{\{\mathbf{W}_{ik}\}} \tilde{\mathcal{L}}_\rho\left(\mathbf{W}^j, \{\mathbf{V}^j_{ik}, \mathbf{W}_{ik}, \mathbf{S}^j_{ik}, \mathbf{T}^j_{ik}\}\right). \tag{22}$$

2) Update the auxiliary variables $\{\mathbf{V}^{l+1}_{ik}\}$:

$$\{\mathbf{V}^{j+1}_{ik}\} = \arg\min_{\{\mathbf{V}_{ik}\}} \tilde{\mathcal{L}}_{\tilde{\rho}}\left(\mathbf{W}^j, \{\mathbf{V}_{ik}, \mathbf{W}^{j+1}_{ik}, \mathbf{S}^j_{ik}, \mathbf{T}^j_{ik}\}\right). \tag{23}$$

3) Update beamforming matrix $\mathbf{W}^{j+1}$:

$$\mathbf{W}^{j+1} = \arg\min_{\mathbf{W}} \tilde{\mathcal{L}}_{\tilde{\rho}}\left(\mathbf{W}, \{\mathbf{V}^{j+1}_{ik}, \mathbf{W}^{j+1}_{ik}, \mathbf{S}^j_{ik}, \mathbf{T}^j_{ik}\}\right). \tag{24}$$

4) Update dual variables $\mathbf{S}^{j+1}_{ik}, \mathbf{T}^{j+1}_{ik}$:

$$\mathbf{S}^{j+1}_{ik} = \mathbf{S}^j_{ik} + \mathbf{W}^{j+1} - \mathbf{W}^{j+1}_{ik}, \tag{25}$$

$$\mathbf{T}^{j+1}_{ik} = \mathbf{T}^j_{ik} + \mathbf{W}^{j+1}\mathbf{D}_{ik} - \mathbf{V}^{j+1}_{ik}. \tag{26}$$

5) Set $j \leftarrow j + 1$.

until convergence

The augmented Lagrangian for problem (20) is given by

$$\begin{aligned}&\tilde{\mathcal{L}}_{\tilde{\rho}}\left(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}\right) \\ &= (\rho+1)\operatorname{tr}\left(\mathbf{W}\mathbf{W}^H\right) + 2\rho\Re\left\{\operatorname{tr}\left(\mathbf{B}^l\mathbf{W}^H\right)\right\} \\ &\quad + \mathbb{I}_{\mathcal{F}}\left(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}\right) + \tilde{\rho}\sum_{i=1}^{G}\sum_{k=1}^{K}\|\mathbf{W}\mathbf{D}_{ik} - \mathbf{V}_{ik} + \mathbf{T}_{ik}\|_F^2 \\ &\quad + \tilde{\rho}\sum_{i=1}^{G}\sum_{k=1}^{K}\|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|_F^2.\end{aligned} \tag{21}$$

where $\{\mathbf{S}_{ik}\}$ and $\{\mathbf{T}_{ik}\}$ are the dual variables associated with the constraints in (19b) and (19c), respectively, and $\tilde{\rho}$ is the penalty parameter.

Based on the above, we construct three ADMM blocks to update these variables iteratively by minimizing $\tilde{\mathcal{L}}_{\tilde{\rho}}(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\})$ w.r.t. $\{\mathbf{W}_{ik}\}$, $\{\mathbf{V}_{ik}\}$, and $\mathbf{W}$, respectively, to solve $\mathcal{P}^{\text{ADMM}}_{\mathbf{W}}$. This consensus-ADMM-based algorithm is summarized in Algorithm 2. The three updating blocks involve solving optimization problems (22), (23), and (24), respectively. These subproblems are simple quadratic convex problems that are either unconstrained or with a single constraint. Below, we derive a closed-form solution for each of these three subproblems. For notation simplicity, we remove the superscript for iteration index, with the understanding that the solutions are used for updates in iteration j.

1) Updating $\mathbf{W}_{ik}$: We first fix $(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\})$ and minimize $\tilde{\mathcal{L}}_{\tilde{\rho}}(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}$ w.r.t. $\{\mathbf{W}_{ik}\}$ in (22). From (21), we see that the problem can be decomposed into $\sum_{i=1}^G K_i$ separate subproblems, one for each k and i:

$$\begin{aligned}\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}: &\min_{\mathbf{W}_{ik}} \|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|_F^2. \\ &\text{s.t. } \Re\left\{\operatorname{tr}\left(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H\right)\right\} \geq \sigma^2\gamma_{ik}\end{aligned}$$

Since $\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}$ is a quadratic convex optimization problem with one linear constraint, we can obtain its solution in closed-form via the KKT conditions. The Lagrangian for $\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}$ is given by

$$\mathcal{L}(\mathbf{W}_{ik},\nu_{ik}) = \|\mathbf{W}-\mathbf{W}_{ik}+\mathbf{S}_{ik}\|_F^2 + \nu_{ik}(-\mathfrak{Re}\left\{\text{tr}\left(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H\right)\right\}+\sigma^2\gamma_{ik}).$$

where $\nu_{ik}\geq 0$ is the Lagrangian multiplier associated with the constraint. Set the gradient $\nabla_{\mathbf{W}^*_{ik}}\mathcal{L}(\mathbf{W}_{ik},\lambda_{ik})=0$, we obtain the optimal $\mathbf{W}^\star_{ik}$ as

$$\mathbf{W}^\star_{ik} = \mathbf{W}+\mathbf{S}_{ik}+\frac{1}{2}\nu_{ik}\mathbf{R}_{ik}\mathbf{V}_{ik}. \tag{27}$$

Substituting the expression of $\mathbf{W}^\star_{ik}$ in (27) into the inequality constraint and based on the complementary slackness condition along with $\nu_{ik}\geq 0$, we obtain the optimal $\nu^\star_{ik}$:

$$\nu^\star_{ik} = \max\left\{\frac{2\left(-\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}(\mathbf{W}+\mathbf{S}_{ik})^H)\}+\sigma^2\gamma_{ik}\right)}{\|\mathbf{R}_{ik}\mathbf{V}_{ik}\|_F^2},0\right\}. \tag{28}$$

2) Updating $\mathbf{V}_{ik}$*:* To update $\mathbf{V}_{ik}$ for fixed $(\mathbf{W},\{\mathbf{S}_{ik},\mathbf{T}_{ik}\})$, the optimization problem (23) can again be decomposed into $\sum_{i=1}^G K_i$ separate subproblems, one for each k and i:

$$\mathcal{P}^{\text{in}}_{\mathbf{V}_{ik}}:\min_{\mathbf{V}_{ik}} \|\mathbf{W}\mathbf{D}_{ik}-\mathbf{V}_{ik}+\mathbf{T}_{ik}\|_F^2.$$
$$\text{s.t. } \mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\}\geq\sigma^2\gamma_{ik}.$$

The form of $\mathcal{P}^{\text{in}}_{\mathbf{V}_{ik}}$ is the same as $\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}$. Thus, we directly have the solution as

$$\mathbf{V}^\star_{ik} = \mathbf{W}\mathbf{D}_{ik}+\mathbf{T}_{ik}+\frac{1}{2}\tilde{\nu}^\star_{ik}\mathbf{R}_{ik}^H\mathbf{W}_{ik} \tag{29}$$

where $\tilde{\nu}^\star_{ik}=\max\left\{\frac{2(-\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}(\mathbf{W}\mathbf{D}_{ik}+\mathbf{T}_{ik})\mathbf{W}_{ik}^H)\}+\sigma^2\gamma_{ik})}{\|\mathbf{W}_{ik}^H\mathbf{R}_{ik}\|_F^2},0\right\}$.

3) Updating $\mathbf{W}$*:* Problem (24) to solve $\mathbf{W}$ for fixed $\{\mathbf{V}_{ik},\mathbf{W}_{ik},\mathbf{S}_{ik},\mathbf{T}_{ik}\}$ is an unconstrained quadratic optimization problem, given by

$$\min_{\mathbf{W}} (\rho+1)\text{tr}\left(\mathbf{W}\mathbf{W}^H\right)+2\rho\mathfrak{Re}\left\{\text{tr}\left(\mathbf{B}^l\mathbf{W}^H\right)\right\} + \tilde{\rho}\sum_{i=1}^G\sum_{k=1}^K\|\mathbf{W}\mathbf{D}_{ik}-\mathbf{V}_{ik}+\mathbf{T}_{ik}\|_F^2 + \tilde{\rho}\sum_{i=1}^G\sum_{k=1}^K\|\mathbf{W}-\mathbf{W}_{ik}+\mathbf{S}_{ik}\|_F^2.$$

Taking the derivative of the objective function and setting it to zero, we obtain

$$\mathbf{W}^\star = \left(\tilde{\rho}\mathbf{C}-\rho\mathbf{B}^l\right)\left(\tilde{\rho}\sum_{i=1}^G\sum_{k=1}^K\mathbf{D}_{ik}^2+\left(\tilde{\rho}\sum_{i=1}^G K_i+\rho+1\right)\mathbf{I}\right)^{-1}. \tag{30}$$

where $\mathbf{C}\triangleq\sum_{i=1}^G\sum_{k=1}^K(\mathbf{W}_{ik}-\mathbf{S}_{ik}+(\mathbf{V}_{ik}-\mathbf{T}_{ik})\mathbf{D}_{ik})$.

C. Updating a

For subproblem (15), we now minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y}^l,\mathbf{Z}^l)$ w.r.t. $\mathbf{a}$. After removing the terms in (14) that are not related to $\mathbf{a}$, we have the following equivalent problem:

$$\min_{\mathbf{a}} \ \zeta|\mathbf{a}^T\mathbf{a}-L| \tag{31a}$$
$$\text{s.t. } \|\mathbf{e}_m^T\mathbf{Y}\|^2\leq a_m P_{\max},\ m\in\mathcal{M}, \tag{31b}$$
$$\mathbf{1}^T\mathbf{a}=L,\quad \mathbf{0}\preccurlyeq\mathbf{a}\preccurlyeq\mathbf{1}. \tag{31c}$$

Note that by (31c), we have $\mathbf{a}^T\mathbf{a}\leq\mathbf{1}^T\mathbf{a}=L$. Thus, minimizing the objective function in (31a) is equivalent to maximizing $\mathbf{a}^T\mathbf{a}$. Furthermore, we can combine (31b) and the inequality constraint in (31c) to determine the lower bound on a_m. Then, problem (31) is equivalent to the following problem:

$$\mathcal{P}_{\mathbf{a}}:\max_{\mathbf{a}} \mathbf{a}^T\mathbf{a}$$
$$\text{s.t. } \mathbf{p}\preccurlyeq\mathbf{a}\preccurlyeq\mathbf{1},\ \mathbf{1}^T\mathbf{a}=L.$$

where $\mathbf{p}\triangleq[p_1,\ldots,p_M]^T$ with $p_m\triangleq\frac{\|\mathbf{e}_m^T\mathbf{Y}\|^2}{P_{\max}}$. Note that $p_m\leq 1$ for $\mathcal{P}_{\mathbf{a}}$ being feasible.

Since $\mathcal{P}_{\mathbf{a}}$ contains only linear constraints, it meets the linearity constraint qualification, under which the KKT conditions is necessary for optimality [39]. Thus, we use the KKT conditions to derive the semi-closed-form solution to $\mathcal{P}_{\mathbf{a}}$. In particular, we show there is a unique solution satisfying the KKT conditions, and thus it is the optimal solution to $\mathcal{P}_{\mathbf{a}}$. The Lagrangian for $\mathcal{P}_{\mathbf{a}}$ is

$$\mathcal{L}(\mathbf{a},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\mu) = -\mathbf{a}^T\mathbf{a}+\boldsymbol{\lambda}_1^T(\mathbf{p}-\mathbf{a})+\boldsymbol{\lambda}_2^T(\mathbf{a}-\mathbf{1}) + \mu(\mathbf{1}^T\mathbf{a}-L) \tag{32}$$

where $\boldsymbol{\lambda}_1\triangleq[\lambda_{11},\ldots,\lambda_{1M}]^T$, $\boldsymbol{\lambda}_2\triangleq[\lambda_{21},\ldots,\lambda_{2M}]^T$, and μ are the Lagrangian multipliers associated with the constraints in $\mathcal{P}_{\mathbf{a}}$, respectively. The KKT conditions are listed below:

C1 : $\nabla_{\mathbf{a}}\mathcal{L}(\mathbf{a},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\mu)=\mathbf{0}\Rightarrow -2\mathbf{a}-\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_2+\mu\mathbf{1}=\mathbf{0}$.
C2 : $\boldsymbol{\lambda}_1\odot(\mathbf{p}-\mathbf{a})=\mathbf{0}$, C3 : $\boldsymbol{\lambda}_2\odot(\mathbf{a}-\mathbf{1})=\mathbf{0}$,
C4 : $\mathbf{p}\preccurlyeq\mathbf{a}\preccurlyeq\mathbf{1}$, C5 : $\mathbf{1}^T\mathbf{a}-L=0$.

From C2–C4, we have the following three situations for a_m, $m\in Mc$:

S1) If $\lambda_{2m}\neq 0$, then $a_m=1$ and $\lambda_{1m}=0$.
S2) If $p_m<a_m<1$, then $\lambda_{1m}=\lambda_{2m}=0$; and by C1, $a_m=\mu/2\triangleq\bar{a}$, where $\bar{a}\in(p_m,1)$.
S3) If $\lambda_{1m}\neq 0$, then $a_m=p_m$, and $\lambda_{2m}=0$.

From the above three possible values for a_m, we partition the set $\{a_m\}$ into three subsets. Define the following index sets: $\mathcal{S}_1\triangleq\{m:\ a_m=1, m\in M\}$, $\mathcal{S}_2=\{m:\ a_m=\bar{a}, m\in M\}$, $\mathcal{S}_3=\{m:\ a_m=p_m, m\in M\}$. Also, denote $N_1\triangleq|\mathcal{S}_1|$ and $N_2\triangleq|\mathcal{S}_2|$. We have $|\mathcal{S}_3|=M-N_1-N_2$. With these three sets, C5 can be rewritten as

$$N_1+N_2\bar{a}+\sum_{m\in\mathcal{S}_3}p_m=L. \tag{33}$$

Similarly, we also rewrite the objective function in $\mathcal{P}_{\mathbf{a}}$ as

$$\mathbf{a}^T\mathbf{a}=N_1+N_2\bar{a}^2+\sum_{m\in\mathcal{S}_3}p_m^2. \tag{34}$$

Thus, to solve $\mathcal{P}_{\mathbf{a}}$, we need to determine N_1 and N_2, *i.e.,* the sizes of $\mathcal{S}_1$ and $\mathcal{S}_2$, and the value $\bar{a}$ for $\mathcal{S}_2$.

Note that in general, if $\sum_{m\in\mathcal{S}_3}p_m\notin\mathbb{Z}^+$, then for (33) to hold, we must have $N_2\geq 1$ ($\mathcal{S}_2\neq\emptyset$). Thus, we have the following two possible cases for $\mathcal{S}_2$: 1) $N_2\geq 1$; 2) $N_2=0$. Note that $N_2\geq 1$ is most likely the case. This is because for any given

$\mathbf{p}$, we typically need $\mathcal{S}_2$ in order to satisfy C5. In what follows, we first focus on $N_2 \geq 1$.

1) When $N_2 \geq 1$: First, assume that $\mathcal{S}_3$ is given. Let $z \triangleq N_1 + N_2$, where $z > N_1$ for $N_2 \geq 1$. Also, since $\mathcal{S}_3$ is given, z is fixed. From (33), we have

$$N_1 + N_2\bar{a} = L - \sum_{m\in\mathcal{S}_3} p_m \triangleq L' \tag{35}$$

where L' is fixed for given $\mathcal{S}_3$. Then, we have $\bar{a} = \frac{L'-N_1}{z-N_1}$. From (34), let $y(N_1) \triangleq N_1 + N_2\bar{a}^2$. Replacing $\bar{a}$ with the expression above, we can write $y(N_1)$ as

$$y(N_1) = N_1 + \frac{(L'-N_1)^2}{z-N_1}. \tag{36}$$

Following this, we can express the objective function as

$$\mathbf{a}^T\mathbf{a} = y(N_1) + \sum_{m\in\mathcal{S}_3} p_m^2 \tag{37}$$

where the second term is fixed for given $\mathcal{S}_3$. Thus, $\mathcal{P}_{\mathbf{a}}$ with given $\mathcal{S}_3$ is equivalent to

$$\max_{N_1:N_1<z} y(N_1), \tag{38}$$

for which we only need to check the difference in $y(N_1)$ w.r.t. N_1. Consider $N_1 + 1 < z$. From (36), we have

$$y(N_1+1) - y(N_1) = \frac{(z-L')^2}{(z-N_1)(z-N_1-1)} > 0. \tag{39}$$

Thus, for any given $z > 0$, $y(N_1)$ is an increasing function of N_1, and therefore for (38), N_1 should be maximized. This leads to $N_1^\star = z - 1$, and $N_2^\star = 1$, for given $z > 0$. Thus, we obtain $N_2^\star$ for $\mathcal{S}_2$, *i.e.,* only one $a_m = \bar{a}$ with inactive inequality constraints.

Next, we determine the optimal $\mathcal{S}_3$ and N_1 for $\mathcal{P}_{\mathbf{a}}$. Note that for $N_2^\star = 1$, the objective function in (34) becomes

$$\mathbf{a}^T\mathbf{a} = N_1 + \bar{a}^2 + \sum_{m\in\mathcal{S}_3} p_m^2 \triangleq g(N_1, \mathbf{p}) \tag{40}$$

where $|\mathcal{S}_3| = M - N_1 - 1$. Now, we examine $g(N_1, \mathbf{p})$ w.r.t. N_1. By (35), we can express N_1 in terms of $\mathcal{S}_3$ as

$$N_1 = L - \left(\bar{a} + \sum_{m\in\mathcal{S}_3} p_m\right). \tag{41}$$

Substituting the above expression into $g(N_1, \mathbf{p})$, we have

$$g(N_1, \mathbf{p}) = L - \bar{a} + \bar{a}^2 - \sum_{m\in\mathcal{S}_3} p_m + \sum_{m\in\mathcal{S}_3} p_m^2$$

which implicitly depends on N_1 through $\mathcal{S}_3$. Note that increasing N_1 to $N_1 + 1$ means removing an element from $\mathcal{S}_3$ and reducing $|\mathcal{S}_3|$ by one. Let m' be the index of the entry that is removed from $\mathcal{S}_3$. Then, we have

$$g(N_1+1, \mathbf{p}) - g(N_1, \mathbf{p}) = p_{m'} - p_{m'}^2 \geq 0, \tag{42}$$

where the last inequality is due to $0 \leq p_m \leq 1, \forall m$. Thus, with $\bar{a}$ given, $g(N_1, \mathbf{p})$ is an increasing function of N_1, and maximizing N_1 leads to the maximum objective value $\mathbf{a}^T\mathbf{a}$ in (40).

Now, consider N_1 in (41), which is obtained from the equality constraint C5. For any given size $|\mathcal{S}_3|$, N_1 in (41) is maximized by choosing elements in $\mathcal{S}_3$ whose p_m's have the lowest values. Hence, for any N_1, $\mathcal{S}_3$ can be determined as follows. Sort $\{p_m\}$ in the ascending order: $p_{m_1} \leq p_{m_2} \leq \cdots \leq p_{m_M}$. Then, $\mathcal{S}_3 = \{m_1, \ldots, m_{M-N_1-1}\}$ and

$$N_1 = L - \left(\bar{a} + \sum_{j=1}^{M-N_1-1} p_{m_j}\right). \tag{43}$$

To determine the optimal $N_1^\star$, we recall that $\bar{a} < 1$. From (43), we have the following bounds: $L - 1 < N_1^\star + \sum_{j=1}^{M-N_1^\star-1} p_{m_j} < L$. This leads to a simple search for the optimal $N_1^\star$: Starting from $N_1 = L$, we reduce N_1 until the above inequalities are met. Then, we obtain the optimal $N_1^\star$ and $\mathcal{S}_3^\star$. Once $N_1^\star$ is obtained, the optimal $\bar{a}^\star$ is obtained as

$$\bar{a}^\star = L - N_1^\star - \sum_{j=1}^{M-N_1^\star-1} p_{m_j}. \tag{44}$$

Algorithm 3 The Optimal Solution $\mathbf{a}^\star$ for $\mathcal{P}_{\mathbf{a}}$.

1) Sort $\mathbf{p}$ in ascending order: $p_{m_1} \leq p_{m_2} \leq \cdots \leq p_{m_M}$.
2) Initialize $N_1 = L - 1$.
while $N_1 + \sum_{j=1}^{M-N_1-1} p_{m_j} \geq L$ **do**
  $N_1 \leftarrow N_1 - 1$.
end while
3) Set $N_1^\star = N_1$.
4) Determine the optimal $\mathbf{a}^\star$:
  a) $a_{m_j}^\star = p_{m_j}$, $j = 1, \ldots, M - N_1^\star - 1$.
  b) $a_{m_{M-N_1^\star}}^\star = \bar{a}^\star$ as in (44).
  c) $a_{m_j}^\star = 1$, $j = M - N_1^\star + 1, \ldots, M$.

2) When $N_2 = 0$: Previously, for general $\mathbf{p}$, we have assumed $N_2 \geq 1$ in (35) and obtain $N_2^\star = 1$. Although less likely, it is possible that $N_2 = 0$ for some $\mathcal{S}_3$, where $\sum_{m\in\mathcal{S}_3} p_m \in \mathbb{Z}^+$. Denote such special set as $\mathcal{S}_3^z$. Then, N_1 in (41) is reduced to $N_1 = L - \sum_{m\in\mathcal{S}_3^z} p_m$. Comparing it with $N_1^\star$ in (43), since $0 < \bar{a} < 1$, it is not difficult to see that unless $\mathcal{S}_3^z$ happens to contain the indices of the lowest values in $\{p_m\}$, $N_1^\star$ in (43) under $N_2^\star = 1$ still has the maximum value and is optimal for $\mathcal{P}_{\mathbf{a}}$. If $\mathcal{S}_3^z$ contains the indices of the lowest values in $\{p_m\}$, *i.e.,* $\mathcal{S}_3^z = \{m_1, \ldots, m_{M-N_1}\}$, then

$$N_1 = L - \sum_{j=1}^{M-N_1} p_{m_j}. \tag{45}$$

Comparing the above expression with that in (43), we see that in fact we can treat the solution above as the special case of (43) for $N_2 \geq 1$, where substituting N_1 in (45) into (44), we have $\bar{a}^\star = p_{m_{M-N_1^\star}}$. In other words, in this special case, $\mathcal{S}_2$ and $\mathcal{S}_3$ are merged into $\mathcal{S}_3$.

We summarize the optimal solution $\mathbf{a}^\star$ to $\mathcal{P}_{\mathbf{a}}$ in Algorithm 3. In this algorithm, only sorting the elements in $\mathbf{p}$ is required; otherwise, $\mathbf{a}^\star$ is provided in closed form.

Remark 1. In the reformulated problem $\mathcal{P}_3^{\text{SINR}}$ (and $\mathcal{P}_3$), the objective function contains an additional penalty term

$\zeta|\mathbf{a}^T\mathbf{a} - L|$, where ζ is the penalty parameter which typically requires tuning to find its appropriate value. However, note that our proposed ADMM-based algorithm in Algorithm 1 does not require tuning ζ. Specifically, the joint optimization in (15) is separated into problem (18) for $\mathbf{W}$ and $\mathcal{P}_{\mathbf{a}}$ for $\mathbf{a}$. The value of ζ does not affect the optimal solution $\mathbf{a}$ for $\mathcal{P}_{\mathbf{a}}$ and can be ignored. Thus, Algorithm 1 does not depend on the penalty parameter ζ and is simple to implement.

D. Updating $\mathbf{Y}$

Subproblem (16) for updating $\mathbf{Y}$ can be decomposed into M separate subproblems, one for each $\mathbf{y}_m \triangleq \mathbf{e}_m^T\mathbf{Y}$, $m \in Mc$:

$$\mathcal{P}_{\mathbf{y}_m}^{\text{in}} : \min_{\mathbf{y}_m} \rho\|\mathbf{e}_m^T\mathbf{W} - \mathbf{y}_m + \mathbf{e}_m^T\mathbf{Z}\|_F^2 \quad \text{s.t. } \|\mathbf{y}_m\|^2 \le a_m P_{\max}.$$

Since $\mathcal{P}_{\mathbf{y}_m}^{\text{in}}$ is a convex QCQP problem with a single constraint, it yields a closed-form solution with two possible forms:

i) $\mathbf{y}_m^\star = \mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})$ if it satisfies constraint in $\mathcal{P}_{\mathbf{y}_m}^{\text{in}}$;

ii) Otherwise, $\mathbf{y}_m^\star = \sqrt{a_m P_{\max}} \frac{\mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})}{\|\mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})\|}$.

We will delay the discussion of the computational complexity of Algorithms 1-3 until Section V-C.

V. Antenna Selection: SLR-Based Approach

A. The SLR-Based Problem Formulation

Our previous approach for antenna selection in $\mathcal{P}_3^{\text{SINR}}$ is based on SINR under the estimated channel covariance matrix. To further reduce the computational complexity, in this section, we propose a different approach to antenna selection that is based on signal-to-leakage-ratio (SLR) under the estimated channel covariance matrix. SLR is a heuristic performance metric often used for downlink multiuser (unicast) beamforming design in the literature [40]. Using SLR is shown to lead to a simpler problem and yet provides an effective solution with good performance, since the SLR metric heuristically resembles the optimal beamforming structure.

In our problem, the SLR for user k in group i is defined as the ratio of the received signal power at user k to the interference that group i causes to all the other groups. Using the estimated channel covariance, we express SLR as

$$\text{SLR}_{ik} = \frac{\mathbf{w}_i^H\mathbf{R}_{ik}\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\sum_{j=1}^{K_l}\mathbf{w}_i^H\mathbf{R}_{lj}\mathbf{w}_i + \sigma^2}. \tag{46}$$

Then, we replace the SINR constraints in $\mathcal{P}_3^{\text{SINR}}$ with the SLR requirements and consider the following problem

$$\begin{aligned}\mathcal{P}_3^{\text{SLR}} : \min_{\mathbf{a},\mathbf{W}} & \sum_{i=1}^G \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| \\ \text{s.t. } & \text{SLR}_{ik} \ge \hat{\gamma}_{ik}, \quad k\in\mathcal{K}_i, i\in\mathcal{G}, \\ & \text{(11b) and (11c)}.\end{aligned}$$

where $\hat{\gamma}_{ik}$ is the minimum SLR target.

Remark 2. Note that due to the difference between SINR and SLR metrics, the value of the SLR target $\hat{\gamma}_{ik}$ may be different from the SINR target γ_{ik}, although we do not expect significant difference. Also, the choice of antenna selection mainly depends on $\{\mathbf{R}_{ik}\}$ but is not sensitive to the value of $\hat{\gamma}_{ik}$. Hence, we chose $\hat{\gamma}_{ik} = \gamma_{ik}$. In our simulation study, we will compare two approaches for antenna selection in Phase 1 and show that they achieve near identical performance.

B. The ADMM-Based Fast Algorithm

Similar to Algorithm 1, we develop an ADMM-based fast algorithm to solve $\mathcal{P}_3^{\text{SLR}}$. The construction procedure is similar to $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$ for $\mathcal{P}_3^{\text{SINR}}$ in Section IV. Thus, we will only describe the procedure briefly. Using auxiliary variables $\overline{\mathbf{Y}} \in \mathbb{C}^{M\times G}$, we have the following problem equivalent to $\mathcal{P}_3^{\text{SLR}}$:

$$\mathcal{P}_{\text{ADMM}}^{\text{SLR}} : \min_{\mathbf{a},\mathbf{W}} \sum_{i=1}^G \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L|$$

$$\text{s.t. } \mathbf{w}_i^H\left(\mathbf{R}_{ik} - \gamma_{ik}\overline{\mathbf{R}}_{-i}\right)\mathbf{w}_i \ge \sigma_{ik}^2\hat{\gamma}_{ik}, \ k\in\mathcal{K}_i, i\in\mathcal{G}, \tag{47a}$$

$$\|\mathbf{e}_m^T\overline{\mathbf{Y}}\|^2 \le a_m P_{\max}, \ j\in\mathcal{M}, \tag{47b}$$

$$\mathbf{1}^T\mathbf{a} = L, \quad \mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}, \tag{47c}$$

$$\overline{\mathbf{Y}} = \mathbf{W} \tag{47d}$$

where $\overline{\mathbf{R}}_{-i} \triangleq \sum_{l\in\mathcal{G}_{-i}}\sum_{j=1}^{K_l}\mathbf{R}_{lj}$. Denote the feasibility set of $(\mathbf{W},\mathbf{a},\overline{\mathbf{Y}})$ satisfying constraints in (47a)–(47c) as $\mathcal{H}$ and define the indicator function $\mathbb{I}_{\mathcal{H}}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}) \triangleq \{0 : \text{if } (\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}) \in \mathcal{H};\ \infty : \text{ otherwise}\}$. We transform $\mathcal{P}_{\text{ADMM}}^{\text{SLR}}$ to the following equivalent problem:

$$\min_{\mathbf{W},\mathbf{a},\overline{\mathbf{Y}}} \sum_{i=1}^G \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| + \mathbb{I}_{\mathcal{H}}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}) \quad \text{s.t., (47d)} \tag{48}$$

for which the augmented Lagrangian is given by

$$\begin{aligned}\bar{\mathcal{L}}_{\bar\rho}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}},\overline{\mathbf{Z}}) = & \sum_{i=1}^G \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| + \mathbb{I}_{\mathcal{H}}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}) \\ & + \bar\rho\|\mathbf{W} - \overline{\mathbf{Y}} + \overline{\mathbf{Z}}\|_F^2\end{aligned}$$

where $\overline{\mathbf{Z}} \in \mathbb{C}^{M\times G}$ is the dual variable associated with constraint (47d) and $\bar\rho$ is the penalty parameter.

Similar to Algorithm 1, we construct two ADMM blocks to minimize $\bar{\mathcal{L}}_{\bar\rho}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}},\overline{\mathbf{Z}})$ alternatingly w.r.t. $\mathbf{a}$, $\mathbf{W}$, and $\overline{\mathbf{Y}}$. Once the convergence is reached, we round each a_m to the nearest 0 or 1. The algorithm is summarized in Algorithm 4. Again, in the first block of Algorithm 4, the joint optimization of $(\mathbf{a},\mathbf{W})$ in subproblem (49) can be decoupled into separate subproblems for $\mathbf{W}$ and $\mathbf{a}$, respectively.[5] In particular, for the first two ADMM blocks:

- Updating $\mathbf{a}$: In subproblem (49), the minimization of $\bar{\mathcal{L}}_{\bar\rho}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}^l,\overline{\mathbf{Z}}^l)$ w.r.t. $\mathbf{a}$ is the same as in problem (31) in Section IV. Thus, Algorithm 3 provides the optimal solution $\mathbf{a}^\star$ for updating $\mathbf{a}$.
- Updating $\overline{\mathbf{Y}}$: Subproblem (50) is the same as subproblem (16). Thus, the solution in Section IV-D follows.

Thus, only the minimization of $\bar{\mathcal{L}}_{\bar\rho}(\mathbf{a},\mathbf{W},\overline{\mathbf{Y}}^l,\overline{\mathbf{Z}}^l)$ w.r.t. $\mathbf{W}$ in subproblem (49) is different from that for $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$ due to

[5]Remark 1 also applies to Algorithm 4 for our SLR-based approach, where no tuning of ζ is needed.

Algorithm 4 The ADMM-Based Algorithm for $\mathcal{P}_3^{\text{SLR}}$ to Obtain Antenna Selection $\mathbf{a}$.

Initialization: Set $\bar{\rho}$; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\overline{\mathbf{Z}}^0 = \mathbf{0}$; Set $l = 0$.
repeat
1) Update optimization variables $\mathbf{W}^{l+1}$, $\mathbf{a}^{l+1}$:

$$(\mathbf{W}^{l+1}, \mathbf{a}^{l+1}) = \arg\min_{\mathbf{a},\mathbf{W}} \bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}^l, \overline{\mathbf{Z}}^l\big). \tag{49}$$

2) Update the auxiliary variable $\overline{\mathbf{Y}}^{l+1}$:

$$\overline{\mathbf{Y}}^{l+1} = \arg\min_{\overline{\mathbf{Y}}} \bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}^{l+1}, \mathbf{W}^{l+1}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}^l\big). \tag{50}$$

3) Update dual variable $\overline{\mathbf{Z}}^{l+1}$:

$$\overline{\mathbf{Z}}^{l+1} = \overline{\mathbf{Z}}^l + \overline{\mathbf{Y}}^{l+1} - \mathbf{W}^{l+1}. \tag{51}$$

4) Set $l \leftarrow l + 1$.
until convergence
Obtain $\mathbf{a}$: Set the L largest elements in $\mathbf{a}^l$ to 1 and the rest to 0.

the SLR constraints in (47a). Below, we show $\mathbf{W}$ update can be computed efficiently using semi-closed-form expressions. In particular, we will show that the SLR metric allows us to further decompose the optimization problem into smaller subproblems, which yields (semi)-closed-form updates with lower computational complexity.

1) Updating $\mathbf{W}$*:* Recall $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_G]$. Similarly, we denote $\overline{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_G]$ and $\overline{\mathbf{Z}} = [\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_G]$. Then, optimizing $\mathbf{W}$ in subproblem (49) can be written as follows:

$$\begin{aligned} \min_{\{\mathbf{w}_i\}} \quad & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \bar{\rho} \sum_{i=1}^{G} \|\mathbf{w}_i - \bar{\mathbf{y}}_i + \bar{\mathbf{z}}_i\|^2 \\ \text{s.t.} \quad & \mathbf{w}_i^H \left(\mathbf{R}_{ik} - \hat{\gamma}_{ik}\overline{\mathbf{R}}_{-i}\right) \mathbf{w}_i \geq \sigma^2 \hat{\gamma}_{ik}, \;\; k \in \mathcal{K}_i, i \in \mathcal{G}. \end{aligned} \tag{52}$$

The above problem can be decomposed into G separate subproblems, one for each $\mathbf{w}_i$, given by

$$\begin{aligned} \mathcal{P}_{\mathbf{w}_i} : \min_{\mathbf{w}_i} \; & \|\mathbf{w}_i\|^2 + \bar{\rho}\|\mathbf{w}_i - \bar{\mathbf{y}}_i + \bar{\mathbf{z}}_i\|^2 \\ \text{s.t.} \; & \mathbf{w}_i^H \left(\mathbf{R}_{ik} - \hat{\gamma}_{ik}\overline{\mathbf{R}}_{-i}\right) \mathbf{w}_i \geq \sigma^2 \hat{\gamma}_{ik}, \;\; k \in \mathcal{K}_i. \end{aligned}$$

Remark 3. Compared with $\mathcal{P}_{\mathbf{W}}$ in Algorithm 1 for solving $\mathcal{P}_3^{\text{SINR}}$, the SLR constraints in $\mathcal{P}_3^{\text{SLR}}$ allow us to decompose problem (52) into G subproblems to solve separately, which leads to simpler subproblems with closed-form solutions.

Problem $\mathcal{P}_{\mathbf{w}_i}$ is a non-convex QCQP problem. We again develop an inner consensus-ADMM-based fast algorithm to solve it. For notation simplicity, in the discussion below, we remove subscript i from all variables in $\mathcal{P}_{\mathbf{w}_i}$, with the understanding that they are all for group i. Using auxiliary variables $\hat{\mathbf{w}}_k \in \mathbb{C}^{M\times 1}$, $k \in \mathcal{K}$, we have the following problem equivalent to $\mathcal{P}_{\mathbf{w}_i}$:

$$\min_{\mathbf{w},\{\hat{\mathbf{w}}_k\}} \; \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 \tag{53a}$$

$$\text{s.t.} \;\; \mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} - \frac{1}{\hat{\gamma}_k}\hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k + \sigma^2 \leq 0, \;\; k \in \mathcal{K}, \tag{53b}$$

$$\hat{\mathbf{w}}_k = \mathbf{w}, \;\; k \in \mathcal{K}. \tag{53c}$$

Algorithm 5 Consensus-ADMM-Based Algorithm for $\mathcal{P}_{\mathbf{w}_i}$.

Initialization: Set $\hat{\rho}$; Set initial $\mathbf{w}^0$, $\hat{\mathbf{z}}_k^0 = \mathbf{0}, \forall k$; Set $q = 0$.
repeat
1) Update the auxiliary variables $\{\hat{\mathbf{w}}_k^{j+1}\}$:

$$\{\hat{\mathbf{w}}_k^{j+1}\} = \arg\min_{\{\hat{\mathbf{w}}_k\}} \hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}^j, \{\hat{\mathbf{w}}_k\}, \{\hat{\mathbf{z}}_k^j\}\big). \tag{55}$$

2) Update beamforming matrix $\mathbf{w}^{j+1}$:

$$\mathbf{w}^{j+1} = \arg\min_{\mathbf{w}} \hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k^{j+1}\}, \{\hat{\mathbf{z}}_k^j\}\big). \tag{56}$$

3) Update dual variables $\{\hat{\mathbf{z}}_k^{j+1}\}$:

$$\hat{\mathbf{z}}_k^{j+1} = \hat{\mathbf{z}}_k^j + \mathbf{w}^{j+1} - \mathbf{w}_k^{j+1}. \tag{57}$$

4) Set $j \leftarrow j + 1$.
until convergence

Define $\mathcal{V}$ as the feasible set satisfying the inequality constraints in the above problem and define the indicator function $\mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) \triangleq \{0 : \text{if } \big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) \in \mathcal{V}; \infty : \text{otherwise}\}$. Then, the above problem is equivalent to

$$\begin{aligned} \min_{\mathbf{w},\{\hat{\mathbf{w}}_k\}} \quad & \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) \\ \text{s.t.} \quad & (53c), \end{aligned} \tag{54}$$

for which the augmented Lagrangian is given by

$$\begin{aligned} \hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}, \{\hat{\mathbf{z}}_k\}\big) &= \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 \\ &\quad + \mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) + \hat{\rho}\sum_{k=1}^{K} \|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 \end{aligned}$$

where $\hat{\mathbf{z}}_k$'s are the dual variables associated with constraints in (53c) and $\hat{\rho}$ is the penalty parameter. Again, we construct two ADMM blocks to minimize $\hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k\}\big)$ alternatingly. We show below that each optimization problem yields a (semi)-closed-form solution. The consensus-ADMM-based algorithm is illustrated in Algorithm 5.

i) Updating $\{\hat{\mathbf{w}}_k\}$: The subproblem (55) can be decomposed into K $(= |\mathcal{K}|)$ separate subproblems, each given by

$$\begin{aligned} \min_{\hat{\mathbf{w}}_k} \; & \|\mathbf{w} + \hat{\mathbf{z}}_k - \hat{\mathbf{w}}_k\|^2 \\ \text{s.t.} \; & \hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k \geq \gamma_k \left(\mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} + \sigma^2\right) \end{aligned} \tag{58}$$

Problem (58) is a QCQP-1 problem with a single concave constraint function. Since Slater's condition is satisfied, the strong duality holds for problem (58), and the optimal solution satisfies the KKT conditions [41, Appendix B]. Thus, we use the KKT conditions to derive the solution and obtain a unique semi-closed-form solution, which is optimal to problem (58). The details are given in Appendix A.

ii) Updating $\mathbf{w}$: The subproblem (56) for $\mathbf{w}$ is given by

$$\begin{aligned} \min_{\mathbf{w}} \; & \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \hat{\rho}\sum_{k=1}^{K} \|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 \\ \text{s.t.} \; & \mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} \leq \min_{k\in\mathcal{K}} \frac{1}{\hat{\gamma}_k}\hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k - \sigma^2. \end{aligned} \tag{59}$$

It is a convex QCQP problem with a single constraint, for which we obtain the closed-form solution. The details are given in Appendix B.

C. Computational Complexity Comparison

1) The SINR-Based Approach via $\mathcal{P}_3^{\text{SINR}}$: Overall, our proposed algorithm to solve $\mathcal{P}_3^{\text{SINR}}$ is a two-layered ADMM-based fast algorithm. Algorithm 1 provides the outer-layer ADMM procedure, where the update for $\mathbf{W}$ in (15) is obtained by the inner-layer consensus-ADMM-based algorithm in Algorithm 2. The dominant computational operation in Algorithm 1 is for updating $\mathbf{W}$ in $\mathcal{P}_{\mathbf{W}}$ (in Section IV-B), which is computed by the consensus-ADMM-based algorithm in Algorithm 2. The updates of $\mathbf{a}$ and $\mathbf{Y}$ both yield closed-form solutions as shown in Algorithm 3 and in Section IV-D, respectively, where only simple matrix-vector operations are involved. Since Algorithm 3 involves sorting with average complexity $\mathcal{O}(M\log(M))$, the overall complexity for $\mathbf{a}$ and $\mathbf{Y}$ updates is $\mathcal{O}(M\log(M)+MG)$ in each iteration. Algorithm 2 contains three ADMM blocks in each iteration for optimizing $\{\mathbf{W}_{ik}\}$, $\{\mathbf{V}_{ik}\}$, and $\mathbf{W}$, with closed-form solutions given in (27), (29), and (30), respectively. Thus, the overall computational complexity in each iteration of Algorithm 2 is $\mathcal{O}(M^2G\sum_{i=1}^{G}K_i)$.[6]

2) The SLR-Based Approach via $\mathcal{P}_3^{\text{SLR}}$: Our algorithm to solve $\mathcal{P}_3^{\text{SLR}}$ is also a two-layered ADMM-based fast algorithm. Algorithm 4 provides the outer-layer ADMM procedure for solving $\mathcal{P}_3^{\text{SLR}}$, where the update for $\mathbf{W}$ in (49) is obtained by the inner-layer ADMM algorithm in Algorithm 5. The main computational operation in Algorithm 4 is for updating $\mathbf{W}$ (in Section V-B1), which is the consensus-ADMM-based algorithm in Algorithm 5. As discussed in Section V-B, the updates of $\mathbf{a}$ and $\mathbf{Y}$ are the same as the SINR-based approach in Algorithm 1. Thus, the only difference between the two approaches is in the computational complexity incurred for updating $\mathbf{W}$. In particular, Algorithm 5 for updating $\mathbf{w}_i$ contains two ADMM blocks in each iteration for optimizing $\{\hat{\mathbf{w}}_k\}$ and $\{\mathbf{w}_i\}$, whose semi-closed-form solutions are given in (61) and (59), respectively. Note that the matrix inversions in (61) and (59) only need to be performed once at the beginning of Algorithm 4. The overall computational complexity in each iteration is $\mathcal{O}(M^2\sum_{i=1}^{G}K_i)$. Comparing this with that for the SINR-based approach above, we see that the SLR-based approach has lower computational complexity in each iteration, leading to faster computation of a solution. Furthermore, in Section VII, we will show in our simulation study that the SLR-based approach also results in faster convergence than the SINR-based approach.

VI. Robust Multicast Beamforming

Given the antenna selection vector $\mathbf{a}$ obtained in Phase 1, in Phase 2, we find the robust beamforming solution for the selected antennas. Denote the channel covariance matrix of the selected antennas by $\mathbf{R}_{ik}^{\text{s}}\in\mathbb{C}^{L\times L}$; it is obtained by selecting the columns and rows from $\mathbf{R}_{ik}$, for which $a_m=1$. Similarly, denote $\mathbf{W}^{\text{s}}\in\mathbb{C}^{L\times G}$ as the beamforming matrix of selected antennas, which corresponds to the L selected rows of $\mathbf{W}$, for which $a_m=1$. Then, removing the constraints (9b)–(9d) on $\mathbf{a}$ from $\mathcal{P}_1$, we have the following equivalent robust multicast beamforming optimization problem:

$$\mathcal{P}_4:\min_{\mathbf{W}^{\text{s}}}\ \text{tr}\left(\mathbf{W}^{\text{s}}\mathbf{W}^{\text{s}H}\right)$$
$$\text{s.t. tr}\left(\mathbf{R}_{ik}^{\text{s}}\mathbf{W}^{\text{s}}\mathbf{D}_{ik}\mathbf{W}^{\text{s}H}\right)-\epsilon_{ik}\|\mathbf{W}^{\text{s}}\mathbf{D}_{ik}\mathbf{W}^{\text{s}H}\|_F\geq\sigma^2\gamma_{ik},\forall i,k.$$

Given the selected antennas, the above problem is the same as the regular robust multicast beamforming optimization problem considered in [22], for which an ADMM-based fast algorithm has been developed to solve $\mathcal{P}_4$. Thus, we can directly apply the algorithm proposed in [22] to solve $\mathcal{P}_4$. We omit the algorithm here and refer the readers to [22] for the details.

1) Discussion on the Convergence of Proposed Algorithms: Problems $\mathcal{P}_3^{\text{SINR}}$ and $\mathcal{P}_3^{\text{SLR}}$ in Phase 1 and $\mathcal{P}_4$ in Phase 2 are non-convex problems. The convergence results of ADMM for the non-convex problems are rather limited in the literature. Typically numerical results are used to study the convergence of ADMM and demonstrate its effectiveness for a specific optimization problem. Nonetheless, a recent work [38] has shown the convergence of ADMM to a wide range of non-convex problems. Based on this result, we have proved in [22, Proposition 1] that the convergence of the proposed ADMM-based algorithm to a stationary point of the robust multicast beamforming problem is guaranteed. That is, the algorithm to solve $\mathcal{P}_4$ is guaranteed to converge to a stationary point.

Furthermore, recall that problems $\mathcal{P}_3^{\text{SINR}}$, $\mathcal{P}_3^{\text{SLR}}$, and $\mathcal{P}_{\mathbf{w}_i}$ are converted into equivalent problems (13), (48), and (54), and then solved by Algorithms 1, 4, and 5, respectively. The structures of problems (13), (48), and (54) are similar to the problem considered in [22, Proposition 1] (both the objective function and the equality constraint)[7]. Also, two ADMM blocks are constructed in Algorithms 1, 4, and 5. Thus, the outer-layer ADMM in Algorithms 1 and 4 and the inner-layer ADMM in Algorithm 5 are also guaranteed to converge to a stationary point of their respective problems, by applying the convergence proof similar to that in [22, Proposition 1]. Specifically, the outer-layer ADMM in Algorithm 1 is guaranteed to converge to a stationary point of $\mathcal{P}_3^{\text{SINR}}$, assuming the subproblems are solved. Algorithm 2 uses three ADMM blocks to compute the solution to the non-convex subproblem, where the convergence guarantee is more challenging to show theoretically. Instead, we resort to the simulation study in Section VII to show the convergence of this algorithm and the overall convergence of Algorithm 1. For Algorithm 4, although both its outer-layer ADMM and the inner-layer ADMM (Algorithm 5) are guaranteed to converge, Algorithm 5 is only guaranteed to solve the subproblem approximately but may not be exactly. This is considered an inexact ADMM in the literature [37], [42], [43],

[6]Note that $\mathbf{D}_{ik}$ is a diagonal matrix. We use this diagonal structure in calculating the required matrix multiplication operation in (27), (29), and (30). Also, the diagonal matrix inversion in (30) only needs to be computed once at the beginning of Algorithm 1.

[7]For example, in problem (13), the objective function can be written as $h(\mathbf{W})+f_o(\mathbf{a},\mathbf{Y},\mathbf{W})$, where $h(\mathbf{W})=\text{tr}(\mathbf{W}\mathbf{W}^h)$ is Lipschitz differentiable and $f_o(\mathbf{a},\mathbf{Y},\mathbf{W})=\zeta|\mathbf{a}^T\mathbf{a}-L|+\mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y})$ is non-smooth. This form is similar to that in [22]. Thus, using the argument similar to [22, Proposition 1], we can verify h, f_o, and the equality constraint satisfy the assumptions for algorithm convergence.

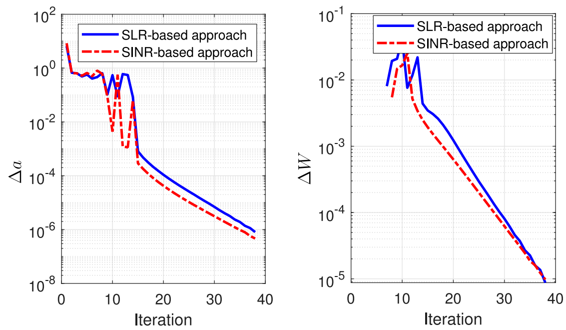


Fig. 1. The convergence behavior of Algorithms 1 and 4. Left: relative residuals difference Δa; right: relative residuals difference ΔW.

[44]. The existing studies show that inexact ADMM may still converge provided certain conditions are satisfied, such as a summable suboptimality gap for the subproblems over iterations [42] or convexity assumptions of the objective function [43], [44]. Note that in our problem, the objective function is smooth and strongly convex, and the constraint functions are smooth, which are often required for convergence. Thus, we expect that such favorable conditions lead to the overall convergence in Algorithm 4. However, with the limited convergence results on inexact ADMM in the literature, we resort to the simulation study in Section VII to show the overall convergence of Algorithm 4.

VII. Simulation Results

We consider a downlink multicast beamforming scenario with G groups and $K_i = K, \forall\, i \in \mathcal{G}$, and $\gamma_{ik} = \gamma, \forall\, k \in \mathcal{K}_i, i \in \mathcal{G}$. Unless otherwise stated, the default setup is $G = 3,\ K = 4,\ M = 100,\ L = 10,\ \text{and}\ \gamma = 3$ dB. We set noise variance to $\sigma^2 = 1$.

For the error bound in (4), we express it as $\epsilon_{ik} = \eta\|\bar{\mathbf{R}}_{ik}\|_F$, where η is the normalized error bound. We use η to study the robust beamforming performance. For modeling the channel covariance matrix, consider the channel vector $\mathbf{h}_{ik}$ between user $k \in \mathcal{K}_i, i \in \mathcal{G}$ and BS, and the estimated channel $\hat{\mathbf{h}}_{ik} = \mathbf{h}_{ik} + \mathbf{e}_{ik}$, where $\mathbf{e}_{ik}$ is the estimation error. We model the channel covariance matrix and estimated channel covariance matrix as $\bar{\mathbf{R}}_{ik} = \mathbf{h}_{ik}\mathbf{h}_{ik}^H$ and $\mathbf{R}_{ik} = \hat{\mathbf{h}}_{ik}\hat{\mathbf{h}}_{ik}^H$, respectively. The error matrix is obtained as $\mathbf{E}_{ik} = \bar{\mathbf{R}}_{ik} - \mathbf{R}_{ik}$. We generate each channel vector independently as $\mathbf{h}_{ik} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and the error vectors independently as $\mathbf{e}_{ik} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$, where σ_e^2 is the corresponding variance. To model the bounded error in (4) more realistically, we obtain $\mathbf{E}_{ik}$ as the above. We then set η as a cut-off bound such that $\|\mathbf{E}_{ik}\|_F \le \eta\|\bar{\mathbf{R}}_{ik}\|_F = \epsilon_{ik}$ holds for 90% of realizations of $\mathbf{e}_{ik}$. Finally, those realizations with $\mathbf{E}_{ik}$ satisfying (4) are considered in the simulation. Unless otherwise stated, we set the default value to $\eta = 0.01$. In our simulation, we set the values of the penalty parameters ρ, $\tilde{\rho}$, $\bar{\rho}$ and $\hat{\rho}$ used in Algorithms 1, 2, 4, and 5 within the range $[0.1, 0.4]$.[8]

A. Convergence Behavior of Algorithms for Phase 1

For the proposed SINR-based and SLR-based approaches for antenna selection in Phase 1, we study the convergence behavior of the ADMM-based algorithm developed for each approach.

For the SINR-based approach, Algorithm 1 provides the outer-layer ADMM procedure for solving $\mathcal{P}_3^{\text{SINR}}$, in which the update for $\mathbf{W}$ in (15) is obtained by the inner-layer ADMM algorithm in Algorithm 2. Similarly, for the SLR-based approach, Algorithm 4 provides the outer-layer ADMM procedure for solving $\mathcal{P}_3^{\text{SLR}}$, where the update for $\mathbf{W}$ in (49) is obtained by the inner-layer ADMM algorithm in Algorithm 5. We first consider the outer-layer ADMM-based algorithm in each approach. We consider both residual convergence and objective convergence. Define the relative residual difference for $\mathbf{a}$ between iterations $l-1$ and l as $\Delta a \triangleq \frac{\|\mathbf{a}^l - \mathbf{a}^{l-1}\|}{\|\mathbf{a}^l\|}$, and the relative residual for the equality constraint in (12d) as $\Delta W \triangleq \frac{\|\mathbf{W}^l - \mathbf{Y}^{l-1}\|_F}{\|\mathbf{W}^l\|_F}$ (for the equality constraint in (47d) as $\Delta W \triangleq \frac{\|\mathbf{W}^l - \overline{\mathbf{Y}}^{l-1}\|_F}{\|\mathbf{W}^l\|_F}$). Fig. 1 shows the trajectories of Δa and ΔW over iterations for $\mathbf{a}$ and $\mathbf{W}$, respectively, by both SINR-based and SLR-based approaches. We see that the convergence behavior for both approaches are similar: the values of Δa and ΔW drop below 10^{-3} after 20 iterations, indicating fast convergence.

Next, we study the convergence of the inner-layer ADMM-based algorithm to update $\mathbf{W}$ in each approach: Algorithm 2 for $\mathcal{P}_3^{\text{SINR}}$ and Algorithm 5 for $\mathcal{P}_3^{\text{SLR}}$. For the SINR-based approach, Algorithm 2 iterates among three update blocks $\mathbf{W}$, $\{\mathbf{W}_{ik}\}$, and $\{\mathbf{V}_{ik}\}$. Define the relative residuals in iteration j for the equality constraints in (19b) and (19c) as $\Delta \overline{W} \triangleq \max_{i,k}\{\frac{\|\mathbf{W}_{ik}^j - \mathbf{W}^j\|_F}{\|\mathbf{W}_{ik}^j\|_F}\}$ and $\Delta V = \max_{i,k}\{\frac{\|\mathbf{V}_{ik}^j - \mathbf{W}^j\mathbf{D}_{ik}\|_F}{\|\mathbf{V}_{ik}^j\|_F}\}$, respectively. For the SLR-based approach, Algorithm 5 for $\mathcal{P}_3^{\text{SLR}}$ has two update blocks $\mathbf{w}_i$ and $\{\mathbf{w}_{ik}\}$ (with subscript i reinstated). Define the relative residual for the equality constraint in (53c) as $\Delta w_i = \max_k\{\frac{\|\mathbf{w}_{ik}^j - \mathbf{w}_i^j\|}{\|\mathbf{w}_{ik}^j\|}\}$. Fig. 2 plots the trajectories of $\Delta\overline{W}$, ΔV, and Δw_i over iterations for $M = 50, 100, 200$, at the first outer-layer iteration. For Δw_i, we select $i \in \mathcal{G}$ with the slowest convergence to show. We see that the values of $\Delta\overline{W}$ and ΔV in Algorithm 2 drop below 10^{-3} within about 200 iterations for all value of M, and the value of Δw_i in Algorithm 5 drops below 10^{-3} after about 100 iterations. Thus, we see that the convergence speed of Algorithm 5 is noticeably faster than that of Algorithm 2.[9] As analyzed in Section V-C, the computational complexity of Algorithm 5 is lower than that of Algorithm 2 per ADMM iteration. Thus, overall the SLR-based approach has a faster algorithm with much lower complexity than the SINR-based approach.

B. Impact of Approximated Approach in Phase 1

To determine antenna selection in Phase 1, we use the approximated problem $\mathcal{P}_3^{\text{SINR}}$, where we ignore the second term of the constraint (11a) in $\mathcal{P}_3$. To study the effect of such approximation as compared to the original problem $\mathcal{P}_3$, we solve $\mathcal{P}_3$ directly

[8] Our experiments show that for different system setups for M and GK, the values of the penalty parameters in this range generally result in good convergence rates.

[9] Our experimental studies also show that the convergence speed of Δw_i in Algorithm 5 becomes faster at the later outer-layer iterations, while that of Algorithm 2 is roughly unchanged.

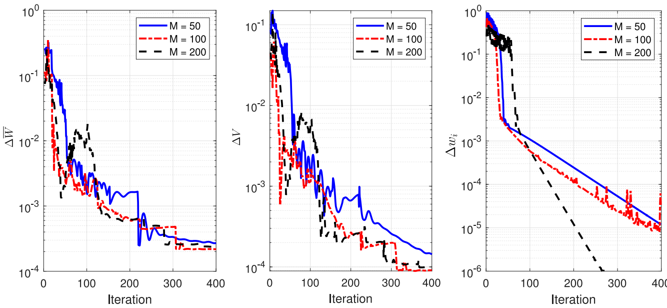


Fig. 2. The convergence behavior of the inner-layer ADMM. Left: $\Delta\overline{W}$ in Algorithm 2 (SINR-based approach); middle: ΔV in Algorithm 2 (SINR-based approach); right: Δw_i in Algorithm 5 (SLR-based approach).

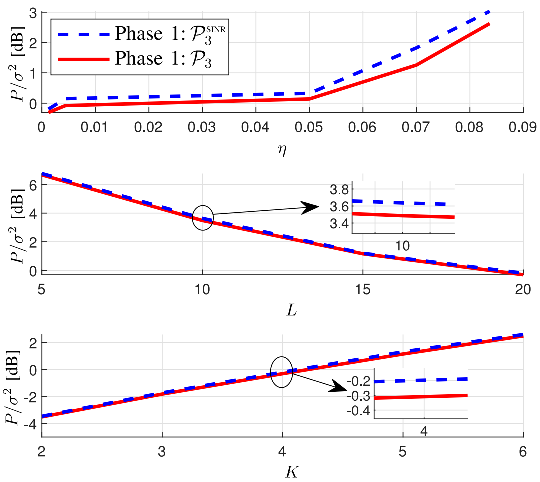


Fig. 3. Top: normalized transmit power $\frac{P}{\sigma^2}$ vs. η; middle: normalized transmit power $\frac{P}{\sigma^2}$ vs. L; bottom: normalized transmit power $\frac{P}{\sigma^2}$ vs. K.

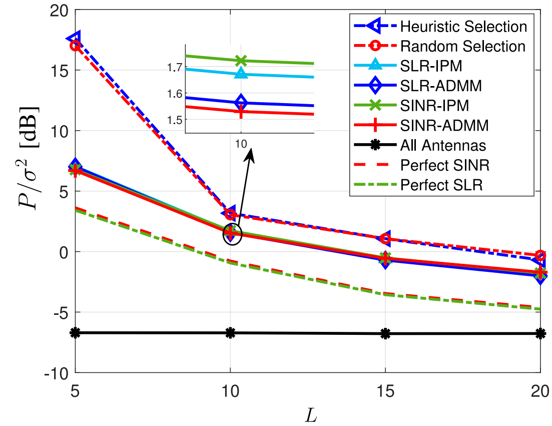


Fig. 4. Normalized transmit power $\frac{P}{\sigma^2}$ vs. L ($M = 100$, $G = 3$, $K = 4$, $\gamma = 3$ dB, $\eta = 0.01$).

TABLE I
AVERAGE COMPUTATION TIME OVER L (SEC)

L	5	10	15	20
SINR-ADMM	98	100	102	104
SLR-ADMM	22.3	25	27.8	30.6
SINR-IPM	10178	12119	12396	13016
SLR-IPM	6069.5	6983.1	7754.7	7906.7

via the IPM to obtain $\mathbf{a}$ and then round its elements to the nearest 0 or 1 for binary selection in Phase 1. Phase 2 is carried the same way as in Section VI. We use the genetic nonlinear program solver `fmincon` in MATLAB for implementing the IPM.[10] We search the value of ζ from 0.005 to 2 and select the best one that leads to the lowest transmit power. Fig. 3-top shows the normalized transmit power P/σ^2 vs. the normalized error bound η for channel covariance. We see that using the approximate version $\mathcal{P}_3^{\text{SINR}}$ for antenna selection results in less than 0.3 dB performance gap to that of directly solving $\mathcal{P}_3$. Fig. 3-middle and Fig. 3-bottom show P/σ^2 vs. L and K, respectively. The performance of our SINR-based approach using $\mathcal{P}_3^{\text{SINR}}$ is nearly identical to that of using $\mathcal{P}_3$ directly in Phase 1. The results indicate that the approximation in our proposed approach for antenna selection causes negligible impact on the overall performance but with significantly less computational complexity, as we will see in the next subsection.

C. Performance Comparison

We now evaluate the performance of our proposed algorithms. For comparison, we consider the following methods for antenna selection in Phase 1: **i) SINR-ADMM and SLR-ADMM**: our proposed SINR-based approach using Algorithm 1 for $\mathcal{P}_3^{\text{SINR}}$ and SLR-based approach using Algorithm 4 for $\mathcal{P}_3^{\text{SLR}}$; **ii) SINR-IPM and SLR-IPM**: solving $\mathcal{P}_3^{\text{SINR}}$ and $\mathcal{P}_3^{\text{SLR}}$ using IPM via the generic nonlinear program solver `fmincon` in MATLAB; **iii) Heuristic selection**: antenna selection based on the average channel strength, where we compute the average channel power gain from each antenna m to all users: $\bar{g}_m = \frac{1}{\sum_{i=1}^{G} K_i} \sum_{i=1}^{G} \sum_{k=1}^{K_i} [\mathbf{R}_{ik}]_{mm}$, and select the antennas with the L largest gains;[11] **iv) Random search-100**: we perform 100 trials of random selection and choose the best selection among them that gives the lowest transmit power; **v) All antennas**: all M antennas are used for multicast beamforming, *i.e.,* only Phase 2 is performed. This benchmark provides a lower bound for all antenna selection methods. Finally, we also consider the performance benchmark under perfect channel knowledge: **vi) Perfect SINR and perfect SLR**: the proposed two-phase approach under perfect knowledge of channel covariance, *i.e.,* $\epsilon_{ik} = 0, \forall k, i$, where Algorithm 1 (SINR) or Algorithm 4 (SLR) is used for Phase 1.

1) Performance vs. Number of Selected Antennas L: Fig. 4 shows the normalized transmit power P/σ^2 vs. the number of selected antennas L for the default setup. As expected, the transmit power decreases as L increases for all the antenna selection methods. Our proposed two-phased solutions (SINR-ADMM and SLR-ADMM) outperforms the heuristic selection and random search-100. The performance gap becomes significant when L becomes smaller, as having a good antenna

[10]We resort to the generic IPM as there is no other existing algorithm to solve the problem.

[11]This heuristic method is based on the intuition from the multiple-input single-output (MISO) channel in the single-user case, where the antennas with the L highest channel power gains are the best choices.

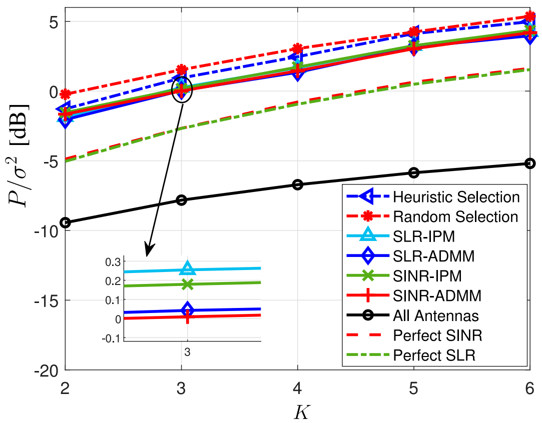


Fig. 5. Normalized transmit power $\frac{P}{\sigma^2}$ vs. K ($M = 100$, $G = 3$, $L = 10$, $\gamma = 3$ dB, $\eta = 0.01$).

TABLE II
AVERAGE COMPUTATION TIME OVER K (SEC)

K	2	3	4	5	6
SINR-ADMM	46	70	100	110	124
SLR-ADMM	11.5	17.5	25	27.5	31
SINR-IPM	6314.9	9566.8	12119	14769	16869
SLR-IPM	4031	5588.3	6983.1	9372.7	11177

selection becomes critical when only a few antennas are selected. Furthermore, comparing the two proposed approaches for antenna selection in Phase 1, from the zoom-in plot, we see the SLR-ADMM shows a negligible loss as compared to SINR-ADMM. Comparing the two approaches with their IPM counterpart, we see that the performance by using our ADMM-based algorithms is nearly identical to that produced by IPM.

The computational advantage of our proposed ADMM-based algorithms is shown in Table I, where we provide the average computation time for the SINR-based and SLR-based approaches shown in Fig. 4. Our proposed ADMM-based algorithms are significantly faster in computing a solution than IPM for both of the SINR-based and SLR-based approaches. Furthermore, SLR-ADMM has lower computational complexity than SINR-ADMM for all L values, with nearly identical performance. This demonstrates the effectiveness of our proposed SLR-based approach.

2) Performance vs. Number of Users K: Fig. 5 shows the normalized transmit power $\frac{P}{\sigma^2}$ vs. the number of users per group K for the default setup. Our proposed two-phased solutions outperform the heuristic selection and random search-100 for all values of K considered. Also, the performance of SLR-based and SINR-based approaches is nearly identical. The performance of our proposed SINR-ADMM and SLR-ADMM is close but slightly better than SINR-IPM and SLR-IPM. Table II shows the computational advantage of our proposed ADMM-based fast algorithms for different values of K. We see that our ADMM-based algorithms are substantially faster than IPM, and the SLR-based approach provides lower complexity than the SINR-based approach, with similar performance.

VIII. CONCLUSION

In this work, we have considered joint antenna selection and robust multi-group multicast beamforming design in massive MIMO systems with a limited number of RF chains. Our design is based on the estimated channel covariance matrices available at the BS with a given error bound. Our design is to ensure that the worst-case SINR meets the target at each user. We reformulate this challenging joint robust optimization problem into a more tractable form, for which we proposed a two-phase approach with an antenna selection phase, followed by multicast beamforming optimization. For the antenna selection phase, we have proposed the SINR-based approach and developed a two-layered ADMM-based fast algorithm to solve the joint optimization problem. In particular, our algorithm contains (semi-)closed-form solutions in each iteration update. To further lower the complexity, we have proposed the SLR-based approach for antenna selection and similarly developed a two-layered ADMM-based fast algorithm with closed-form updates in each iteration. With the selected antennas, we can apply the fast algorithm in [22] for robust multicast beamforming design. The simulation results show that our proposed approaches outperform other antenna selection methods, and our ADMM-based algorithms provide fast computation of both antenna selection and beamforming solution.

APPENDIX A
THE SOLUTION TO PROBLEM (58)

There are two possible solutions:
i) $\hat{\mathbf{w}}_k^\circ = \mathbf{w} + \hat{\mathbf{z}}_k$ is the solution if it satisfies constraint (58).
ii) Otherwise, the constraint in (58) holds with equality. Let $b_k \triangleq \hat{\gamma}_k(\mathbf{w}^H\bar{\mathbf{R}}\mathbf{w} + \sigma^2)$. The Lagrangian of problem (58) is

$$\mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k) = \|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 + \hat{\lambda}_k(-\hat{\mathbf{w}}_k^H\mathbf{R}_k\hat{\mathbf{w}}_k + b_k)$$

where $\hat{\lambda}_k$ is Lagrange multiplier associated with the constraint in (58). Set the gradient of $\mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k)$ to 0, we have

$$\nabla_{\hat{\mathbf{w}}_k^*}\mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k) = -\mathbf{w} + \hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k - \hat{\lambda}_k\mathbf{R}_k\hat{\mathbf{w}}_k = 0, \quad (60)$$

which leads to

$$\hat{\mathbf{w}}_k = (\hat{\lambda}_k\mathbf{R}_k + \mathbf{I})^{-1}(\mathbf{w} + \hat{\mathbf{z}}_k). \quad (61)$$

Since $-\hat{\mathbf{w}}_k^H\mathbf{R}_k\hat{\mathbf{w}}_k + b_k = 0$, we obtain $\hat{\lambda}_k$ as the solution of the following equality:

$$(\mathbf{w} + \hat{\mathbf{z}}_k)^H(\hat{\lambda}_k\mathbf{R}_k + \mathbf{I})^{-1}\mathbf{R}_k(\hat{\lambda}_k\mathbf{R}_k + \mathbf{I})^{-1}(\mathbf{w} + \hat{\mathbf{z}}_k) = b_k.$$

Consider eigenvalue decomposition $\mathbf{R}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{U}_k^H$, where $\mathbf{\Sigma}_k \triangleq \text{diag}(\tilde{\sigma}_{1k}^2, \ldots, \tilde{\sigma}_{Mk}^2)$. Then, the above equation is equivalent to $\sum_{m=1}^M \frac{\tilde{\sigma}_{mk}^2|\beta_{mk}|^2}{(1+\hat{\lambda}_k\tilde{\sigma}_{mk}^2)^2} = b_k$, where $\mathbf{U}_k^H(\mathbf{w} + \hat{\mathbf{z}}_k) \triangleq [\beta_{1k}, \ldots, \beta_{Mk}]^T$. Since $\hat{\lambda}_k \geq 0$, the expression at the LHS of this equation is a decreasing function of $\hat{\lambda}_k$. Thus, the solution $\hat{\lambda}_k$ can be found via bi-section search. Since the solution is unique, it is the optimal solution to problem (58).

Appendix B
The Solution to Problem (59)

There are two possible solutions:

i) $\mathbf{w}^\circ = \frac{1}{1+\rho+K\bar{\rho}}(\bar{\rho}\sum_{k=1}^{K}(\hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k) + \rho(\bar{\mathbf{y}} - \bar{\mathbf{z}}_k))$ is the solution if it satisfies the constraint in (59).

ii) Otherwise, the constraint in (59) holds with equality. We solve it by KKT conditions. Define $d \triangleq \min_k \frac{1}{\bar{\gamma}_k}\hat{\mathbf{w}}_k^H\mathbf{R}_k\hat{\mathbf{w}}_k - \sigma^2$. The Lagrangian for problem (59) is $\mathcal{L}(\mathbf{w},\bar{\lambda}) = \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \hat{\rho}\sum_{k=1}^{K}\|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 + \bar{\lambda}(\mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} - d)$, where $\bar{\lambda} \geq 0$ is the Lagrange multiplier for the constraint in (59). Set the gradient $\nabla_{\hat{\mathbf{w}}^*}\mathcal{L}(\mathbf{w},\bar{\lambda}) = \left(\bar{\lambda}\overline{\mathbf{R}} + (1+\bar{\rho}+K\hat{\rho})\mathbf{I}\right)\mathbf{w} - \mathbf{c} = 0$, where $\mathbf{c} \triangleq \hat{\rho}\sum_{k=1}^{K}(\hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k) + \bar{\rho}(\bar{\mathbf{y}} - \bar{\mathbf{z}})$. This leads to

$$\mathbf{w} = (\bar{\lambda}\overline{\mathbf{R}} + (1+\bar{\rho}+K\hat{\rho})\mathbf{I})^{-1}\mathbf{c}. \tag{62}$$

Since $\mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} - d = 0$, we obtain $\bar{\lambda}$ as the solution of the following inequality.

$$\mathbf{c}^H(\bar{\lambda}\overline{\mathbf{R}} + (1+\bar{\rho}+K\hat{\rho})\mathbf{I})^{-1}\overline{\mathbf{R}}(\bar{\lambda}\overline{\mathbf{R}} + (1+\bar{\rho}+K\hat{\rho})\mathbf{I})^{-1}\mathbf{c} = b.$$

Again, applying eigenvalue decomposition on $\overline{\mathbf{R}} = \overline{\mathbf{U}}\,\overline{\mathbf{\Sigma}}\,\overline{\mathbf{U}}^H$ with $\overline{\mathbf{\Sigma}} \triangleq \mathrm{diag}(\bar{\sigma}_1^2,\ldots,\bar{\sigma}_M^2)$, we can rewrite the above equation as

$$\sum_{m=1}^{M}\frac{\bar{\sigma}_m^2|\bar{c}_m|^2}{(1+\rho+K\bar{\rho}+\bar{\lambda}\bar{\sigma}_m^2)^2} = b$$

where $[\bar{c}_1,\ldots,\bar{c}_M]^T \triangleq \bar{\mathbf{U}}^H\mathbf{c}$. The solution $\bar{\lambda} \geq 0$ to this equation can be found via bi-section search.

References

[1] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, 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, no. 3, pp. 1268–1279, Mar. 2008.

[3] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Oct. 2014.

[4] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, Jan. 2013.

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

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

[7] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2685–2698, Jun. 2017.

[8] 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, no. 5, pp. 2963–2975, May 2017.

[9] J. Yu and M. Dong, "Low-complexity weighted MRT multicast beamforming in massive MIMO cellular networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3849–3853.

[10] J. Yu and M. Dong, "Distributed low-complexity multi-cell coordinated multicast beamforming with large-scale antennas," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2018.

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

[12] S. Mohammadi, M. Dong, and S. ShahbazPanahi, "Fast algorithm for joint unicast and multicast beamforming for large-scale massive MIMO," *IEEE Trans. Signal Process.*, vol. 70, pp. 5413–5428, 2022.

[13] C. Zhang, M. Dong, and B. Liang, "Fast first-order algorithm for large-scale max-min fair multi-group multicast beamforming," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1560–1564, Aug. 2022.

[14] C. Zhang, M. Dong, and B. Liang, "Ultra-low-complexity algorithms with structurally optimal multi-group multicast beamforming in large-scale systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 1626–1641, 2023.

[15] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[16] O. Tervo, L.-N. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, Sep. 2018.

[17] O. T. Demir and T. E. Tuncer, "Antenna selection and hybrid beamforming for simultaneous wireless information and power transfer in multi-group multicasting systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6948–6962, Oct. 2016.

[18] M. S. Ibrahim, A. Konar, and N. D. Sidiropoulos, "Fast algorithms for joint multicast beamforming and antenna selection in massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 1897–1909, 2020.

[19] Y. Huang, Q. Li, W.-K. Ma, and S. Zhang, "Robust multicast beamforming for spectrum sharing-based cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 527–533, Jan. 2012.

[20] M.-C. Yue, S. X. Wu, and A. M.-C. So, "A robust design for MISO physical-layer multicasting over line-of-sight channels," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 939–943, Jul. 2016.

[21] G. Zang et al., "Low complexity algorithms for robust multigroup multicast beamforming," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1409–1412, Aug. 2019.

[22] N. Mohamadi, M. Dong, and S. ShahbazPanahi, "Low-complexity ADMM-based algorithm for robust multi-group multicast beamforming in large-scale systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 2046–2061, 2022.

[23] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[24] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[25] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.

[26] A. Decurninge, M. Guillaud, and D. T. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, 2015, pp. 1–6.

[27] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.

[28] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.

[29] A. d'Aspremont and S. Boyd, "Relaxations and randomized methods for nonconvex QCQPs," EE392o Class Notes, vol. 1, Stanford University, Stanford, CA, USA, 2003, pp. 1–16.

[30] L. You, A. Liu, W. Wang, and X. Gao, "Outage constrained robust multigroup multicast beamforming for multi-beam satellite communication systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 352–355, Apr. 2019.

[31] Q.-D. Vu, K.-G. Nguyen, and M. Juntti, "Weighted max–min fairness for C-RAN multicasting under limited fronthaul constraints," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1534–1548, Apr. 2018.

[32] Y. Chen, S. He, Y. Huang, J. Ren, and L. Yang, "Robust multigroup multicast beamforming design for backhaul-limited cloud radio access network," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 189–193, Jan. 2019.

[33] M. Bengtsson and B. Ottersten, “Optimal and suboptimal transmit beamforming,” in *Handbook of Antennas in Wireless Communications,* Lal C. Godara, Ed., Boca Raton, FL, USA: CRC Press, 2001, pp. 18-1–18-33.

[34] K. L. Law, I. Wajid, and M. Pesavento, “Robust downlink beamforming in multi-group multicasting using trace bounds on the covariance mismatches,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2012, pp. 3229–3232.

[35] W. Zhang, Z. Chen, Q. Gong, and G. Wei, “Robust multicast beamforming using covariance channel state information,” in *Proc. IEEE Int. Conf. Commun. China (ICCC)*, 2012, pp. 636–641.

[36] K. Huang and N. D. Sidiropoulos, “Consensus-ADMM for general quadratically constrained quadratic programming,” *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, Oct. 2016.

[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[38] Y. Wang, W. Yin, and J. Zeng, “Global convergence of ADMM in nonconvex nonsmooth optimization,” *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, 2019.

[39] A. F. Izmailov and M. V. Solodov, “Karush-Kuhn-Tucker systems: Regularity conditions, error bounds and a class of newton-type methods,” *Math. Program.*, vol. 95, no. 3, pp. 631–650, 2003.

[40] E. Björnson, M. Bengtsson, and B. Ottersten, “Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [Lecture Notes],” *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

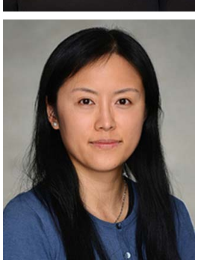
[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[42] J. Eckstein and D. P. Bertsekas, “On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Math. Program.*, vol. 55, pp. 293–318, Apr. 1992.

[43] T.-H. Chang, M. Hong, and X. Wang, “Multi-agent distributed optimization via inexact consensus ADMM,” *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 482–497, Jan. 2015.

[44] J. Bai, W. W. Hager, and H. Zhang, “An inexact accelerated stochastic ADMM for separable convex optimization,” *Comput. Optim. Appl.*, vol. 81, pp. 479–518, Jan. 2022.

Niloofar Mohamadi received the B.Eng. and M.A.Sc. degrees in electrical engineering from the University of Kurdistan, Sanandaj, Iran, in 2014 and 2017, respectively. She received her Ph.D. degree in electrical engineering from Ontario Tech University, Canada in 2023. Her research interests include statistical signal processing, massive multi-input multi-output communication system, with emphasis on 5G technology.



Min Dong (Fellow, IEEE) received the B.Eng. degree from Tsinghua University, Beijing, China, in 1998, and the Ph.D. degree in electrical and computer engineering with a minor in applied mathematics from Cornell University, Ithaca, NY, USA, in 2004. From 2004 to 2008, she was with Qualcomm Research, Qualcomm Inc., San Diego, CA, USA. Since 2008, she has been with Ontario Tech University, where she is currently a Professor with the Department of Electrical, Computer, and Software Engineering. Her research interests include wireless communications, statistical signal processing, learning techniques, optimization, and control applications in cyber-physical systems. She received the Early Researcher Award from the Ontario Ministry of Research and Innovation in 2012, the Best Paper Award at IEEE ICCC in 2012, and the 2004 IEEE Signal Processing Society Best Paper Award. She is a Co-Author of Best Student Paper at IEEE SPAWC 2021 and Best Student Paper of Signal Processing for Communications and Networking at IEEE ICASSP 2016. She is an Associate Editor for IEEE OPEN JOURNAL OF SIGNAL PROCESSING. She served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2018–2023), and as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING (2010–2014) and IEEE SIGNAL PROCESSING LETTERS (2009–2013). She served on the Steering Committee of IEEE TRANSACTIONS ON MOBILE COMPUTING (2019–2021). She was an Elected Member of the Signal Processing for Communications and Networking (SP-COM) Technical Committee of IEEE Signal Processing Society (2013–2018). She was the Lead Co-Chair of the Communications and Networks to Enable the Smart Grid Symposium at the IEEE International Conference on Smart Grid Communications in 2014.

Shahram ShahbazPanahi (Senior Member, IEEE) was born in Sanandaj, Kurdistan, Iran. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1992, 1994, and 2001, respectively. From September 1994 to September 1996, he was an Instructor with the Department of Electrical Engineering, Razi University, Kermanshah, Iran. From July 2001 to March 2003, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. From April 2003 to September 2004, he was a Visiting Researcher with the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. From September 2004 to April 2005, he was a Lecturer and an Adjunct Professor with the Department of Electrical and Computer Engineering, McMaster University. In July 2005, he joined the Faculty of Engineering and Applied Science, Ontario Tech University, Oshawa, ON, Canada, where he currently holds a Professor position. His research interests include statistical and array signal processing; space-time adaptive processing; detection and estimation; multiantenna, multiuser, and cooperative communications; spread spectrum techniques; DSP programming; and hardware/real-time software design for telecommunication systems. He has served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and for IEEE SIGNAL PROCESSING LETTERS. He has also served as a Senior Area Editor for IEEE SIGNAL PROCESSING LETTERS. He was an Elected Member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He has received several awards, including the Early Researcher Award from Ontario's Ministry of Research and Innovation, the NSERC Discovery Grant (three awards), the Research Excellence Award from the Faculty of Engineering and Applied Science, the University of Ontario Institute of Technology, and the Research Excellence Award, Early Stage, from Ontario Tech University. In 2019, he received an Inter-Mobility Award from the Luxembourg National Research Fund. With the support of this award, he held a Visiting Professor position at the Interdisciplinary Centre for Security, Reliability and Trust (SnT) of the University of Luxembourg.