Robust Group Sparse Beamforming for Multicast Green Cloud-RAN With Imperfect CSI

Yuanming Shi, *Student Member, IEEE*, Jun Zhang, *Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

Abstract—In this paper, we investigate the network power minimization problem for the multicast cloud radio access network (Cloud-RAN) with imperfect channel state information (CSI). The key observation is that network power minimization can be achieved by adaptively selecting active remote radio heads (RRHs) via controlling the group-sparsity structure of the beamforming vector. However, this yields a non-convex combinatorial optimization problem, for which we propose a three-stage robust group sparse beamforming algorithm. In the first stage, a quadratic variational formulation of the weighted mixed ℓ_1/ℓ_2-norm is proposed to induce the group-sparsity structure in the aggregated beamforming vector, which indicates those RRHs that can be switched off. A perturbed alternating optimization algorithm is then proposed to solve the resultant non-convex group-sparsity inducing optimization problem by exploiting its convex substructures. In the second stage, we propose a *PhaseLift* technique based algorithm to solve the feasibility problem with a given active RRH set, which helps determine the active RRHs. Finally, the semidefinite relaxation (SDR) technique is adopted to determine the robust multicast beamformers. Simulation results will demonstrate the convergence of the perturbed alternating optimization algorithm, as well as, the effectiveness of the proposed algorithm to minimize the network power consumption for multicast Cloud-RAN.

Index Terms—Cloud-RAN, multicast beamforming, green communications, group-sparsity, robust optimization, alternating optimization, PhaseLift, semidefinite relaxation.

I. INTRODUCTION

NETWORK densification has been recognized as an effective way to meet the exponentially growing mobile data traffic and to accommodate increasingly diversified mobile applications. Cooperative transmission/reception among multiple base stations is a well-known approach to improve the spectral efficiency and energy efficiency of dense wireless networks [1]–[3], which is driving the development of novel collaborative architectures for cellular networks. Cloud radio access networks (Cloud-RAN) [4]–[6] have recently been proposed as a cost-effective and flexible way to exploit the cooperation gains by moving the baseband units (BBUs) into a single cloud data center, i.e., forming a BBU pool with powerful shared computing resources. As a result, with efficient hardware utilization at the BBU pool, both the CAPEX (e.g., via low-cost site construction) and OPEX (e.g., via centralized cooling) can be reduced significantly. Furthermore, the conventional base stations are replaced by the light and low-cost remote radio heads (RRHs) with basic functionalities of signal transmission and reception, which are then connected to the BBU pool by high-capacity and low-latency optical fronthaul links. The capacity of Cloud-RAN thus can be significantly improved through network densification and centralized signal processing at the BBU pool.

However, the new architecture of Cloud-RAN also brings new design and operating challenges, e.g., high-capacity and low-latency requirements for the optical fronthaul links [7], virtualization techniques for resource management in the BBU pool [4], and massive CSI acquisition for cooperative interference management [8], [9]. In particular, energy efficiency is an important aspect for operating such a dense wireless network, and it is among the major design objectives for 5G networks [10]. Conventionally, the energy efficiency oriented design only takes into account the transmit power [11] and the circuit power [12] at the base stations. Nevertheless, in such dense collaborative networks as Cloud-RAN, a holistic view is needed when measuring network power consumption, which should also include the power consumption of the additional optical fronthaul links [5]. Observing that the mobile data traffic would vary temporally and spatially, it was proposed in [5] to adaptively switch off some fronthaul links and the corresponding RRHs to minimize the network power consumption, which is achieved by a new beamforming technique, called *group sparse beamforming*.

The effectiveness of group sparse beamforming has been demonstrated in [5], but with certain limitations in the network model, e.g., perfect CSI is assumed at the BBU pool, and only unicast services are considered. In practice, inevitably there will be uncertainty in the available CSI, originating from various sources, e.g., limited feedback [13], channel estimation errors [14], partial CSI acquisition [8], [9] and delay in the obtained CSI [15], [16]. In terms of transmission services from the RRHs, it has been well recognized that the physical layer integration technique [17] can effectively improve the network performance. In particular, the RRHs should not only transmit data to individual users [18] (i.e., broadcast/unicast services) but also integrate additional multicast services [19], where the RRHs transmit a common message in such a way that all the MUs in the same group can decode it. Such multigroup multicast transmission is promising to provide high capacity services and content-aware applications in next generation

wireless networks. For instance, with physical layer caching for wireless video delivery [20], it is common that multiple users are interested in the same video stream, which creates multicast groups.

In this paper, we will thus focus on the design of green Cloud-RAN by jointly minimizing the RRH power consumption and transport link power consumption, considering the practical scenarios with imperfect CSI and multigroup multicast services. We adopt the robust optimization approach to address the CSI uncertainty, such that the QoS requirements are satisfied for *any* realization of the uncertainty in a predefined set [21]. The unique challenges of the network power minimization problem arise from both the infinite number of the non-convex quadratic QoS constraints (due to the robust design criteria and multicast transmission) and the combinatorial composite objective function (due to the consideration of both the relative fronthaul link power consumption and the RRH transmit power consumption).

A. Related Works

1) Robust Multicast Beamforming: Although the integration of multicast, individual services and cooperative transmission can significantly improve the capacity of wireless networks [17], it will bring significant challenges from both the information theoretic [22] and signal processing perspectives [19], [23]. In particular, the physical-layer multicast beamforming problem is in general NP-hard due to the non-convex quadratic QoS constraints [19]. Furthermore, to address the CSI uncertainty, one may either adopt the stochastic optimization formulation [24] or the robust optimization formulation [25]. However, the stochastic optimization formulations often yield highly intractable problems, e.g., the stochastic coordinated beamforming problem based on the chance constrained programming [9]. The worst-case based robust optimization, on the other hand, has the advantage of computational tractability [21]. Although the original robust and/or multicast beamforming design problems may be non-convex due to the infinite number of non-convex quadratic QoS constraints [26], the convex optimization based SDR technique [27] with S-lemma [28] has recently been applied to provide a principled way to develop polynomial time complexity algorithms to find an approximate solution [29].

However, we cannot directly apply such SDR technique to solve the network power minimization problem due to the non-convex combinatorial composite objective function, which represents the network power consumption.

2) Group Sparse Beamforming: The convex sparsity-inducing penalty approach [30] has recently been widely used to develop polynomial time complexity algorithms for the mixed combinatorial optimization problems in wireless networks, e.g., joint base station clustering and transmit beamforming [31], joint antenna [32] or RRH [5] selection and transmit beamforming. The main idea of this approach is that the sparsity pattern of the beamforming vector, which can be induced by minimizing a sparsity penalty function (e.g., the mixed ℓ_1/ℓ_2-norm minimization can induce the group-sparsity), can provide guidelines for, e.g., antenna selection [32], where the antennas with smaller beamforming coefficients (measured

by the ℓ_∞-norm) have a higher priority to be switched off. However, most works only consider the ideal scenario (e.g., perfect CSI and broadcast services [5]), which usually yield convex constraints (e.g., second-order cone constraints [5]).

Unfortunately, we cannot directly adopt the *non-smooth* weighted mixed ℓ_1/ℓ_2-norm developed in [5] to induce the group-sparsity for the robust multicast beamforming vector. This is because the resultant group-sparsity inducing optimization problem will be highly intractable, due to the non-smooth sparsity-inducing objective function and the infinite number of non-convex quadratic QoS constraints.

Based on above discussion and in contrast to the previous work [5] on group sparse beamforming with a non-convex combinatorial composite objective function but convex QoS constraints in the unicast Cloud-RAN, we need to address the following coupled challenges in order to solve the network power minimization problem for multicast green Cloud-RAN with imperfect CSI:

- An infinite number of non-convex quadratic QoS constraints;
- The combinatorial composite objective function.

Thus, to apply the computationally efficient group sparse beamforming approach [5] to more practical scenarios, unique challenges arise. We need to redesign the group-sparsity inducing norm, and then deal with the non-convex group-sparsity inducing optimization problem with an infinite number of non-convex quadratic QoS constraints. We should also develop efficient algorithms for non-convex feasibility problems for the adaptive RRH selection, and for non-convex robust multicast beamforming design after determining the active RRHs.

B. Contributions

In this paper, we provide a convex relaxation based robust group sparse beamforming framework for network power minimization in multicast Cloud-RAN with imperfect CSI. The major contributions are summarized as follows:

1) A group sparse beamforming formulation is proposed to minimize the network power consumption for Cloud-RAN. It will simultaneously control the group-sparsity structure and the magnitude of the beamforming coefficients, thereby minimizing the relative fronthaul link power consumption and the transmit power consumption, respectively. The group sparse beamforming modeling framework lays the foundation for developing the three-stage robust group sparse beamforming algorithm based on the convex relaxation.

2) In the first stage, a novel quadratic variational formulation of the weighted mixed ℓ_1/ℓ_2-norm is proposed to induce the group-sparsity structure for the robust multicast beamforming vector, thereby guiding the RRH selection. The main motivation for such a quadratic form formulation is to make the group-sparsity inducing penalty function compatible with the quadratic QoS constraints. Based on the SDR technique, a perturbed alternating optimization algorithm with convergence guarantee is then proposed to solve the resultant non-convex quadratic form group-sparsity inducing optimization problem by exploiting its convex substructures.

3) In the second stage, a *PhaseLift* approach based algorithm is proposed to solve the non-convex feasibility problems, based on which the active RRHs can be determined with a binary search. Finally, the SDR technique is adopted to solve the non-convex robust multicast beamforming optimization problem to determine the transmit beamformers for the active RRHs.

4) Simulation results will demonstrate the effectiveness of the proposed robust group sparse beamforming algorithm to minimize the network power consumption.

C. Organization

The remainder of the paper is organized as follows. Section II presents the system model and problem formulation, followed by the problem analysis. In Section III, the group sparse beamforming modeling framework is proposed to formulate the network power minimization problem. The semidefinite programming (SDP) based robust group sparse beamforming algorithm is developed in Section IV. Simulation results will be illustrated in Section V. Finally, conclusions and discussions are presented in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider a multicast Cloud-RAN with L RRHs and K single-antenna mobile users (MUs), where the l-th RRH is equipped with N_l antennas, as shown in Fig. 1. The centralized signal processing is performed at the baseband unit (BBU) pool [4], [5]. Define $\mathcal{S} = \{1, \ldots, K\}$ as the set of all the MUs and $\mathcal{L} = \{1, \ldots, L\}$ as the set of all the RRHs. We focus on the downlink transmission, for which the signal processing is more challenging. Assume that there are M $(1 \leq M \leq K)$ multicast groups, i.e., $\{\mathcal{G}_1, \ldots, \mathcal{G}_M\}$, where $\mathcal{G}_m$ is the set of MUs in the multicast group m with $1 \leq m \leq M$. Let $\mathcal{M} = \{1, \ldots, M\}$ be the set of the multicast groups. Each MU only belongs to a single multicast group, i.e., $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$ such that $\cup_i \mathcal{G}_i = \mathcal{S}$ and $\sum_i |\mathcal{G}_i| = K$.

Let $\mathbf{v}_{lm} \in \mathbb{C}^{N_l}$ be the transmit beamforming vector from the l-th RRH to the k-th MU in group $\mathcal{G}_m$. The encoded transmission information symbol of the multicast group m is denoted as $s_m \in \mathbb{C}$ with $\mathbb{E}[|s_m|^2] = 1$. The channel propagation between MU k and RRH l is denoted as $\mathbf{h}_{kl} \in \mathbb{C}^{N_l}$. Therefore, the received signal $y_{k,m} \in \mathbb{C}$ at MU k in the multicast group m is given by

$$y_{k,m} = \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathsf{H}} \mathbf{v}_{lm} s_m + \sum_{i \neq m} \sum_{l=1}^{L} \mathbf{h}_{kl}^{\mathsf{H}} \mathbf{v}_{li} s_i + n_k,$$
$$\forall k \in \mathcal{G}_m, \quad (1)$$

where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the additive Gaussian noise at MU k. We assume that s_m's and n_k's are mutually independent and all the MUs apply single user detection. The signal-to-interference-plus-noise ratio (SINR) for MU $k \in \mathcal{G}_m$ is given by

$$\Gamma_{k,m} = \frac{|\mathbf{h}_k^{\mathsf{H}} \mathbf{v}_m|^2}{\sum_{i \neq m} |\mathbf{h}_k^{\mathsf{H}} \mathbf{v}_i|^2 + \sigma_k^2}, \quad \forall k \in \mathcal{G}_m, \quad (2)$$
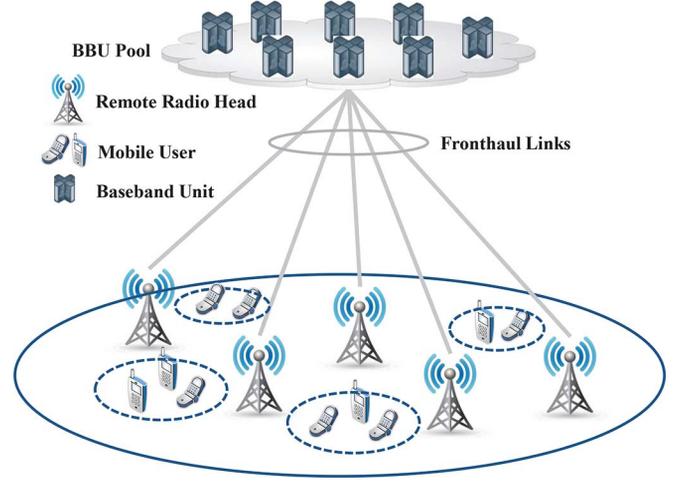
Fig. 1. The architecture of the multicast Cloud-RAN, in which, all the RRHs are connected to a BBU pool through high-capacity and low-latency optical fronthaul links. All the MUs in the same dashed circle form a multicast group and request the same message.

where $\mathbf{h}_k \triangleq [\mathbf{h}_{k1}^T, \ldots, \mathbf{h}_{kL}^T]^T \in \mathbb{C}^N$ with $N = \sum_{l=1}^{L} N_l$, and $\mathbf{v}_m \triangleq [\mathbf{v}_{1m}^T, \mathbf{v}_{2m}^T, \ldots, \mathbf{v}_{Lm}^T]^T \in \mathbb{C}^N$ is the aggregative beamforming vector for the multicast group m from all the RRHs. The transmit signal at RRH l is given by

$$\mathbf{x}_l = \sum_{m=1}^{M} \mathbf{v}_{lm} s_m, \quad \forall l. \quad (3)$$

Each RRH has its own transmit power constraint, i.e.,

$$\sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 \leq P_l, \quad \forall l, \quad (4)$$

where $P_l > 0$ is the maximum transmit power of RRH l.

B. Problem Formulation

1) Imperfect CSI: In practice, the CSI at the BBU pool will be imperfect, which may originate from a variety of sources. For instance, in frequency-division duplex (FDD) systems, the CSI imperfection may originate from downlink training based channel estimation [14] and uplink limited feedback [13]. It could also be due to the hardware deficiencies, partial CSI acquisition [8], [9] and delays in CSI acquisition [15], [16]. In this paper, we adopt the following additive error model [26], [33], [34] to model the channel imperfection from all the RRHs to MU k, i.e.,

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{e}_k, \quad \forall k, \quad (5)$$

where $\hat{\mathbf{h}}_k$ is the estimated channel vector and $\mathbf{e}_k$ is the estimation error vector. There are mainly two ways to model the CSI uncertainty: one is the stochastic modeling based on the probabilistic description, and the other is the deterministic and set-based modeling. However, the stochastic CSI uncertainty modeling will yield probabilistic QoS constraints. The resulting chance constrained programming problems are highly intractable in general [9]. Therefore, to seek a computationally

tractable formulation, we further assume that the error vectors satisfy the following elliptic model [26], [33], [34]:

$$\mathbf{e}_k^{\mathsf{H}} \boldsymbol{\Theta}_k \mathbf{e}_k \leq 1, \quad \forall k, \tag{6}$$

where $\boldsymbol{\Theta}_k \in \mathbb{H}^{N \times N}$ with $\boldsymbol{\Theta}_k \succeq \mathbf{0}$ is the shape of the ellipsoid. This model is motivated by viewing the channel estimation as the main source of CSI uncertainty ([34] Section 4.1).

2) Network Power Consumption: In Cloud-RAN, it is vital to minimize the network power consumption, consisting of RRH transmit power and relative fronthaul network power [5], in order to design a green wireless network. RRH selection will be adopted for this purpose. Specifically, let $\mathcal{A}$ be the set of active RRHs, the network power consumption is given by

$$p(\mathcal{A}) = \sum_{l \in \mathcal{A}} P_l^c + \sum_{l \in \mathcal{A}} \sum_{m=1}^{M} \frac{1}{\eta_l} \|\mathbf{v}_{lm}\|_2^2, \tag{7}$$

where $P_l^c \geq 0$ is the relative fronthaul link power consumption [5] (i.e., the static power saving when both the fronthaul link and the corresponding RRH are switched off) and $\eta_l > 0$ is the drain inefficiency coefficient of the radio frequency power amplifier. The typical values are $P_l^c = 5.6$ W and $\eta_l = 25\%$ [5], respectively.

Given the QoS thresholds $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$, in this paper, we aim at minimizing the network power consumption while guaranteeing the worst-case QoS requirements in the presence of CSI uncertainty and the per-RRH power constraints, i.e., we will consider the following non-convex mixed combinatorial robust multicast beamforming optimization problem,

$$\mathscr{P}: \underset{\mathbf{v}, \mathcal{A}, \mathcal{Z}}{\text{minimize}} \quad \sum_{l \in \mathcal{A}} P_l^c + \sum_{l \in \mathcal{A}} \sum_{m=1}^{M} \frac{1}{\eta_l} \|\mathbf{v}_{lm}\|_2^2 \tag{8}$$

$$\text{subject to} \quad \sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 \leq P_l, \quad \forall l \in \mathcal{A} \tag{9}$$

$$\sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 = 0, \quad \forall l \in \mathcal{Z} \tag{10}$$

$$\frac{|(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_m|^2}{\sum_{i \neq m} |(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k \tag{11}$$

$$\mathbf{e}_k^{\mathsf{H}} \boldsymbol{\Theta}_k \mathbf{e}_k \leq 1, \quad \forall k \in \mathcal{G}_m, \quad m \in \mathcal{M}, \tag{12}$$

where $\mathcal{Z}$ is the set of inactive RRHs such that $\mathcal{A} \cup \mathcal{Z} = \mathcal{L}$ and $\mathbf{v} = [\mathbf{v}_{lm}]$ is the aggregated beamforming vector from all the RRHs to all the MUs. The constraints in (10) indicate that the transmit powers of the inactive RRHs are enforced to be zero. That is, the beamforming coefficients at the inactive RRHs are set to be zero simultaneously. Constraints (11) and (12) indicate that all the QoS requirements in (11) should be satisfied for *all* realizations of the errors $\mathbf{e}_k$'s within the feasible set formed by the constraint (12).

The network power minimization problem $\mathscr{P}$ imposes the following challenges:
1) For a given set of CSI error vectors $\mathbf{e}_k$'s, the corresponding network power minimization problem is highly intractable,

due to the combinatorial composite objective function (8) and the non-convex quadratic constraints (10) and (11).
2) There are an infinite number of non-convex quadratic QoS constraints due to the worst-case design criterion.

To efficiently address the above unique challenges in a unified fashion, in this paper, we will propose a systematic convex relaxation approach based on SDP optimization to solve problem $\mathscr{P}$. In particular, the combinatorial challenge will be addressed by the sparsity-inducing penalty approach in Section IV.A, based on the quadratic variational formulation for the weighted mixed ℓ_1/ℓ_2-norm. The convex optimization technique based on PhaseLift, SDR and S-lemma will be adopted to cope with the infinite number of non-convex quadratic constraints in Sections IV.B and IV.C.

In the next subsection, we will provide a detailed analysis of problem $\mathscr{P}$. In particular, the connections with the formulations in existing literatures will be discussed, which will reveal the generality of the formulation $\mathscr{P}$ for practical design problems in Cloud-RAN.

C. Problem Analysis

While problem $\mathscr{P}$ incorporates most of the practical elements in Cloud-RAN, i.e., imperfect CSI and multigroup multicast transmission, it raises unique challenges compared with the existing works. Following is a list of key aspects of the difficulty of problem $\mathscr{P}$, accompanied with potential solutions.

- *Robust Beamforming Design*: Suppose that all the RRHs are active, i.e., $\mathcal{A} = \mathcal{L}$, with broadcast/unicast transmission, i.e., $|\mathcal{G}_m| = 1, \forall m$ and $M = K$. Then problem $\mathscr{P}$ reduces to the conventional worst-case non-convex robust beamforming design problems [26], [33]. For this special case, the SDR technique [27] combined with the S-lemma [28] is proven to be powerful to find good approximation solutions to such problems.

- *Multicast Beamforming Design*: Physical-layer multicast beamforming design problems [19] prove to be non-convex quadratically constrained problems (QCQP) [28], even with perfect CSI and all the RRHs active. Again, the SDR technique can relax this problem to a convex one, yielding efficient approximation solutions.

- *Quadratically Constrained Feasibility Problem*: Suppose that the inactive RRH set $\mathcal{Z}$ with $|\mathcal{Z}| > 0$ is fixed, then we have the quadratic equation constraints (10) in problem $\mathscr{P}$. PhaseLift [35] is a convex programming technique to relax the non-convex feasibility problem with such quadratic equation constraints to a convex one by lifting the problem to higher dimensions and relaxing the rank-one constraints by the convex surrogates, i.e., the trace norms or nuclear norms.

- *Non-convex Mixed-integer Nonlinear Programming Optimization Problem*: Problem $\mathscr{P}$ can be easily reformulated as a mixed-integer non-linear programming (MINLP) problem as shown in [5]. However, the MINLP problem has exponential complexity [36]. Therefore, such a reformulation cannot bring algorithmic design advantages. One thus has to resort to some global optimization techniques [37], [38] (e.g., branch-and-bound method) or greedy algorithms [5]. Instead, the group-sparsity inducing penalty

approach has recently received enormous attention to seek effective convex relaxation for the MINLP problems, e.g., for jointly designing transmit beamformers and selecting bases stations [31], transmit antennas [32], or RRHs [5]. However, with multicast transmission and imperfect CSI, we cannot directly adopt the group-sparsity inducing penalty developed in [5] with the weighted mixed ℓ_1/ℓ_2-norm, as we have seen that we need to lift the problem $\mathscr{P}$ to higher dimensions to cope with the non-convexity of the robust multicast beamforming problem. This requires to develop a new group-sparsity inducing penalty function, which needs to be compatible with quadratic forms, as the beamforming coefficients will be lifted to higher dimensions.

The above discussions show that problem $\mathscr{P}$ cannot be directly solved by existing methods. Thus, we will propose a new robust group sparse beamforming algorithm in this paper, to solve the highly intractable problem $\mathscr{P}$. Specifically, in Section III, we will propose a group sparse beamforming modeling framework to reformulate the original problem $\mathscr{P}$. The algorithmic advantages of working with the group sparse beamforming formulation will be revealed in Section IV, where a robust group sparse beamforming algorithm will be developed.

III. A GROUP SPARSE BEAMFORMING MODELING FRAMEWORK

In this section, we propose a group sparse beamforming modeling framework to reformulate the network power minimization problem $\mathscr{P}$ by controlling the group-sparsity structure and the magnitude of the beamforming coefficients simultaneously. The main advantage of such a modeling framework is the capability of enabling polynomial time complexity algorithm design via convex relaxation.

A. Network Power Consumption Modeling

We observe that the network power consumption (7) can be modeled by a composite function parameterized by the aggregative beamforming coefficients $\mathbf{v} \in \mathbb{C}^{NM}$, which can be written as a partition

$$\mathbf{v} = [\underbrace{\mathbf{v}_{11}^T, \ldots, \mathbf{v}_{1M}^T}_{\tilde{\mathbf{v}}_1^T}, \ldots, \underbrace{\mathbf{v}_{L1}^T, \ldots, \mathbf{v}_{LM}^T}_{\tilde{\mathbf{v}}_L^T}]^T, \qquad (13)$$

where all the coefficients in a given vector $\tilde{\mathbf{v}}_l = [\mathbf{v}_{l1}^T, \ldots, \mathbf{v}_{lM}^T]^T \in \mathbb{C}^{MN_l}$ form a beamforming coefficient group. Specifically, observe that the optimal aggregative beamforming vector $\mathbf{v}$ in problem $\mathscr{P}$ should have the group-sparsity structure. That is, when the RRH l is switched off, the corresponding coefficients in the beamforming vector $\tilde{\mathbf{v}}_l$ will be set to zero simultaneously. Overall there may be multiple RRHs being switched off and the corresponding beamforming vectors will be set to zero, yielding a group-sparsity structure in the beamforming vector $\mathbf{v}$.

Define the support of the beamforming vector $\mathbf{v}$ as

$$\mathcal{T}(\mathbf{v}) = \{i | v_i \neq 0\}, \qquad (14)$$

where $\mathbf{v} = [v_i]$ is indexed by $i \in \mathcal{V}$ with $\mathcal{V} = \{1, \ldots, MN\}$. Furthermore, define the sets $\mathcal{V}_l = \{M \sum_{i=1}^{l-1} N_i + 1, \ldots, M$

$\sum_{i=1}^{l} N_i\}, l = 1, \ldots, L$, as a partition of $\mathcal{V}$, such that $\tilde{\mathbf{v}}_l = [v_i]$ is indexed by $i \in \mathcal{V}_l$. The network power consumption in the first term of (7) thus can be defined by the following *combinatorial function* with respect to the support of the beamforming vector, i.e.,

$$F(\mathcal{T}(\mathbf{v})) = \sum_{l=1}^{L} P_l^c I(\mathcal{T}(\mathbf{v}) \cap \mathcal{V}_l \neq \emptyset), \qquad (15)$$

where $I(\mathcal{T} \cap \mathcal{V}_l \neq \emptyset)$ is an indicator function that takes value 1 if $\mathcal{T} \cap \mathcal{V}_l \neq \emptyset$ and 0 otherwise. Therefore, the total relative fronthaul link power consumption can be reduced by encouraging the group-sparsity structure of the beamforming vector $\mathbf{v}$.

Furthermore, the total transmit power consumption in the second term of (7) can be defined by the *continuous function* with respect to the ℓ_2-norms of the beamforming vector, i.e.,

$$T(\mathbf{v}) = \sum_{l=1}^{L} \sum_{m=1}^{M} \frac{1}{\eta_l} \|\mathbf{v}_{lm}\|_2^2, \qquad (16)$$

which implicates that the transmit powers of the inactive RRHs are zero, i.e., the corresponding beamforming coefficients are zero. Therefore, the transmit power consumption can be minimized by controlling the magnitude of the beamforming coefficients. As a result, the network power consumption in (7) can be rewritten as the following combinatorial composite function parameterized by the beamforming vector coefficients $\mathbf{v}$, i.e.,

$$P(\mathbf{v}) = F(\mathcal{T}(\mathbf{v})) + T(\mathbf{v}). \qquad (17)$$

Thus, it requires to *simultaneously* control both the combinatorial function F and the continuous function T to minimize the network power consumption. Such a composite function in (17) captures the unique property of the network power consumption that involves two parts (i.e., relative fronthaul network power consumption and transmit power consumption) only through the beamforming coefficients $\mathbf{v}$.

B. Group Sparse Beamforming Modeling

Based on (17), problem $\mathscr{P}$ can be reformulated as the following robust group sparse beamforming problem

$$\mathscr{P}_{\text{sparse}} : \underset{\mathbf{v}}{\text{minimize}} \quad F(\mathcal{T}(\mathbf{v})) + T(\mathbf{v})$$

$$\text{subject to} \quad \sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 \leq P_l, \forall l \in \mathcal{L}$$

$$\frac{|(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_m|^2}{\sum_{i \neq m} |(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k$$

$$\mathbf{e}_k^{\mathsf{H}} \mathbf{\Theta}_k \mathbf{e}_k \leq 1, \quad \forall k \in \mathcal{G}_m, \quad m \in \mathcal{M}, \qquad (18)$$

via optimizing the beamforming coefficients $\mathbf{v}$. We will show that the special structure of the objective function in $\mathscr{P}_{\text{sparse}}$ yields computationally efficient algorithm design. In particular, the weighted mixed ℓ_1/ℓ_2-norm will be derived as a convex surrogate to control both parts in (17) by inducing the group-sparsity structure for the robust multicast beamforming vector $\mathbf{v}$, thereby providing guidelines for RRH selection.
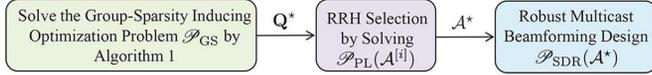
Fig. 2.   The proposed three-stage robust group sparse beamforming framework.

IV. A SEMIDEFINITE PROGRAMMING BASED ROBUST GROUP SPARSE BEAMFORMING ALGORITHM

In this section, we will present the semidefinite programming technique for the robust group sparse beamforming problem $\mathscr{P}_{\mathrm{sparse}}$ by lifting the problem to higher dimensions. The general idea is to relax the combinatorial composite objective function by the quadratic variational formulation of the weighted mixed ℓ_1/ℓ_2-norm to induce the group-sparsity structure for the beamforming vector $\mathbf{v}$. Unfortunately, the resultant group sparse inducing optimization problem is still non-convex. We thus propose a perturbed alternating optimization algorithm to find a stationary point to it, thereby providing the information on determining the priority for the RRHs that should be switched off. Based on the ordering result, a selection procedure is then performed to determine active RRH sets, followed by the robust multicast coordinated beamforming for the active RRHs in the final stage. The proposed three-stage robust group sparse beamforming framework is presented in Fig. 2.

A. Stage One: Group-Sparsity Inducing Penalty Minimization

In this section, we describe a systematic way to address the combinatorial challenge in problem $\mathscr{P}_{\mathrm{sparse}}$ by deriving a convex surrogate to approximate the composite objective function in problem $\mathscr{P}_{\mathrm{sparse}}$. Specifically, we first derive the tightest convex positively homogeneous lower bound for the network power consumption function (17) in the following proposition.

Proposition 1: The tightest convex positively homogeneous lower bound of the objective function in problem $\mathscr{P}_{\mathrm{sparse}}$ is given by

$$\Omega(\mathbf{v}) = 2 \sum_{l=1}^{L} \sqrt{\frac{P_l^c}{\eta_l}} \|\tilde{\mathbf{v}}_l\|_2, \qquad (19)$$

which is a group-sparsity inducing norm for the aggregative robust multicast beamformer vector $\mathbf{v}$.

Proof: Please refer to [5, Appendix A] for the proof.   ∎

Based on proposition 1, we propose to minimize the weighted mixed ℓ_1/ℓ_2-norm to induce the group-sparsity structure for the aggregative robust multicast beamforming vector $\mathbf{v}$:

$$\mathscr{P}_{\mathrm{GSBF}} : \underset{\mathbf{v}}{\mathrm{minimize}} \quad \Omega(\mathbf{v})$$

$$\text{subject to} \quad \sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 \leq P_l, \forall l \in \mathcal{L}$$

$$\frac{|(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_m|^2}{\sum_{i \neq m} |(\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k$$
$$\qquad (20)$$

$$\mathbf{e}_k^{\mathsf{H}} \boldsymbol{\Theta}_k \mathbf{e}_k \leq 1, \forall k \in \mathcal{G}_m, m \in \mathcal{M}. \quad (21)$$

This is, however, a non-convex optimization problem due to the non-convex worst-case QoS constraints (20) and (21).

To seek computationally efficient algorithms to solve the non-convex problem $\mathscr{P}_{\mathrm{GSBF}}$, we propose to lift the problem to higher dimensions with optimization variables as $\mathbf{Q}_m = \mathbf{v}_m \mathbf{v}_m^{\mathsf{H}} \in \mathbb{C}^{N \times N}, \forall m$. To achieve this goal, in Section IV.A1, a variational formulation is proposed to turn the non-smooth group-sparsity inducing norm $\Omega(\mathbf{v})$ into a smooth one with quadratic forms, thereby extracting the variables $\mathbf{Q}_m$'s. We then "linearize" the non-convex worst-case QoS constraints with the S-lemma in Section IV.A2. In Section IV.A3, the perturbed alternating optimization algorithm is proposed to solve the resultant non-convex group-sparsity inducing optimization problem by exploiting its convex substructures.

1) Quadratic Variational Formulation of the Weighted Mixed ℓ_1/ℓ_2-Norm: In order to extract the variables $\mathbf{Q}_m$'s from the weighted mixed ℓ_1/ℓ_2-norm, we introduce the following lemma to obtain an equivalent expression for the square norm $\Omega^2(\mathbf{v})$, which has the same capability of inducing group-sparsity as the non-smooth one $\Omega(\mathbf{v})$ [30] and is widely used in multiple kernel learning [39].

Lemma 1: [30]: Let $\mathbf{x} = (x_1, \ldots, x_L) \in \mathbb{R}_+^L$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L) \in \mathbb{R}_+^L$, then

$$\left(\sum_{l=1}^{L} \omega_l x_l \right)^2 = \inf_{\boldsymbol{\mu} \in \mathbb{R}_+^L} \sum_{l=1}^{L} \frac{\omega_l^2 x_l^2}{\mu_l}, \text{ s.t. } \sum_{l=1}^{L} \mu_l = 1. \quad (22)$$

Proof: This can be obtained directly through the Cauchy-Schwarz inequality

$$\sum_{l=1}^{L} \omega_l x_l = \sum_{l=1}^{L} \frac{\omega_l x_l}{\sqrt{\mu_l}} \cdot \sqrt{\mu_l}$$

$$\leq \left(\sum_{l=1}^{L} \frac{\omega_l^2 x_l^2}{\mu_l} \right)^{1/2} \left(\sum_{l=1}^{L} \mu_l \right)^{1/2}, \quad (23)$$

where $\omega_l \geq 0$ and the equality is met when $\sqrt{\mu_l}$ is proportional to $(\omega_l x_l)/\sqrt{\mu_l}$, i.e.,

$$\mu_l = \frac{\omega_l x_l}{\sum_{l=1}^{L} \omega_l x_l}, \quad (24)$$

which leads to the conclusion (22).   ∎

Based on Lemma 1, the square of the weighted mixed ℓ_1/ℓ_2-norm (19) can be rewritten as

$$\Omega^2(\mathbf{v}) = \inf_{\boldsymbol{\mu} \in \mathcal{X}} \mathcal{R}(\boldsymbol{\mu}, \mathbf{Q}), \quad (25)$$

where $\mathcal{X} = \{\mu_l | \mu_l > 0, \sum_{l=1}^{L} \mu_l = 1\}$ is a simplex set and

$$\mathcal{R}(\boldsymbol{\mu}, \mathbf{Q}) = 4 \sum_{l=1}^{L} \frac{P_l^c}{\eta_l \mu_l} \left(\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m) \right), \quad (26)$$

where $\boldsymbol{\mu} = [\mu_l]$, $\mathbf{Q} = [\mathbf{Q}_l]$ and $\mathbf{C}_{lm} \in \mathbb{R}^{N \times N}$ is a block diagonal matrix with the identity matrix $\mathbf{I}_{N_l}$ as the l-th main diagonal block square matrix and zeros elsewhere. Therefore, the group-sparsity structure of the beamformer $\mathbf{v}$ can be extracted from the trace of $\mathbf{Q}_m$'s, as will be shown in (36). This procedure is known as *the quadratic variational formulation* of norms [30].

2) Linearize the Non-Convex Worst-Case QoS Constraints:
Define $\mathbf{G}_m = (\mathbf{Q}_m - \gamma_k \sum_{i \neq m} \mathbf{Q}_i)$, and then the worst-case QoS constraints (20) and (21) can be rewritten as

$$\min_{\mathbf{e}_k^{\mathsf{H}} \boldsymbol{\Theta}_k \mathbf{e}_k \leq 1} (\hat{\mathbf{h}}_k + \mathbf{e}_k)^{\mathsf{H}} \mathbf{G}_m (\hat{\mathbf{h}}_k + \mathbf{e}_k) \geq \gamma_k \sigma_k^2, \quad \forall k \in \mathcal{G}_m. \tag{27}$$

As the number of choices of $\mathbf{e}_k$'s in the worst-case QoS constraint (27) is infinite, there are an infinite number of such "linearized" QoS constraints. Fortunately, using the S-lemma [28, Appendix B.2], the worst-case QoS constraints (27) can be equivalently written as the following finite number of convex constraints:

$$\mathcal{C}_1 : \left[\begin{array}{c|c} \mathbf{G}_m & \mathbf{G}_m \hat{\mathbf{h}}_k \\ \hline \hat{\mathbf{h}}_k^H \mathbf{G}_m & \hat{\mathbf{h}}_k^H \mathbf{G}_m \hat{\mathbf{h}}_k - \gamma_k \sigma_k^2 \end{array} \right] + \lambda_k \left[\begin{array}{c|c} \boldsymbol{\Theta}_k & \mathbf{0} \\ \hline \mathbf{0}^H & -1 \end{array} \right] \succeq \mathbf{0}, \tag{28}$$

where $\lambda_k \geq 0$ and $k \in \mathcal{G}_m$ with $m \in \mathcal{M}$.

Based on the above discussions and utilizing the principle of SDR technique [27] by dropping the rank-one constraints for $\mathbf{Q}_k$'s, we propose to solve the following problem to induce the group-sparsity structure for the beamforming vector $\mathbf{v}$

$$\mathscr{P}_{\text{GS}} : \underset{\mathbf{Q}, \boldsymbol{\lambda}, \boldsymbol{\mu} \in \mathcal{X}}{\text{minimize}} \quad \mathcal{R}(\boldsymbol{\mu}, \mathbf{Q})$$
$$\text{subject to} \quad \mathcal{C}_1, \mathcal{C}_2(\mathcal{L}), \lambda_k \geq 0, \mathbf{Q}_m \succeq \mathbf{0}$$
$$\forall k \in \mathcal{G}_m, m \in \mathcal{M}, \tag{29}$$

where $\boldsymbol{\lambda} = [\lambda_k]$ and $\mathcal{C}_2(\mathcal{A})$ is the set of linearized per-RRH transmit power constraints,

$$\mathcal{C}_2(\mathcal{A}) : \sum_{m=1}^{M} \text{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m) \leq P_l, l \in \mathcal{A}. \tag{30}$$

Problem $\mathscr{P}_{\text{GS}}$ is still non-convex, as the objective function $\mathcal{R}(\boldsymbol{\mu}, \mathbf{Q})$ is not jointly convex in the variables $(\boldsymbol{\mu}, \mathbf{Q})$. Nevertheless, the objective function is biconvex [40], i.e., function $\mathcal{R}$ is convex with respect to $\boldsymbol{\mu}$ for fixed $\mathbf{Q}$ and vice versa. In the next subsection, we thus exploit the convex substructures of problem $\mathscr{P}_{\text{GS}}$ to develop a perturbed alternating optimization algorithm to find an efficient sub-optimal solution.

3) Perturbed Alternating Optimization Algorithm: The general idea of the alternating optimization algorithm is that problem $\mathscr{P}_{\text{GS}}$ is first optimized with respect to $(\mathbf{Q}, \boldsymbol{\lambda})$ with a fixed $\boldsymbol{\mu}$, then the variables μ_l's are chosen to minimize $\mathcal{R}(\boldsymbol{\mu}, \mathbf{Q})$ with a fixed $\mathbf{Q}$. However, to avoid singularity when μ_l's approach to zeros during the alternating procedure as discussed in [39], we instead adopt the perturbed version of the alternating optimization algorithm [41] to solve problem $\mathscr{P}_{\text{GS}}$. Specifically, define the perturbed objective function of problem $\mathscr{P}_{\text{GS}}$ as

$$\mathcal{R}_\epsilon(\boldsymbol{\mu}, \mathbf{Q}) = 4 \sum_{l=1}^{L} \frac{P_l^c}{\eta_l \mu_l} \left(\sum_{m=1}^{M} \text{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m + \epsilon \mathbf{I}_N) \right), \tag{31}$$

where $\epsilon > 0$. Let $\mathscr{P}_{\text{GS}}(\epsilon)$ be the problem by replacing the objective function in problem $\mathscr{P}_{\text{GS}}$ with the perturbed function $\mathcal{R}_\epsilon(\boldsymbol{\mu}, \mathbf{Q})$. We thus solve problem $\mathscr{P}_{\text{GS}}(\epsilon)$ via alternatively solving the following two problems:

- *Fixing $\boldsymbol{\mu}$, Optimizing $\mathbf{Q}$ and $\boldsymbol{\lambda}$:* Given $\boldsymbol{\mu} = \boldsymbol{\mu}^{[i]}$ at the i-th iteration, we need to solve the following problem

$$\mathscr{P}_{\text{GS}}^{[i]}(\epsilon; \boldsymbol{\mu}^{[i]}) : \underset{\mathbf{Q}, \boldsymbol{\lambda}}{\text{minimize}} \quad \mathcal{R}_\epsilon(\boldsymbol{\mu}^{[i]}, \mathbf{Q})$$
$$\text{subject to} \quad \mathcal{C}_1, \mathcal{C}_2(\mathcal{L}), \lambda_k \geq 0, \mathbf{Q}_m \succeq \mathbf{0}$$
$$\forall k \in \mathcal{G}_m, m \in \mathcal{M}, \tag{32}$$

to obtain $(\mathbf{Q}^{[i]}, \boldsymbol{\lambda}^{[i]})$. This is an SDP problem and can be solved efficiently using the interior-point method [28].

- *Fixing $\mathbf{Q}$ and $\boldsymbol{\lambda}$, Optimizing $\boldsymbol{\mu}$:* Given $\mathbf{Q} = \mathbf{Q}^{[i]}$ at the i-th iteration, we need to optimize $\boldsymbol{\mu}$ over the simplex set $\mathcal{X}$, i.e.,

$$\mathscr{P}_{\text{GS}}^{[i]}(\epsilon; \mathbf{Q}^{[i]}) : \underset{\boldsymbol{\mu} \in \mathcal{X}}{\text{minimize}} \ \mathcal{R}_\epsilon(\boldsymbol{\mu}, \mathbf{Q}^{[i]}), \tag{33}$$

which has the following optimal solution based on Lemma 1:

$$\mu_l^{[i]} = \frac{\sqrt{(P_l^c/\eta_l) \cdot \sum_{m=1}^{M} \text{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m^{[i]} + \epsilon \mathbf{I}_N)}}{\sum_{l=1}^{L} \sqrt{(P_l^c/\eta_l) \cdot \sum_{m=1}^{M} \text{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m^{[i]} + \epsilon \mathbf{I}_N)}}, \tag{34}$$

for any $l \in \mathcal{L}$.

As the objective function in problem $\mathscr{P}_{\text{GS}}(\epsilon)$ is bounded and non-increasing at each iteration, the sequence $\{\mathcal{R}_\epsilon(\boldsymbol{\mu}^{[i]}, \mathbf{Q}^{[i]})\}$ generated by this algorithm, clearly, converges monotonically to a sub-optimal value [40]. Since we will use the solution of the problem $\mathscr{P}_{\text{GS}}(\epsilon)$ to predicate the group-sparsity pattern for the beamformer $\mathbf{v}$, we thus are also interested in investigating the convergence of the sequence $\{\mathbf{Q}^{[i]}\}$ itself generated by this algorithm when $\epsilon \to 0$. This is presented in the following Theorem.

Theorem 1: The sequence $\{\boldsymbol{\mu}^{[i]}(\epsilon), \mathbf{Q}^{[i]}(\epsilon), \boldsymbol{\lambda}^{[i]}(\epsilon)\}$ generated by the perturbed alternating optimization algorithm converges to a stationary point of problem $\mathscr{P}_{\text{GS}}(\epsilon)$. Furthermore, when $\epsilon \to 0$, we have

$$\lim_{\epsilon \searrow 0} \mathbb{D}(\Lambda(\epsilon), \Lambda_0) = 0, \tag{35}$$

where Λ_0 (45) and $\Lambda(\epsilon)$ (46) denote the set of stationary points of problem $\mathscr{P}_{\text{GS}}$ and $\mathscr{P}_{\text{GS}}(\epsilon)$, respectively; and $\mathbb{D}(A_1, A_2)$, defined in (52), denotes the deviation of the set A_1 from the set A_2.

Proof: Please refer to Appendix A for details. ∎

The perturbed alternating optimization algorithm is presented in Algorithm 1.

Algorithm 1: Perturbed Alternating Optimization Algorithm

---

input: Initialize $\boldsymbol{\mu}^{[0]} = (1/L, \ldots, 1/L)$; I (the maximum number of iterations)

Repeat
 1) Solve problem $\mathscr{P}_{\text{GS}}^{[i]}(\epsilon; \boldsymbol{\mu}^{[i]})$ (32). If it is feasible, **go to** 2); otherwise, **stop** and return **output 2**.
 2) Calculate $\boldsymbol{\mu}^{[i]} = (\mu_1^{[i]}, \ldots, \mu_L^{[i]})$ according to (34).
Until convergence or attain the maximum iterations and return **output 1**.

---

 output 1: $\mathbf{Q}_1^{\star}, \ldots, \mathbf{Q}_M^{\star}$; **output 2:** Infeasible.

Based on the solutions $\mathbf{Q}_m^\star$'s generated by the perturbed alternating optimization algorithm, in the next subsection, we will present how to extract the group-sparsity pattern information for the beamformer $\mathbf{v}$, thereby providing information on the RRH ordering, i.e., determine the priority of the RRHs that should be switched off.

B. Stage Two: RRH Selection

Given the solution $\mathbf{Q}^\star$ to the group sparse inducing optimization problem $\mathscr{P}_{\mathrm{GS}}$, the group-sparsity structure information for the beamformer $\mathbf{v}$ can be extracted from the following relation:

$$\|\tilde{\mathbf{v}}_l\|_{\ell_2} = \left(\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{C}_{lm}\mathbf{Q}_m) \right)^{1/2}, \quad \forall l. \qquad (36)$$

Based on the (approximated) group-sparsity information in (36), the following ordering criterion [5] incorporating the key system parameters is adopted to determine which RRHs should be switched off, i.e.,

$$\theta_l = \sqrt{\frac{\kappa_l \eta_l}{P_l^c}} \left(\sum_{m=1}^{M} \mathrm{Tr}(\mathbf{C}_{lm}\mathbf{Q}_m^\star) \right)^{1/2}, \quad \forall l \in \mathcal{L}, \qquad (37)$$

where $\kappa_l = \sum_{k=1}^{K} \|\hat{\mathbf{h}}_{kl}\|_2^2$ is the channel gain for the estimated channel coefficients between RRH l and all the MUs. Therefore, the RRH with a smaller parameter θ_l will have a higher priority to be switched off. Note that most previous works applying the idea of sparsity inducing norm minimization approach directly map the sparsity pattern to their applications. For instance, in [32], the transmit antenna with smaller coefficients in the beamforming coefficient group (measured by the ℓ_∞-norm) will have a higher priority to be switched off. In [5], however, we show that the ordering rule (37), which incorporates the key system parameters, yields much better performance than the pure sparsity pattern based selection rule in terms of network power minimization.

In this paper, we adopt a simple RRH selection procedure, i.e., binary search, due to its low-complexity. Specifically, based on the ordering rule (37), we sort the coefficients in the ascending order: $\theta_{\pi_1} \le \theta_{\pi_2} \le \cdots \le \theta_{\pi_L}$ to determine the active RRH set. Denote J_0 as the maximum number of RRHs that can be switched off. That is, problem $\mathscr{F}(\mathcal{A}^{[i]})$ is feasible for any $i \le J_0$,

$$\mathscr{F}(\mathcal{A}^{[i]}) : \text{find } \mathbf{v}$$
$$\text{subject to } (9), (10), (11), (12), \qquad (38)$$

where $\mathcal{A}^{[i]} \cup \mathcal{Z}^{[i]} = \mathcal{L}$ with $\mathcal{Z}^{[i]} = \{\pi_0, \pi_1, \ldots, \pi_i\}$ and $\pi_0 = \emptyset$. Likewise, problem $\mathscr{F}(\mathcal{A}^{[i]})$ with $\mathcal{A}^{[i]} = \{\pi_{i+1}, \ldots, \pi_L\}$ is infeasible for any $i > J_0$. A binary search procedure can be adopted to determine J_0, which only needs to solve no more than $(1 + \lceil \log(1 + L) \rceil)$ feasibility problems (38) as will be presented in Algorithm 2. Denote $\mathcal{A}^{[J_0]}$ as the final active RRH set, we thus need to solve the following transmit power minimization problem

$$\mathscr{P}(\mathcal{A}) : \underset{\mathbf{v}}{\text{minimize}} \sum_{l \in \mathcal{A}} \left(\frac{1}{\eta_l} \sum_{m=1}^{M} \|\mathbf{v}_{lm}\|_2^2 + P_l^c \right)$$
$$\text{subject to } (9), (10), (11), (12), \qquad (39)$$

with the fixed active RRH set $\mathcal{A} = \mathcal{A}^{[J_0]}$ to determine the transmit beamformer coefficients for the active RRHs. Unfortunately, both problems $\mathscr{F}(\mathcal{A})$ and $\mathscr{P}(\mathcal{A})$ are non-convex and intractable. Thus, in the paper, we resort to the computationally efficient semidefinite programming technique to find approximate solutions to feasibility problem $\mathscr{F}(\mathcal{A})$ and optimization problem $\mathscr{P}(\mathcal{A})$.

Notice that, with perfect CSI assumptions as in [5], [32], given the active RRH set $\mathcal{A}$, the size of the corresponding optimization problem $\mathscr{P}(\mathcal{A})$ (e.g., [5, (12)] and [32, (13)] will be reduced. The key observation is that we only need to consider the channel links from the active RRHs. However, with imperfect CSI, we still need to consider the channel links from all the RRHs due to the lack of the knowledge of the exact values of the CSI errors $\mathbf{e}_k$'s. As a result, the sizes of corresponding optimization problems $\mathscr{P}(\mathcal{A}^{[i]})$'s cannot be reduced with imperfect CSI.

1) Phaselift To the Non-Convex Feasibility Problem: In this subsection, we use the *PhaseLift* technique [35] to find approximate solutions to the non-convex feasibility problem $\mathscr{F}(\mathcal{A})$. Specifically, we first lift the problem to higher dimensions such that the feasibility problem $\mathscr{F}(\mathcal{A})$ can be reformulated as

$$\begin{aligned}
\text{find} \quad & \mathbf{Q}_1, \ldots, \mathbf{Q}_M \\
\text{subject to} \quad & \mathcal{C}_1, \mathcal{C}_2(\mathcal{A}), \mathcal{C}_3(\mathcal{Z}), \lambda_k \ge 0, \mathbf{Q}_m \succeq \mathbf{0} \\
& \mathrm{rank}(\mathbf{Q}_m) = 1, \forall k \in \mathcal{G}_m, m \in \mathcal{M}, \quad (40)
\end{aligned}$$

where

$$\mathcal{C}_3(\mathcal{Z}) : \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{C}_{lm}\mathbf{Q}_m) = 0, \forall l \in \mathcal{Z}. \qquad (41)$$

The main idea of the PhaseLift technique is to approximate the non-convex rank functions in problem (40) using the convex surrogates, yielding the following convex feasibility problem

$$\begin{aligned}
\mathscr{P}_{\mathrm{PL}}(\mathcal{A}) : \text{find} \quad & \mathbf{Q}_1, \ldots, \mathbf{Q}_M \\
\text{subject to} \quad & \mathcal{C}_1, \mathcal{C}_2(\mathcal{A}), \mathcal{C}_3(\mathcal{Z}), \lambda_k \ge 0, \mathbf{Q}_m \succeq \mathbf{0} \\
& \forall k \in \mathcal{G}_m, m \in \mathcal{M}, \quad (42)
\end{aligned}$$

which is an SDP problem and can be solved using the interior-point method [28] efficiently. In general, the solution of problem $\mathscr{P}_{\mathrm{PL}}(\mathcal{A})$ may not be rank-one. If this happens, to yield a feasible solution for problem $\mathscr{F}(\mathcal{A})$, the Gaussian randomization procedure [27] will be applied to obtain a feasible rank-one approximate solution for problem $\mathscr{F}(\mathcal{A})$ from the solution of problem $\mathscr{P}_{\mathrm{PL}}(\mathcal{A})$.

Remark 1: The PhaseLift technique, serving as one promising application of the SDR method, was proposed in [35] to solve the phase retrieval problem [42], which is mathematically a feasibility problem with multiple quadratic equation constraints. Various conditions are presented in [35], [42] for the phase retrieval problem, under which the corresponding solution of the PhaseLift relaxation problem yields a rank-one solution with a high probability. However, for our problem $\mathscr{P}_{\mathrm{PL}}$ with additional complicated constraints, it is challenging to perform such rank-one solution analysis. Thus, in this paper, we only focus on developing computationally efficient approximation algorithms based on the SDR technique.

C. Stage Three: SDR To the Robust Multicast Beamforming Problem

Once we have selected active RRHs, i.e., fix the set $\mathcal{A}$, we need to finalize the beamforming vector by solving problem $\mathscr{P}(\mathcal{A})$. We lift the non-convex optimization problem $\mathscr{P}(\mathcal{A})$ to higher dimensions and adopt the SDR technique by dropping the rank-one constraints, yielding the following convex relaxation problem

$$\mathscr{P}_{\mathrm{SDR}}(\mathcal{A}): \underset{\mathbf{Q}, \boldsymbol{\lambda}}{\operatorname{minimize}} \quad \sum_{l \in \mathcal{A}}\left(\frac{1}{\eta_l} \sum_{m=1}^{M} \operatorname{Tr}(\mathbf{C}_{lm} \mathbf{Q}_m) + P_l^c\right)$$
$$\text{subject to} \quad \mathcal{C}_1, \mathcal{C}_2(\mathcal{A}), \mathcal{C}_3(\mathcal{Z}), \lambda_k \geq 0, \mathbf{Q}_m \succeq \mathbf{0}$$
$$\forall k \in \mathcal{G}_m, m \in \mathcal{M}, \quad (43)$$

which is an SDP problem and can be solved using the interior-point method [28]. It is important to investigate whether the solution of problem $\mathscr{P}_{\mathrm{SDR}}(\mathcal{A})$ yields a rank-one solution $\{\mathbf{Q}_m^\star\}$. This is, however, an on-going research topic and some preliminary results were presented in [26], [29]. In this paper, if $\operatorname{rank}(\mathbf{Q}_m^\star) = 1, \forall m$, we can write $\mathbf{Q}_m^\star = \mathbf{v}_m^\star \mathbf{v}_m^{\star\mathsf{H}}, \forall m$ and $\{\mathbf{v}_m^\star\}$ is a feasible (in fact optimal) solution to problem $\mathscr{P}(\mathcal{A})$. Otherwise, if the rank-one solution is failed to be obtained, the Gaussian randomization method [27] will be employed to obtain a feasible rank-one approximate solution to problem $\mathscr{P}(\mathcal{A})$.

Finally, we arrive at the robust group sparse beamforming algorithm as shown in Algorithm 2.

Algorithm 2. Robust Group Sparse Beamforming Algorithm

Step 0: Solve the group-sparsity inducing optimization problem $\mathscr{P}_{\mathrm{GS}}$ (29) using Algorithm 1.
  1) **If it is infeasible, go to End**.
  2) **If it is feasible**, obtain the solutions $\mathbf{Q}_m^\star$'s, calculate the ordering criterion (37), and sort them in the ascending order: $\theta_{\pi_1} \leq \ldots \leq \theta_{\pi_L}$, **go to Step 1**.

Step 1: Initialize $J_{\mathrm{low}} = 0, J_{\mathrm{up}} = L, i = 0$.

Step 2: Repeat
  1) Set $i \leftarrow \lfloor \frac{J_{\mathrm{low}} + J_{\mathrm{up}}}{2} \rfloor$.
  2) Solve problem $\mathscr{P}_{\mathrm{PL}}(\mathcal{A}^{[i]})$ (42): if it is infeasible, set $J_{\mathrm{up}} = i$; otherwise, set $J_{\mathrm{low}} = i$.

Step 3: Until $J_{\mathrm{up}} - J_{\mathrm{low}} = 1$, obtain $J_0 = J_{\mathrm{low}}$ and obtain the optimal active RRH set $\mathcal{A}^\star$ with $\mathcal{A}^\star \cup \mathcal{J} = \mathcal{L}$ and $\mathcal{J} = \{\pi_1, \ldots, \pi_{J_0}\}$.

Step 4: Solve problem $\mathscr{P}_{\mathrm{SDR}}(\mathcal{A}^\star)$ (43), obtain the robust multicast beamforming coefficients for the active RRHs.

End

Remark 2: The proposed robust group sparse beamforming algorithm consists of three stages. In the first stage, we observe that the perturbed alternating optimization algorithm converges in 20 iterations on average in all the simulated settings in this paper, while it is interesting to analyze the convergence rate for this algorithm. In the second stage, to find the set of active RRHs, we only need to solve no more than $(1 + \lceil \log(1 + L) \rceil)$ convex feasibility problems (42) using the bi-section method.
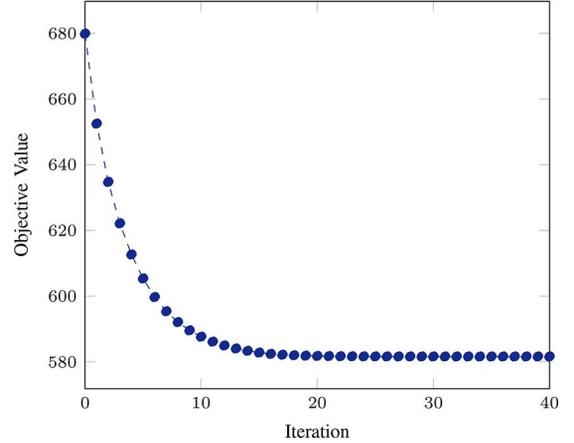
Fig. 3. Convergence of the perturbed alternating optimization algorithm.

Finally, we need to solve problem (43) to determine the transmit beamforming coefficients for the fixed active RRHs.

V. SIMULATION RESULTS

In this section, we analyze the performance of the proposed robust group sparse beamforming algorithm. For illustration purposes, all the estimated channels $\hat{\mathbf{h}}_k$'s are modeled as spatially uncorrelated Rayleigh fading and the CSI errors are modeled as the elliptic model (6) with $\mathbf{Q}_k = \varepsilon_k^{-2} \mathbf{I}_N, \forall k$. We assume that each multicast group has the same number of MUs, i.e., $|\Omega_1| = |\Omega_2| = \ldots = |\Omega_M|$. The power amplifier efficiency coefficients are set to be $\eta_l = 25\%, \forall l$. The perturbed parameter ϵ in the perturbed alternating optimization algorithm is set to be 10^{-3} and the algorithm will stop if either the difference between the objective values of consecutive iterations is less than 10^{-3} or it exceeds the predefined maximum iterations 20. Each point of the simulation results is averaged over 50 randomly generated channel realizations, except for Fig. 3, where we only report one typical channel realization.

A. Convergence of the Perturbed Alternating Optimization Algorithm

Consider a network with $L = 10$ 2-antennas RRHs and 3 multicast groups with 2 single-antenna MUs in each group, i.e., $|\Omega_m| = 2, \forall m$. All error radii ε_k's are set to be 0.05. The convergence of the perturbed alternating optimization algorithm is demonstrated in Fig. 3 for a typical channel realization. This figure shows that the proposed alternating optimization algorithm converges very fast (less 20 iterations) in the simulated network size.

B. Network Power Minimization

1) Scenario One: We first consider a network with $L = 5$ 2-antenna RRHs and $M = 2$ multicast groups each has 2 single-antenna MUs, i.e., $|\Omega_m| = 2, \forall m$. The relative fronthaul links power consumption are set to be $P_l^c = 5.6\,W, \forall l$. All error radii ε_k's are set to be 0.01. Fig. 4 demonstrates the average network power consumption with different target SINRs. The corresponding average number of active RRHs and average total transmit power consumption are showed in Table I and Table II, respectively.
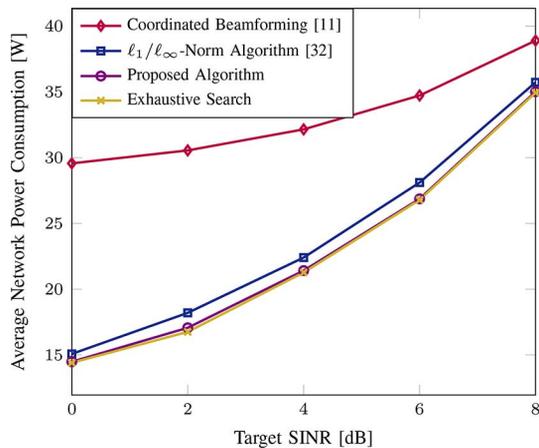
Fig. 4. Average network power consumption versus target SINR for scenario one.

Fig. 5. Average network power consumption versus target SINR for scenario two.

TABLE I
THE AVERAGE NUMBER OF ACTIVE RRHS WITH DIFFERENT
ALGORITHMS FOR SCENARIO ONE

Target SINR [dB]	0	2	4	6	8
Coordinated Beamforming	5.00	5.00	5.00	5.00	5.00
ℓ_1/ℓ_∞-Norm Algorithm	2.00	2.33	2.73	3.30	4.10
Proposed Algorithm	2.00	2.13	2.63	3.13	4.00
Exhaustive Search	2.00	2.07	2.60	3.10	4.00

TABLE II
THE AVERAGE TOTAL TRANSMIT POWER CONSUMPTION WITH DIFFERENT
ALGORITHMS FOR SCENARIO ONE

Target SINR [dB]	0	2	4	6	8
Coordinated Beamforming	1.56	2.55	4.15	6.72	10.89
ℓ_1/ℓ_∞-Norm Algorithm	3.88	5.13	7.10	9.63	12.76
Proposed Algorithm	3.28	5.12	6.67	9.32	12.61
Exhaustive Search	3.20	5.18	6.71	9.43	12.54

TABLE III
THE AVERAGE RELATIVE FRONTHAUL LINKS POWER CONSUMPTION WITH
DIFFERENT ALGORITHMS FOR SCENARIO TWO

Target SINR [dB]	0	2	4	6	8
Coordinated Beamforming	72.80	72.80	72.80	72.80	72.80
ℓ_1/ℓ_∞-Norm Algorithm	36.08	43.76	52.36	60.16	69.56
Proposed Algorithm	30.40	38.08	45.56	56.76	70.48

TABLE IV
THE AVERAGE TOTAL TRANSMIT POWER CONSUMPTION WITH DIFFERENT
ALGORITHMS FOR SCENARIO TWO

Target SINR [dB]	0	2	4	6	8
Coordinated Beamforming	3.02	5.16	8.84	15.05	25.41
ℓ_1/ℓ_∞-Norm Algorithm	8.54	10.96	14.43	19.87	27.42
Proposed Algorithm	8.03	11.25	16.32	20.03	26.28

Specifically, Fig. 4 shows that the proposed robust group sparse beamforming algorithm achieves near-optimal values of network power consumption compared with the ones obtained by the exhaustive search algorithm via solving a sequence of problems (43). Furthermore, it is observed that the proposed algorithm outperforms the square of ℓ_1/ℓ_∞-norm based algorithm with sparsity pattern ordering rule in [32] in terms of network power minimization. Specifically, the objective function of the group-sparsity inducing optimization problem (29) will be replaced by $\mathcal{R} = \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \max_m \max_{n_{l_1}} \max_{n_{l_2}} |\mathbf{Q}_m(n_{l_1}, n_{l_2})|$ with $\mathbf{Q}_m(i,j)$ being the entry indexed by (i,j) in $\mathbf{Q}_m$. Then the RRH with smaller beamforming coefficients measured by the ℓ_∞-norm will have a higher priority to be switched off. In particular, Table I shows that the proposed algorithm can switch off more RRHs than the ℓ_1/ℓ_∞-norm based algorithm, which is almost the same as the exhaustive search algorithm. Besides, this table also verifies the group-sparsity assumption for the aggregative transmit beamformer $\mathbf{v}$, i.e., the beamforming coefficients of the switched off RRHs are set to be zeros simultaneously. Meanwhile, Table II shows that the proposed algorithm can achieve higher transmit beamforming gains, yielding lower total transmit power consumption compared with the ℓ_1/ℓ_∞-norm based algorithm. The coordinated beamforming algorithm [11], which aims at only minimizing the total transmit power consumption with all the RRHs active, achieves the highest beamforming gain but with the highest relative fronthaul links power consumption.

Overall, Fig. 4, Table I and Table II show the effectiveness of the proposed robust group sparse beamforming algorithm to minimize the network power consumption.

2) Scenario Two: We then consider a larger-sized network with $L = 8$ 2-antenna RRHs and $M = 5$ multicast groups each has 2 single-antenna MUs, i.e., $|\Omega_m| = 2, \forall m$. The relative fronthaul links power consumption are set to be $P_l^c = [5.6 + (l - 1)]W, \forall l$. All error radii ε_k's are set to be 0.05. Due to the high computational cost of the exhaustive search algorithm, we only simulate the ℓ_1/ℓ_∞-norm based algorithm and the proposed robust group sparse beamforming algorithm. Fig. 5, Tables III and IV show the average network power consumption, the average relative fronthaul link power consumption and the average total transmit power consumption versus SINRs with different algorithms, respectively. From Fig. 5, we see that the proposed robust beamforming algorithm achieves lower network power consumption compared with the ℓ_1/ℓ_∞-norm

algorithm and the coordinated beamforming algorithm. In particular, Table III shows that proposed algorithm achieves much lower relative fronthaul links power consumption, thought with a little higher transmit power consumption at the moderate target SINR regimes. Compared with the ℓ_1/ℓ_∞-norm algorithm, the performance gain of the proposed algorithm is more prominent with low target SINRs.

Overall, all the simulation results illustrate the effectiveness of the proposed robust group sparse beamforming algorithm to control both the relative fronthaul power consumption and the RRH transmit power consumption with different network configurations.

VI. CONCLUSIONS AND FUTURE WORKS

This paper described a systematic way to develop computationally efficient algorithms based on the group-sparsity inducing penalty approach for the highly intractable network power minimization problem for multicast Cloud-RAN with imperfect CSI. A novel quadratic variational formulation of the weighted mixed ℓ_1/ℓ_2-norm was proposed to induce the group-sparsity structure for the robust multicast beamformer, thereby guiding the RRH selection. The perturbed alternating optimization, PhaseLift method, and SDR technique based algorithms were developed to solve the group-sparsity inducing optimization problem, the feasibility problems in RRH selection procedure and the transmit beamformer design problem in the final stage, respectively. Simulation results illustrated the effectiveness of the proposed robust group sparse beamforming algorithm to minimize the network power consumption.

Several future directions of interest are listed as follows:

- Although the proposed SDP based robust group sparse beamforming algorithm has a polynomial time complexity, the computational cost of the interior-point method will the prohibitive when the dimensions of the SDP problems are large, such as in dense wireless networks. One may use the first-order method, e.g., the alternating direction method of multipliers (ADMM) [43]–[46] to seek modest accuracy solutions within reasonable time for the large-scale SDP problems [47].

- It is desirable to lay the theoretical foundations for the tightness of the group-sparsity inducing penalty approach for finding approximate solutions to the network power minimization problem as a mixed-integer non-linear optimization problem, and also for the tightness of PhaseLift method and SDR technique.

- It is interesting to apply the sparsity modeling framework to more mixed-integer nonlinear optimization problems, i.e., the joint user scheduling or admission and beamforming problems, which are essentially required to control the sparsity structure and the magnitude of the beamforming coefficients.

APPENDIX A
PROOF OF THEOREM 1

We first consider problem $\mathscr{P}_{\mathrm{GS}}(\epsilon)$ with a fixed ϵ. Based on ([40] Theorem 4.9), we know that the accumulation point $(\boldsymbol{\mu}^\star(\epsilon), \mathbf{Q}^\star(\epsilon), \boldsymbol{\lambda}^\star(\epsilon))$ of the sequence $\{\boldsymbol{\mu}^{[i]}(\epsilon), \mathbf{Q}^{[i]}(\epsilon), \boldsymbol{\lambda}^{[i]}(\epsilon)\}$

converges to a stationary point of problem $\mathscr{P}_{\mathrm{GS}}(\epsilon)$, provided that the optimal solution (34) is unique with $\mathbf{Q} = \mathbf{Q}^\star(\epsilon)$. This can be easily justified by the strict convexity of $\mathcal{R}_\epsilon(\boldsymbol{\mu}, \mathbf{Q})$ with respect to $\boldsymbol{\mu}$ for a fixed $\mathbf{Q}$.

Next, we will prove the relationship (35) between Λ_0 and $\Lambda(\epsilon)$. For convenience, we define the feasible region of problems $\mathscr{P}_{\mathrm{GS}}$ and $\mathscr{P}_{\mathrm{GS}}(\epsilon)$ as $\mathcal{C}$. Then problem $\mathscr{P}_{\mathrm{GS}}(\epsilon)$ can be rewritten as

$$\mathscr{P}_{\mathrm{GS}}(\epsilon) : \underset{\mathbf{x} \in \mathcal{C}}{\mathrm{minimize}} \ \mathcal{R}_\epsilon(\mathbf{x}), \tag{44}$$

where $\mathbf{x} = (\boldsymbol{\mu}, \mathbf{Q}, \boldsymbol{\lambda})$. Let Λ_0 and $\Lambda(\epsilon)$ denote the sets of the stationary points (or Karush-Kuhn-Tucker (KKT) pairs) of problems $\mathscr{P}_{\mathrm{GS}}$ and $\mathscr{P}_{\mathrm{GS}}(\epsilon)$ as

$$\Lambda_0 = \{\mathbf{x} \in \mathcal{C} : -\nabla_{\mathbf{x}} \mathcal{R}(\mathbf{x}) \in \mathcal{N}_{\mathcal{C}}(\mathbf{x})\}, \tag{45}$$
$$\text{and}$$
$$\Lambda(\epsilon) = \{\mathbf{x} \in \mathcal{C} : -\nabla_{\mathbf{x}} \mathcal{R}_\epsilon(\mathbf{x}) \in \mathcal{N}_{\mathcal{C}}(\mathbf{x})\}, \tag{46}$$

respectively, where $\mathcal{N}_{\mathcal{C}}(\mathbf{x})$ is the normal cone [48] to the convex set $\mathcal{C}$ at $\mathbf{x}$, i.e.,

$$\mathcal{N}_{\mathcal{C}}(\mathbf{x}) = \{\mathbf{v} | \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \le 0, \forall \mathbf{y} \in \mathcal{C}\}. \tag{47}$$

We first prove that

$$\limsup_{\epsilon \searrow 0} \Lambda(\epsilon) \subset \Lambda_0. \tag{48}$$

Assuming that for any $\mathbf{x}^\star \in \limsup_{\epsilon \searrow 0} \Lambda(\epsilon)$, there exists $\epsilon_k \searrow 0$ and $\mathbf{x}_k \in \Lambda(\epsilon_k)$ such that $\mathbf{x}_k \to \mathbf{x}^\star$. Based on [48, Proposition 6.6], we have that

$$\limsup_{\mathbf{x}_k \to \mathbf{x}^\star} \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k) = \mathcal{N}_{\mathcal{C}}(\mathbf{x}^\star). \tag{49}$$

Furthermore, we have

$$-\nabla_{\mathbf{x}} \mathcal{R}_{\epsilon_k}(\mathbf{x}_k) \in \mathcal{N}_{\mathcal{C}}(\mathbf{x}_k), \tag{50}$$

and

$$\begin{aligned}
\lim_{k \to +\infty} \nabla_{\mathbf{x}} \mathcal{R}_{\epsilon_k}(\mathbf{x}_k) &= \lim_{\epsilon_k \searrow 0} \lim_{\mathbf{x}_k \to \mathbf{x}^\star} \nabla_{\mathbf{x}} \mathcal{R}_{\epsilon_k}(\mathbf{x}_k) \\
&= \lim_{\epsilon_k \searrow 0} \nabla_{\mathbf{x}} \mathcal{R}_{\epsilon_k}(\mathbf{x}^\star) \\
&= \nabla_{\mathbf{x}} \mathcal{R}(\mathbf{x}^\star).
\end{aligned} \tag{51}$$

Therefore, taking $k \to +\infty$ in equation (50), we obtain that $\mathbf{x}^\star \in \Lambda_0$. We thus complete the proof for (48).

Define the deviation of a given set A_1 from another set A_2 as [24]

$$\mathbb{D}(A_1, A_2) = \sup_{x_1 \in A_1} \left(\inf_{x_2 \in A_2} \|x_1 - x_2\| \right). \tag{52}$$

Based on the conclusion (48) and [49, Theorem 4], we complete the proof for the conclusion (35).

ACKNOWLEDGMENT

REFERENCES

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, pp. 56–61, Aug. 2006.

[2] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1910–1921, Apr. 2009.

[3] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, pp. 1380–1408, Sep. 2010.

[4] China Mobile, "C-RAN: The road towards green RAN," White Paper, ver. 3.0, Dec. 2013.

[5] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 2809–2823, May 2014.

[6] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen, "Large-scale convex optimization for ultra-dense Cloud-RAN," *IEEE Wireless Commun. Mag.*, 2015, to be published.

[7] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, pp. 5646–5658, Nov. 2013.

[8] Y. Shi, J. Zhang, and K. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," presented at the IEEE Int. Conf. Commun. (ICC), Sydney, Australia, Jun. 2014.

[9] Y. Shi, J. Zhang, and K. Letaief, "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty," *IEEE Trans. Signal Process.*, vol. 63, pp. 960–973, Feb. 2015.

[10] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1065–1082, June 2014.

[11] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1748–1759, Sep. 2010.

[12] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative mimo techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1089–1098, Aug. 2004.

[13] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1341–1365, Oct. 2008.

[14] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, pp. 2939–2948, Oct. 2010.

[15] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, pp. 4418–4431, Jul. 2012.

[16] J. Zhang, R. W. Heath, M. Kountouris, and J. G. Andrews, "Mode switching for the multi-antenna broadcast channel based on delay and channel quantization," *EURASIP J. Adv. Signal Process. (Special Issue Multiuser Limited Feedback)*, vol. 2009, p. 15, 2009, Article ID 802548.

[17] R. Schaefer and H. Boche, "Physical layer service integration in wireless networks: Signal processing challenges," *IEEE Signal Process. Mag.*, vol. 31, pp. 147–156, May 2014.

[18] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

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

[20] N. Golrezaei, A. Molisch, A. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, pp. 142–149, Apr. 2013.

[21] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, pp. 464–501, Aug. 2011.

[22] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 5011–5023, Nov. 2009.

[23] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Trans. Signal Process.*, vol. 59, pp. 1128–1141, Mar. 2011.

[24] A. Shapiro, D. Dentcheva, and A. P. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2009, vol. 9.

[25] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[26] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, pp. 2988–3003, June 2012.

[27] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20–34, May 2010.

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

[29] E. Song, Q. Shi, M. Sanjabi, R.-Y. Sun, and Z.-Q. Luo, "Robust SINR-constrained MISO downlink beamforming: When is semidefinite programming relaxation tight?," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–11, 2012.

[30] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, pp. 1–106, Jan. 2012.

[31] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 226–240, Feb. 2013.

[32] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, pp. 2660–2674, May 2013.

[33] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, pp. 2508–2523, May 2012.

[34] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Found. Trends Commun. Inf. Theory*, vol. 9, pp. 113–381, Jan. 2013.

[35] E. J. Candes and X. Li, "Solving quadratic equations via phaselift when there are about as many equations as unknowns," *Found. Comput. Math.*, pp. 1–10, Junr 2012.

[36] S. Leyffer, *Mixed Integer Nonlinear Programming*. New York, NY, USA: Springer, 2012, vol. 154.

[37] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Signal Process.*, vol. 61, pp. 3972–3987, Aug. 2013.

[38] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud radio access networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 4635–4640.

[39] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, 2008.

[40] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.

[41] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Mach. Learn.*, vol. 73, no. 3, pp. 243–272, 2008.

[42] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 199–225, 2013.

[43] 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, pp. 1–122, Jul. 2011.

[44] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," 2013, arXiv Preprint arXiv:1312.3039 [Online]. Available: http://arxiv.org/abs/1312.3039

[45] Y. Shi, J. Zhang, and K. B. Letaief, "Scalable coordinated beamforming for dense wireless cooperative networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 3603–3608.

[46] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, 2015, DOI: 10.1109/TSP.2015.2443731, preprint.

[47] J. Cheng, Y. Shi, B. Bai, W. Chen, J. Zhang, and K. Letaief, "Group sparse beamforming for multicast green Cloud-RAN via parallel semidefinite programming," presented at the IEEE Int. Conf. Commun. (ICC), London, U.K., 2015.

[48] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. : Springer, 1998, vol. 317.
[49] L. J. Hong, Y. Yang, and L. Zhang, "Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach," *Oper. Res.*, vol. 59, pp. 617–630, May–June 2011.
[50] G. Obozinski and F. Bach, "Convex relaxation for combinatorial penalties," 2012, arXiv Preprint arXiv:1205.1240 [Online]. Available: http://arxiv.org/abs/1205.1240
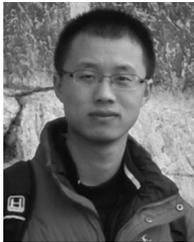
Yuanming Shi (S'13) received his B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011. He is currently working towards the Ph.D. degree in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST). His research interests include large-scale convex optimization and analysis, Riemannian optimization, stochastic optimization, mobile edge computing, and dense wireless networking.

Jun Zhang (S'06–M'10) received the B.Eng. degree in electronic engineering from the University of Science and Technology of China in 2004, the M.Phil. degree in information engineering from the Chinese University of Hong Kong in 2006, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2009. He is currently a Research Assistant Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST). Dr. Zhang co-authored the book *Fundamentals of LTE* (Prentice-Hall, 2010). He received the 2014 Best Paper Award for the EURASIP Journal on Advances in Signal Processing, and the PIMRC 2014 Best Paper Award. He served as a MAC track co-chair for IEEE WCNC 2011. His research interests include wireless communications and networking, green communications, and signal processing.

Khaled B. Letaief (S'85–M'86–SM'97–F'03) received the BS degree *with distinction* in electrical engineering (1984) from Purdue University, USA. He has also received the MS and Ph.D. degrees in electrical engineering from Purdue University in 1986 and 1990, respectively.

From January 1985 and as a Graduate Instructor at Purdue, he has taught courses in communications and electronics. From 1990 to 1993, he was a faculty member at the University of Melbourne, Australia. Since 1993, he has been with the Hong Kong University of Science and Technology (HKUST) where he is currently Chair Professor and the Dean of Engineering with expertise in wireless communications and networks. In these areas, he has over 500 journal and conference papers and given invited keynote talks as well as courses all over the world. He has 13 patents including 11 US patents.

Dr. Letaief served as consultants for different organizations and is the founding Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has served on the editorial board of other prestigious journals including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS-WIRELESS SERIES (as Editor-in-Chief). He has been involved in organizing a number of major international conferences. These include WCNC'07 in Hong Kong; ICC'08 in Beijing; ICC'10 in Cape Town, TTM'11 in Hong Kong, and ICCC'12 in Beijing.

Professor Letaief has been a dedicated teacher committed to excellence in teaching and scholarship. He received the *Mangoon Teaching Award* from Purdue University in 1990; HKUST Engineering Teaching Excellence Award; and Michael Gale Medal for Distinguished Teaching (*Highest university-wide teaching award* at HKUST). He is also the recipient of many other distinguished awards including 2007 IEEE Communications Society Publications Exemplary Award; 2009 IEEE Marconi Prize Award in Wireless Communications; 2010 Purdue University Outstanding Electrical and Computer Engineer Award; 2011 IEEE Communications Society Harold Sobol Award; 2011 IEEE Wireless Communications Technical Committee Recognition Award; and 12 IEEE Best Paper Awards.

Dr. Letaief is a Fellow of IEEE and a Fellow of HKIE. He served as an elected member of the IEEE Communications Society (ComSoc) Board of Governors, IEEE Distinguished lecturer, IEEE ComSoc Treasurer, and IEEE ComSoc Vice-President for Conferences.

He is currently serving as IEEE ComSoc Vice-President for Technical Activities, member of the IEEE Product Services and Publications Board, and member of the IEEE Fellow Committee. He is also recognized by Thomson Reuters as an *ISI Highly Cited Researcher*.