

# Enhanced Group Sparse Beamforming for Green Cloud-RAN: A Random Matrix Approach

Yuanming Shi<sup>1</sup>, Member, IEEE, Jun Zhang<sup>2</sup>, Senior Member, IEEE, Wei Chen<sup>3</sup>, Senior Member, IEEE, and Khaled B. Letaief, Fellow, IEEE

**Abstract**—Group sparse beamforming is a general framework to minimize the network power consumption for cloud radio access networks, which, however, suffers high computational complexity. In particular, a complex optimization problem needs to be solved to obtain the remote radio head (RRH) ordering criterion in each transmission block, which will help to determine the active RRHs and the associated fronthaul links. In this paper, we propose innovative approaches to reduce the complexity of this key step in group sparse beamforming. Specifically, we first develop a smoothed  $\ell_p$ -minimization approach with the iterative reweighted- $\ell_2$  algorithm to return a Karush–Kuhn–Tucker (KKT) point solution, as well as enhance the capability of inducing group sparsity in the beamforming vectors. By leveraging the Lagrangian duality theory, we obtain closed-form solutions at each iteration to reduce the computational complexity. The well-structured solutions provide opportunities to apply the large-dimensional random matrix theory to derive deterministic approximations for the RRH ordering criterion. Such an approach helps to guide the RRH selection only based on the statistical channel state information, which does not require frequent update, thereby significantly reducing the computation overhead. Simulation results shall demonstrate the performance gains of the proposed  $\ell_p$ -minimization approach, as well as the effectiveness of the large system analysis-based framework for computing the RRH ordering criterion.

**Index Terms**—Cloud-RAN, green communications, sparse optimization, smoothed  $\ell_p$ -minimization, Lagrangian duality, and random matrix theory.

Manuscript received March 20, 2017; revised July 21, 2017 and November 19, 2017; accepted January 14, 2018. Date of publication February 6, 2018; date of current version April 8, 2018. This work was supported in part by the National Nature Science Foundation of China under Grant 61601290, in part by the Shanghai Sailing Program under Grant 16YF1407700, in part by the Hong Kong Research Grant Council under Grant 16200214, in part by the National Natural Science Foundation of China under Project 61671269 and Project 61621091, and in part by the Chinese National 973 Program under Project 2013CB336600. This paper was presented at the IEEE International Symposium on Information Theory, Barcelona, Spain, July 2016 [1]. The associate editor coordinating the review of this paper and approving it for publication was L. Song. (Corresponding author: Yuanming Shi.)

Y. Shi is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: shiym@shanghaitech.edu.cn).

J. Zhang is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eejzhang@ust.hk).

W. Chen is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wchen@tsinghua.edu.cn).

K. B. Letaief is with Hamad Bin Khalifa University, Doha, Qatar, and also with the Hong Kong University of Science and Technology, Hong Kong (e-mail: eekhaled@ust.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2018.2797203

## I. INTRODUCTION

NETWORK densification [2]–[4] has been proposed as a promising way to provide ultra-high data rates, achieve low latency, and support ubiquitous connectivity for the upcoming 5G networks [5]. However, to fully harness the benefits of dense wireless networks, formidable challenges arise, including interference management, radio resource allocation, mobility management, as well as high capital expenditure and operating expenditure. Cloud-RAN emerges as a disruptive technology to deploy cost-effective dense wireless networks [6], [7]. It can significantly improve both energy and spectral efficiency, by leveraging recent advances in cloud computing and network function virtualization [8]. With shared computation resources at the cloud data center and distributed low-cost low-power remote radio heads (RRHs), Cloud-RAN provides an ideal platform to achieve the benefits of network cooperation and coordination [9]. There is a unique characteristic in Cloud-RANs, namely, the high-capacity fronthaul links are required to connect the cloud center and RRHs [10], [11]. Such links will consume power comparable to that of each RRH, and thus brings new challenges to design green Cloud-RANs. To address this issue, a new performance metric, i.e., the *network power consumption*, which consists of both the fronthaul link power consumption and RRH transmit power consumption, has been identified in [7] for the green Cloud-RAN design.

Unfortunately, the network power minimization problem turns out to be a mixed-integer nonlinear programming problem [12], which is highly intractable. Specifically, the combinatorial composite objective function involves a *discrete* component, indicating which RRHs and the corresponding fronthaul links should be switched on, and a *continuous* component, i.e., beamformers to reduce RRH transmit power. To address this challenge, a unique group sparsity structure in the optimal beamforming vector has been identified in [7] to unify RRH selection and beamformer optimization. Accordingly, a novel three-stage *group sparse beamforming* framework was proposed to promote group sparsity in the solution. Specifically, a mixed  $\ell_1/\ell_2$ -minimization approach was first proposed to induce group sparsity in the solution, thereby guiding the RRH selection via solving a sequence of convex feasibility problems, followed by coordinated beamforming to minimize the transmit power for the active RRHs.

Although the group sparse beamforming framework provides polynomial-time complexity algorithms via convex

optimization, it has the limited capability to enhance group sparsity compared to the non-convex approaches [13], [14]. In particular, the smoothed  $\ell_p$ -minimization supported by the iterative reweighted- $\ell_2$  algorithm was developed in [15] to enhance sparsity for multicast group sparse beamforming. However, the computation burden of this iterative algorithm is still prohibitive in dense wireless networks with a large number of RRHs and mobile users. In [16], a generic two-stage approach was proposed to solve large-scale convex optimization problems in dense wireless cooperative networks via matrix stuffing and operator splitting, i.e., the alternating direction method of multipliers (ADMM) [17], [18]. This approach also enables parallel computing and infeasibility detection. However, as the proposed solutions need to be accomplished for each channel realization, and also due to the iterative procedures of the group sparse beamforming framework, it is still computationally expensive.

In this paper, we improve the performance of group sparse beamforming [7] via  $\ell_p$ -minimization, and a special emphasis is on reducing the computational complexity, with two ingredients: closed-form solutions for each iteration, and the deterministic equivalents of the optimal Lagrange multipliers. Specifically, we first develop an iterative reweighted- $\ell_2$  algorithm with closed-form solutions at each iteration via leveraging the principles of the majorization-minimization (MM) algorithm [19] and the Lagrangian duality theory [20]. It turns out that the proposed iterative reweighted- $\ell_2$  algorithm can find a KKT point for the smoothed non-convex  $\ell_p$ -minimization problem. Furthermore, this reveals the explicit structures of the optimal solutions to each subproblem in the iterative reweighted- $\ell_2$  algorithm, while the convex optimization approach in [7], [15], and [21] fails to obtain the closed-form solutions. Thereafter, the well-structured closed-form solutions provide opportunities to leverage the large-dimensional random matrix theory [22]–[24] to perform the asymptotic analysis in the large system regimes [25], [26]. Specifically, the deterministic equivalents of the optimal Lagrange multipliers are derived based on the recent results of large random matrix analysis [27] to decouple the dependency of system parameters. These results are further used to perform an asymptotic analysis for computing the RRH ordering criterion, which only depends on long-term channel attenuation, thereby significantly reducing the computation overhead compared with previous algorithms that heavily depends on instantaneous CSI [7], [15], [21], [28]–[31].

Based on the above proposal, we provide a three-stage enhanced group sparse beamforming framework for green Cloud-RAN via random matrix theory. Specifically, in the first stage, we compute the enhanced RRH ordering criterion only based on the statistical CSI. This thus avoids frequent updates, thereby significantly reducing computational complexity, which is a sharp difference compared to the original proposal in [8]. This new algorithm is based on the principles of the MM algorithm and the Lagrangian duality theory, followed by the random matrix theory. With the obtained deterministic equivalents of RRH ordering criterion, a two-stage large-scale convex optimization framework [16] is adopted to solve a sequence of convex feasibility problems to determine

the active RRHs in the second stage, as well as solving the transmit power minimization problem for the active RRHs in the final stage. Note that global instantaneous CSI is required for these two stages. Simulation results are provided to demonstrate the improvement of the enhanced group sparse beamforming framework. Moreover, the deterministic approximations turn out to be accurate even in the finite-sized systems.

#### A. Related Works

1) *Sparse Optimization in Wireless Networks*: The sparse optimization paradigm has recently been popular for complicated network optimization problems in wireless network design, e.g., the group sparse beamforming framework for green Cloud-RAN design [7], [15], [30]–[32], wireless caching networks [33], [34], user admission control [15], [35], as well as computation offloading [36]. In particular, the convex relaxation approach provides a principled way to induce sparsity via the  $\ell_1$ -minimization [35] for individual sparsity inducing and the mixed  $\ell_1/\ell_2$ -minimization for group sparsity inducing [7]. The reweighted- $\ell_1$  algorithm [13] and the reweighted- $\ell_2$  algorithm [15] were further developed to enhance sparsity. To enable parallel and distributed computing, the first-order method ADMM algorithm was adopted to solve the group sparse beamforming problems [37]. A generic large-scale convex optimization framework was further proposed to solve general large-scale convex programs in dense wireless networks to enable scalability, parallel computing and infeasibility detection [16].

However, all the above algorithms need to be computed for each channel realization, which is computationally expensive. In this paper, we adopt large system analysis to compute the RRH ordering criterion for network adaptation only based on statistical CSI.

2) *Large System Analysis via Random Matrix Theory*: Random matrix theory [22] has been proven powerful for performance analysis, and the understanding and improving algorithms in wireless communications [23], [38], signal processing [39], and machine learning [40], [41], especially in large dimensional regimes for applications in the era of big data [42]. In dense wireless networks, random matrix theory provides a powerful way for performance analysis and algorithm design. In particular, the large system analysis was performed for simple precoding schemes, e.g., regularized zero-forcing in MISO broadcast channels with imperfect CSI [24]. A random matrix approach to the optimal coordinated multicell beamforming for massive MIMO was presented in [43] without closed-form expressions for the optimal Lagrange multipliers. A simple channel model with two cells under different coordination levels was considered in [25].

However, all of the above results cannot be directly applied in the group sparse beamforming framework in dense Cloud-RAN due to the general channel models with heterogeneous pathloss and the complicated beamformer structures with fully cooperative transmission. To determine the RRH ordering criterion based on statistical CSI, we adopt the technique in [26] and [27] to compute the closed-forms for the optimal Lagrange multipliers for each iteration in the procedure of reweighted- $\ell_2$  minimization for sparsity inducing.

## B. Organization

The remainder of the paper is organized as follows. Section II presents the system model and problem formulation, followed by performance analysis with the proposed three-stage enhanced group sparse beamforming framework for green Cloud-RAN. In Section III, the iterative reweighted- $\ell_2$  algorithm for group sparse beamforming is presented via the principle of MM algorithm and duality theory. The large system analysis for RRH ordering is performed in Section IV. Simulation results will be demonstrated in Section V. Finally, conclusions and discussions are presented in Section VI. To keep the main text clean and free of technical details, we divert most of the proofs to the Appendix.

*Notations:* Throughout this paper,  $\|\cdot\|_p$  is the  $\ell_p$ -norm.  $|\cdot|$  stands for either the size of a set or the absolute value of a scalar, depending on the context.  $\|\mathbf{M}\|$  is the spectral radius of the Hermitian matrix  $\mathbf{M}$ . Boldface lower case and upper case letters represent vectors and matrices, respectively.  $(\cdot)^{-1}$ ,  $(\cdot)^T$ ,  $(\cdot)^H$  and  $\text{Tr}(\cdot)$  denote the inverse, transpose, Hermitian and trace operators, respectively. We use  $\mathbb{C}$  to represent complex domain.  $\mathbb{E}[\cdot]$  denotes the expectation of a random variable. We denote  $\mathbf{A} = \text{diag}\{x_1, \dots, x_N\}$  and  $\mathbf{I}_N$  as a diagonal matrix of order  $N$  and the identity matrix of order  $N$ , respectively.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

Consider a Cloud-RAN with  $L$  RRHs and  $K$  single-antenna mobile users (MUs), where each RRH is equipped with  $N$  antennas. In Cloud-RANs, the BBU pool will perform the centralized signal processing and is connected to all the RRHs via high-capacity and low-latency fronthaul links. In this paper, we will focus on the downlink signal processing. Specifically, let  $\mathbf{v}_{lk} \in \mathbb{C}^N$  be the transmit beamforming vector from the  $l$ -th RRH to the  $k$ -th MU. The received signal  $y_k \in \mathbb{C}$  at MU  $k$  is given by

$$y_k = \sum_{l=1}^L \mathbf{h}_{kl}^H \mathbf{v}_{lk} s_k + \sum_{i \neq k} \sum_{l=1}^L \mathbf{h}_{kl}^H \mathbf{v}_{li} s_i + n_k, \quad (1)$$

where  $\mathbf{h}_{kl} \in \mathbb{C}^N$  is the channel propagation between MU  $k$  and RRH  $l$ ,  $s_k \in \mathbb{C}$  with  $\mathbb{E}[|s_k|^2] = 1$  is the encoded transmission information symbol for MU  $k$ , and  $n_k \sim \mathcal{CN}(0, \sigma_k^2)$  is the additive Gaussian noise at MU  $k$ .

We assume that  $s_k$ '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  $k$  is given by

$$\text{sinr}_k = \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2}, \quad (2)$$

where  $\mathbf{h}_k = [\mathbf{h}_{k1}^T, \dots, \mathbf{h}_{kL}^T]^T \in \mathbb{C}^{LN}$  is the vector consisting of the channel coefficients from all the RRHs to MU  $k$ , and  $\mathbf{v}_k = [\mathbf{v}_{1k}^T, \dots, \mathbf{v}_{Lk}^T]^T \in \mathbb{C}^{LN}$  is the aggregative beamforming vector for the MU  $k$  from all the RRHs.

### B. Problem Formulation

In this paper, we aim at designing a green Cloud-RAN by minimizing the network power consumption, which consists of the fronthaul links power consumption, as well as the RRH transmit power consumption [7]. Specifically, let  $\mathbf{v} = [\mathbf{v}_{lk}] \in \mathbb{C}^{KLN}$  be the aggregative beamforming vector from all the RRHs to all the MUs. Define the support of the beamforming vector  $\mathbf{v}$  as  $\mathcal{T}(\mathbf{v}) = \{i | v_i \neq 0\}$ , where  $\mathbf{v} = [v_i]$  is indexed by  $i \in \mathcal{V}$  with  $\mathcal{V} = \{1, \dots, KLN\}$ . The relative fronthaul network power consumption is given by

$$f_1(\mathbf{v}) = \sum_{l=1}^L P_l^c I(\mathcal{T}(\mathbf{v}) \cap \mathcal{V}_l \neq \emptyset), \quad (3)$$

where  $P_l^c \geq 0$  is the relative fronthaul link power consumption [7] (i.e., the static power saving when both the fronthaul link and the corresponding RRH are switched off),  $\mathcal{V}_l = \{(l-1)KN + 1, \dots, lKN\}$ ,  $\forall l \in \mathcal{L}$ , is a partition of  $\mathcal{V}$ , and  $I(\mathcal{T} \cap \mathcal{V}_l \neq \emptyset)$  is an indicator function that takes value 1 if  $\mathcal{T} \cap \mathcal{V}_l \neq \emptyset$  and 0 otherwise. Note that  $f_1$  is a non-convex combinatorial function.

Furthermore, the total transmit power consumption is given by

$$f_2(\mathbf{v}) = \sum_{l=1}^L \sum_{k=1}^K \frac{1}{\zeta_l} \|\mathbf{v}_{lk}\|_2^2, \quad (4)$$

where  $\zeta_l > 0$  is the drain inefficiency coefficient of the radio frequency power amplifier [7]. Therefore, the network power consumption is represented by the combinatorial composite function

$$f(\mathbf{v}) = f_1(\mathbf{v}) + f_2(\mathbf{v}). \quad (5)$$

Given the QoS thresholds  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$  for all the MUs, in this paper, we aim at solving the following network power consumption minimization problem with the QoS constraints:

$$\begin{aligned} \mathcal{P}: \quad & \underset{\mathbf{v}}{\text{minimize}} \quad f_1(\mathbf{v}) + f_2(\mathbf{v}) \\ & \text{subject to} \quad \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k. \end{aligned} \quad (6)$$

Let  $\tilde{\mathbf{v}}_l = [\mathbf{v}]_{\mathcal{V}_l} = [\mathbf{v}_{l1}^T, \dots, \mathbf{v}_{lK}^T]^T \in \mathbb{C}^{KN}$  form the beamforming coefficient group from RRH  $l$  to all the MUs. Note that, when the RRH  $l$  is switched off, all the beamforming coefficients in  $\tilde{\mathbf{v}}_l$  will be set to zeros simultaneously. Observing that there may be multiple RRHs being switched off to minimize the network power consumption, the optimal beamforming vector  $\mathbf{v} = [\tilde{\mathbf{v}}_1^T, \dots, \tilde{\mathbf{v}}_L^T]^T \in \mathbb{C}^{KLN}$  should have a group-sparsity structure. Therefore, problem  $\mathcal{P}$  is called a *group sparse beamforming problem* [7]. Note that, to simplify the presentation, we only impose the QoS constraints in problem  $\mathcal{P}$ . However, the proposed instantaneous CSI based iterative reweighted- $\ell_2$  algorithm can be extended to the scenario with per-RRH transmit power constraints, following the principles in [20]. More technical efforts are required to derive the asymptotic results using the random matrix method with more complicated structures of the optimal beamformers.

### C. Problem Analysis

Although the constraints in problem  $\mathcal{P}$  can be reformulated as convex second-order cone constraints, the non-convex objective function makes it highly intractable. To address this challenge, a weighted mixed  $\ell_1/\ell_2$ -norm minimization approach was proposed in [7] to convexify the objective and induce the group sparsity in beamforming vector  $\mathbf{v}$ , thereby guiding the RRH ordering to enable adaptively RRH selection. Specifically, we will first solve the  $\ell_1/\ell_2$ -norm minimization problem, and denote  $\tilde{\mathbf{v}}_1^*, \dots, \tilde{\mathbf{v}}_L^*$  as the induced (approximated) group sparse beamforming vectors. Then the following RRH ordering criterion is adopted to determine which RRHs should be switched off [7]:

$$\theta_l = \kappa_l \|\tilde{\mathbf{v}}_l^*\|_2^2, \quad \forall l = 1, \dots, L, \quad (7)$$

where  $\kappa_l = \sum_{k=1}^K \|\mathbf{h}_{kl}\|_2^2 / \nu_l$ . In particular, the group sparsity structure information for beamforming vector  $\mathbf{v}$  is extracted from the squared  $\ell_2$ -norm of the beamforming vectors  $\tilde{\mathbf{v}}_l$ 's, i.e.,  $\|\tilde{\mathbf{v}}_1\|_2^2, \dots, \|\tilde{\mathbf{v}}_L\|_2^2$ . The RRH with a smaller  $\theta_l$  will have a higher priority to be switched off. Based on the RRH ordering result  $\theta^*$  in (7), a bi-section search approach can be used to find the optimal active RRHs [7] via solving a sequence of the following feasibility problems:

$$\begin{aligned} \mathcal{F}(\mathcal{A}^{[i]}): \quad & \text{find } \mathbf{v}_1, \dots, \mathbf{v}_K \\ & \text{subject to } \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k, \end{aligned} \quad (8)$$

where  $\mathbf{v}_k = [\mathbf{v}_{kl}] \in \mathbb{C}^{|\mathcal{A}|N}$  and  $\mathbf{h}_k = [\mathbf{h}_{kl}] \in \mathbb{C}^{|\mathcal{A}|N}$ . Problem  $\mathcal{F}(\mathcal{A}^{[i]})$  turns out to be convex via reformulating the QoS constraints as second-order cone constraints [7]. To further enhance the sparsity as well as to seek the quadratic forms of the beamforming vectors in the multicast transmission setting, a smoothed  $\ell_p$ -minimization approach was proposed in [15]. To scale to large problem sizes in dense Cloud-RANs, a two-stage large-scale parallel convex optimization framework was developed in [16] with the capability of infeasibility detection.

However, all of the above developed algorithms bear high computation overhead. In particular, while the feasibility problem (8) for RRH selection can be efficiently solved with the large-scale optimization algorithm in [8], the RRH ordering criterion in (7) may be highly complicated to obtain, especially with the non-convex formulation as in [15]. Moreover, the ordering criterion needs to be recomputed for each transmission slot, and depends on instantaneous CSI. Observing that the statistical CSI normally changes much slower than the instantaneous CSI, to reduce the computational burden, we propose to compute the RRH ordering criterion (7) only based on statistical CSI, i.e., the long-term channel attenuation. This is achieved by first developing a group sparsity penalty with quadratic forms in the aggregative beamforming vector  $\mathbf{v}$ , followed by an iterative reweighted- $\ell_2$  algorithm with closed-form solutions at each iteration via Lagrangian duality theory, as will be presented in Section III. Then asymptotic analysis is performed to obtain the RRH ordering criterion (7) based only on statistical CSI by leveraging the large-dimensional random matrix theory [23], [24], [27], as will be presented in Section IV.

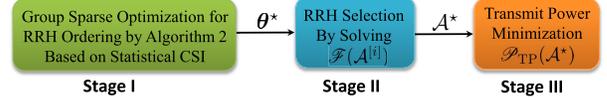


Fig. 1. The proposed three-stage enhanced group sparse beamforming framework for dense green Cloud-RAN. In the first stage, the RRH ordering criterion  $\theta^*$  is computed only based on the statistical CSI via large system analysis. The optimal active RRHs  $\mathcal{A}^*$  in the second stage and the optimal coordinated beamforming for transmit power minimization in the third stage are computed based on the instantaneous CSI via the large-scale convex optimization algorithm in [16].

Overall, the proposed three-stage enhanced group sparse beamforming framework is presented in Fig. 1. Specifically, in the first stage, the RRH ordering criterion  $\theta^*$  is calculated only based on statistical CSI using Algorithm 2. In the second stage, the set of active RRHs  $\mathcal{A}^*$  is obtained based on instantaneous CSI via solving a sequence of convex feasibility problems  $\mathcal{F}(\mathcal{A}^{[i]})$  using the large-scale convex optimization framework in [16]. In the third stage, the transmit power is minimized by solving the convex program (6) with the fixed active RRHs  $\mathcal{A}^*$  using the large-scale convex optimization algorithm in [16]. Overall, the proposed enhanced group sparse beamforming framework is scalable to large network sizes. This paper will focus on developing an effective algorithm for the first stage.

*Remark 1:* In this paper, we assume that, in the first stage, problem  $\mathcal{P}$  (6) is feasible for developing Algorithm 1 and Algorithm 2 to determine the RRH ordering criterion based on the instantaneous CSI and statistical CSI, respectively. We enable the capability of handling the infeasibility in the second stage based on the instantaneous CSI, using the large-scale convex optimization algorithm [16]. Furthermore, when problem  $\mathcal{F}$  (8) is infeasible with all RRHs active, we adopt the user admission algorithm to find the maximum number of admitted users [15].

### III. GROUP SPARSE BEAMFORMING VIA ITERATIVE REWEIGHTED- $\ell_2$ ALGORITHM

In this section, we develop a group sparse beamforming approach based on the smoothed  $\ell_p$ -minimization, supported by an iterative reweighted- $\ell_2$  algorithm, thereby enhancing the group sparsity in the beamforming vectors. Instead of reformulating the QoS constraints in problem  $\mathcal{P}$  as second-order cone constraints [7], we use the Lagrangian duality theory to reveal the structures of the optimal solutions at each iteration. The results will assist the large system analysis in Section IV.

#### A. Group Sparsity Inducing Optimization via Smoothed $\ell_p$ -Minimization

To induce the group sparsity structure in the beamforming vector  $\mathbf{v}$ , thereby guiding the RRH ordering, the weighted mixed  $\ell_1/\ell_2$ -norm was proposed in [7]. However, the non-smooth mixed  $\ell_1/\ell_2$ -norm fails to introduce the quadratic forms in the beamforming vector  $\mathbf{v}$ , in order to be compatible with the quadratic QoS constraints so that the Lagrangian

duality theory can be applied [20]. To address this challenge, we adopt the following smoothed  $\ell_p$ -minimization ( $0 < p \leq 1$ ) approach to induce group sparsity [14], [15]:

$$\begin{aligned} \mathcal{P}_{\text{GSBF}}: \underset{\mathbf{v}}{\text{minimize}} \quad & g_p(\mathbf{v}; \epsilon) := \sum_{l=1}^L v_l \left( \|\tilde{\mathbf{v}}_l\|_2^2 + \epsilon^2 \right)^{p/2} \\ \text{subject to} \quad & \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k, \end{aligned} \quad (9)$$

where  $\epsilon > 0$  is some fixed regularizing parameter and  $v_l > 0$  is the weight for the beamforming coefficient group  $\tilde{\mathbf{v}}_l$  by encoding the prior information of system parameters [7]. Compared with the mixed  $\ell_1/\ell_2$ -norm minimization approach, the  $\ell_p$ -minimization approach can induce sparser solutions based on the fact  $\|z\|_0 = \lim_{p \rightarrow 0} \|z\|_p^p = \lim_{p \rightarrow 0} \sum_i |z_i|^p$ . Unfortunately, problem (9) is non-convex due to the non-convexity of both the objective function and the QoS constraints.

### B. Iterative Reweighted- $\ell_2$ Algorithm

We use the MM algorithm and the Lagrangian duality theory to solve problem (9) with closed-form solutions. Specifically, this approach generates the iterates  $\{\mathbf{v}^{[n]}\}$  by successively minimizing upper bounds  $Q(\mathbf{v}; \mathbf{v}^{[n]})$  of the objective function  $g_p(\mathbf{v}; \epsilon)$ . We adopt the following upper bounds to approximate the smoothed  $\ell_p$ -norm  $g(\mathbf{v}; \epsilon)$  by leveraging the results of the expectation-maximization (EM) algorithm [44].

*Proposition 1:* Given the value of  $\mathbf{v}^{[n]}$  at the  $n$ -th iteration, an upper bound for the objective function  $g_p(\mathbf{v}; \epsilon)$  can be constructed as follows:

$$Q(\mathbf{v}; \boldsymbol{\omega}^{[n]}) := \sum_{l=1}^L \omega_l^{[n]} \|\tilde{\mathbf{v}}_l\|_2^2, \quad (10)$$

where

$$\omega_l^{[n]} = \frac{pv_l}{2} \left[ \|\tilde{\mathbf{v}}_l^{[n]}\|_2^2 + \epsilon^2 \right]^{\frac{p}{2}-1}, \quad \forall l = 1, \dots, L. \quad (11)$$

*Proof:* The proof is mostly based on [15, Proposition 1]. ■

Therefore, at the  $n$ -th iteration, we need to solve the following optimization problem:

$$\begin{aligned} \mathcal{P}_{\text{GSBF}}^{[n]}: \underset{\mathbf{v}}{\text{minimize}} \quad & \sum_{l=1}^L \omega_l^{[n]} \|\tilde{\mathbf{v}}_l\|_2^2 \\ \text{subject to} \quad & \frac{|\mathbf{h}_k^H \mathbf{v}_k|^2}{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k, \end{aligned} \quad (12)$$

which is non-convex due to the non-convex QoS constraints. Although the QoS constraints can be reformulated as second-order cone constraints as in [7], in this paper, we leverage the Lagrangian duality theory to obtain explicit structures of the optimal solution to problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ , thereby reducing the computational complexity and further aiding the large system analysis in next section. The success of applying the Lagrangian duality approach is based on the fact that strong duality holds for problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$  [45], i.e., the gap between the primal optimal objective and the dual optimal objective is zero.

1) *Simple Solution Structures:* As the strong duality holds for problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$  [45], we solve it using duality theory. Specifically, let  $\lambda_k/(NL) \geq 0$  denote the Lagrange multipliers corresponding to the QoS constraints in problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ . Define the Lagrangian function  $\mathcal{L}(\{\mathbf{v}_k\}, \boldsymbol{\lambda})$  with  $\boldsymbol{\lambda} = \{\lambda_i\}$  as

$$\begin{aligned} \mathcal{L}(\{\mathbf{v}_k\}, \boldsymbol{\lambda}) = \sum_{k=1}^K \frac{\lambda_k}{LN} \left( \sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2 - \frac{1}{\gamma_k} |\mathbf{h}_k^H \mathbf{v}_k|^2 \right) \\ + \sum_{k=1}^K \mathbf{v}_k^H \mathbf{Q}^{[n]} \mathbf{v}_k, \end{aligned} \quad (13)$$

where  $\mathbf{Q}^{[n]} \in \mathbb{C}^{LN \times LN}$  is a block diagonal matrix with the scaled identity matrix  $\omega_l^{[n]} \mathbf{I}_N$  as the  $l$ -th main diagonal block square matrix.

To find the optimal  $\mathbf{v}_k$ 's, we take the gradient of the Lagrangian function  $\mathcal{L}(\mathbf{v}, \boldsymbol{\lambda})$  with respect to  $\mathbf{v}_k$  and set it to zero, which implies

$$\mathbf{Q}^{[n]} \mathbf{v}_k + \sum_{i \neq k} \frac{\lambda_k}{LN} \mathbf{h}_i \mathbf{h}_i^H \mathbf{v}_k - \frac{\lambda_k}{LN \gamma_k} \mathbf{h}_k \mathbf{h}_k^H \mathbf{v}_k = \mathbf{0}. \quad (14)$$

By adding  $\lambda_k/(LN) \times \mathbf{h}_k \mathbf{h}_k^H \mathbf{v}_k$  to both sides of (14), we have

$$\left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \frac{\lambda_i}{LN} \mathbf{h}_i \mathbf{h}_i^H \right) \mathbf{v}_k = \frac{\lambda_k}{LN} \left( 1 + \frac{1}{\gamma_k} \right) \mathbf{h}_k \mathbf{h}_k^H \mathbf{v}_k, \quad (15)$$

which implies  $\mathbf{v}_k = \left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \lambda_i/(LN) \mathbf{h}_i \mathbf{h}_i^H \right)^{-1} \mathbf{h}_k \times (\lambda_k/(LN)) (1 + 1/\gamma_k) \mathbf{h}_k^H \mathbf{v}_k$ . As  $(\lambda_k/(LN)) (1 + 1/\gamma_k) \mathbf{h}_k^H \mathbf{v}_k$  is a scalar, the optimal  $\mathbf{v}_k$  must be parallel to the beamforming direction

$$\tilde{\mathbf{v}}_k = \left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \frac{\lambda_i}{LN} \mathbf{h}_i \mathbf{h}_i^H \right)^{-1} \mathbf{h}_k, \quad \forall k. \quad (16)$$

Therefore, the optimal beamforming vectors  $\mathbf{v}_1, \dots, \mathbf{v}_K$  can be written as

$$\mathbf{v}_k = \sqrt{\frac{p_k}{LN}} \frac{\tilde{\mathbf{v}}_k}{\|\tilde{\mathbf{v}}_k\|_2}, \quad \forall k, \quad (17)$$

where  $p_k$  denotes the optimal beamforming power. Observing that the beamforming powers  $p_k$ 's need to satisfy the SINR constraints with equality [20], i.e.,

$$\frac{p_k}{\gamma_k LN} \frac{|\mathbf{h}_k^H \tilde{\mathbf{v}}_k|^2}{\|\tilde{\mathbf{v}}_k\|_2^2} - \sum_{i \neq k} \frac{p_i}{LN} \frac{|\mathbf{h}_k^H \tilde{\mathbf{v}}_i|^2}{\|\tilde{\mathbf{v}}_i\|_2^2} = \sigma_k^2, \quad \forall k, \quad (18)$$

the optimal powers thus can be obtained via solving the following linear equation:

$$\begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} = \mathbf{M}^{-1} \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_K^2 \end{bmatrix}, \quad (19)$$

where

$$[\mathbf{M}]_{ij} = \begin{cases} \frac{1}{\gamma_i LN} \frac{|\mathbf{h}_i^H \tilde{\mathbf{v}}_i|^2}{\|\tilde{\mathbf{v}}_i\|_2^2}, & i = j, \\ -\frac{1}{LN} \frac{|\mathbf{h}_i^H \tilde{\mathbf{v}}_j|^2}{\|\tilde{\mathbf{v}}_j\|_2^2}, & i \neq j. \end{cases}$$

Here,  $[\mathbf{M}]_{ij}$  denotes the  $(i, j)$ -th element of the matrix  $\mathbf{M} \in \mathbb{R}^{K \times K}$ .

To find the optimal  $\lambda$ -parameter, by multiplying both sides by  $\mathbf{h}_k^H \left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \frac{\lambda_i}{LN} \mathbf{h}_i \mathbf{h}_i^H \right)^{-1}$  in (15), we obtain  $\mathbf{h}_k^H \mathbf{v}_k$  as

$$\frac{\lambda_k}{LN} \left( 1 + \frac{1}{\gamma_k} \right) \mathbf{h}_k^H \left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \frac{\lambda_i}{LN} \mathbf{h}_i \mathbf{h}_i^H \right)^{-1} \mathbf{h}_k \mathbf{h}_k^H \mathbf{v}_k, \quad (20)$$

which implies

$$\lambda_k = LN \left[ \left( 1 + \frac{1}{\gamma_k} \right) \mathbf{h}_k^H \left( \mathbf{Q}^{[n]} + \sum_{i=1}^K \frac{\lambda_i}{LN} \mathbf{h}_i \mathbf{h}_i^H \right)^{-1} \mathbf{h}_k \right]^{-1}. \quad (21)$$

These fixed-point equations can be computed using iterative function evaluation.

Based on (17), (19) and (21), we obtain the squared  $\ell_2$ -norm of the optimal solution to problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$  as follows:

$$\|\tilde{\mathbf{v}}_l^{[n]}\|_2^2 = \sum_{k=1}^K \mathbf{v}_k^H \mathbf{Q}_{lk} \mathbf{v}_k, \quad (22)$$

where  $\mathbf{Q}_{lk} \in \mathbb{C}^{NL \times NL}$  is a block diagonal matrix with the identity matrix  $\mathbf{I}_N$  as the  $l$ -th main diagonal block square matrix and zeros elsewhere. Note that the optimal powers  $p_k$ 's (19) and Lagrangian multipliers  $\lambda_k$ 's (21) should depend on the weights  $\omega_l^{[n]}$ 's (11) at each iteration  $n$ .

The instantaneous CSI based iterative reweighted- $\ell_2$  algorithm for group sparse beamforming is presented in Algorithm 1. We have the following result for its performance and convergence.

---

**Algorithm 1** Instantaneous CSI Based Iterative Reweighted- $\ell_2$  Algorithm for Problem  $\mathcal{P}_{\text{GSBF}}$

---

**input:** Initialize  $\omega^{[0]} = (1, \dots, 1)$ ;  $I$  (the maximum number of iterations)

Repeat

1) Compute the squared  $\ell_2$ -norm of the solution to problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ ,  $\|\tilde{\mathbf{v}}_1^{[n]}\|_2^2, \dots, \|\tilde{\mathbf{v}}_L^{[n]}\|_2^2$ , using (17), (19), (21) and (22).

2) Update the weights  $\omega_l^{[n+1]}$  using (11).

Until convergence or attain the maximum iterations and return **output**.

**output:** RRH ordering criterion  $\theta$  (7).

---

*Theorem 1:* Let  $\{\mathbf{v}^{[n]}\}_{n=1}^\infty$  be the sequence generated by the iterative reweighted- $\ell_2$  algorithm. Then, every limit point  $\tilde{\mathbf{v}}$  of  $\{\mathbf{v}^{[n]}\}_{n=1}^\infty$  has the following properties:

- 1)  $\tilde{\mathbf{v}}$  is a KKT point of problem  $\mathcal{P}_{\text{GSBF}}$  (9);
- 2)  $g_p(\mathbf{v}^{[n]}; \epsilon)$  converges monotonically to  $g_p(\mathbf{v}^*; \epsilon)$  for some KKT point  $\mathbf{v}^*$ .

*Proof:* Please refer to Appendix A for details. ■

Compared with the algorithm for the smoothed- $\ell_p$  minimization in [15], the main novelty of the proposed iterative reweighted- $\ell_2$  algorithm is revealing the explicit structures of the solutions at each iteration in Algorithm 1. Instead of using

the interior-point algorithm to solve the convex subproblems at each iteration [46], Algorithm 1 with closed-form solutions helps reduce the computational cost. The proposed iterative reweighted- $\ell_2$  algorithm can only guarantee to converge to a KKT point, which may be a local minimum or the other stationary point (e.g., a saddle point and local maximum).

*Remark 2:* The main contributions of the developed Algorithm 1 include finding the closed-form solutions in the iterations, in comparison with the work [15] using the interior-point algorithm, and proving the convergence of the low-complexity closed-form iterative algorithm for the non-convex group sparse inducing problem  $\mathcal{P}_{\text{GSBF}}$  (9), instead of the convex coordinated beamforming problem [45].

Unfortunately, computing the RRH ordering criterion  $\theta_l$ 's in (7) requires to run Algorithm 1 for each channel realization, which brings a heavy computation burden. In next section, we will find deterministic approximations for  $\theta_l$ 's to determine the RRH ordering only based on statistical CSI, which changes much more slowly than instantaneous channel states, and thus requires less frequent update.

#### IV. GROUP SPARSE BEAMFORMING VIA LARGE SYSTEM ANALYSIS

In this section, we present the large system analysis for the iterative reweighted- $\ell_2$  algorithm in Algorithm 1, thereby enabling RRH ordering only based on statistical CSI. In this way, the ordering criterion will change only when the long-term channel attenuation is updated, and thus it can further reduce the computational complexity of group sparse beamforming. The main novelty of this section is providing the closed-forms for the asymptotic analysis of the optimal Lagrangian multipliers based on the recent results in [27], thereby providing explicit expressions for the asymptotic RRH ordering results.

##### A. Deterministic Equivalent of Optimal Parameters

In this subsection, we provide asymptotic analysis for the optimal beamforming parameters  $p_k$ 's (19) and  $\lambda_k$ 's (21) when  $N \rightarrow \infty$ . In a Cloud-RAN with distributed RRHs, the channel can be modeled as  $\mathbf{h}_k = \Theta_k^{1/2} \mathbf{g}_k, \forall k$ , where  $\mathbf{g}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NL})$  is the small-scale fading and  $\Theta_k = \text{diag}\{d_{k1}, \dots, d_{kL}\} \otimes \mathbf{I}_N$  with  $d_{kl}$  as the path-loss from RRH  $l$  to MU  $k$ . With this channel model, we have the following result for the deterministic equivalent of the  $\lambda$ -parameter.

*Lemma 1 (Asymptotic Results for  $\lambda$ -Parameter):* Assume

$$0 < \liminf_{N \rightarrow \infty} K/N \leq \limsup_{N \rightarrow \infty} K/N < \infty.$$

Let  $\{d_{kl}\}$  and  $\{\gamma_k\}$  satisfy  $\limsup_N \max_{k,l} \{d_{kl}\} < \infty$  and  $\limsup_N \max_k \gamma_k < \infty$ , respectively. We have  $\max_{1 \leq k \leq K} |\lambda_k - \lambda_k^\circ| \xrightarrow{N \rightarrow \infty} 0$  almost surely, where  $\lambda_k^\circ$  is given by

$$\lambda_k^\circ = \gamma_k \left( \frac{1}{L} \sum_{l=1}^L d_{kl} \eta_l \right)^{-1}. \quad (23)$$

Here,  $\eta_l$  is the unique solution of the following set of equations:

$$\eta_l = \left( \frac{1}{NL} \sum_{i=1}^K \frac{d_{il}}{\frac{1}{L} \sum_{j=1}^L d_{ij} \eta_j} \frac{\gamma_i}{1 + \gamma_i} + \omega_l^{[n]} \right)^{-1}. \quad (24)$$

*Proof:* Please refer to Appendix B for details. ■

Based on the above results, we further have the following asymptotic result for the optimal powers  $p_k$ 's.

**Lemma 2 (Asymptotic Results for Optimal Powers):** Let  $\Delta \in \mathbb{C}^{K \times K}$  be such that

$$[\Delta]_{k,i} := \frac{1}{NL} \frac{\gamma_i}{(1 + \gamma_i)^2} \frac{\psi'_{ik}}{\psi_i^{\circ 2}}. \quad (25)$$

If and only if  $\limsup_K \|\Delta\|_2 < 1$ , then  $\max_k |p_k - p_k^\circ| \xrightarrow{N \rightarrow \infty} 0$  almost surely, where  $p_k^\circ$  is given by

$$p_k^\circ = \gamma_k \frac{\psi'_k}{\psi_k^{\circ 2}} \left( \frac{\tau_k}{(1 + \gamma_k)^2} + \sigma_k^2 \right). \quad (26)$$

Here  $\psi_k^\circ$ ,  $\psi'_k$  and  $\psi'_{ik}$  are given as follows:

$$\psi_k^\circ = \frac{1}{L} \sum_{l=1}^L d_{kl} \eta_l, \quad (27)$$

and

$$\psi'_k = \frac{1}{L} \sum_{l=1}^L d_{kl} \eta_l^2 + \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_j}{(1 + \gamma_j)^2} \frac{1}{L} \sum_{l=1}^L d_{il} d_{jl} \eta_l^2, \quad (28)$$

and

$$\psi'_{ik} = \frac{1}{L} \sum_{l=1}^L d_{il} d_{kl} \eta_l^2 + \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_{jk}}{(1 + \gamma_j)^2} \frac{1}{L} \sum_{l=1}^L d_{il} d_{jl} \eta_l^2, \quad (29)$$

respectively; and  $\boldsymbol{\tau} = [\tau_1, \dots, \tau_K]^T$  is given as

$$\boldsymbol{\tau} = \sigma^2 (\mathbf{I}_K - \Delta)^{-1} \boldsymbol{\delta}, \quad (30)$$

where  $\boldsymbol{\delta} = [\delta_1, \dots, \delta_K]^T$  with

$$\delta_k = \frac{1}{NL} \sum_{i=1}^K \gamma_i \frac{\psi'_{ik}}{\psi_i^{\circ 2}}. \quad (31)$$

*Proof:* Please refer to Appendix C for details. ■

### B. Statistical CSI Based Group Sparse Beamforming Algorithm

Based on the deterministic equivalents of the optimal beamforming parameters  $\lambda_k$ 's and  $p_k$ 's in Lemma 1 and Lemma 2, respectively, we have the following theorem on the squared  $\ell_2$ -norm of the solution to problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ , i.e.,  $\|\tilde{\mathbf{v}}_1^{[n]}\|_2^2, \dots, \|\tilde{\mathbf{v}}_L^{[n]}\|_2^2$ .

**Theorem 2 (Asymptotic Results for the Beamformers):** At the  $n$ -th iteration, for the squared  $\ell_2$ -norm of the solution to problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ ,  $\|\tilde{\mathbf{v}}_1^{[n]}\|_2^2, \dots, \|\tilde{\mathbf{v}}_L^{[n]}\|_2^2$ , we have  $\max_l \left| \|\tilde{\mathbf{v}}_l^{[n]}\|_2^2 - \bar{\chi}_l \right| \xrightarrow{N \rightarrow \infty} 0$  almost surely with

$$\bar{\chi}_l = \frac{1}{NL} \sum_{k=1}^K p_k^\circ \frac{\psi_{kl}}{\psi_k'}, \quad (32)$$

where  $\psi_{kl} = \frac{1}{NL} d_{kl} \eta_l^2 + \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi_{jl}}{(1 + \gamma_j)^2} \frac{1}{L} \sum_{l=1}^L d_{il} d_{jl} \eta_l^2$ .

*Proof:* Please refer to Appendix D for details. ■

We thus have the following deterministic equivalent of the weights  $\omega_l^{[n]}$ 's (11):

$$\bar{\omega}_l^{[n]} = \frac{p_{vl}}{2} \left[ \bar{\chi}_l + \epsilon^2 \right]^{\frac{v}{2} - 1}, \quad \forall l = 1, \dots, L. \quad (33)$$

Note that the asymptotic results of the powers  $p_k^\circ$ 's (26) and Lagrangian multipliers  $\lambda_k^\circ$ 's (23) should depend on the weights  $\bar{\omega}_l^{[n]}$ 's (33) at each iteration. Based on (7), we have the following asymptotic result for the RRH ordering criterion:

$$\bar{\theta}_l = \kappa_l \bar{\chi}_l^*, \quad (34)$$

where  $\bar{\chi}_l^*$  is the deterministic equivalent of the squared  $\ell_2$ -norm of the solution to problem  $\mathcal{P}_{\text{GSBF}}$  using the iterative reweighted- $\ell_2$  algorithm. Therefore, the RRH with a smaller  $\bar{\theta}_l$  will have a higher priority to be switched off. This ordering criterion will change only when the long-term channel attenuation is updated. Note that global instantaneous CSI is still needed in stage II to find the active RRHs, i.e., solving a sequence of convex feasibility problems  $\mathcal{F}(\mathcal{A}^{[i]})$  (8).

The statistical CSI based iterative reweighted- $\ell_2$  algorithm for group sparse beamforming is presented in Algorithm 2.

---

#### Algorithm 2 Statistical CSI Based Iterative Reweighted- $\ell_2$ Algorithm for Problem $\mathcal{P}_{\text{GSBF}}$

---

**input:** Initialize  $\boldsymbol{\omega}^{[0]} = (1, \dots, 1)$ ;  $I$  (the maximum number of iterations)

Repeat

1) Compute deterministic equivalent of the squared  $\ell_2$ -norm of the solution to  $\mathcal{P}_{\text{GSBF}}^{[n]}$ ,  $\|\tilde{\mathbf{v}}_1^{[n]}\|_2^2, \dots, \|\tilde{\mathbf{v}}_L^{[n]}\|_2^2$ , using  $\bar{\chi}_l^{[n]}$  (32).

2) Update the weights  $\bar{\omega}_l^{[n+1]}$  using (33).

Until convergence or attain the maximum iterations and return **output**.

**output:** Asymptotic result of the RRH ordering criterion  $\bar{\boldsymbol{\theta}}$  (34).

---

**Remark 3:** Although stage 2 and stage 3 still require instantaneous CSI, the RRH ordering criterion computed by Algorithm 2 in stage 1 is only based on statistical CSI and thus will be updated only when the long-term channel propagation is changed. Therefore, Algorithm 2 serves the purpose of further reducing the computation complexity of Algorithm 1 for RRH ordering. A promising future research direction is to select RRHs in stage 2 only based on statistical CSI. However, the main challenge is the infeasibility issue if only statistical CSI is available, as discussed in Remark 1.

## V. SIMULATION RESULTS

In this section, we will simulate the proposed algorithms for network power minimization for Cloud-RANs. In all the realizations, we only account the channel realizations making the original problem  $\mathcal{P}$  feasible. If problem  $\mathcal{P}$  is infeasible, further works on user admission are required [15]. We set

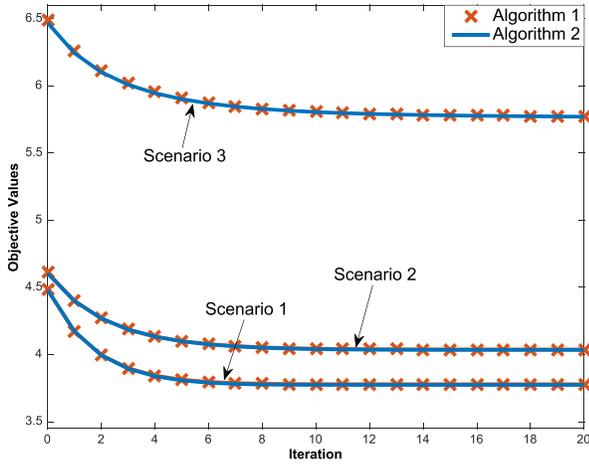


Fig. 2. Convergence of the iterative reweighted- $\ell_2$  algorithm.

$p = 1$ ,  $\epsilon = 10^{-3}$  and  $\omega^{[0]} = (1, \dots, 1)$  for all the simulations. The proposed iterative reweighted- $\ell_2$  algorithms (Algorithm 1 and Algorithm 2) will terminate if either the number of iterations exceeds 30 or the difference between the objective values of consecutive iterations is less than  $10^{-3}$ .

#### A. Convergence of the Iteratively Reweighted- $\ell_2$ Algorithm

Consider a network with  $L = 5$  30-antenna RRHs and 5 single antenna MUs uniformly and independently distributed in the square region  $[-2000, 2000] \times [-2000, 2000]$  meters. Fig. 2 shows the convergence of the iterative reweighted- $\ell_2$  algorithm in different scenarios with different realizations of RRH and MU positions. For each scenario, the numerical results are obtained by averaging over 100 small-scale fading realizations. The large-scale fading (i.e., statistical CSI) is randomly generated and fixed during simulations. In the two curves of each scenario, Algorithm 1 and Algorithm 2 are applied for Stage I of group sparse beamforming, respectively. It demonstrates that the large system analysis based Algorithm 2 provides accurate approximation even in a small system. Fast convergence is observed in the simulated setting.

#### B. Network Power Minimization

Consider a network with  $L = 5$  10-antenna RRHs and 6 single antenna MUs uniformly and independently distributed in the square region  $[-1000, 1000] \times [-1000, 1000]$  meters. The relative transport link power consumption are set to be  $P_l^c = (5.6 + 2l)W$ ,  $l = 1, \dots, L$ . We average over 200 small-scale channel realizations. Fig. 3 demonstrates the average network power consumption using different algorithms. This figure shows that the proposed iterative reweighted- $\ell_2$  algorithm outperforms the weighted mixed  $\ell_1/\ell_2$ -norm algorithm in [7]. Furthermore, it illustrates that the large system analysis provides accurate approximations for the iterative reweighted- $\ell_2$  algorithm in finite systems with reduced computation overhead. In Table I, for the simulated scenarios, we see that the iterative reweighted- $\ell_2$  algorithms achieve similar performance with different values of parameter  $p$  and the large

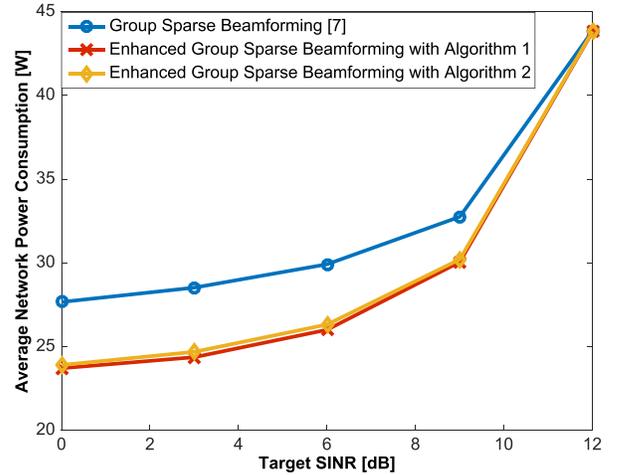


Fig. 3. Average network power consumption versus target SINR with different algorithms.

TABLE I  
NETWORK POWER CONSUMPTION USING ALGORITHM 1  
AND ALGORITHM 2 WITH DIFFERENT PARAMETERS  $p$

Target SINR [dB]	0	3	6	9	12
GSBF [7]	27.671	28.528	29.924	32.759	43.855
Alg. 1 with $p = 1$	23.716	24.374	26.011	30.057	43.810
Alg. 1 with $p = 0.5$	23.789	24.463	26.092	30.500	43.783
Alg. 2 with $p = 1$	23.907	24.695	26.326	30.214	43.827
Alg. 2 with $p = 0.5$	23.905	24.689	26.315	30.205	43.821

system analysis provides accurate approximations. Although the simulated results demonstrate that the performance is robust to different values of parameter  $p$ , it is very interesting to theoretically identify the typical scenarios, where smaller values of parameter  $p$  will yield much lower network power consumption.

## VI. CONCLUSIONS AND DISCUSSIONS

In this paper, we developed an enhanced three-stage group sparse beamforming framework with reduced computation overhead in Cloud-RANs. In particular, closed-form solutions for the group sparse optimization problem were obtained by developing the iterative reweighted- $\ell_2$  algorithm based on the MM algorithm and Lagrangian duality theory. This is the first effort to reduce the computation cost for RRH ordering in the first stage of group sparse beamforming. Based on the developed structured iterative algorithm, we further provided a large system analysis for the optimal Lagrangian multipliers at each iteration via random matrix theory, thereby computing the RRH ordering criterion only based on the statistical CSI. This is the second effort to enable computation scalability for RRH ordering in the first stage.

Several future directions of interest are listed as follows:

- Although computing the RRH ordering criterion in the first stage with Algorithm 2 only needs statistical CSI, the second stage still requires global instantaneous CSI

to find the active RRHs by solving a sequence of convex feasibility problems. It is thus particularly interesting to investigate efficient algorithms to select the active RRHs only based on the statistical CSI.

- It is desirable to establish the optimality of the iterative reweighted- $\ell_2$  algorithm for the network power minimization problem  $\mathcal{P}$ . However, considering the complicated problem structures of problem  $\mathcal{P}$ , this becomes challenging. It is also interesting to apply the developed statistical CSI based iterative reweighted- $\ell_2$  algorithm for more complicated network optimization problems, e.g., the wireless caching problem [33], [34], the computation offloading problem [36], and beamforming problems with CSI uncertainty [32].

#### APPENDIX A PROOF OF THEOREM 1

1) Observing that the phases of  $\mathbf{v}_k$  will not change the objective function and constraints of problem  $\mathcal{P}_{\text{GSBF}}$ , the QoS constraints thus can be equivalently transformed to the second-order cone constraints:  $\mathcal{C} = \left\{ \mathbf{v} : \sqrt{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{v}_i|^2 + \sigma_k^2} \leq \Re(\mathbf{h}_k^H \mathbf{v}_k) / \gamma_k, \forall k \right\}$ , which are convex cones. The KKT points of problem  $\mathcal{P}_{\text{GSBF}}$  should satisfy:

$$0 \in \nabla_{\mathbf{v}} g_p(\mathbf{v}; \epsilon) + \mathcal{N}_{\mathcal{C}}(\mathbf{v}), \quad (35)$$

where  $\mathcal{N}_{\mathcal{C}}(\mathbf{v})$  is the normal cone of the second-order cone at point  $\mathbf{v}$  [47]. We shall show that any convergent subsequence  $\{\mathbf{v}^{[n_k]}\}_{k=1}^{\infty}$  of  $\{\mathbf{v}^{[n]}\}_{n=1}^{\infty}$  satisfies (35). Specifically, let  $\mathbf{v}^{[n_k]} \rightarrow \bar{\mathbf{v}}$  be one such convergent subsequence with

$$\lim_{k \rightarrow \infty} \mathbf{v}^{[n_k+1]} = \lim_{k \rightarrow \infty} \mathbf{v}^{[n_k]} = \bar{\mathbf{v}}. \quad (36)$$

Based on the strong duality result for problem  $\mathcal{P}_{\text{GSBF}}^{[n]}$ , the following KKT condition holds at  $\mathbf{v}^{[n_k+1]}$ :

$$0 \in \nabla_{\mathbf{v}} Q(\mathbf{v}^{[n_k+1]}; \boldsymbol{\omega}^{[n_k]}) + \mathcal{N}_{\mathcal{C}}(\mathbf{v}^{[n_k+1]}). \quad (37)$$

Based on (36) and (11), we have  $\lim_{k \rightarrow \infty} \nabla_{\mathbf{v}} Q(\mathbf{v}^{[n_k+1]}; \boldsymbol{\omega}^{[n_k]}) = \nabla_{\mathbf{v}} g_p(\bar{\mathbf{v}}; \epsilon)$ . Furthermore, based on [47, Proposition 6.6] and (36), we have  $\limsup_{\mathbf{v}^{[n_k+1]} \rightarrow \bar{\mathbf{v}}} \mathcal{N}_{\mathcal{C}}(\mathbf{v}^{[n_k+1]}) = \mathcal{N}_{\mathcal{C}}(\bar{\mathbf{v}})$ . Therefore, by taking  $k \rightarrow \infty$  in (37), we have  $0 \in \nabla_{\mathbf{v}} Q(\bar{\mathbf{v}}; \bar{\boldsymbol{\omega}}) + \mathcal{N}_{\mathcal{C}}(\bar{\mathbf{v}})$ , which indicates that  $\bar{\mathbf{v}}$  is a KKT point of problem  $\mathcal{P}_{\text{GSBF}}$ . We thus complete the proof.

2) We first conclude the following fact:

$$\begin{aligned} & g_p(\mathbf{v}^{[n+1]}; \epsilon) \\ &= Q(\mathbf{v}^{[n+1]}; \boldsymbol{\omega}^{[n]}) + g_p(\mathbf{v}^{[n+1]}; \epsilon) - Q(\mathbf{v}^{[n+1]}; \boldsymbol{\omega}^{[n]}) \\ &\leq Q(\mathbf{v}^{[n+1]}; \boldsymbol{\omega}^{[n]}) + g_p(\mathbf{v}^{[n]}; \epsilon) - Q(\mathbf{v}^{[n]}; \boldsymbol{\omega}^{[n]}) \\ &\leq Q(\mathbf{v}^{[n]}; \boldsymbol{\omega}^{[n]}) + g_p(\mathbf{v}^{[n]}; \epsilon) - Q(\mathbf{v}^{[n]}; \boldsymbol{\omega}^{[n]}) \\ &= g_p(\mathbf{v}^{[n]}; \epsilon), \end{aligned} \quad (38)$$

where the first inequality is based on the fact that function  $(g_p(\mathbf{v}; \epsilon) - Q(\mathbf{v}; \boldsymbol{\omega}^{[n]}))$  attains its maximum at  $\mathbf{v} = \mathbf{v}^{[n]}$  [15, Proposition 1], and the second inequality follows from (37). Furthermore, as  $g_p(\mathbf{v}; \epsilon)$  is continuous and  $\mathcal{C}$  is compact, the limit of the sequence  $g_p(\mathbf{v}^{[n]}; \epsilon)$  is finite. Based on the convergence results in 1), we thus complete the proof.

#### APPENDIX B PROOF OF LEMMA 1

The proof technique here is mainly based on the work [26]. However, we have to modify their proof as we have different weighting matrices  $\mathbf{Q}^{[n]}$  in vectors  $\bar{\mathbf{v}}_k$  (16) at the  $n$ -th iteration and such modification is non-trivial. Based on the following matrix inversion lemma [48],  $\mathbf{x}^H (\mathbf{U} + c\mathbf{x}\mathbf{x}^H)^{-1} = \frac{\mathbf{x}^H \mathbf{U}^{-1}}{1 + c\mathbf{x}^H \mathbf{U}^{-1} \mathbf{x}}$ , (21) can be rewritten as

$$\frac{\gamma_k}{\lambda_k} = \frac{1}{LN} \mathbf{h}_k^H \left( \frac{1}{LN} \sum_{i \neq k} \lambda_i \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q} \right)^{-1} \mathbf{h}_k, \quad (39)$$

which can be further rewritten as

$$\gamma_k \rho_k = \frac{\lambda_k^\circ}{LN} \mathbf{h}_k^H \left( \frac{1}{LN} \sum_{i \neq k} \frac{\lambda_i^\circ}{\rho_i} \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q} \right)^{-1} \mathbf{h}_k, \quad (40)$$

where  $\rho_k = \lambda_k^\circ / \lambda_k$ .

Assume that  $0 \leq \rho_1 \leq \rho_2 \leq \dots \leq \rho_K$ . From (40), replacing  $\rho_i$  with  $\rho_K$  and using monotonicity arguments, we have

$$\gamma_K \rho_K \leq \frac{\lambda_K^\circ}{LN} \mathbf{h}_K^H \left( \frac{1}{LN} \sum_{i \neq K} \frac{\lambda_i^\circ}{\rho_K} \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q} \right)^{-1} \mathbf{h}_K, \quad (41)$$

or, equivalently

$$\gamma_K \leq \frac{\lambda_K^\circ}{LN} \mathbf{h}_K^H \left( \frac{1}{LN} \sum_{i \neq K} \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + \rho_K \mathbf{Q} \right)^{-1} \mathbf{h}_K. \quad (42)$$

Assume now that  $\rho_K$  is infinitely often larger than  $(1+\ell)$  with  $\ell$  being some positive value [27]. Let us restrict ourselves to such a subsequence. From (42), using monotonicity arguments we obtain

$$\gamma_K \leq \frac{\lambda_K^\circ}{LN} \mathbf{h}_K^H \left( \frac{1}{LN} \sum_{i \neq K} \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + (1+\ell) \mathbf{Q} \right)^{-1} \mathbf{h}_K. \quad (43)$$

Define

$$m_k(\ell) = \frac{1}{LN} \mathbf{h}_k^H \left( \frac{1}{LN} \sum_{i \neq k} \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + (1+\ell) \mathbf{Q} \right)^{-1} \mathbf{h}_k. \quad (44)$$

Now we investigate the deterministic equivalents for  $m_k(\ell)$ . We first introduce the following important lemma:

*Lemma 3 ([23, Lemma 14.2]):* Let  $\mathbf{A}_1, \mathbf{A}_2, \dots$ , with  $\mathbf{A}_N \in \mathbb{C}^{N \times N}$  be a series of random matrices generated by the probability space  $(\Omega, \mathcal{F}, P)$  such that, for  $\omega \in A \subset \Omega$ , with  $P(A) = 1$ ,  $\|\mathbf{A}_N(\omega)\| < K(\omega) < \infty$ , uniformly on  $N$ . Let  $\mathbf{x}_1, \mathbf{x}_2, \dots$ , with  $\mathbf{x}_N \in \mathbb{C}^N$ , be random vectors of i.i.d. entries with zero mean, variance  $1/N$ , and the eighth-order moment of order  $O(1/N^4)$ , independent of  $\mathbf{A}_N$ . Then

$$\mathbf{x}_N^H \mathbf{A}_N \mathbf{x}_N - \frac{1}{N} \text{Tr}(\mathbf{A}_N) \xrightarrow{N \rightarrow \infty} 0, \quad (45)$$

almost surely.

Therefore, based on Lemma 3, we have

$$m_K(\ell) - \frac{1}{LN} \text{Tr} \Theta_K \left( \frac{1}{LN} \sum_{i \neq K} \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + (1+\ell) \mathbf{Q} \right)^{-1} \xrightarrow{N \rightarrow \infty} 0. \quad (46)$$

We further introduce the following important lemma to derive the deterministic equivalent for the  $m_k(\ell)$ .

*Lemma 4 ([24, Th. 1]):* Let  $\mathbf{B}_N = \mathbf{X}_N^H \mathbf{X}_N + \mathbf{S}_N$  with  $\mathbf{S}_N \in \mathbb{C}^{N \times N}$  Hermitian nonnegative definite and  $\mathbf{X}_N \in \mathbb{C}^{n \times N}$  random. The  $i$ -th column  $\mathbf{x}_i$  of  $\mathbf{X}_N^H$  is  $\mathbf{x}_i = \Psi_i \mathbf{y}_i$ , where the entries of  $\mathbf{y}_i \in \mathbb{C}^i$  are i.i.d. of zero mean, variance  $1/N$  and have eighth-order moment of order  $O(1/N^4)$ . The matrices  $\Psi_i \in \mathbb{C}^{N \times r_i}$  are deterministic. Furthermore, let  $\Theta_i = \Psi_i \Psi_i^H \in \mathbb{C}^{N \times N}$  and define  $\mathbf{Q}_N \in \mathbb{C}^{N \times N}$  deterministic. Assume  $\limsup_{N \rightarrow \infty} \sup_{1 \leq i \leq n} \|\Theta_i\| < \infty$  and let  $\mathbf{Q}_N$  have uniformly bounded spectral norm (with respect to  $N$ ). Define

$$m_{\mathbf{B}_N, \mathbf{Q}_N}(z) := \frac{1}{N} \text{Tr} \mathbf{Q}_N (\mathbf{B}_N - z \mathbf{I}_N)^{-1}. \quad (47)$$

Then, for  $z \in \mathbb{C} \setminus \mathbb{R}^+$ , as  $n, N$  grow large with ratios  $\beta_{N,i} := N/r_i$  and  $\beta_N := N/n$  such that  $0 < \liminf_N \beta_N \leq \limsup_N \beta_N < \infty$  and  $0 < \liminf_N \beta_{N,i} \leq \limsup_N \beta_{N,i} < \infty$ , we have that

$$m_{\mathbf{B}_N, \mathbf{Q}_N}(z) - m_{\mathbf{B}_N, \mathbf{Q}_N}^\circ(z) \xrightarrow{N \rightarrow \infty} 0 \quad (48)$$

almost surely, with  $m_{\mathbf{B}_N, \mathbf{Q}_N}^\circ(z)$  given by

$$m_{\mathbf{B}_N, \mathbf{Q}_N}^\circ(z) = \frac{1}{N} \text{Tr} \mathbf{Q}_N \left( \frac{1}{N} \sum_{j=1}^n \frac{\Theta_j}{1 + e_{N,j}(z)} + \mathbf{S}_N - z \mathbf{I}_N \right)^{-1}, \quad (49)$$

where the functions  $e_{N,1}(z), \dots, e_{N,n}(z)$  from the unique solution of

$$e_{N,i}(z) = \frac{1}{N} \text{Tr} \Theta_i \left( \frac{1}{N} \sum_{j=1}^n \frac{\Theta_j}{1 + e_{N,j}(z)} + \mathbf{S}_N - z \mathbf{I}_N \right)^{-1}, \quad (50)$$

which is the Stieltjes transform of a nonnegative finite measure on  $\mathbb{R}^+$ . Moreover, for  $z < 0$ , the scalars  $e_{N,1}(z), \dots, e_{N,n}(z)$  are unique nonnegative solutions to (50).

Therefore, based on Lemma 4, we have

$$\psi_K(\ell) - \psi_K^\circ(\ell) \xrightarrow{N \rightarrow \infty} 0 \quad (51)$$

almost surely with  $\psi_K^\circ(\ell)$  being the unique positive solution to

$$\psi_K^\circ(\ell) = \frac{1}{NL} \text{Tr} (\Theta_K \mathbf{A}(\ell)) \quad (52)$$

where

$$\mathbf{A}(\ell) = \left( \frac{1}{NL} \sum_{i=1}^K \frac{\lambda_i^\circ \Theta_i}{1 + \lambda_i^\circ \psi_i^\circ(\ell)} + (1+\ell) \mathbf{Q} \right)^{-1}. \quad (53)$$

From (51), recalling (43) yields

$$\liminf_{K \rightarrow \infty} \psi_K^\circ(\ell) \geq \gamma_K / \lambda_K^\circ \quad (54)$$

Using the fact that  $\psi_K^\circ(\ell)$  is a decreasing function of  $\ell$ , it can be proved [23] that for any  $\ell > 0$ , we have

$$\lim_{K \rightarrow \infty} \sup \psi_K^\circ(\ell) < \gamma_K / \lambda_K^\circ. \quad (55)$$

However, this is against the former condition and creates a contradiction on the initial hypothesis that  $\rho_K < 1+\ell$  infinitely often. Therefore, we must admit that  $\rho_K \leq 1+\ell$  for all large values of  $K$ . Reverting all inequalities and using similar arguments yields  $\rho_K \geq 1-\ell$  for all large values of  $K$ .

We eventually obtain that  $1-\ell \leq \rho_K \leq 1+\ell$  from which we may write  $\max |\rho_K - 1| \leq \ell$  for all large values of  $K$ . Taking a countable sequence of  $\ell$  going to zero yields  $\max |\rho_K - 1| \rightarrow 0$  from which using  $\rho_K = \lambda_K^\circ / \lambda_K$  and assuming  $\lim_{K \rightarrow \infty} \sup \gamma_K < \infty$ , we obtain

$$\max |\lambda_K - \lambda_K^\circ| \xrightarrow{N \rightarrow \infty} 0, \quad (56)$$

almost surely. From (40), we have  $\gamma_K / \lambda_K^\circ = \psi_K^\circ(0) = \psi_K^\circ$ . Therefore, we obtain  $\lambda_K^\circ = \gamma_K / \psi_K^\circ$ , where  $\psi_K^\circ$  is the solution of the following equation:

$$\psi_K^\circ = \frac{1}{NL} \text{Tr} \Theta_K \left( \frac{1}{NL} \sum_{i=1}^K \frac{\Theta_i}{\psi_K^\circ} \frac{\gamma_i}{1 + \gamma_i} + \mathbf{Q} \right)^{-1}. \quad (57)$$

Following the same steps for  $k = 1, \dots, K$  yields the following desired result:

$$\lambda_k^\circ = \gamma_k / \psi_k^\circ, \quad \forall k. \quad (58)$$

As  $\Theta_k = \text{diag}\{d_{k1}, \dots, d_{kL}\} \otimes \mathbf{I}_N$ , we have

$$\psi_k^\circ = \frac{1}{LN} \text{Tr} (\Theta_k \mathbf{A}) = \frac{1}{L} \sum_{l=1}^L d_{kl} \eta_l, \quad (59)$$

where  $\mathbf{A} = \mathbf{A}(0)$  in (53) and  $\{\eta_l\}$  is the unique positive solution to the following set of equations

$$\eta_l = \left( \frac{1}{NL} \sum_{i=1}^K \frac{d_{il}}{1/L \sum_{j=1}^L d_{kj} \eta_j} \frac{\gamma_i}{1 + \gamma_i} + \omega_l^{[n]} \right)^{-1}. \quad (60)$$

We thus complete the proof for the asymptotic results of the optimal Lagrange multipliers in Lemma 1.

## APPENDIX C PROOF OF LEMMA 2

The SINR for MU  $k$  (2) can be rewritten as

$$\text{sinr}_k = \frac{\frac{p_k}{NL} \frac{\|\mathbf{h}_k^H \tilde{\mathbf{v}}_k\|^2}{\|\tilde{\mathbf{v}}_k\|_2^2}}{\sum_{i \neq k} \frac{p_i}{NL} \frac{\|\mathbf{h}_i^H \tilde{\mathbf{v}}_i\|^2}{\|\tilde{\mathbf{v}}_i\|_2^2} + \sigma_k^2}, \quad (61)$$

where  $\tilde{\mathbf{v}}_k = \left( \sum_{i=1}^K \frac{\lambda_i^\circ}{LN} \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q}^{1/2} \right)^{-1} \mathbf{h}_k$ . The interference part of the denominator of  $\Gamma_k$  in (61) can be rewritten as

$$\begin{aligned} \frac{1}{NL} \sum_{i \neq k} p_i \frac{\|\mathbf{h}_i^H \tilde{\mathbf{v}}_i\|^2}{\|\tilde{\mathbf{v}}_i\|_2^2} &= \frac{1}{NL} \sum_{i \neq k} p_i \mathbf{h}_i^H \left( \frac{\tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^H}{\|\tilde{\mathbf{v}}_i\|_2^2} \right) \mathbf{h}_k \\ &= \frac{1}{NL} \mathbf{h}_k^H \mathbf{V} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \end{aligned} \quad (62)$$

with  $\mathbf{V} = \left( \frac{1}{NL} \sum_{i=1}^K \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q} \right)^{-1}$ ,  $\mathbf{H}^{[k]} := [\mathbf{h}_1, \dots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \dots, \mathbf{h}_K] \in \mathbb{C}^{NL \times (K-1)}$  and  $\mathbf{P}^{[k]} := \text{diag} \left\{ \frac{p_1}{NL \|\tilde{\mathbf{v}}_1\|_2^2}, \dots, \frac{p_{k-1}}{NL \|\tilde{\mathbf{v}}_{k-1}\|_2^2}, \frac{p_{k+1}}{NL \|\tilde{\mathbf{v}}_{k+1}\|_2^2}, \dots, \frac{p_K}{NL \|\tilde{\mathbf{v}}_K\|_2^2} \right\}$ .

In order to eliminate the dependence between  $\mathbf{h}_k$  and  $\mathbf{V}$ , rewrite (62) as

$$\begin{aligned} & \frac{1}{NL} \mathbf{h}_k^H \mathbf{V} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \\ &= \frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^{[k]} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \\ &+ \frac{1}{NL} \mathbf{h}_k^H (\mathbf{V} - \mathbf{V}^{[k]}) \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k, \end{aligned} \quad (63)$$

where  $\mathbf{V}^{[k]} = \left( \frac{1}{NL} \sum_{i \neq k} \lambda_i^\circ \mathbf{h}_i \mathbf{h}_i^H + \mathbf{Q} \right)^{-1}$ . Using the resolvent identity [48] (i.e.,  $\mathbf{U}^{-1} - \mathbf{V}^{-1} = -\mathbf{U}^{-1}(\mathbf{U} - \mathbf{V})\mathbf{V}^{-1}$  with  $\mathbf{U}$  and  $\mathbf{V}$  as two invertible complex matrices of size  $N \times N$ ), we have

$$\mathbf{V} - \mathbf{V}^{[k]} = -\mathbf{V}(\mathbf{V}^{-1} - \mathbf{V}^{[k]-1})\mathbf{V}^{[k]}. \quad (64)$$

Then, observing that  $\mathbf{V}^{-1} - \mathbf{V}^{[k]-1} = \frac{\lambda_k^\circ}{NL} \mathbf{h}_k \mathbf{h}_k^H$ . From (63), one gets

$$\begin{aligned} & \frac{1}{NL} \mathbf{h}_k^H \mathbf{V} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \\ &= \frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^{[k]} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \\ &- \frac{\tilde{\lambda}_k}{NL} \mathbf{h}_k^H \mathbf{V} \mathbf{h}_k \left[ \frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^{[k]} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \right]. \end{aligned} \quad (65)$$

*Lemma 5 ([24, Lemma 7]):* Let  $\mathbf{U}, \mathbf{V}, \Theta \in \mathbb{C}^{N \times N}$  be of uniformly bounded spectral norm with respect to  $N$  and let  $\mathbf{V}$  be invertible. Further, define  $\mathbf{x} := \Theta^{1/2} \mathbf{z}$  and  $\mathbf{y} := \Theta^{1/2} \mathbf{q}$ , where  $\mathbf{z}, \mathbf{q} \in \mathbb{C}^N$  have i.i.d. complex entries of zero mean, variance  $1/N$ , and finite eighth-order moment and be mutually independent as well as independent of  $\mathbf{U}, \mathbf{V}$ . Define  $c_0, c_1, c_2 \in \mathbb{R}^+$  such that  $c_0 c_1 - c_2^2 \geq 0$  and let  $\mu := \frac{1}{N} \text{Tr} \Theta \mathbf{V}^{-1}$  and  $\mu' := \frac{1}{N} \text{Tr}(\Theta \mathbf{U} \mathbf{V}^{-1})$ . Then, we have

$$\begin{aligned} & \mathbf{x}^H \mathbf{U} (\mathbf{V} + c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{x} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H)^{-1} \mathbf{x} \\ & - \frac{\mu'(1 + c_1 \mu)}{(c_0 c_1 - c_2^2) \mu^2 + (c_0 + c_1) \mu + 1} \xrightarrow{N \rightarrow \infty} 0 \end{aligned} \quad (66)$$

almost surely. Furthermore,

$$\begin{aligned} & \mathbf{x}^H \mathbf{U} (\mathbf{V} + c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{x} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H)^{-1} \mathbf{y} \\ & - \frac{c_2 \mu \mu'}{(c_0 c_1 - c_2^2) \mu^2 + (c_0 + c_1) \mu + 1} \xrightarrow{N \rightarrow \infty} 0 \end{aligned} \quad (67)$$

almost surely.

Therefore, applying Lemma 5, we obtain that

$$\frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^{[k]} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k - \frac{\mu'}{1 + \lambda_k^\circ \mu} \xrightarrow{N \rightarrow \infty} 0 \quad (68)$$

almost surely, where  $\mu = \frac{1}{NL} \text{Tr}(\Theta_k \mathbf{V}^{[k]})$  and  $\mu' = \frac{1}{NL} \text{Tr} \left( \frac{1}{NL} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \mathbf{V}^{[k]} \Theta_k \mathbf{V}^{[k]} \mathbf{H}^{[k]} \right)$ . Furthermore, we have

$$\begin{aligned} & \frac{\lambda_k^\circ}{NL} \mathbf{h}_k^H \mathbf{V} \mathbf{h}_k \left[ \frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^{[k]} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k \right] \\ & - \frac{\lambda_k^\circ \mu \mu'}{(1 + \lambda_k^\circ \mu)^2} \xrightarrow{N \rightarrow \infty} 0 \end{aligned} \quad (69)$$

almost surely. Based on the results in (65), (68), (69), we have

$$\frac{1}{NL} \mathbf{h}_k^H \mathbf{V} \left( \frac{1}{NL} \mathbf{H}^{[k]} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \right) \mathbf{V} \mathbf{h}_k - \frac{\mu'}{(1 + \lambda_k^\circ \mu)^2} \xrightarrow{N \rightarrow \infty} 0 \quad (70)$$

almost surely.

From Lemma 1, we have  $\mu - \psi_k^\circ \xrightarrow{N \rightarrow \infty} 0$  almost surely.

Applying [24, Lemma 6], we have

$$\begin{aligned} & \frac{1}{NL} \text{Tr} \left( \frac{1}{NL} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \mathbf{V}^{[k]} \Theta_k \mathbf{V}^{[k]} \mathbf{H}^{[k]} \right) \\ & - \frac{1}{NL} \text{Tr} \left( \frac{1}{NL} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \mathbf{V} \Theta_k \mathbf{V} \mathbf{H}^{[k]} \right) \xrightarrow{N \rightarrow \infty} 0 \end{aligned} \quad (71)$$

almost surely. We further rewrite

$$\begin{aligned} & \frac{1}{NL} \text{Tr} \left( \frac{1}{NL} \mathbf{P}^{[k]} \mathbf{H}^{[k]H} \mathbf{V} \Theta_k \mathbf{V} \mathbf{H}^{[k]} \right) \\ &= \frac{1}{NL} \sum_{i \neq k} p_i \frac{\frac{1}{NL} \mathbf{h}_i^H \mathbf{V} \Theta_k \mathbf{V} \mathbf{h}_i}{\frac{1}{NL} \mathbf{h}_i^H \mathbf{V}^2 \mathbf{h}_i}. \end{aligned} \quad (72)$$

Applying [24, Lemma 1, 4 and 6], we obtain almost surely

$$\frac{1}{NL} \mathbf{h}_i^H \mathbf{V} \Theta_k \mathbf{V} \mathbf{h}_i - \frac{\frac{1}{NL} \text{Tr}(\Theta_i \mathbf{V} \Theta_k \mathbf{V})}{\left[ 1 + \frac{1}{NL} \lambda_i^\circ \text{Tr}(\Theta_i \mathbf{V}) \right]^2} \rightarrow 0. \quad (73)$$

To derive a deterministic equivalent for  $\frac{1}{NL} \text{tr}(\Theta_i \mathbf{V} \Theta_k \mathbf{V})$ , we write

$$\frac{1}{NL} \text{Tr}(\Theta_i \mathbf{V} \Theta_k \mathbf{V}) = \frac{1}{NL} \frac{\partial}{\partial z} \text{Tr} \left( \Theta_i (\mathbf{V}^{-1} - z \Theta_k)^{-1} \right) \Big|_{z=0}. \quad (74)$$

Observe now that [24, Theorem 1]

$$\text{Tr} \left( \Theta_i (\mathbf{V}^{-1} - z \Theta_k)^{-1} \right) - \psi_{ik}(z) \rightarrow 0, \quad (75)$$

almost surely, where  $\psi_{ik}(z)$  is given by  $\psi_{ik}(z) = \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{T}_k(z))$  and  $\mathbf{T}_k(z)$  is computed as

$$\mathbf{T}_k(z) = \left( \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^\circ \Theta_j}{1 + \lambda_j^\circ \psi_{jk}(z)} + \mathbf{Q} - z \Theta_k \right)^{-1}. \quad (76)$$

By differentiating along  $z$ , we have  $\psi'_{ik}(z) = \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{T}'_k(z))$ , where  $\mathbf{T}'_k(z) = \frac{\partial \mathbf{T}_k(z)}{\partial z}$  is given by

$$\mathbf{T}'_k(z) = \mathbf{T}_k(z) \left( \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_{jk}(z) \Theta_j}{(1 + \lambda_k^\circ \psi_{jk}(z))^2} + \Theta_k \right) \mathbf{T}_k(z). \quad (77)$$

Setting  $z = 0$  yields

$$\psi'_{ik}(0) = \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{T}'_k(0)), \quad (78)$$

where

$$\mathbf{T}'_k(0) = \mathbf{T}_k(0) \left( \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_{jk}(0) \Theta_k}{(1 + \gamma_k)^2} + \Theta_k \right) \mathbf{T}_k(0). \quad (79)$$

In writing the above result, we have taken into account that  $\mathbf{A} = \mathbf{T}_j(0)$ ,  $\psi_{jk}(0) = \frac{1}{NL} \text{Tr}(\Theta_j \mathbf{T}_k(0)) = \frac{1}{NL} \text{Tr}(\Theta_j \mathbf{A}) = \psi_j^\circ$  and  $\gamma_j = \lambda_j^\circ \psi_j^\circ$ . Plugging (79) into (78) and neglecting the functional dependence from  $z = 0$ ,  $\boldsymbol{\psi}'_k = [\psi'_{1k}, \dots, \psi'_{Kk}]^T$  is found as the unique solution of

$$\begin{aligned} \psi'_{ik} &= \frac{1}{NL} \text{Tr} \left( \Theta_i \mathbf{A} \left( \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_{jk} \Theta_j}{(1 + \gamma_j)^2} + \Theta_k \right) \mathbf{A} \right) \\ &= \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{A} \Theta_k \mathbf{A}) \\ &\quad + \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi'_{jk}}{(1 + \gamma_j)^2} \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{A} \Theta_j \mathbf{A}). \end{aligned} \quad (80)$$

Observing that

$$\frac{1}{NL} \text{Tr}(\Theta_i \mathbf{A} \Theta_j \mathbf{A}) = \frac{1}{L} \sum_{l=1}^L d_{il} d_{jl} \eta_l^2. \quad (81)$$

Let  $\boldsymbol{\psi}'_k = [\psi'_{1k}, \dots, \psi'_{Kk}]^T \in \mathbb{C}^K$  and  $\mathbf{J} = [J_{ij}] \in \mathbb{C}^{K \times K}$  with

$$J_{ij} = \frac{\lambda_j^{\circ 2}}{NL(1 + \gamma_j)^2} \left( \frac{1}{L} \sum_{l=1}^L d_{il} d_{jl} \eta_l^2 \right). \quad (82)$$

We can rewrite the above system of equations in the compact form as

$$\boldsymbol{\psi}'_k = \mathbf{c}_k + \mathbf{J} \boldsymbol{\psi}'_k, \quad (83)$$

where  $\mathbf{c}_k = [c_{ik}]$  with  $c_{ik} = \frac{1}{L} \sum_{l=1}^L d_{il} d_{kl} \eta_l^2$ .

On the basis of above results, using  $\frac{1}{NL} \lambda_i^\circ \text{tr}(\Theta_i \mathbf{A}) - \gamma_i \rightarrow 0$ , we eventually obtain that

$$\frac{1}{NL} \mathbf{h}_i^H \mathbf{V} \Theta_k \mathbf{V} \mathbf{h}_i - \frac{\psi'_{ik}}{(1 + \gamma_i)^2} \xrightarrow{N \rightarrow \infty} 0 \quad (84)$$

almost surely. Following similar arguments of above yields

$$\frac{1}{NL} \mathbf{h}_i^H \mathbf{V}^2 \mathbf{h}_i - \frac{\psi'_i}{(1 + \gamma_i)^2} \xrightarrow{N \rightarrow \infty} 0 \quad (85)$$

almost surely with  $\boldsymbol{\psi}' = [\psi'_1, \dots, \psi'_K]^T = (\mathbf{I}_K - \mathbf{J})^{-1} \mathbf{c}$  where  $\mathbf{c} \in \mathbb{C}^K$  has elements  $[c]_i = \frac{1}{L} \sum_{l=1}^L d_{il} \eta_l^2$ .

Putting the results in (84) and (85) together, we obtain almost surely

$$\frac{1}{NL} \sum_{i \neq k} p_i \frac{\frac{1}{NL} \mathbf{h}_i^H \mathbf{V} \Theta_k \mathbf{V} \mathbf{h}_i}{\frac{1}{NL} \mathbf{h}_i^H \mathbf{V}^2 \mathbf{h}_i} - \frac{1}{NL} \sum_{i \neq k} p_i \frac{\psi'_{ik}}{\psi'_i} \xrightarrow{N \rightarrow \infty} 0. \quad (86)$$

The deterministic equivalent of the numerator of the SINR in (61) is now easily obtained as

$$p_k \frac{|\frac{1}{NL} \mathbf{h}_k^H \mathbf{v}_k|^2}{\frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^2 \mathbf{h}_k} - p_k \frac{\psi_k^{\circ 2}}{\psi'_k} \xrightarrow{N \rightarrow \infty} 0, \quad (87)$$

since  $\frac{1}{NL} \mathbf{h}_k^H \mathbf{v}_k - \frac{\psi_k^\circ}{1 + \gamma_k} \xrightarrow{N \rightarrow \infty} 0$  and  $\frac{1}{NL} \mathbf{h}_k^H \mathbf{V}^2 \mathbf{h}_k - \frac{\psi'_k}{(1 + \gamma_k)^2} \xrightarrow{N \rightarrow \infty} 0$  almost surely.

Therefore, the deterministic equivalent SINR is given by

$$\text{sinr}_k^\circ = \frac{\psi_k^{\circ 2}}{\psi'_k} \frac{p_k}{I_k^\circ + \sigma_k^2}, \quad (88)$$

where

$$I_k^\circ := \frac{1}{(1 + \gamma_k)^2} \left( \frac{1}{NL} \sum_{i=1}^K p_i \frac{\psi'_{ik}}{\psi'_i} \right). \quad (89)$$

From (88), it follows that  $p_k^\circ$  such that  $\text{sinr}_k^\circ = \gamma_k$  is obtained as

$$p_k^\circ = \gamma_k \frac{\psi'_k}{\psi_k^{\circ 2}} (I_k^\circ + \sigma_k^2), \quad (90)$$

which can be further rewritten as

$$p_k^\circ \frac{\psi_k^{\circ 2}}{\psi'_k} \frac{1}{\gamma_k} = I_k^\circ + \sigma_k^2 = \frac{1}{(1 + \gamma_k)^2} \left( \frac{1}{NL} \sum_{i=1}^K p_i \frac{\psi'_{ik}}{\psi'_i} \right) + \sigma_k^2. \quad (91)$$

Therefore, the deterministic equivalent for the powers is given by

$$\begin{bmatrix} p_1^\circ \\ \vdots \\ p_K^\circ \end{bmatrix} = \mathbf{M}^{\circ -1} \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_K^2 \end{bmatrix},$$

where

$$[\mathbf{M}^\circ]_{ij} = \begin{cases} \frac{1}{\gamma_i} \frac{\psi_i^{\circ 2}}{\psi'_i}, & i = j, \\ -\frac{1}{(1 + \gamma_i^2) LN} \frac{\psi'_{ji}}{\psi'_j}, & i \neq j. \end{cases}$$

We thus complete the proof for the asymptotic results for optimal powers in Lemma 2.

#### APPENDIX D PROOF OF THEOREM 2

We rewrite  $\|\tilde{\mathbf{v}}_l\|_2^2$  as

$$\begin{aligned} \|\tilde{\mathbf{v}}_l\|_2^2 &= \sum_{k=1}^K \mathbf{v}_k^H \mathbf{Q}_{lk} \mathbf{v}_k = \frac{1}{NL} \sum_{k=1}^K p_k \frac{\tilde{\mathbf{v}}_k^H \mathbf{Q}_{lk} \tilde{\mathbf{v}}_k}{\|\tilde{\mathbf{v}}_k\|_2^2} \\ &= \frac{1}{NL} \sum_{k=1}^K p_k \frac{\mathbf{h}_k^H \mathbf{V} \mathbf{Q}_{lk} \mathbf{V} \mathbf{h}_k}{\mathbf{h}_k^H \mathbf{V}^2 \mathbf{h}_k}, \end{aligned} \quad (92)$$

where  $\mathbf{Q}_{lk} \in \mathbb{C}^{NL \times NL}$  is a block diagonal matrix with the identity matrix  $\mathbf{I}_N$  as the  $l$ -th main diagonal block square matrix and zeros elsewhere.

Based on the similar arguments in Appendix C, we have the deterministic equivalents

$$\mathbf{h}_k^H \mathbf{V} \mathbf{Q}_{lk} \mathbf{V} \mathbf{h}_k - \frac{\psi_{kl}}{(1 + \gamma_k)^2} \xrightarrow{N \rightarrow \infty} 0, \quad (93)$$

where

$$\begin{aligned} \psi_{kl} &= \frac{1}{NL} \text{Tr}(\Theta_k \mathbf{A} \mathbf{Q}_{kl} \mathbf{A}) \\ &= \frac{1}{NL} \sum_{j=1}^K \frac{\lambda_j^{\circ 2} \psi_{jl}}{(1 + \gamma_j)^2} \frac{1}{NL} \text{Tr}(\Theta_i \mathbf{A} \Theta_j \mathbf{A}). \end{aligned} \quad (94)$$

Observing that

$$\frac{1}{NL} \text{Tr}(\Theta_k \mathbf{A} \mathbf{Q}_{kl} \mathbf{A}) = \frac{1}{L} d_{kl} \eta_l^2, \quad (95)$$

we obtain the final result. We thus complete the proof for the asymptotic results for  $\|\tilde{\mathbf{v}}_l^{[n]}\|_2^2$ 's in Theorem 2.

#### REFERENCES

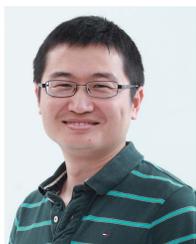
- [1] Y. Shi, J. Zhang, and K. B. Letaief, "Statistical group sparse beamforming for green Cloud-RAN via large system analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 870–874.
- [2] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
- [3] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen, "Large-scale convex optimization for ultra-dense cloud-RAN," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 84–91, Jun. 2015.
- [4] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 184–190, Oct. 2016.
- [5] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [6] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 126–135, Dec. 2014.
- [7] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.
- [8] D. Wubben *et al.*, "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.
- [9] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [10] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.
- [11] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.
- [12] S. Leyffer, *Mixed Integer Nonlinear Programming*, vol. 154. New York, NY, USA: Springer, 2012.
- [13] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.
- [14] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.
- [15] Y. Shi, J. Cheng, J. Zhang, B. Bai, W. Chen, and K. B. Letaief, "Smoothed  $L_p$ -minimization for green cloud-RAN with user admission control," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1022–1036, Apr. 2016.
- [16] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.
- [17] 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, Jan. 2011.
- [18] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016.
- [19] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.
- [20] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
- [21] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, Oct. 2014.
- [22] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [23] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [24] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
- [25] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.
- [26] L. Sanguinetti, R. Couillet, and M. Debbah, "Large system analysis of base station cooperation for power minimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5480–5496, Aug. 2016.
- [27] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Multivariate Anal.*, vol. 131, pp. 99–120, Oct. 2014.
- [28] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.
- [29] 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, no. 2, pp. 226–240, Feb. 2013.
- [30] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.
- [31] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.
- [32] Y. Shi, J. Zhang, and K. Letaief, "Robust group sparse beamforming for multicast green cloud-RAN with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4647–4659, Sep. 2015.
- [33] X. Peng, Y. Shi, J. Zhang, and K. B. Letaief, "Layered group sparse beamforming for cache-enabled green wireless networks," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5589–5603, Dec. 2017.
- [34] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.
- [35] J. Zhao, T. Q. S. Quek, and Z. Lei, "Heterogeneous cellular networks using wireless backhaul: Fast admission control and large system analysis," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2128–2143, Oct. 2015.
- [36] J. Cheng, Y. Shi, B. Bai, and W. Chen, "Computation offloading in cloud-ran based mobile cloud computing system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
- [37] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.
- [38] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [39] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24–39, Jan. 2013.
- [40] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, pp. 1–230, May 2015.
- [41] R. Couillet, G. Wainrib, H. T. Ali, and H. Sevi, "A random matrix approach to echo-state neural networks," in *Proc. 33th Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 517–525.
- [42] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.
- [43] S. Lakshminarayana, M. Assaad, and M. Debbah, "Coordinated multicell beamforming for massive MIMO: A random matrix approach," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3387–3412, Jun. 2015.
- [44] K. Lange and J. S. Sinsheimer, "Normal/independent distributions and their applications in robust regression," *J. Comput. Graph. Statist.*, vol. 2, no. 2, pp. 175–198, 1993.
- [45] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
- [46] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

- [47] R. T. Rockafellar, *Convex Analysis*, vol. 28. Princeton, NJ, USA: Princeton Univ. Press, 1997.
- [48] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.



**Yuanming Shi** (S'13–M'15) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology in 2015. Since 2015, he has been with the School of Information Science and Technology, ShanghaiTech University, as a tenure-track Assistant Professor. He visited the University of California at Berkeley, Berkeley, CA, USA, from 2016 to 2017. He was a recipient of the 2016 IEEE Marconi Prize Paper

Award in Wireless Communications, and the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society. He was involved in optimization theory and algorithms, deep and machine learning, high-dimensional probability and statistics, with applications in intelligent IoT, large-scale data analysis, and distributed edge computing.



**Jun Zhang** (S'06–M'10–SM'15) 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 with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology. He has co-authored the

book *Fundamentals of LTE* (Prentice-Hall, 2010). His research interests include dense wireless cooperative networks, mobile edge computing, cloud computing, and big data analytics systems.

Dr. Zhang was a co-recipient of the 2016 Marconi Prize Paper Award in Wireless Communications, the 2014 Best Paper Award for the *EURASIP Journal on Advances in Signal Processing*, the IEEE GLOBECOM Best Paper Award in 2017, the IEEE ICC Best Paper Award in 2016, and the IEEE PIMRC Best Paper Award in 2014. A paper he co-authored received the 2016 Young Author Best Paper Award of the IEEE Signal Processing Society. He also received the 2016 IEEE ComSoc Asia-Pacific Best Young Researcher Award. He frequently serves on the technical program committees of major IEEE conferences in wireless communications, such as ICC, Globecom, WCNC, VTC, and so on, and served as a MAC track co-chair for the IEEE WCNC 2011. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and was a Guest Editor of the special section on Mobile Edge Computing for Wireless Networks in the IEEE ACCESS.



**Wei Chen** (S'05–M'07–SM'13) received the B.S. and Ph.D. degrees (Hons.) from Tsinghua University in 2002 and 2007, respectively. From 2005 to 2007, he was also a visiting Ph.D. student at The Hong Kong University of Science & Technology. Since 2007, he has been on the faculty at Tsinghua University, where he is a tenured Full Professor, the Director of the Academic Degree Office of Tsinghua University, and a member of the University Council. From 2014 to 2016, he served as the Deputy Head of the Department of Electronic Engineering. He has also held visiting appointments at several universities, including most recently at Princeton. His research interests are in the areas of wireless networks and information theory.

Dr. Chen is a member of the National 10000-Talent Program, a Cheung Kong Young Scholar, and a Chief Scientist of the National 973 Youth Project. He is also supported by the NSFC Excellent Young Investigator Project, the New Century Talent Program of Ministry of Education, and the Beijing Nova Program. He received the first prize of the 14th Henry Fok Ying-Tung Young Faculty Award, the 17th Yi-Sheng Mao Science & Technology Award for Beijing Youth, the 2017 Young Scientist Award of the China Institute of Communications, and the 2015 Information Theory New Star Award of the China Institute of Electronics. He received the IEEE Marconi Prize Paper Award and the IEEE Comsoc Asia Pacific Board Best Young Researcher Award in 2009 and 2011, respectively. He is a winner of the National May 1st Medal and Beijing Youth May 4th Medal. He also serves as the Executive Chairman of the Youth Forum of the China Institute of Communications. He serves as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON EDUCATION.



**Khaled B. Letaief** (S'85–M'86–SM'97–F'03) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1984, 1986, and 1990, respectively. From 1990 to 1993, he was a Faculty Member with The University of Melbourne, Melbourne, VIC, Australia. Since 1993, he has been with The Hong Kong University of Science and Technology (HKUST), where he is known as one of HKUST's most distinguished professors. While at HKUST, he has held numerous administrative

positions, including the Head of the Electronic and Computer Engineering Department, the Director of the Huawei Innovation Laboratory, and the Director of the Hong Kong Telecom Institute of Information Technology. While at HKUST, he has also served as the Chair Professor and the Dean of Engineering. Under his leadership, the school has also dazzled in international rankings (rising from No. 26 in 2009 to 14 in the world in 2015 according to QS World University Rankings). In 2015, he joined Hamad Bin Khalifa University as a Provost to help establish a research-intensive university in Qatar in partnership with esteemed universities that include Northwestern University, Cornell University, and Texas A&M. He has over 580 journal and conference papers and has given keynote talks and courses all over the world. He has 15 patents, including 11 U.S. patents. He is an internationally recognized leader in wireless communications with research interests in green communications, Internet of Things, cloud-RANs, and 5G systems. He served as a consultant for different organizations, including Huawei, ASTRI, ZTE, Nortel, PricewaterhouseCoopers, and Motorola. In addition to his active research and professional activities, he has been a dedicated teacher committed to excellence in teaching and scholarship. He is a fellow of HKIE. He received the Michael G. Gale Medal for Distinguished Teaching (highest university-wide teaching award and only one recipient/year is honored for his/her contributions). He was a recipient of many other distinguished awards and honors, including the 2007 IEEE Joseph LoCicero Publications Exemplary Award, the 2009 IEEE Marconi Prize Award in Wireless Communications, the 2010 Purdue University Outstanding Electrical and Computer Engineer Award, the 2011 IEEE Harold Sobol Award, the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, and over 11 IEEE best paper awards. He is recognized as a longtime volunteer with dedicated service to professional societies and in particular IEEE, where he has served in many leadership positions including as Treasurer, a Vice President for Conferences, and a Vice President for Technical Activities of the IEEE Communications Society. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher (which places him among the top 250 preeminent individual researchers in the field of computer science and engineering) and is currently the President of the IEEE Communications Society in 2018. He has served on the editorial board of other prestigious journals. He is the founding Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.