Distributed Fronthaul-Constrained Joint Transmission Design and Selection Using Augmented Consensus-based Dual Decomposition

Mykola Servetnyk and Carrson C. Fung

Abstract—User-centric coordinated multipoint (CoMP) joint transmission (JT) is a novel technique to manage interference and enhance system performance with single frequency reuse, where user equipment (UE) communicates with their closest transmission points (TPs). Unfortunately, in coherent JT, requirement of strict network synchronization accuracy makes it difficult and expensive to be practically deployed. The noncoherent JT has therefore received growing attention since it requires less strict network synchronization accuracy compared to its coherent counterpart as it does not require the signal to be phase-aligned at the receiver. Moreover, cell-free massive MIMO, which is regarded as combining CoMP with massive MIMO systems, has recently been touted as a solution for avoiding intercell interference and provide uniform coverage over a large area. However, the operational costs of CoMP, such as the associated control signaling and communication overhead and the increase of network complexity, could prevent the practical implementation of CoMP. A distributed joint transmission CoMP (JT-CoMP) scheme is proposed herein that allows distributed design and selection of cooperating transmission nodes. To maximize user capacity, the proposed distributed consensus optimization problem assumes spectrum underlay transmission is used so that the solution can achieve non-orthogonal multiple access (NOMA) for cell-free user centric JT-CoMP systems. The proposed algorithm is different from others in the literature because it solves a design problem that involves a coupling constraint that no existing algorithm can solve. Analytical results based on spectral graph theory are given to prove its convergence and characterize its rate of convergence. The more practical scenario is further considered, where limited fronthaul capacity is also included in the problem. A successive convex approximation (SCA) method is used to solve the resulting nonconvex problem, which is shown to maximize spectral efficiency. Simulation results are provided to show that the performance of both proposed distributed algorithms (that address problem without and with fronthaul constraint) is comparable to its centralized counterpart.

Index Terms—Distributed consensus algorithm, fronthaul constraint, NOMA, transmission point selection, user centric cell-free cooperative MIMO.

I. INTRODUCTION

AS spectral efficiency of point-to-point transmission in cellular networks approaches its theoretical limits, to

further improve network capacity and ease network management, traditional cellular networks are evolving into cell-free massive MIMO systems, regarded as a combination between user-centric coordinated multipoint (CoMP) [1] and traditional massive multiple-input multiple-output (MIMO) systems [2]. In this scenario, smarter resource coordination among transmission points (TPs) for efficient interference management and transmission coordination can provide substantial gains in throughput and user experience. CoMP allows joint transmission (JT) from cooperating TPs, which gives the ability to simultaneously transmit data to each user equipment (UE) by sharing the same channel resources, such as time and frequency, to boost reception performance. Qualcomm has implemented a 5G CoMP testbed [3] which demonstrated that system capacity can increase by four-fold when CoMP spatial multiplexing was used, thus proving CoMP's potential. However, finding an optimal solution for joint TP selection and transmission design problem is known to be NP-hard, because it requires search for the determination of the cooperating sets. Besides, the operational costs of JT-CoMP, prior phase alignment and tight synchronization to the target user requirements often hinder its implementation in industrial scale [4], not to mention communication overhead and the increase of network complexity.

Coordinated beamforming/scheduling and dynamic point selection have been investigated as alternatives which require less communication overhead compared to JT, but without achieving the same transmission performance as JT. There are two possible approaches to tackle the joint TP selection and transmission design problem: *centralized* and *distributed*. Among the advantages of the centralized approach is its relative simplicity in algorithm implementation and the possibility to directly use the interior point method (IPM) to solve the joint design problem. However, to take full advantage, the centralized server is required to handle high amount of computations with high throughput communication backhaul to every TPs. Among the data transmitted is the channel state information (CSI) from the TPs to the server and precoder coefficients from the server to the TPs. It is possible to satisfy such a requirement for small networks but this is not practical as the system scales up. Several joint transmit and TP selection design strategies have been proposed in the literature to reduce signaling overhead. [5] tackled the joint precoder and clustering problem by formulating the problem as a mixed-integer convex problem which minimizes the transmit energy subject to signal-to-interference-plus-noise quality-of-service (SINR QoS) constraint, where convex formulation was

obtained upon a reformulation of the constraint. However, the approach only applies to the case of a single receive antenna. Two algorithms were proposed in [6] with similar problem formulation, where one is based on iterative reweighed ℓ_1-norm minimization, and the other is based on solving the ℓ_2-norm relaxed problem and then iteratively removing the links that correspond to the smallest transmit power. [7] extended aforementioned designs by including per TP backhaul constraint and solved the design problem by customized branch and the bound algorithm applied to discrete monotonic optimization. [8] proposed to minimize a sum transmit signal power subject to the SINR QoS constraint similar to [5] with an additional fronthaul capacity constraint. Proposed problem formulation falls into a category of the mixed integer second order cone problem and can be solved with continuous relaxation, or the proposed inflation algorithm.

In addition, none of the aforementioned works are appropriate for large-sized networks due to computational complexity at the central unit that increases proportionally to the number of TPs in the network as was mentioned before. [9] solved the distributed precoder design problem by maximizing the weighted sum rate via minimization of the weighted sum mean-squared-error (WMMSE) proposed in [10], [11]. [12] developed a CSI signaling technique for coordinated beamforming for time division duplex (TDD) multi-cell system by also solving the weighted sum rate problem by reformulating the problem into minimizing the weighted sum mean-squared-error, making implementation of this CoMP method feasible. A solution for intercell interference level management for weighted sum rate problem was proposed in [13] using primal decomposition for multiple-input single-output (MISO) systems with SINR constraints. The nonconvex problem was decomposed into a master problem with several subproblems which, unlike the scheme in [9], only involves coordination between cooperating base stations. However, solving for the solution of the master problem requires synchronous base stations. [14] proposed to use dual decomposition method to decentralize interference management among adjacent cells, which was extended by [15] using the alternating direction of method of multiplier based decentralized beamforming design. [16] used the WMMSE technique in [9] and extend the linear transceiver design problem for JT-CoMP. [17] has embedded rate QoS constraint into the problem formulation and was able to solve the problem in a way similar to [9], [16] via successive convex approximation to turn one of the subproblems into convex form. [18] proposed a distributed precoding design scheme for JT-CoMP that requires coordination among cooperating base stations using backhaul signaling where correlation information among base stations, known as cross-term information, are exchanged. Unfortunately, the signaling scheme does not scale well with system size. [19] proposed distributed precoding scheme for cell-free massive MIMO system by minimizing weighted sum mean-squared-error, where a new over-the-air signaling scheme was proposed to exchange cross-term information that scales well with system size compared to the backhaul signaling scheme in [18]. Since the quality of over-the-air channel will usually be worse than that of the backhaul, this may adversely affect the quality of the precoder. Table I summarizes the methods above.

During the distributed design, agents in the network coordinate as a swarm to perform a certain task and reach *consensus*. [20] provided a theoretical framework for a wide range of consensus algorithms for multiagent networked systems. [21] and [22] have developed distributed algorithms based on consensus-based primal-dual decomposition and [23] proposed algorithm based on Newton's method. However, [21]–[23] have limitation in the form of the constraint to be $\sum_q g^q(x^q) \leq 0$, where q is the node index, x^q is the variable at node q and $g^q(x^q)$ is local constraint function at the qth node. Recently, [24] has extended the algorithm in [22] to be applied to a constraint of the form $\sum_q g^q(x^q) = L$, where author interprets L as a resource shared among nodes. Equality constraint implies that after optimization all resources must be used. However, it is more meaningful to replace equality with inequality $\sum_q g^q(x^q) \leq L$ to emphasize the amount of used resource can be flexible. Recently, [25] has considered a problem with the coupling constraint being an inequality and proposed an algorithm that converges to a fixed point. The main focus of that work is to solve a problem where the constraint graph does not necessarily have to equal to the physical network graph. However, global constraint functions $g^q(x^q)$ are limited to be an affine transformation. Moreover, problem formulation in [25] does not include local constraint, which significantly simplifies algorithm and convergence analysis.

In this work, a distributed consensus algorithm using dual decomposition, which circumvents the previously mentioned limitations using an adaptive "resource" allocation scheme, is proposed. The proposed scheme can be effectively applied to the distributed joint transmission and selection problem where each TP is modeled as a processing node in the network. Moreover, the amount of communication overhead does not scale with the number of transmission points. The problem of CSI overhead and data sharing among transmission points is not considered herein. The internode communications are modeled as an undirected graph, where the resulting graph Laplacian is used to prove the convergence of the proposed technique. This proof augments the ones from [21], [22], [24], [26], [27] and extends them with the incorporation of the analysis of the resource allocation scheme. In addition, none of the distributed CoMP schemes above considered non-orthogonal multiple access (NOMA) where transmitters can transmit simultaneously using the same time, frequency and space. This allows more users to be served and enhanced spectrum efficiency. Spectrum underlay techniques [28]–[30] are used herein where transceivers behave altruistically such that transmission between one pair of transceiver does not cause noticeable interference to other users (considered as victims). The interference can be managed by constraining the transmit power emitted by the transmitter toward its victims such that victims are oblivious to other transmissions. Finally, to relax the stringent requirement of coherent JT, noncoherent JT is considered. The main contributions and novelties of this work are as follows:

- This work attempts to jointly optimize the coordination clusters and linear precoders in the network by solving a nonconvex programming problem. Solving the problem

Table I
COMPARISON OF DIFFERENT COMP DESIGNS.

	Objective	Limitations
[5]	Transmit power	Limited to single antenna receivers
[6]	Reweighted ℓ_1 and ℓ_2 relaxed of transmit power	Limited to single antenna receivers
[7]	Sum rate	Requires branch and bound, slow convergence
[8]	Total power consumption	Use of MISOCP solver (high computational complexity) or continuous relaxation
[9]	Weighted sum rate	Design for coordinated beamforming, misestimation of received signal covariance matrix can degrade performance
[12]	Weighted sum rate	Design for coordinated beamforming, guarantees to only obtain a stationary solution of the WMMSE problem
[13]	Weighted sum rate	Design for coordinated beamforming, requires synchronous base stations to solve master problem
[14]	Total transmit power	Requires consistency among interference at different base stations, requires single antenna receivers
[15]	Weighted total transmit power	Design for coordinated beamforming, signaling between base stations may be high for large system
[16]	Weighted sum rate, proportional fairness	Guarantees to only obtain a stationary solution of the WMMSE problem
[17]	Weighted sum-rate	Local optimal solution is obtained using SCA. QoS constraint may not be satisfied for fast-fading channels
[18]	Weighted sum rate	Requires backhaul signaling that does not scale well with system size
[19]	Weighted sum mean squared error	Quality of over-the-air channel may adversely impact precoder design

allows users to transmit non-orthogonally, enabling cell-free user centric JT-CoMP systems [2], thus increasing user capacity compared to orthogonal multiple access techniques. The resulting solution maximizes received signal power at the UEs, which in return increases the overall system sum rate.

- The transformed problem can be regarded as a multiagent convex problem with coupling constraints including one of the form $\sum_q g^q(x^q) \leq L$. An efficient distributed algorithm that computes a stationary solution to the transformed problem is proposed.
- The proposed distributed algorithm generalizes the works in [21], [22], [24] as they failed to handle $\sum_q g^q(x^q) \leq L$ type of constraints. Analytical results prove the convergence property of the proposed algorithm to a stationary solution of the transformed problem. It is important to solve the problem distributively with inequality constraints as the considered problem with equality constraints is often infeasible.
- Algorithm is further extended to incorporate the fronthaul constraint as part of the design. Proposed consensus based primal dual decomposition method is augmented with successive convex approximation (SCA) and alternating direction method of multipliers (ADMM) and allows for sophisticated fronthaul resource allocation.
- The convergence and effectiveness of the proposed algorithm have been theoretically validated and evaluated via extensive simulations.

The rest of the paper is organized as follows. In Section II, system model is presented. The joint TP selection and precoder design problem is formulated and then transformed into convex form in Section III. The proposed efficient distributed algorithm is proposed in Section IV to solve the transformed problem. The algorithm convergence properties and underlying assumptions are stated in Section V. Extension of the proposed distributed algorithm to fronthaul-constrained transmission nodes is presented in Section VI. Discussion about proposed algorithms' complexity and system overhead is given in Section VII, followed by numerical examples and analytical results to validate the proposed algorithms in Section IX. The paper is concluded in Section X and proofs of the theorems are given in the Appendix.

Table II
NOTATIONS.

Symbol	Definition
B	Number of spatial streams.
$c^{(n)}$	Dual consensus/objective balance coefficient.
C^q	qth TP's fronthaul capacity threshold.
$\mathcal{C}^q/\widetilde{\mathcal{C}}^q$	Local constraint set of qth TP.
$f^q(\cdot)$	qth TP's objective function.
$\mathbf{F}_i^q$	Precoder from qth TP to ith UE.
$\mathbf{g}^q(\cdot)$	qth TP's coupling constraint term.
$\mathbf{H}_i^q$	Channel from qth TP to ith UE.
i,j	UEs subscript indices.
I	Total number of users
I_{th}	Instantaneous interference leakage power constraint threshold.
$\mathbf{Q}_i^q$	Precoder variable from qth TP to ith UE.
n_T/n_R	Number of Tx/Rx antennas.
$L_{ij}^{q(n)}$	Instantaneous leakage interference power at nth iteration.
$\mathcal{N}^q$	qth TP's neighborhood set.
$\mathcal{Q}$	TPs cooperating set, i.e., all TPs.
q,r	TPs superscript indices.
S_{ij}^q	Slack leakage interference power for qth TP on ij interference link.
$\mathbf{W}$	Consensus matrix.
$[\mathbf{W}]_{rq}$	Weight between TPs q and r.
α	Fronthaul penalty parameter.
$\boldsymbol{\lambda}^q$	Dual variable for interference leakage constraint.
$\boldsymbol{\ell}^q$	Consensus dual variable for interference leakage constraint.
σ^2	Noise power
φ	Number of consensus iterations in $\mathbf{W}^\varphi$

Notations: Uppercase (lowercase) bold face letters indicate matrices (column vectors). Superscript ^{H} denotes Hermitian, ^{T} denotes transposition. $\mathrm{Diag}(\mathbf{a})$ creates an $N \times N$ diagonal matrix using the elements of $\mathbf{a} \in \mathbb{C}^N$. $\mathbf{A} \succeq \mathbf{0}$ designates $\mathbf{A}$ as a symmetric positive semidefinite matrix, $\mathbf{1}_M$ denotes an $M \times 1$ vector, containing 1 in all of its entries. $[\mathbf{A}]_{mn}$ denotes the (m, n)th element of $\mathbf{A}$. $\|\cdot\|_0$, $\|\cdot\|_2$, and $\|\cdot\|_F$ denote ℓ_0, ℓ_2 and Frobenius norm, respectively. $|\mathbf{A}|$ denotes the elementwise magnitude value of $\mathbf{A}$. $\mathcal{P}_+[\cdot]$ denotes the projection operator that projects the argument onto $\mathbb{R}_+$. The superscripts $^{q(n)}, ^{r(n)}$ on $\mathbf{Q}_i^{q(n)}, \mathbf{Q}_j^{r(n)}$ are used to denote the matrix $\mathbf{Q}$ at the q, rth TP and at the nth global iteration.
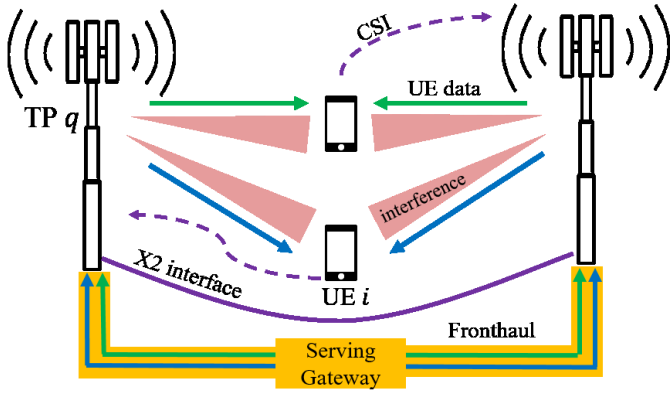
Fig. 1. System model.

II. SYSTEM MODEL

A depiction of the system model under consideration is shown in Fig. 1 and described below. Assume the network consists of a set of TPs $\mathcal{Q} = \{1, \cdots, Q\}$, known as the cooperating set, each having n_T transmit antennas. $\mathcal{I} = \{1, \cdots, I\}$ denotes a set of UEs that should be served by subset of $\mathcal{Q}$, with each user equipped with n_R receive antennas. Data from UEs are delivered to each TP from a serving gate through a fronthaul link (labeled in yellow) with limited capacity and transmitted through a serving link (green and blue arrows). The fundamental idea of JT-CoMP is to simultaneously serve multiple users over the same spectrum resources (time, frequency, and space) at the expense of inter-user interference (red beams). It is assumed that TPs can estimate channel through uplink pilot sequence and share it with cooperating set through X2 interface [31]. Denote channel between the qth TP and the ith user as $\mathbf{H}_i^q \in \mathbb{C}^{n_R \times n_T}$. Furthermore, precoder matrix from the qth TP to the ith user is defined as $\mathbf{F}_i^q \in \mathbb{C}^{n_T \times B}$, where $B \leq n_T$ denotes number of spatial streams. The received signal for user i can thus be written as $\mathbf{y}_i = \sum_q \mathbf{H}_i^q \mathbf{F}_i^q \mathbf{s}_i + \sum_{j \neq i} \sum_q \mathbf{H}_i^q \mathbf{F}_j^q \mathbf{s}_j + \mathbf{n}_i$, where the second term represents interference and $\mathbf{n}_i$ denotes additive white Gaussian noise (AWGN) with known variance σ_i^2. It is assumed that the noise is uncorrelated with the data signal $\mathbf{s}_i \in \mathbb{C}^B$ with $\mathbb{E}[\mathbf{s}_i^H \mathbf{s}_i] = 1$. Major notations used throughout this paper can be found in Table II.

In coherent JT, based on the CSI shared among all cooperating set, all TPs transmit signals that are jointly precoded with prior phase alignment and tight synchronization to the target user [32]. The requirement of strict network synchronization accuracy makes it difficult and expensive to be practically deployed. The noncoherent JT has therefore received growing attention since it requires less strict network synchronization accuracy compared to its coherent counterpart. It, however, requires more complex operation at the receiver because the UEs need to successfully decode all received signals [33] and combine them such that the overall received signal power is equal to the sum of received signal powers from each TP. Nevertheless, due to its advantages, the noncoherent JT scenario is considered, which led to the problem formulation below.

III. PROBLEM FORMULATION

A. Joint TP Selection and Precoder Design Objective

In the system model considered hereafter, a user receives a noncoherent sum of multiple data streams of the useful signal transmitted by the cooperating TPs (i.e., noncoherent JT). Therefore, the UEs' rate equals to a sum of rates from individual TPs. Due to the monotonic relationship between received signal power and SINR, direct dependence of achievable throughput on SINR, and the use of spectrum underlay transmission, the problem is formulated as maximization of a sum of the received signal power subject to the instantaneous leakage interference power being constrained below the threshold parameter I_{th}, which is usually obtained from long-term statistical measures or link budget to guarantee successful communication between TP and UE [34] for all users. Notably, for noncoherent JT-CoMP, received signal power is represented as sum of powers of each TP-UE link. In contrast, for the coherent version, the power would be written as power of sum of the signals as signals are phase-aligned. However, the leakage interference constraint is still not amenable to distributed design, as will be shown in the sequel. Finally, the transmit power of the $q \in \mathcal{Q}$ TP is limited below P^q. However, an increasing number of cooperating TPs will impact operational complexity and create extra load on the fronthaul infrastructure, which has to deliver each UEs' data to each TP in the cooperating set. This can preclude usage of the cooperative scheme in practice. The number of precoding elements at all TPs selected to be in the cooperation set can be done by including a sparsity inducing ℓ_0 regularization term in the objective. Hence, the problem can be formulated as

$$\max_{\mathbf{F}_i^q, i \in \mathcal{I}, q \in \mathcal{Q}} \sum_i \sum_q \|\mathbf{H}_i^q \mathbf{F}_i^q\|^2 - \alpha \|\mathbf{F}_i^q\|_0^2 \quad (1a)$$

$$s.t. \sum_q \|\mathbf{H}_j^q \mathbf{F}_i^q\|^2 \leq I_{th}, \ i, j \in \mathcal{I} : i \neq j \quad (1b)$$

$$\sum_i \|\mathbf{F}_i^q\|^2 \leq P^q, q \in \mathcal{Q}, \quad (1c)$$

where $\alpha \in \mathbb{R}_+$ is a regularization parameter that represents the associated cost for assigning TP to a UE and shall be referred to as the fronthaul penalty parameter. An increase in α will undoubtedly promote sparsity in the precoder vector, thus lowering operating cost. Although many NOMA techniques have been proposed in literature, only spectrum underlay is considered as the focus of this work is on distributed user-centric noncoherent JT-CoMP.

B. Problem Reformulation

It can be observed that (1) is not a convex problem because the objective consists of sum of convex and nonconvex terms and optimal solution cannot be found in polynomial time. This can be easily circumvented by defining $\mathbf{Q}_i^q \triangleq \mathbf{F}_i^q \mathbf{F}_i^{qH}$ such that received signal power for user i becomes $\sum_q tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH})$ and all terms inside the constraints can be rewritten in similar fashion. Finally, relaxing the ℓ_0 norm by replacing it with the ℓ_1 norm and using rank relaxation on $\mathbf{Q}_i^q$ [35], i.e., removing the constraint $\text{rank}(\mathbf{Q}_i^q) = B$, the complete design problem can be written as

$$\max_{\mathbf{Q}} \sum_i \sum_q tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}) - \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T} \quad (2a)$$

Algorithm 1: Distributed consensus-based dual decomposition joint TP design and selection.

---

0. *Initialization* Choose $\boldsymbol{\lambda}^{q(0)} \in \mathbb{R}^+$, $\alpha \in \mathbb{R}^+$, γ, sequence $c^{(n)} = \gamma/(n+1), \varphi \in \mathbb{Z}_+$. Choose ϵ_{glo} as stop criteria parameter. Set $n = 0$ as iteration index.

while $\|\mathbf{Q}^{(n)} - \mathbf{Q}^{(n-1)}\|_F \le \epsilon_{glo}$ **do**

    1. *Local primal optimization*: Compute $\mathbf{Q}_i^{q(n+1)}$ by solving (9); Compute $\boldsymbol{\lambda}^{q(n+1)}$ using (12);

    2. *Communication*: Compute $S_{ij}^{q(n+1)}$ and send it with $\boldsymbol{\lambda}^{q(n+1)}$ to neighbors;

    3. *Local dual optimization*: Compute $\boldsymbol{\ell}^{q(n+1)}$ by (13) and obtain $I_{th,ij}^{q(n+1)}$ using (14); $n = n + 1$;

4. *Obtain precoder* $\mathbf{F}_i^q$: Apply randomization procedure on $\mathbf{Q}_i^{q(n)}$ as described in Section VIII.

$$s.t. \sum_q tr\left(\mathbf{H}_j^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}\right) \le I_{th}, \; i \in \mathcal{I} : j \ne i \quad (2b)$$

$$\sum_i tr\left(\mathbf{Q}_i^q\right) \le P^q, \; \mathbf{Q}_i^q \succeq \mathbf{0}, \; q \in \mathcal{Q}, i \in \mathcal{I}, \quad (2c)$$

where maximization with respect to $\mathbf{Q}$ denotes maximization with respect to $\mathbf{Q}_i^q, \forall i \in \mathcal{I}, \forall q \in \mathcal{Q}$. (2) is a convex programming problem and global optimal solution to it can be found using interior-point method [35]. After $\mathbf{Q}_i^q$ is obtained, $\mathbf{F}_i^q$ can be obtained via randomization procedure as described in Section VIII, which is asymptotically optimal approach for the rank recovery.

Notice that (2b) contains $|\mathcal{I}|(|\mathcal{I}| - 1)$ number of constraints and $|\mathcal{Q}|$ variables, with the former usually greater than the latter. Thus, if these constraints are converted to equality constraints, there is no way to guarantee feasibility in (2), hence, this is not considered in this work.

IV. DISTRIBUTED CONSENSUS BASED ALGORITHM WITH UNLIMITED FRONTHAUL

The proposed distributed algorithm is summarized in Algorithm 1 and developed based on a combination of dual decomposition and proximal minimization with inequality-coupled constraints that are different from the one in [21] and [22], which is of the form $\sum_q g^q(x^q) \le 0$, where only dual variables are exchanged between agent (TP) q and its one-hop neighbors, i.e., members of $\mathcal{N}^q \subseteq \mathcal{Q}$. This increases privacy protection as it is difficult, in the context of the current problem, to convert the dual variables to the primal ones even if someone is eavesdropping on the network. Local copies of the dual variable and auxiliary variable will have to be created and exchanged to achieve consensus amongst all the nodes.

Note that the leakage interference power constraint (2b) can only be written in this form $\sum_q tr(\mathbf{H}_j^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}) - I_{th} \le 0$, if processing is done in a centralized manner, but not in a distributed manner, as there is no apparent way to optimally distribute I_{th} amongst all the agents without impacting system performance such as sum rate.

It is also possible to argue, since I_{th} is the interference threshold, each agent should have *a priori* information about I_{th} at each node. However, the goal here is to obtain a solution that is close, or identical, to the one obtained by centralized processing. Hence, being able to obtain the optimal interference threshold for each node is required. Moreover, even though it is possible for each node to determine its interference leakage power threshold, this will most likely require an exchange of location information amongst cooperating TPs, which will raise privacy concerns.

The proposed algorithm is separated into three steps, namely, local processing, communications, and a consensus step. All steps are summarized in Algorithm 1 and are described in detail in following sections.

A. Local Processing - Computing $\mathbf{Q}_i^q$, $\boldsymbol{\lambda}^q$, and $\boldsymbol{\ell}^q$

Define $\mathcal{C}^q \triangleq \{\mathbf{Q}_i^q : \sum_i tr(\mathbf{Q}_i^q) \le P^q, \mathbf{Q}_i^q \succeq \mathbf{0}, \; i \in \mathcal{I}, q \in \mathcal{Q}\}$, then (2) can be written as

$$\max_{\mathbf{Q}} \sum_q \sum_i tr\left(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}\right) - \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T} \quad (3a)$$

$$s.t. \sum_q tr\left(\mathbf{H}_j^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}\right) \le I_{th}, \; i, j \in \mathcal{I} : j \ne i \quad (3b)$$

$$\mathbf{Q}_i^q \in \mathcal{C}^q, \; i \in \mathcal{I}, \; q \in \mathcal{Q}. \quad (3c)$$

(3) is a convex optimization problem, but the existence of the coupling constraints (3b) makes it impossible to solve (3) in a distributive manner as solution at each node depends on precoder from other nodes, resulting in an inequality-coupled problem [22]. In such a problem, each agent aims to optimize a local performance criterion subject to local constraints and yet, the decision variables from each agent need to agree. In the case of (3), I_{th} is regarded as a resource that is available to all agents but its exact value is unknown at each node. [21] and [22] considered a similar problem, however, assumed resource is equally distributed. For problem (3), it means that I_{th} is evenly distributed amongst all agents as $I_{th,ij}^q = I_{th}/Q, \forall q, i \ne j$, which represents the leakage interference power threshold for the qth agent that is used to bound the instantaneous leakage interference power to the jth user when signal is transmitted to the ith user. However, equal distribution of the resource is not optimal as shown in Fig. 3 of [36]. Therefore, an adaptive algorithm for interference leakage resource allocation is proposed herein to obtain a suboptimal $I_{th,ij}^q$, which is presented in detail in Section IV-B.

Before presenting distributed optimization framework, (3) shall be rewritten in shorter form for clarity. First, denote the objective function as

$$f^q\left(\mathbf{Q}^q\right) \triangleq \sum_i tr\left(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}\right) - \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T}$$

with $\mathbf{Q}^q \triangleq \texttt{Diag}([\mathbf{Q}_1^q \cdots \mathbf{Q}_I^q])$. Second, the constraint function rewritten as

$$\sum_q tr(\mathbf{H}_j^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}) - I_{th,ij}^q \triangleq \sum_q g_{ij}^q\left(\mathbf{Q}_i^q\right) \le 0.$$

Now, (3) can be written as

$$\max_{\mathbf{Q}} \sum_q f^q\left(\mathbf{Q}^q\right) \tag{4a}$$

$$s.t. \sum_q g_{ij}^q\left(\mathbf{Q}_i^q\right) \leq 0, \forall i,j \in \mathcal{I},\ j \neq i \tag{4b}$$

$$\mathbf{Q}_i^q \in \mathcal{C}^q,\ i \in \mathcal{I},\forall q \in \mathcal{Q}. \tag{4c}$$

To begin with distributed algorithm for (4), the problem with coupling constraint (4b) should be circumvented. This is done by forming a Lagrangian function that includes the objective and the coupling constraint. Dual variables for user pairs (i,j) of node q are denoted as $\boldsymbol{\lambda}^q \triangleq [\lambda_{12}^q,\cdots,\lambda_{1I}^q,\cdots,\lambda_{ij}^q,\cdots,\lambda_{I(I-1)}^q]^T$ and correspond to constraint functions $\mathbf{g}^q\left(\mathbf{Q}^q\right) \triangleq [g_{12}^q, g_{13}^q,\cdots,g_{1I}^q,\cdots,g_{ij}^q,\cdots,g_{I(I-1)}^q]^T \in \mathbb{R}^{I(I-1)}$. Therefore, the Lagrangian can be written as

$$\mathcal{L}(\mathbf{Q},\boldsymbol{\lambda}) = \sum_q f^q\left(\mathbf{Q}^q\right) + \boldsymbol{\lambda}^{qT}\mathbf{g}^q\left(\mathbf{Q}^q\right)$$
$$= \sum_q \mathcal{L}^q\left(\mathbf{Q}^q,\boldsymbol{\lambda}^q\right) \tag{5}$$

with requirement that $\boldsymbol{\lambda}_{ij}^q = \boldsymbol{\lambda}_{ij}^r, \forall q,r \in \mathcal{Q}$ for each (i,j) user pair, so that constraint (4b) is satisfied. Notably, terms inside (5) are independent for different q and can be computed by different node independently and in parallel. In order for all nodes to reach consensus in the Lagrangian function above, $\boldsymbol{\lambda}^q$ will be broadcasted from each node $q \in \mathcal{Q}$, to its neighboring set $\mathcal{N}_q$. An auxiliary dual variable $\boldsymbol{\ell}^q$, computed as $\boldsymbol{\ell}^q \triangleq \sum_{r \in \mathcal{N}^q}[\mathbf{W}^\varphi]_{rq}\boldsymbol{\lambda}^r$, is used to achieve consensus on dual variable $\boldsymbol{\lambda}^q$. First, $\boldsymbol{\ell}^q$ is used as a surrogate for $\boldsymbol{\lambda}^q$ when solving for $\mathbf{Q}_i^q$. It is then used as a proximal term when solving for $\boldsymbol{\lambda}^q$ to promote consensus. These two steps will be elaborated clearly later. The update equation for $\boldsymbol{\ell}^q$ is a weighted sum of $\boldsymbol{\lambda}^q$, where the weights are contained inside the *consensus matrix* [26] $\mathbf{W} \in \mathbb{R}^{Q \times Q}$, with properties

$$[\mathbf{W}]_{rq} = 0 \text{ if } r \notin \mathcal{N}^q,\ \mathbf{W} = \mathbf{W}^T,\ \mathbf{W}\mathbf{1}_Q = \mathbf{1}_Q, \tag{6}$$

$$\lim_{\varphi \to \infty} \mathbf{W}^\varphi = \frac{\mathbf{1}_Q\mathbf{1}_Q^T}{Q},\ \text{and } \rho\left(\mathbf{W} - \frac{\mathbf{1}_Q\mathbf{1}_Q^T}{Q}\right) \leq \nu < 1, \tag{7}$$

where $\rho(\cdot)$ is the spectral radius operator and ν is spectral radius upper bound value. φ denotes the number of consensus iterations, which reflects how far messages from one node are passed to next-hop nodes. $\mathbf{W}$ is used to model the TP network as a connected undirected graph. As indicated in Algorithm 1, the update for $\boldsymbol{\ell}^q$ is carried out after that of $\boldsymbol{\lambda}^q$.

Since consensus for $\boldsymbol{\lambda}^{q(n)}$ is obtained through $\boldsymbol{\ell}^q$, hence, the Lagrangian function in (5) can be approximated at point $\boldsymbol{\ell}^q$ as

$$\mathcal{L}^q(\mathbf{Q}^q,\boldsymbol{\ell}^q) = f^q\left(\mathbf{Q}^q\right) + \boldsymbol{\ell}^{qT}\mathbf{g}^q\left(\mathbf{Q}^q\right) \tag{8}$$

and it will be used to find the precoder matrix. Before proceeding, it should be indicated that when iteration index $^{(n)}$ appears above certain mathematical quantity, it means that this variable is fixed and obtained after corresponding optimization iteration n. When the index is absent, the corresponding quantity is a variable or contains a variable that should be optimized. For instance, $\mathbf{g}^q$ in (9) contains $\mathbf{Q}^q$, while $\mathbf{g}^{q(n)}$ in (11) is a quantity evaluated at iteration n at $\mathbf{Q}^{q(n+1)}$.

The local processing step is stated as follows. At the nth iteration of the proposed algorithm, part of the local processing stage will be solving for $\mathbf{Q}_i^q$ in the next iteration and will require maximizing $\mathcal{L}^q(\mathbf{Q}^q,\boldsymbol{\ell}^q)$ at a fixed point $\boldsymbol{\ell}^{q(n)}$ to obtain

$$\mathbf{Q}^{q(n+1)} = \arg\max_{\mathbf{Q}^q \in \mathcal{C}^q} f^{q(n)}\left(\mathbf{Q}^q\right) + \boldsymbol{\ell}^{q(n)T}\mathbf{g}^q\left(\mathbf{Q}^q\right). \tag{9}$$

In the simulation, $\boldsymbol{\ell}^{q(0)}$ was initialized to be 4. The other part of the local processing involves solving for the dual variable $\boldsymbol{\lambda}^q$ which can be done by minimizing the Lagrange dual function $d\left(\boldsymbol{\lambda}^q\right) \triangleq \max_{\mathbf{Q}^q \in \mathcal{C}^q} \mathcal{L}^q(\mathbf{Q}^q,\boldsymbol{\lambda}^q)$ with respect to $\boldsymbol{\lambda}^q$. This dual function will be augmented with a proximal term to incentivize the consensus of the dual variables. Specifically, the new dual problem is formulated as

$$\min_{\boldsymbol{\lambda}^q \geq 0} \max_{\mathbf{Q}^q \in \mathcal{C}^q} \left\{\mathcal{L}^q(\mathbf{Q}^q,\boldsymbol{\lambda}^q) + \frac{\|\boldsymbol{\lambda}^q - \boldsymbol{\ell}^{q(n)}\|_2^2}{2c^{(n)}}\right\} \tag{10}$$

for each TP, where the second term in the inner objective is a proximal term. $c^{(n)}$ is a scalar that balances emphasis between objective and consensus on $\boldsymbol{\lambda}^q$. The update of $\boldsymbol{\lambda}^{q(n)}$, denoted as $\boldsymbol{\lambda}^{q(n+1)}$, can be obtained as

$$\boldsymbol{\lambda}^{q(n+1)} = \arg\min_{\boldsymbol{\lambda}^q \geq 0} \boldsymbol{\lambda}^{qT}\mathbf{g}^{q(n)} + \frac{\|\boldsymbol{\lambda}^q - \boldsymbol{\ell}^{q(n)}\|_2^2}{2c^{(n)}}, \tag{11}$$

which requires the computation of $\mathbf{g}^{q(n)} \triangleq tr(\mathbf{H}^q\mathbf{Q}^{q(n+1)}\mathbf{H}^{qH}) - I_{th,ij}^{q(n)}$. Using the fact that $\boldsymbol{\lambda}^q \succeq \mathbf{0}$, then $\boldsymbol{\lambda}^q$ for the next iteration is obtained via differentiation as

$$\boldsymbol{\lambda}^{q(n+1)} = \mathcal{P}_+\left[\boldsymbol{\ell}^{q(n)} - c^{(n)}\mathbf{g}^{q(n)}\right], \tag{12}$$

where $c^{(n)}$ can be regarded as an (adaptive) step size at the nth iteration and $\mathbf{g}^{q(n)}$ is a subgradient of the objective function in (11). Finally, $\boldsymbol{\ell}^{q(n)}$ is updated according to

$$\boldsymbol{\ell}^{q(n+1)} = \sum_{r \in \mathcal{N}^q}[\mathbf{W}^\varphi]_{rq}\boldsymbol{\lambda}^{r(n+1)}. \tag{13}$$

The local updates of $\mathbf{Q}_i^q$, $\boldsymbol{\lambda}_i^q$ and $\boldsymbol{\ell}_i^q$ are summarized in Step 1 and 3 of Algorithm 1, respectively.

B. Adaptive Strategy for Computing $I_{th,ij}^q$

Since the coupling constraint $\sum_q g_{ij}^q\left(\mathbf{Q}_i^q\right) \leq 0$ must have prior knowledge about $I_{th,ij}^q$ and even distribution of I_{th} is not optimal, adaptive resource allocation algorithm is proposed to allocate I_{th} across all agents which achieves better performance than the even distribution scheme in [21], [22], [24]. The issue of distributing resource for the coupling sum constraint was considered in [37], in which knowledge about the lower and upper bound of the second-order derivative of the objective is required. However, this approach cannot be applied to (4) as the objective is not differentiable. Even if the penalty term in (3a) is removed, the second-order derivative equals to zero. Besides, it is unclear how the scheme in [37] can be applied with additional coupling constraints as is the case in Section VI.

To address problem above, novel resource allocation proposed in [36] is used, while the scheme's optimality is proven in this work. The procedure is illustrated as follows. Initialization is done with $I_{th,ij}^{q(0)} = I_{th}/Q$, i.e., equal distribution. Define the *slack leakage interference power* as $S_{ij}^{q(n+1)} \triangleq I_{th,ij}^{q(n)} - L_{ij}^{q(n+1)}$

where $L_{ij}^{q(n+1)} \triangleq tr(\mathbf{H}_j^q \mathbf{Q}_i^{q(n+1)} \mathbf{H}_j^{qH})$ is the *instantaneous leakage interference power*. $S_{ij}^{q(n+1)}$ can be viewed as the amount of excess leakage interference power (resource) that is not used by the qth TP while serving the ith user. Then a new value for the leakage threshold can be updated as

$$I_{th,ij}^{q(n+1)} \triangleq L_{ij}^{q(n)} + \sum_{r \in \mathcal{N}^q}[\mathbf{W}]_{rq} S_{ij}^{r(n+1)}, \qquad (14)$$

which corresponds to Step 3 of Algorithm 1. Notice that the slack leakage interference power $S_{ij}^{q(n+1)}$ is transmitted from the qth agent to all members of $\mathcal{N}^q$ in Step 2 in Algorithm 1. After $I_{th,ij}^{q(n+1)}$ is computed at the nth iteration, it will be used in (9) to compute $g_{ij}^q(\mathbf{Q}_i^q) = tr\left(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}\right) - I_{th,ij}^{q(n+1)}$, which will be used to obtain $\mathbf{Q}^{q(n+2)}$ at the next iteration. Notice that all nodes $q \in \mathcal{Q}$ will be running the above steps in parallel. Define $\mathcal{L}_{na}^{(n)} \triangleq \left\{S_{ij}^{q(n)}|S_{ij}^{q(n)} > 0\right\}$ and $\mathcal{L}_a^{(n)} \triangleq \left\{S_{ij}^{q(n)}|S_{ij}^{q(n)} = 0\right\}$. The optimality of the proposed adaptive resource allocation scheme is stated in the theorem below.

Theorem 1. *Adaptive interference leakage allocation scheme converges, i.e.,*

$$\lim_{n \to \infty} \sum_q S_{ij}^{q(n)} = S_{ij}^\infty \qquad \forall i, j \in \mathcal{I}, j \neq i.$$

Moreover, if leakage interference constraint is active at the optimal solution of (3), then

$$\lim_{n \to \infty} \sum_q S_{ij}^{q(n)} = 0 \qquad \forall i, j \in \mathcal{I}, j \neq i.$$

That is, $\lim_{n \to \infty} \mathcal{L}_{na}^{(n)}$ *will be empty.*

See Appendix A for proof.

V. ALGORITHM CONVERGENCE

In this section, convergence of the proposed primal-dual decomposition algorithm is presented. Theorems 2–4 shown herein are similar to those in [21] and state the dual convergence and optimality (Theorem 2 and Theorem 3, respectively), and then, primal objective optimality (Theorem 4). Notably, these optimality bounds are different from [21] due to two reasons. First, the proposed algorithm uses an adaptive step size $c^{(n)}$ instead of fixed step size. Second, in [21], consensus step is done prior to projection, which requires computation of sets D_μ and $D_\mathbf{G}$ over which projections are done. For definitions of these sets, please refer to the original work. On the contrary, in the proposed algorithm, projection is done prior to consensus and hence avoids the difficulty and requires a new proof.

Convergence results described in a sequel hold under the assumptions below.

Assumption 1. *Between any two nodes $q, r \in \mathcal{Q}$ in a network there exists a path of edges.*

Assumption 2. $\{c^{(n)}\}_{n \geq 0}$ *is a nonincreasing sequence of positive reals and satisfies*

$$\lim_{n \to \infty} c^{(n)} = 0; \quad \sum_{n=0}^\infty c^{(n)} = \infty; \quad \sum_{n=0}^\infty (c^{(n)})^2 < \infty.$$

One possible choice for $c^{(n)}$ satisfying Assumption 2 is $c^{(n)} = \gamma/(n+1)$ for some $\gamma > 0$. Notice that choice of $c^{(n)}$ that satisfies Assumption 2 guarantees $\lim_{n \to \infty} \|\boldsymbol{\lambda}^{(n)} - \boldsymbol{\ell}^{(n)}\|_2^2 = 0$.

Assumption 3. *The difference between consensus dual variables after the first iteration of Algorithm 1 is bounded as* $\|\boldsymbol{\ell}^{q(1)} - \boldsymbol{\ell}^{r(1)}\|_2 \leq \beta^{(1)}, \forall q, r \in \mathcal{Q}.$

Assumption 4. *Constraint functions $g_{ij}^q(\mathbf{Q}_i^q), \forall \mathbf{Q}_i^q$, are finite, i.e.,*

$$\|g_{ij}^q(\mathbf{Q}_i^q)\|_2 \leq G, \text{ for any } \mathbf{Q}_i^q \in \mathcal{C}^q.$$

Assumption 4 defines a bound on the subgradient of the qth Lagrange dual function in (5), that exists due to convexity nature of the dual problem. From (3), the value of G can be upper bounded by $P^q\|\mathbf{H}_j^q\|_F^2$.

Assumption 5. *Lagrangian multipliers and related variables are bounded as following*

$$\|\boldsymbol{\lambda}^q\|_2^2 \leq \Lambda, \ \|\boldsymbol{\ell}^q\|_2^2 \leq \Lambda, \|\mathbf{y}\|_2^2 \leq \Lambda, \qquad (15)$$

for some nonnegative constant Λ. This assumption follows naturally from the assumption of the bounded dual function. $\mathbf{y}$ is an auxiliary dual variable and is defined in Appendix D to be used for the convergence proof.

Theorem 2. *Dual variable convergence*
Define $\bar{\boldsymbol{\ell}}^{(n)} \triangleq \frac{1}{Q}\sum_q \boldsymbol{\ell}^{q(n)}$ as the mean value of the consensus dual variables. Assume Assumptions 1-5 hold, $\mathbf{W}$ satisfies the conditions in (6) and ν is defined as in (7). Let $p := \frac{\nu^\delta \beta^{(1)}}{\beta^{(1)} + c^{(n)}G}$. There exists a number of consensus iterations $\bar{\varphi}$, such that if $\varphi \geq \bar{\varphi} + \delta, \delta \geq 0$, then $\boldsymbol{\ell}^q$ reaches consensus as

$$\|\boldsymbol{\ell}^{q(n)} - \bar{\boldsymbol{\ell}}^{(n)}\|_2 \leq 2p^{n-1}\nu^\delta \sigma\beta^{(1)} + 2pc^{(n)}G\frac{1-p^{n-1}}{1-p}, \forall q \in \mathcal{Q}.$$

Moreover, $\bar{\varphi} = \frac{\log(\beta^{(1)}) - \log(4Q(\beta^{(1)} + c^{(n)}G))}{\log(\nu)} + \delta \triangleq \bar{\varphi} + \delta.$

Theorem 3. *Dual objective optimality*
Let $\boldsymbol{\lambda}$ be generated with Algorithm 1, which corresponds to ϵ-subgradient algorithm [38]. Given that the Lagrange dual function $d(\boldsymbol{\lambda}) \triangleq \arg\max_\mathbf{Q} \mathcal{L}(\mathbf{Q}, \boldsymbol{\lambda})$ is bounded, then

$$\liminf_{n \to \infty} d(\boldsymbol{\lambda}^{(n)}) = d^\star + Q\zeta,$$

where $d^\star$ denotes the optimal value of $d(\boldsymbol{\lambda})$ and ζ is a constant inside the ϵ-subgradient algorithm.

Theorem 4. *Primal objective convergence*
(a) A lower bound on the primal cost is given by

$$f(\mathbf{Q}^{q(n)}) \geq f^\star - \frac{\Lambda^2}{2nc^{(n)}/Q} - e^{(n)};$$

(b) An upper bound on the primal cost is given by

$$f(\mathbf{Q}^{q(n)}) \leq f^\star + \frac{5\Lambda^2}{2nc^{(n)}/Q} + e^{(n)}, where$$

$$e^{(n)} = \frac{c^{(n)}Q(G+\tau)^2}{2} + Q\tau\Lambda$$
$$+ Q(\beta^{(1)}(4G + 2\tau) + \zeta).$$

Proofs of Theorems 1, 2, 3, and 4 can be found in Appendices A, C, D, and E, respectively.

Algorithm 2: Distributed augmented consensus-based dual decomposition for joint fronthaul-constrained TP design and selection.

---

0. *Initialization* Choose $\boldsymbol{\lambda}^{q(0)} \in \mathbb{R}^+$, $\alpha \in \mathbb{R}^+$, γ, sequence $c^{(n)} = \gamma/(n+1)$, $\varphi \in \mathbb{Z}_+$. Choose ϵ_{glo} and ϵ_{SCA} as stop criteria parameters. Set $n = 0$ as iteration index.

while $\|\mathbf{Q}^{(n)} - \mathbf{Q}^{(n-1)}\|_F \leq \epsilon_{glo}$ **do**

    1. *Local primal optimization*: Compute $\mathbf{Q}_i^{q(n+1)}$ using SCA:

    while $\|\mathbf{Q}^{(m)} - \mathbf{Q}^{(m-1)}\|_F \leq \epsilon_{SCA}$ **do**

        Compute $\mathbf{Q}_i^{q(m+1)}$ by solving (22);

    Compute $\boldsymbol{\lambda}^{q(n+1)}$ using (12);

    2. *Communication*: Compute $S_{ij}^{q(n+1)}$ and send it with $\lambda^{q(n+1)}$ to neighbors;

    3. *Local dual optimization*: Compute $\ell^{q(n+1)}$ by (13) and obtain $I_{th,ij}^{q(n+1)}$ using (14);

    4. *Consensus Step*: Compute $\mathbf{Q}_d^{q(n+1)}$ using (23); $n = n + 1$;

5. *Obtain precoder* $\mathbf{F}_i^q$: Apply randomization procedure on $\mathbf{Q}_i^{q(n)}$ as described in Section VIII.

VI. EXTENSION TO LIMITED FRONTHAUL SCENARIO

The distributed design algorithm described below extends the one proposed in [39] for the case of centralized joint precoder and selection design. Wireless and fiber optics are expected to serve as fronthaul between TPs and backhaul between the cloud processing units, with the fronthaul capacity usually lower than that of the backhaul, which will bound the transmission rate to the users. Thus, it is imperative to include this constraint in the joint TP selection and precoder design problem. The proposed distributed algorithm is summarized in Algorithm 2 and derivations are explained below. Note that the main differences between Algorithm 1 and 2 are that Step 1 is modified and an extra step is added to solve the limited-fronthaul problem. Fronthaul capacity constraint can be mathematically formulated as constraining the sum of rates offered to different users by the fronthaul capacity bound C^q. That is, $\sum_i \log\left(1 + \mathrm{SINR}_i^q\right) \leq C^q$, where SINR_i^q denotes the signal-to-interference-plus-noise ratio in the link between the qth TP and ith user. Writing the SINR explicitly, the constraint becomes

$$\sum_i \log\left(1 + \frac{tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH})}{\sigma^2 + \sum_{j \neq i} \sum_r tr(\mathbf{H}_i^r \mathbf{Q}_j^r \mathbf{H}_i^{rH})}\right) \leq C^q, \quad (16)$$

$\forall q \in \mathcal{Q}$. To shorten notations, define $R_i^q \triangleq tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH})$ and $I_{ij}^r \triangleq tr(\mathbf{H}_i^r \mathbf{Q}_j^r \mathbf{H}_i^{rH})$, so that (16) becomes

$$\sum_i \log\left(1 + R_i^q / (\sigma^2 + \sum_{j \neq i} \sum_r I_{ij}^r)\right) \leq C^q, \quad (17)$$

$\forall q \in \mathcal{Q}$. The SINR term in (17) can be rewritten using exponential function as

$$\sum_i \log\left\{1 + \exp\left[\log(R_i^q) - \log(\sigma^2 + \sum_{j \neq i} \sum_r I_{ij}^r)\right]\right\} \leq C^q, \quad (18)$$

where exp and log are inserted into the numerator. Notice that the expression is still nonconvex because the exponential function is nondecreasing and it requires its argument to be convex for the constraint to be convex, which is not the case. To circumvent this problem, affine approximation around an arbitrary point x_0 of $\log(x)$ is used so that $\log(x) \approx \log(x_0) + (x - x_0)/x_0$ can be used as a convex upper bound of the original constraint. x_0 can be chosen to

be $\mathbf{Q}_i^{q(m-1)}$, i.e., the variable $\mathbf{Q}_i^q$ at the previous $(m-1)$th SCA iteration, which can be regarded as a constant in the mth SCA iteration. Hence, the fronthaul constraint at the mth SCA iteration becomes

$$\sum_i \log\left\{1 + \exp\left[\log(\epsilon + R_i^{q(m-1)}) + \frac{(R_i^q - R_i^{q(m-1)})}{\epsilon + R_i^{q(m-1)}}\right.\right.$$
$$\left.\left. - \log(\sigma^2 + \sum_{j \neq i} \sum_r I_{ij}^r)\right]\right\} \leq C^q, \quad (19)$$

where $R_i^{q(m-1)} = tr\left(\mathbf{H}_i^q \mathbf{Q}_i^{q(m-1)} \mathbf{H}_i^{q\,T}\right)$ and $\epsilon = 10^{-8}$ is added to avoid numerical problem since the algorithm initializes with $\mathbf{Q}_i^{q(0)} = \mathbf{0}, \forall i \in \mathcal{I}, \forall q \in \mathcal{Q}$. Since the argument inside exp is convex, the function is log-convex, hence, the left-hand side of (19) is convex. Moreover, the proposed SCA results in an upper bound for the original constraint, hence, it is a valid substitute. The proposed scheme converges because at each iteration, the feasible region of the problem shrinks, leading to monotonic increase in the objective.

However, in the distributed optimization scenario, (19) is a coupling constraint since the interference term in (19) forces the constraint for the qth TP to be dependent on the rth precoder, for $\forall r \neq q$. This problem can be circumvented by replacing the interference term with the one from the previous (global) $(n-1)$th iteration. To ensure feasibility and convergence, the ADMM scheme is applied, and the constraint $\mathbf{Q}_i^q = \mathbf{Q}_i^{q(n-1)}$ is added, thus converting the problem into

$$\max_{\mathbf{Q}} \sum_q \sum_i tr\left(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}\right) - \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T}$$

$$s.t. \sum_q tr\left(\mathbf{H}_j^q \mathbf{Q}_i^q \mathbf{H}_j^{qH}\right) \leq I_{th}, \ i, j \in \mathcal{I} : j \neq i$$

$$\sum_i \log\left\{1 + \exp\left[\log(\epsilon + R_i^{q(m-1)}) + \frac{(R_i^q - R_i^{q(m-1)})}{\epsilon + R_i^{q(m-1)}}\right.\right.$$
$$\left.\left. - \log\left(\sigma^2 + \sum_{j \neq i} \sum_r I_{ij}^{r(n-1)}\right)\right]\right\} \leq C^q,$$

$$\mathbf{Q}_i^q = \mathbf{Q}_i^{q(n-1)}, \ \mathbf{Q}_i^q \in \mathcal{C}^q, \ i \in \mathcal{I}, \ q \in \mathcal{Q}, \quad (20)$$

where $I_{ij}^{r(n-1)} \triangleq tr(\mathbf{H}_i^q \mathbf{Q}_j^{r(n-1)} \mathbf{H}_i^{q\,H})$ denotes the interference from the rth precoder to the ith user while transmitting to

the jth user at the $(n-1)$th consensus iteration. Similar to [36], the above problem can be modified by adding a quadratic penalty term into the objective and express (20) as

$$\max_{\mathbf{Q}} \sum_q f^q(\mathbf{Q}^q) + \rho \|\mathbf{Q}_i^q - \mathbf{Q}_i^{q(n-1)}\|_F^2$$
$$s.t. \sum_q g_{ij}^q(\mathbf{Q}_i^q) \leq 0, \forall i,j \in \mathcal{I}, \ j \neq i \qquad (21)$$
$$\mathbf{Q}_i^q \in \widetilde{\mathcal{C}}^q, \mathbf{Q}_i^q = \mathbf{Q}_i^{q(n-1)}, \ i \in \mathcal{I}, \forall q \in \mathcal{Q}.$$

Notice that the fronthaul constraint in (19) has now be absorbed into the local constraint set $\mathcal{C}^q$, which is denoted as $\widetilde{\mathcal{C}}^q$. (21) can be solved using a scheme similar to Algorithm 1. That is, $\mathbf{Q}^{q(n+1)}$ can be computed as

$$\mathbf{Q}^{q(n+1)} = \arg\max_{\mathbf{Q}^q \in \widetilde{\mathcal{C}}^q} f^q(\mathbf{Q}^q) + \boldsymbol{\ell}^{q(n)T}\mathbf{g}^{q(n)}(\mathbf{Q}^q) +$$
$$\sum_i tr(\mathbf{Q}_{i\ d}^{qH}(\mathbf{Q}_i^q - \mathbf{Q}_i^{q(n-1)})) + \rho\|\mathbf{Q}_i^q - \mathbf{Q}_i^{q(n-1)}\|_F^2 \qquad (22)$$

using the SCA method, where $\mathbf{Q}_{i\ d}^q \succeq \mathbf{0}$ is the Lagrangian multiplier associated with equality constraint in (20). The Lagrange multiplier can be updated as [40]

$$\mathbf{Q}_{i\ d}^{q(n+1)} = \mathbf{Q}_{i\ d}^{q(n)} + \rho(\mathbf{Q}_i^{q(n)} - \mathbf{Q}_i^{q(n-1)}). \qquad (23)$$

The overall proposed scheme is summarized in Algorithm 2.

VII. ALGORITHM OVERHEAD AND COMPLEXITY

To justify the proposed distributed design schemes outlined in Algorithm 1 and 2, a description of complexity and system overhead is provided. For the proposed iterative schemes, it is assumed that the channels can remain static during a transmission time interval (TTI), which can be expected for static environments. The timing bottleneck for Algorithm 1 is processing delay in the local optimization step, which consists of computation time of SDP convex programming problem of size equals to the number of transmit antennas multiplied by the number of UEs. The feasibility of time complexity of the convex programming techniques for signal processing and mobile communications is further supported in [42], which states that it is possible to solve modest-sized convex optimization problems on microsecond time scales. Despite strong dependence on the solver implementation, it is important to understand the advantages of the proposed method over the centralized one proposed in [39]. For centralized approach, the computational complexity of the IPM for SDP grows at complexities $O(n^3) \sim O(n^6)$ [41] per iteration depending on the problem structure. The amount of transferred data is $O(n)$, where $n = n_T^2 QI$ with n_T being the number of transmit antennas, Q is the number of TPs, and I is the number of UEs. Using the distributed method, the computational complexity per iteration is $O(m^3) \sim O(m^6)$, with $m = n_T^2 I$, and the amount of data transferred is $O(I)$. As a result, the computational complexity and amount of transferred data of the algorithm do not scale with the number of TPs. Even though the distributed algorithm may take more iterations to converge than IPM, this difference in practice is less than an order of magnitude (30–40 iterations for centralized IPM, 80–300 iterations of the distributed algorithm). Therefore, there is a clear tradeoff

between the two, with distributed algorithm suited more for cell-free cooperative MIMO systems.

The aforementioned analysis also holds for Algorithm 2 with the added complexity that an SDP problem needs to be solved during each SCA iteration. Hence, the complexity of local optimization in Algorithm 2 is multiplied by the number of SCA approximations being done compared with Algorithm 1.

Notably, the proposed distributed algorithms require nodes exchange information in Step 2 of Algorithm 1 and 2, which creates some overhead. Particularly, each node sends a real-valued dual variable of size $I(I-1)$ to its' neighbors. Given that the X2 infrastructure existing between TPs being a high throughput optical cable, which is a must to support a feasible CoMP system, expected transmission delay is negligible compared to other steps of the algorithm.

VIII. RANDOMIZATION PROCEDURE

The randomization procedure described herein is invoked in Step 4 and 5 in Algorithm 1 and 2, respectively. For Algorithm 1, $\mathbf{Q}_i^q$ is obtained from (4), $\mathbf{F}_i^q$ with rank B is recovered distributively by using the randomization procedure similar to the one in [28]. Specifically, $Nrand$ number of $\mathbf{F}_i^q = 1/B\mathbf{U}_{\mathbf{Q}_i^q}\mathbf{\Lambda}_{\mathbf{Q}_i^q}^{1/2}\mathbf{E}$ are generated. $\mathbf{U}_{\mathbf{Q}_i^q}$ and $\mathbf{\Lambda}_{\mathbf{Q}_i^q}^{1/2}$ are the eigenvector and eigenvalue matrix of $\mathbf{Q}_i^q$, respectively. $[\mathbf{E}]_{ml} = e^{j\theta_{ml}}$ and θ_{ml} is independent and identically distributed uniformly on $[0, 2\pi]$, with $m = 1, \cdots, n_T$ and $l = 1, \cdots, B$. Only those $\mathbf{F}_i^q$'s that satisfy the leakage interference constraints, i.e., $\|\mathbf{H}^q\mathbf{F}^{q(n)}\| - I_{th,ij}^{q(n)} \leq 0$, are retained. The remaining $\mathbf{F}_i^q$'s are compared, and the one which maximizes the objective is the optimal precoder.

It is, however, more difficult to retrieve $\mathbf{F}_i^q$ distributively from $\mathbf{Q}_i^{q(n)}$ in Algorithm 2 after convergence because $\mathbf{F}_i^q$ has to satisfy the limited fronthaul constraint, which requires knowing the interference power terms from of all agents at each agent. This procedure can be done distributively similar to Algorithm 1. It, however, requires additional system overhead between TPs to exchange fronthaul constraint related data. Fortunately, in simulation, the limited fronthaul capacity constraint will often be satisfied after $\mathbf{F}_i^q$ is obtained because the SINR will remain relatively unchanged since both received signal and interference power will decrease by relatively the same amount. Although extra communication overhead will be incurred by the exchange of precoders to satisfy the limited fronthaul constraint, only a few trials of randomization are required.

IX. NUMERICAL RESULTS

A. Simulation Environment

An example of the simulated network is shown in Fig. 2. Each service sector contains a TP (green triangle) in the middle, surrounded by three or five UEs (red squares). A service sector herein looks similar to a cell, but it is an area where users can be serviced by multiple TPs in a cooperating set, and a single TP can serve multiple users in different sectors. UEs are placed at the edge of each of these service sectors to promote JT and to emulate cell-free users in 5G NR networks using JT-CoMP. The
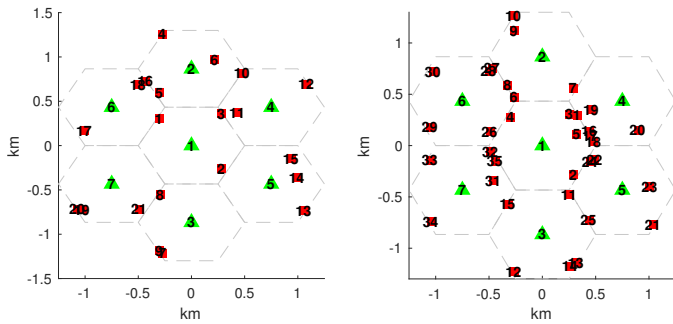
Fig. 2. Example of simulated networks: 3 (left) and 5 (right) UEs per sector.

Table III
SIMULATED NETWORK PARAMETERS.

# of TPs/UEs	$Q = 7, I = 21, 30$
# of Antennas	$n_T = 4, n_R = 2$
I_{th}	10^{-4} W
P^q	1 W
σ^2	-33 dB
L_0 Ref loss (dB)	60
d_0 Ref distance (m)	10
PL exponent	3.76
Shadowing	10 dB
Tx antenna gain	10 dB
Carrier frequency	2.5 GHz

Fig. 3. Two instances of Algorithm 1 convergence for different connection density with unlimited fronthaul.

channel is constructed as $\mathbf{H}_i^q = \mathbf{H}_w^i \mathbf{G}_q^i$, where $\mathbf{H}_w^i \in \mathbb{C}^{n_T \times n_T}$ denotes the channel matrix with i.i.d. zero-mean complex Gaussian entries, $\mathbf{G}_q^i$ is an $n_T \times n_T$ diagonal matrix whose nth diagonal element equals $\sqrt{10^{L_0/10} (d/d_0)^{-\alpha} 10^{\xi_n/10}}$, with L_0 denoting the reference path loss, d and d_0 denoting the distance between the transmitter and receiver and the reference distance, respectively, α is the path loss exponent, and the shadow fading $10^{\xi_n/10}$ is independent of the path loss $(d/d_0)^{-\alpha}$ and is a log-normal distributed random variable, with ξ_n being a normal distributed random variable with zero mean and variance equals to 10 dB as indicated in Table III. The transmit power and instantaneous interference leakage power threshold were chosen similar to [39] and [28]. The choice of $I_{th} = 10^{-4}$ W is because the average SINR in cellular systems varies from -5 to 20 dB. The average power received from a reference signal has a typical range from -40 dBm to -140 dBm, which equals 10^{-4} to 10^{-14} W. Therefore, choosing $I_{th} = 10^{-4}$ W and having received signal power around 10^{-4} W will make the SINR to be around 0 dB, which is typical SINR QoS value in the references. Other network and algorithm parameters are summarized in Table III and are selected according to the 3GPP 5G NR user equipement radio transmission and reception specification document [43]. CVX [47] with MOSEK [48] was used to obtain solution to (3) and (20) on an Intel Core i7 4790 4GHz machine with 32GB RAM. Centralized algorithm in [39] is used as a baseline for the proposed algorithm. In addition, joint precoder design and user selection works in [22], [24], and [46] are used for performance comparison.

With the exception of Fig. 6, all simulation results below are based on the network with 3 UEs per sector shown in Fig. 2.
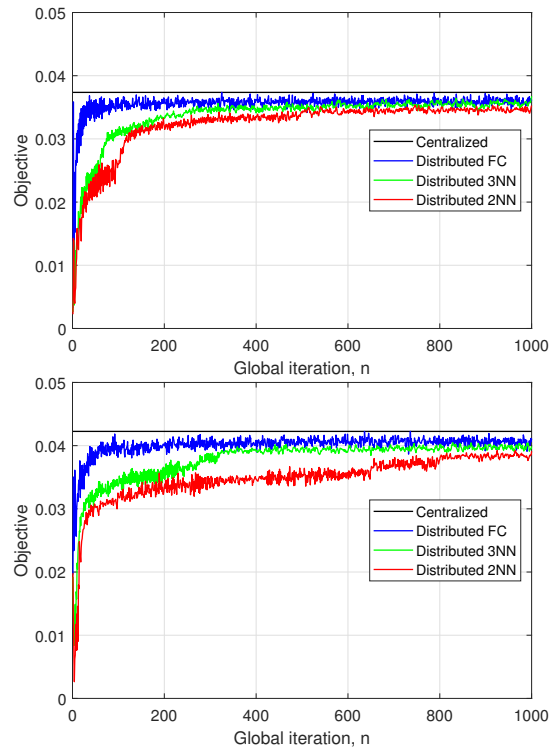
B. Convergence Behavior

In particular interest of this work is to show the impact that network connectivity and the number of message passing have on algorithm convergence for unlimited fronthaul capacity. Fig. 3 shows convergence behavior when each TP is connected to two and three nearest neighbors, labeled as 2NN and 3NN, respectively. A fully connected network is used as a performance benchmark, and is labeled as FC in Fig. 3. For each node q, weights $[\mathbf{W}]_{qr}, r \in \mathcal{N}^q$ are selected to be equal $1/|\mathcal{N}^q|$. Number of consensus iterations was set to $\varphi = 1$.

It can be seen from the results that the algorithm converges faster for a more densely connected network, as stated in Theorem 1. To highlight the effect of limited fronthaul capacity, $C^q = 1, 2, 3, 4$ bits/Hz/sec were simulated. The first value corresponds to worst-case microwave fronthaul capacity in [44] and the last corresponds to the situation when the fronthaul constraint becomes looser than the leakage interference power constraint so that the latter becomes active. Convergence results are shown in Fig. 4. Comparing to the unlimited fronthaul capacity scenario, it can be observed from Fig. 4 that the objective value of the distributed design, regardless of the density of the connection, will exceed that of the centralized in the first few iterations, which does not occur in the unlimited fronthaul case. This is due to the fact that before convergence is reached, interference power tends to oscillate a lot due to the use of the SCA in the interference term in (20), making fronthaul approximation imprecise and producing infeasible solutions. However, as the algorithm proceeds, a solution close to the centralized one can be achieved. Next, algorithm convergence was tested for a different number of consensus iterations. Fig. 5
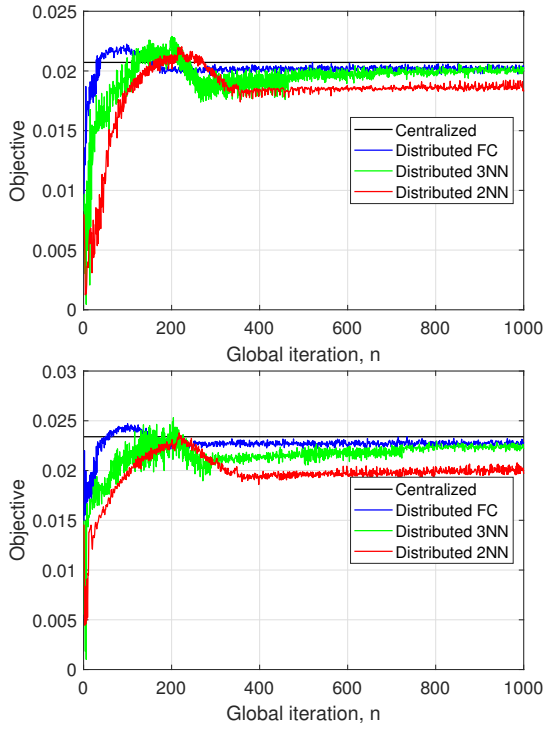
Fig. 4. Two instances of Algorithm 2 convergence for different connection density, $C^q = 2$ bits/sec/Hz.
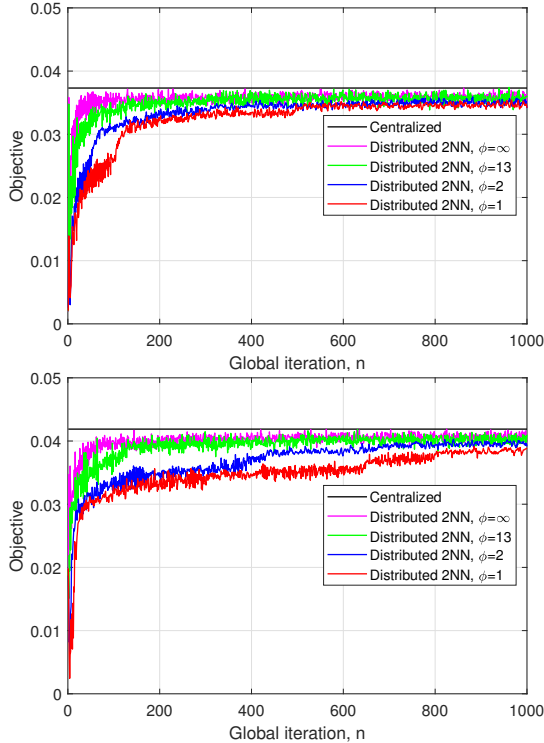

Fig. 5. Two instances of Algorithm 1 convergence for different number of consensus iterations (message passing).

shows convergence behavior when 1, 2, and 13 number of consensus iterations are used with different connection density. The latter case corresponds to the situation where information of each node is passed to all other nodes in the network. The centralized and distributed design with $\varphi = \infty$ are used as performance benchmarks. The latter is equivalent to the FC case (with equal edge weights) in the previous experiment, which holds from (7). From experimental data, it can be observed that convergence becomes faster as the number of consensus iterations increases.

C. Performance with Different α and C^q

Next, the performance of the proposed algorithms in terms of average objective value and sum rate (average over the number of channel realizations) are compared to the previously proposed centralized design in [39]. Moreover, in order to show performance degradation of the proposed algorithms due to unknown interference leakage threshold distribution per TP, i.e., $I_{th,ij}^q, q \in \mathcal{Q}, i, j \in \mathcal{I} : i \neq j$, the proposed algorithm with *a priori* knowledge of $I_{th,ij}^q$ is simulated and is labeled as "GNDTR" (ground truth) in the subsequent figures. The ground truth value of $I_{th,ij}^q$ is obtained from the centralized solution. Fig. 6 shows the performance difference for FC, 2NN, and 3NN topologies and FC with ground truth $I_{th,ij}^q$. Also, performance for different network densities is assessed, for 3 and 5 UEs per sector, respectively. Each point is obtained based on 500 channel realizations with random user placement in each iteration. It can be observed that the average performance degrades for less densely connected networks and there is also some degradation due to the use of the adaptive resource allocation scheme proposed in Section IV-B, which supports the theoretical results above. It can also be seen that the proposed adaptive resource adaptive scheme performs well compared to the ground truth case and causes relatively small degradation. Even though ground truth $I_{th,ij}^q$ is obtained using the solution from the centralized design, it slightly underperforms the centralized design due to the suboptimality of the primal solution as stated in Theorem 4. It is also important to mention that when the number of users per sector increases, overall performance degrades, as there are more constraints regarding interference that need to be satisfied. Next, Fig. 7 shows performance gain in terms of average objective and normalized sum rate for the proposed adaptive resource allocation in comparison to existing schemes in [22] and [24], that considered equal resource allocation. The gain obtained in the proposed method is due to the fact that $I_{th,i,j}^q$ is computed more precisely such that "victims" that can sustain higher leakage interference will lead to higher received signal power allocated to the desired users, resulting in higher rate.

Performance results for the fronthaul-constrained network are shown in Figs. 8 and 9. The performance loss is noticeably bigger for a weakly connected network, due to the combination of dual decomposition, resource allocation, and ADMM algorithms. The average sum rate performance vs. C^q with different values of α of the proposed algorithm is also compared with the joint beamforming-semidefinite relaxation (JB-SDR)
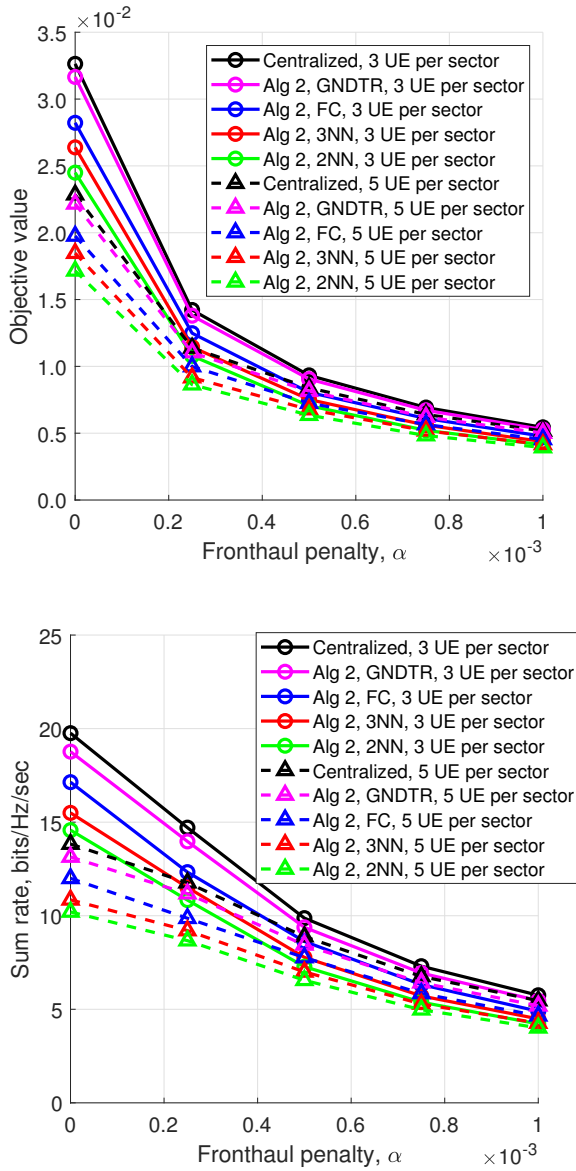
Fig. 6.   Average objective . Average normalized rate. Average objective value and average normalized rate vs. α of Algorithm 1 without fronthaul constraint for different connection density.

Fig. 7.   Average objective. Average normalized rate. Average objective value and average normalized rate vs. α of Algorithm 1 without fronthaul constraint comparison for optimal resource resource allocation, proposed resource allocation scheme and equal resource allocation in [22] and [24].

algorithm in [46] in Fig. 9, which shows the proposed algorithm can outperform JB-SDR when C^q is relatively small.

Finally, the run time performance vs. number of transmit antennas for the proposed distributed algorithms are compared to its centralized counterpart and [22] with and without limited fronthaul in Fig. 10. As expected, the run time grows as the system size increases. Also, the run time for the algorithms that need to solve problem with limited fronthaul is higher than those which consider unlimited fronthaul because SCA needs to be employed in both the centralized and distributed cases when fronthaul constraint exists.
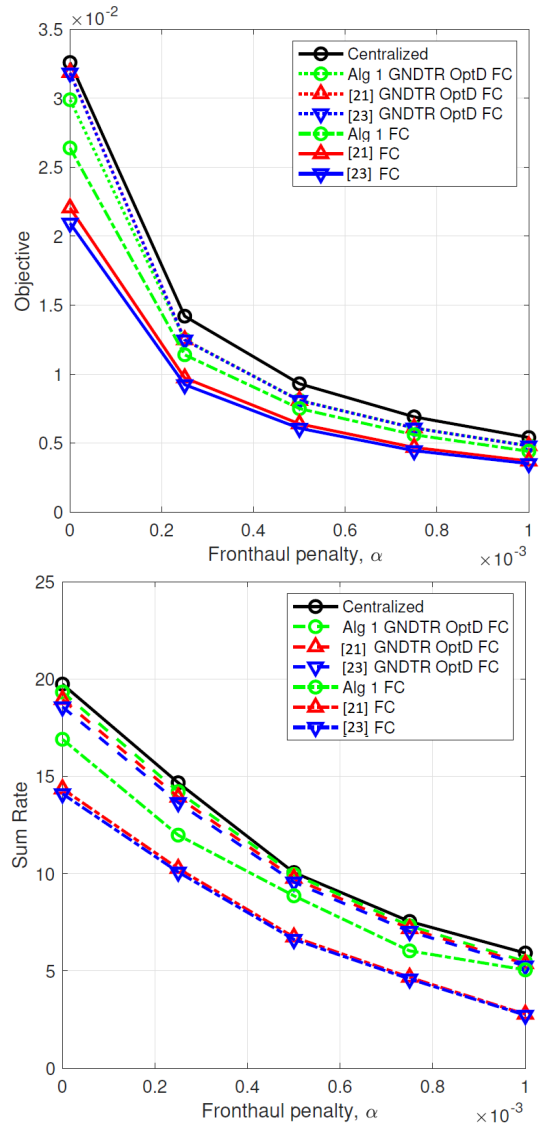
X. CONCLUSION

A joint TP selection and precoding scheme for cell-free cooperative MIMO using underlay transmission is proposed. The original nonconvex problem is converted into a convex one such that a unique global solution can be found efficiently. Novel consensus-based dual decomposition distributed algorithm for unconstrained fronthaul has been proposed to obtain a solution to the reformulated problem where only a small amount of information needs to be exchanged between different agents. Also, data privacy is ensured as it is difficult to reconstruct the primal solution from the exchanged information. This algorithm extended previous works by not requiring *a priori* knowledge about the interference constraint threshold, i.e., the availability of an oracle is not needed. Convergence has been proven and the convergence rate equation has been shown. Next, the proposed scheme is extended to a limited fronthaul capacity scenario. A
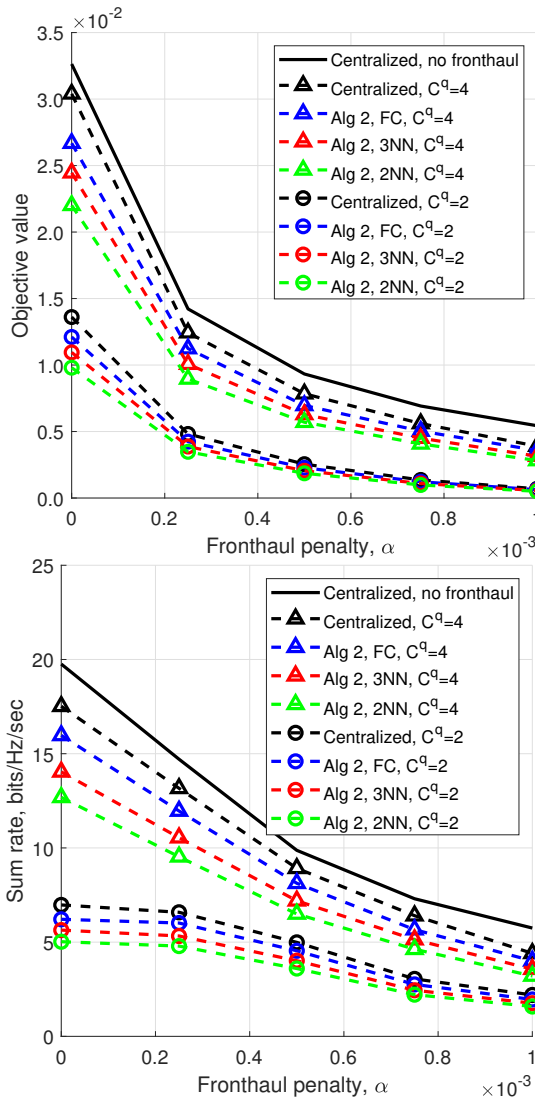
Fig. 8. Average objective. Average sum rate. Average objective value and average normalized rate vs. α of Algorithm 2 with fronthaul constraint for different values of C^q.
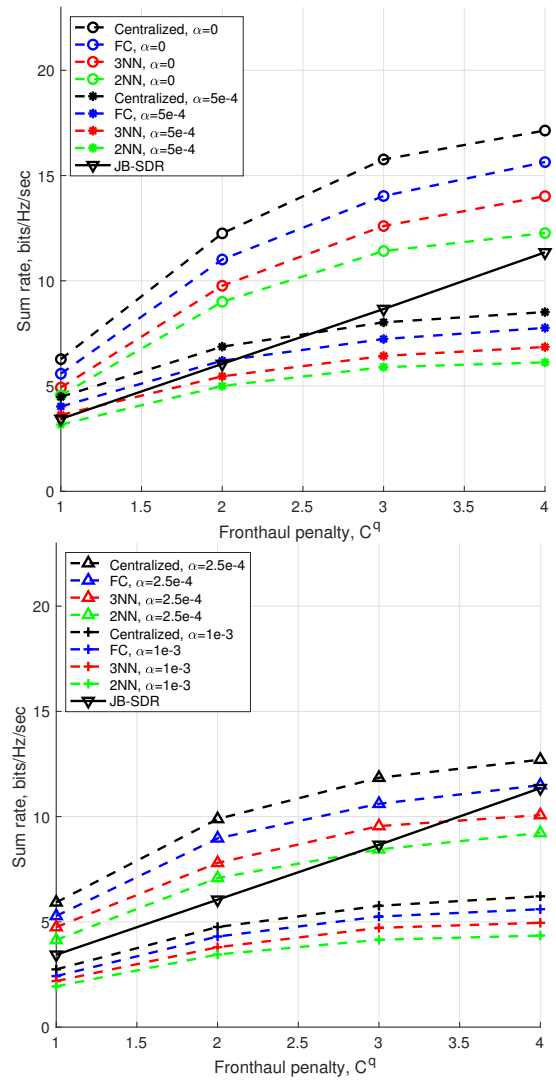

Fig. 9. Average sum rate for $\alpha = 0$ and $\alpha = 5e-4$. Average sum rate for $\alpha = 2.5e-4$ and $\alpha = 1e-3$. Average sum rate vs. C^q of Algorithm 2 with fronthaul constraint for different α.

combination of successive convex approximation and ADMM are used to resolve the nonconvexity and constraint coupling problems to make distributed optimization possible.

APPENDIX A
PROOF OF THEOREM 1

Proof. First, from the proposed resource allocation scheme, the sequence $L_{ij}^{q(n)}$ is nonincreasing. This is true because if the constraint is active, the threshold will not decrease. Denote the overall amount of slack for pair ij as $S_{ij}^{(n)} = \sum_q S_{ij}^{q(n)}$. Using (14), $S_{ij}^{(n)}$ can be written as

$$S_{ij}^{(n)} = \sum_q \left(I_{th,ij}^{q(n)} - L_{ij}^{q(n)} \right)$$
$$= \sum_q \left(L_{ij}^{q(n-1)} + \sum_{r \in \mathcal{N}^q} [\mathbf{W}]_{rq} S_{ij}^{r(n-1)} - L_{ij}^{q(n)} \right)$$
$$= \sum_q \left(L_{ij}^{q(n-1)} - L_{ij}^{q(n)} + \sum_{r \in \mathcal{L}^q} [\mathbf{W}]_{rq} S_{ij}^{r(n-1)} \right)$$

$$= S_{ij}^{(n-1)} + \sum_q \underbrace{\left(L_{ij}^{q(n-1)} - L_{ij}^{q(n)} \right)}_{\substack{\text{some slack that has been consumed} \\ \text{by nodes with active constraints},<0}} ,$$

where the last equality is obtained using (6). As explained at the beginning of the proof $L_{ij}^{q(n)} \leq L_{ij}^{q(n-1)}$, hence $S_{ij}^{(n)}$ is also non-increasing and first statement of the theorem holds. Further, if constraint (3b) is active at optimal point, $\lim_{n \to \infty} S_{ij}^{(n)} = 0$ and it can be concluded that $\inf S_{ij}^{(n)} = 0$ due to the fact that if a sequence of real numbers is decreasing and bounded below, then its infimum is the limit. $\square$

APPENDIX B
BASIC RELATIONS FOR THEOREMS PROOFS

The first two lemmas are general results from consensus theory in [26] and are repeated here for completeness. Other lemmas and theorems are new and are written and proven to establish the convergence properties of the proposed distributed algorithm.
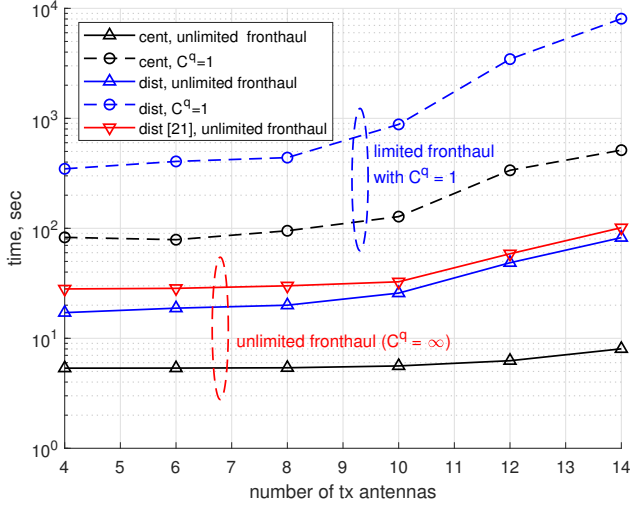
Fig. 10. Run time (in sec) vs. number of transmit antennas for centralized precoder design algorithm, proposed distributed precoder design algorithm and [22], with ($C^q = 1$) and without limited fronthaul. Results are averaged over 20 Monte Carlo experiments.

Lemma 1 ([26] Lemma 1).
Define $\mathbf{x}^q \in \mathbb{R}^{I(I-1)}$ as a subvector of $\mathbf{x}^{(n)} \in \mathbb{R}^{I(I-1)Q}$, i.e., $\mathbf{x}^{(n)} \triangleq [\mathbf{x}^{1T} \cdots \mathbf{x}^{QT}]^T$. Denote $\overline{\mathbf{x}}$ as the mean of $\mathbf{x}^q$, i.e., $\overline{\mathbf{x}} = \frac{1}{Q}\sum_{q=1}^Q \mathbf{x}^q$. Then for $q, r = 1, \cdots, Q$ and $q \neq r$

(a) if $\|\mathbf{x}^q - \mathbf{x}^r\|_2 \leq \beta$, then $\|\mathbf{x}^q - \overline{\mathbf{x}}\|_2 \leq \frac{Q-1}{Q}\beta$,
(b) if $\|\mathbf{x}^q - \overline{\mathbf{x}}\|_2 \leq \beta$, then $\|\mathbf{x}^q - \mathbf{x}^r\|_2 \leq 2\beta$.

Proof.

(a) $\|\mathbf{x}^q - \overline{\mathbf{x}}\|_2 = \frac{1}{Q}\left\| Q\mathbf{x}^q - \sum_q \mathbf{x}^q \right\|_2$

$= \frac{1}{Q}\left\| \sum_q \mathbf{x}^q - Q\mathbf{x}^q \right\|_2$

$= \frac{1}{Q}\left\| \sum_{q \neq r} (\mathbf{x}^q - \mathbf{x}^r) \right\|_2$

$\leq \frac{1}{Q}\sum_{q \neq r} \|\mathbf{x}^q - \mathbf{x}^r\|_2 \leq \frac{Q-1}{Q}\beta.$

(b) $\|\mathbf{x}^q - \mathbf{x}^r\|_2 = \|\mathbf{x}^q - \overline{\mathbf{x}} + \overline{\mathbf{x}} - \mathbf{x}^r\|_2$

$\leq \|\mathbf{x}^q - \overline{\mathbf{x}}\|_2 + \|\overline{\mathbf{x}} - \mathbf{x}^r\|_2 \leq 2\beta. \qquad \square$

Lemma 2 ([26] Lemma 2). *Suppose $\mathbf{x}^{(n+1)} = \left(\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)}\right)\mathbf{x}^{(n)}$, with $\mathbf{W}$ fulfilling the conditions in (6) and (7). Let $\|\mathbf{x}^{q(n)} - \mathbf{x}^{r(n)}\|_2 \leq \sigma \leq \infty$, $q, r = 1, \cdots, Q$. Then*

$$\|\mathbf{x}^{q(n+1)} - \mathbf{x}^{r(n+1)}\|_2 \leq 2\nu^\varphi Q\sigma. \qquad (24)$$

Proof. Let $\mathbf{x}^{(n)} = \widetilde{\mathbf{x}}^{(n)} + \mathbf{a}^{(n)}$ with $\widetilde{\mathbf{x}}^{(n+1)} = \left(\frac{\mathbf{1}_Q \mathbf{1}_Q^T}{Q} \otimes \mathbf{I}_{I(I-1)}\right)\mathbf{x}^{(n)}$ and $\mathbf{1}_{I(I-1)Q}^T \mathbf{a}^{(n)} = 0$. $\mathbf{a}^{(n)} \in \mathbb{R}^{I(I-1)Q}$ also contains subvectors $\mathbf{a}^{q(n)}$, similarly to $\mathbf{x}^{(n)}$ and $\mathbf{x}^{q(n)}$. The results of Lemma 1 shows that $\|\mathbf{a}^{q(n)}\|_2 \leq Q - 1/Q\sigma < \sigma$, for $Q > 1$. Hence, $\|\mathbf{a}^{(n)}\|_2 \leq Q\sigma$. Furthermore, $\mathbf{x}^{(n+1)} = \left(\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)}\right)\left(\overline{\mathbf{x}}^{(n)} + \mathbf{a}^{(n)}\right) = \overline{\mathbf{x}}^{(n)} + \left(\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)}\right)\mathbf{a}^{q(n)}$. Using this relationship,

$$\left\|\mathbf{x}^{q(n+1)} - \overline{\mathbf{x}}^{(n+1)}\right\|_2 \leq \left\|\mathbf{x}^{(n+1)} - \widetilde{\mathbf{x}}^{(n+1)}\right\|_2$$

$= \left\|(\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)})(\mathbf{a}^{(n)} - \mathbf{0}_Q)\right\|_2$

$= \left\|\left[\left(\mathbf{W}^\varphi - \mathbf{1}_Q \mathbf{1}_Q^T/Q\right) \otimes \mathbf{I}_{I(I-1)}\right]\mathbf{a}^{(n)}\right\|_2$

$\leq \left\|\left[\left(\mathbf{W}^\varphi - (\mathbf{1}_Q \mathbf{1}_Q^T/Q)^\varphi\right) \otimes \mathbf{I}_{I(I-1)}\right]\mathbf{a}^{(n)}\right\|_2$

$\leq \left\|\left(\mathbf{W} - \mathbf{1}_Q \mathbf{1}_Q^T/Q\right)\right\|_2^\varphi \|\mathbf{I}_{I(I-1)}\|_2 \|\mathbf{a}^{(n)}\|_2 \leq \nu^\varphi Q\sigma,$

where the last inequality is true because $\|\mathbf{W} - \mathbf{1}_Q \mathbf{1}_Q^T/Q\|_2 = \rho\left(\mathbf{W} - \mathbf{1}_Q \mathbf{1}_Q^T/Q\right) \leq \nu$ according to (7). Then, according to part (b) of Lemma 1,

$$\left\|\mathbf{x}^{q(n+1)} - \mathbf{x}^{r(n+1)}\right\|_2 \leq 2\nu^\varphi Q\sigma.$$

$\square$

Before presenting the next lemma, define $\boldsymbol{\ell} \triangleq \left[\boldsymbol{\ell}^{1T} \cdots \boldsymbol{\ell}^{QT}\right]^T$ and $\mathbf{g} \triangleq \left[\mathbf{g}^{1T} \cdots \mathbf{g}^{QT}\right]^T$, then from (13) and (12), the update of $\boldsymbol{\ell}$ becomes

$$\boldsymbol{\ell}^{(n+1)} = (\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)})\mathcal{P}_+[\boldsymbol{\ell}^{(n)} - c^{(n)}\mathbf{g}^{(n+1)}]. \qquad (25)$$

Also define $\overline{\boldsymbol{\ell}} \triangleq \frac{1}{Q}\sum_q \boldsymbol{\ell}^q$, i.e., average of all $\boldsymbol{\ell}^q$s $\forall q \in \mathcal{Q}$. Lemma below establishes important bound for iterates of $\boldsymbol{\ell}^q$. Also, in order to shorten notations, $\overset{(xx)}{\leq}$ is used to denote "inequality follows from equation (xx)".

Lemma 3.
Suppose $\boldsymbol{\ell}$, is generated according to (25) and ν is defined in (7). Assuming $\|\boldsymbol{\ell}^{q(n)} - \overline{\boldsymbol{\ell}}^{(n)}\|_2 \leq \beta$, there exists a $\overline{\varphi} \geq 1$, such that if $\varphi \geq \overline{\varphi} + \delta$ with $\overline{\varphi} \geq 0$ and $\delta \geq 0$, then $\forall q \in \mathcal{Q}$

$$\|\boldsymbol{\ell}^{q(n+1)} - \overline{\boldsymbol{\ell}}^{(n+1)}\|_2 \leq \nu^\delta \beta,$$

Proof. Let $\mathbf{u}^q \in \mathbb{R}^{I(I-1)}$ and $\mathbf{v}^q \in \mathbb{R}^{I(I-1)}$, $\forall q \in \mathcal{Q}$, at the nth iteration, be defined as

$$\mathbf{u}^{q(n)} \triangleq \boldsymbol{\ell}^{q(n)} - c^{(n)}\mathbf{g}^{q(n)}, \qquad (26)$$

with $\mathbf{g}^{q(n)}$ being the subgradient of the objective function in (9) at the nth global iteration after $\mathbf{Q}_i^{q(n+1)}$ is obtained. Then the distance between iterates of $\mathbf{u}^{q(n)}$ can be bounded as

$\|\mathbf{u}^{q(n)} - \mathbf{u}^{r(n)}\|_2 = \|\boldsymbol{\ell}^{q(n)} - c^{(n)}\mathbf{g}^{q(n)} - \boldsymbol{\ell}^{r(n)} + c^{(n)}\mathbf{g}^{r(n)}\|_2$

$\leq \|\boldsymbol{\ell}^{q(n)} - \boldsymbol{\ell}^{r(n)}\|_2 + \|c^{(n)}(\mathbf{g}^{r(n)} - \mathbf{g}^{q(n)})\|_2$

$\leq \|\boldsymbol{\ell}^{q(n)} - \boldsymbol{\ell}^{r(n)}\|_2 + 2c^{(n)}G$

$\overset{\text{Lemma1}(b)}{\leq} 2(\beta + c^{(n)}G). \qquad (27)$

Next, the non-expansive property of $\mathcal{P}_+$ can be used to obtain

$\|\mathcal{P}_+[\mathbf{u}^{q(n)}] - \mathcal{P}_+[\mathbf{u}^{r(n)}]\|_2 \leq \|\mathbf{u}^{q(n)} - \mathbf{u}^{r(n)}\|_2$

$\leq 2(\beta + c^{(n)}G). \qquad (28)$

Next, (25) implies that $\ell^{q(n+1)} = \sum_{p \in \mathcal{N}^q} [\mathbf{W}^\varphi]_{pq} \mathcal{P}_+[\mathbf{u}^{p(n+1)}]$, then after consensus step

$$\|\ell^{q(n+1)} - \ell^{r(n+1)}\|_2 =$$

$$\left\| \sum_{p \in \mathcal{N}^q} [\mathbf{W}^\varphi]_{pq} \mathcal{P}_+[\mathbf{u}^{p(n+1)}] - \sum_{p \in \mathcal{N}^r} [\mathbf{W}^\varphi]_{pr} \mathcal{P}_+[\mathbf{u}^{p(n+1)}] \right\|_2 =$$

$$\left\| \left[\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)} \mathcal{P}_+[\mathbf{u}^{(n+1)}] \right]_q - \left[\mathbf{W}^\varphi \otimes \mathbf{I}_{I(I-1)} \mathcal{P}_+[\mathbf{u}^{(n+1)}] \right]_r \right\|_2$$

$$\overset{(28),(24)}{\leq} 4\nu^\varphi (Q-1)(\beta + c^{(n)}G) \leq 4\nu^\varphi Q(\beta + c^{(n)}G), \quad (29)$$

where $\mathbf{u} \triangleq \left[\mathbf{u}^{1T} \cdots \mathbf{u}^{QT} \right]^T$. Furthermore, by Lemma 1 (a)

$$\|\ell^{q(n+1)} - \bar{\ell}^{(n+1)}\|_2 \leq 4\nu^\varphi Q(\beta + c^{(n)}G).$$

Next, substituting

$$\varphi \geq \frac{\log(\beta) - \log(4Q(\beta + c^{(n)}G))}{\log(\nu)} + \delta \triangleq \overline{\varphi} + \delta, \delta \geq 0 \quad (30)$$

into previous equation produces $\|\ell^{q(n+1)} - \bar{\ell}^{(n+1)}\|_2 \leq \nu^\delta \beta$ and the lemma is proven. $\quad\square$

APPENDIX C
PROOF OF THEOREM 2

The quantity $\|\ell^{q(1)} - \bar{\ell}^{(1)}\|_2$ is upper bounded by $\beta^{(1)}$ from Assumption 3. Choose $\varphi \geq \overline{\varphi} + \delta, \delta \geq 0$, with $\overline{\varphi}$ stated in Theorem 2. Then, by Lemma 3 it follows that,

$$\|\ell^{q(2)} - \bar{\ell}^{(2)}\|_2 \leq \nu^\delta \beta^{(1)};$$

$$\|\ell^{q(3)} - \bar{\ell}^{(3)}\|_2 \leq 4\nu^\varphi Q(\nu^\delta \beta^{(1)} + c^{(n)}G) \overset{(30)}{=}$$
$$\nu^\delta \beta^{(1)} \frac{\nu^\delta \beta^{(1)} + c^{(n)}G}{\beta^{(1)} + c^{(n)}G};$$

$$\|\ell^{q(4)} - \bar{\ell}^{(4)}\|_2 \leq$$
$$\frac{\nu^\delta \beta^{(1)}}{\beta^{(1)} + c^{(n)}G} \left(\nu^\delta \beta^{(1)} \frac{\nu^\delta \beta^{(1)} + c^{(n)}G}{\beta^{(1)} + c^{(n)}G} + c^{(n)}G \right);$$

$$\|\ell^{q(n)} - \bar{\ell}^{(n)}\|_2 \leq \nu^\delta \beta^{(1)} \left(\frac{\nu^\delta \beta^{(1)}}{\beta^{(1)} + c^{(n)}G} \right)^{n-1}$$
$$+ c^{(n)}G \left(-1 + \sum_{t=0}^{n-1} \left(\frac{\nu^\delta \beta^{(1)}}{\beta^{(1)} + c^{(n)}G} \right)^t \right).$$

Defining $p \triangleq \frac{\nu^\delta \beta^{(1)}}{\beta^{(1)} + c^{(n)}G}$, since $p < 1$, then sum of geometric series formula can be used to obtain

$$\|\ell^{q(n)} - \bar{\ell}^{(n)}\|_2 \leq p^{n-1} \nu^\delta \beta^{(1)} + p c^{(n)} G \frac{1 - p^{n-1}}{1 - p} \triangleq \beta^{(n)} \quad (31)$$

and the theorem is proved. $\quad\square$

APPENDIX D
PROOF OF THEOREM 3

Definition 1. *A vector* $\mathbf{g}$ *is subgradient of a convex function* d *at* $\mathbf{x}$ *if*

$$d(\mathbf{y}) \geq d(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \forall \mathbf{y} \in \mathbf{dom}\, f. \quad (32)$$

Definition 2. *The set of all subgradients of a convex function* d *is called subdifferential of* d *at* $\mathbf{x}$*, and is denoted by* $\partial d(\mathbf{x})$*:*

$$\partial d(\mathbf{x}) = \left\{ \mathbf{g} | d(\mathbf{y}) \geq d(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \forall \mathbf{y} \in \mathbf{dom}\, f \right\}. \quad (33)$$

Definition 3. *The* ϵ*-subdifferential set for* $\epsilon \geq 0$ *of a convex function* d *at* $\mathbf{x}$ *is the collection of* ϵ*-subgradients:*

$$\partial_\epsilon d(\mathbf{x}) = \left\{ \mathbf{g} | d(\mathbf{y}) \geq d(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}) - \epsilon, \forall \mathbf{y} \in \mathbf{dom}\, f \right\}. \quad (34)$$

Next lemma shows important property of ϵ-subdifferential.

Lemma 4. *Let* $d(\mathbf{x}) : X \to \mathbb{R}$ *be a convex function. Let the set* $X \subset \mathbb{R}^n$ *be convex and compact and in particular* $\max_{\mathbf{x} \in X} \|\mathbf{x}\|_2 \leq \eta$*. There exist two finite scalars* $\epsilon > 0$ *and* $\tau > 0$ *such that,* $\forall \mathbf{x} \in X$*,* $\forall \mathbf{g}(\mathbf{x}) \in \partial d(\mathbf{x})$*, and* $\forall \boldsymbol{\nu}$*, with* $\|\boldsymbol{\nu}\|_2 \leq \tau$*,*

$$\mathbf{g}(\mathbf{x}) + \boldsymbol{\nu} \in \partial_\epsilon d(\mathbf{x})$$

holds.

Proof. The claim is proven by directly using the definition of ϵ-subgradient. Since q is convex function $\forall \mathbf{x}, \mathbf{y} \in X$, by (32),

$$d(\mathbf{y}) - d(\mathbf{x}) \geq \langle \mathbf{g}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$
$$= \langle \mathbf{g}(\mathbf{x}) + \boldsymbol{\nu}, \mathbf{y} - \mathbf{x} \rangle - \langle \boldsymbol{\nu}, \mathbf{y} - \mathbf{x} \rangle$$
$$\geq \langle \mathbf{g}(\mathbf{x}) + \boldsymbol{\nu}, \mathbf{y} - \mathbf{x} \rangle - \|\boldsymbol{\nu}\|_2 \|\mathbf{y} - \mathbf{x}\|_2$$
$$\geq \langle \mathbf{g}(\mathbf{x}) + \boldsymbol{\nu}, \mathbf{y} - \mathbf{x} \rangle - 2\tau\eta$$

holds, which has the form of (34) for $\tau \leq \epsilon/(2\eta)$. $\quad\square$

Next, define $\mathbf{y}^{(n)} \in \mathbb{R}^{I(I-1)}$ and $\mathbf{d}^{(n)} \in \mathbb{R}^{I(I-1)}$ as

$$\mathbf{y}^{(n)} \triangleq \mathcal{P}_+[\bar{\mathbf{u}}^{(n-1)}] \text{ and } \mathbf{d}^{(n)} \triangleq \bar{\ell}^{(n)} - \mathbf{y}^{(n)}, \quad (35)$$

where $\bar{\mathbf{u}}^{(n-1)} \triangleq 1/Q \sum_q \mathbf{u}^{q(n-1)}$, with $\mathbf{u}^{q(n)}$ defined in (26). To prove convergence, $\|\bar{\ell}^{(n)} - \mathbf{y}^{(n)}\|_2$ will be shown to be bounded, where $\mathbf{y}$ is updated via an ϵ-subgradient method. Therefore, Proposition 4.1 in [45] can be used to establish a basis for the convergence rate. Then the combination of $\|\bar{\ell}^{(n)} - \mathbf{y}^{(n)}\|_2$ and convergence rate of the ϵ-subgradient method can then directly prove Theorem 3. To see this, the following lemmas need to be established.

Lemma 5. *Let* $\mathbf{y}^{(n)}$ *be defined as in (35). Under the same assumptions as in Theorem 2,* $\forall n \geq 1$
(a) The quantity $\|\mathbf{d}^{(n)}/c^{(n)}\|$ *is upper bounded by* $\beta^{(n)}/c^{(n)}$*.*
(b) The following inequalities are true for all $q \in \mathcal{Q}$

$$d(\ell^{q(n)}) \geq d(\mathbf{y}^{(n)}) - 2QG\beta^{(n)}, \quad (36)$$

$$d^q(\mathbf{y}) \geq d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \mathbf{y}^{(n)} \rangle - \frac{\epsilon^{(n)}}{Q}, \; \mathbf{y} \in \mathbb{R}^+, \quad (37)$$

where $d(\mathbf{y}^{(n)}) \triangleq \sum_q d^q(\mathbf{y}^{(n)})$ *and* $\epsilon^{(n)} = Q(\beta^{(n)}(4G + 2\tau) + \zeta)$*.*
(c) The quantity $\mathbf{h}^{(n)} \triangleq \sum_q \left(\mathbf{g}^{q(n)} + \frac{\mathbf{d}^{(n)}}{c^{(n)}} \right)$ *is an* ϵ*-subgradient of* $d(\mathbf{y})$ *with respect to* $\mathbf{y}$*.*
(d) The variable $\mathbf{y}^{(n)}$ *is updated according to the* ϵ*-subgradient method*

$$\mathbf{y}^{(n+1)} = \mathcal{P}_+ \left[\mathbf{y}^{(n)} - \frac{c^{(n)}}{Q} \mathbf{h}^{(n)} \right]. \quad (38)$$

Proof.

(a) The bound can be proven by proving $\|\mathbf{d}^{(n)}\|_2 < \beta^{(n)}$. Using non-expansive property of $\mathcal{P}_+$,

$$
\begin{aligned}
\|\mathbf{d}^{(n)}\|_2 &= \frac{1}{Q} \left\| \sum_q \mathcal{P}_+[\mathbf{u}^{q(n)}] - \mathcal{P}_+\Big[\sum_q \bar{\mathbf{u}}^{(n)}\Big] \right\|_2 \\
&\leq \frac{1}{Q} \sum_q \|\mathbf{u}^{q(n)} - \bar{\mathbf{u}}^{(n)}\|_2 \leq \beta^{(n)}. \quad \square
\end{aligned}
$$

(b) By the convexity of $d^q(\boldsymbol{\ell}^{q(n)})$ for $q, r \in \mathcal{Q}$ it can be written

$$
\begin{aligned}
d^q(\boldsymbol{\ell}^{q(n)}) &\overset{(33)}{\geq} d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}, \boldsymbol{\ell}^{q(n)} - \mathbf{y}^{(n)} \rangle, \text{ where } \mathbf{g} \in \partial_{\epsilon^{(n)}} d(\mathbf{y}^{(n)}) \\
&= d^q(\mathbf{y}^{(n)}) - \langle \mathbf{g}, \mathbf{y}^{(n)} - \boldsymbol{\ell}^{q(n)} \rangle \\
&= d^q(\mathbf{y}^{(n)}) - \langle \mathbf{g}, \mathbf{y}^{(n)} - \bar{\boldsymbol{\ell}} + \bar{\boldsymbol{\ell}} - \boldsymbol{\ell}^{q(n)} \rangle \\
&\geq d^q(\mathbf{y}^{(n)}) - \|\mathbf{g}\|_2 \|\mathbf{y}^{(n)} - \boldsymbol{\ell}^{q(n)}\|_2 \\
&\geq d^q(\mathbf{y}^{(n)}) - G\left(\|\bar{\boldsymbol{\ell}}^{(n)} - \boldsymbol{\ell}^{q(n)}\|_2 + \|\mathbf{d}^{(n)}\|_2 \right) \\
&\geq d^q(\mathbf{y}^{(n)}) - G(\beta^{(n)} + \beta^{(n)}) \\
&= d^q(\mathbf{y}^{(n)}) - 2G\beta^{(n)}, \quad\quad\quad\quad\quad (39)
\end{aligned}
$$

where the last inequality is obtained from (31). Then, summing the last equation over $q \in Q$ equals (36). Moreover, for any $\mathbf{y} \in \mathbb{R}^+$, from Lemma 4 and (39), $d^q(\mathbf{y})$ is upper bounded as

$$
\begin{aligned}
d^q(\mathbf{y}) &\overset{(33)}{\geq} d^q(\boldsymbol{\ell}^{q(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \boldsymbol{\ell}^{q(n)} \rangle - \zeta \\
&\overset{(39)}{\geq} d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \boldsymbol{\ell}^{q(n)} \rangle - 2G\beta^{(n)} - \zeta \\
&= d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \mathbf{y}^{(n)} + \mathbf{y}^{(n)} - \boldsymbol{\ell}^{q(n)} \rangle \\
&\quad - 2G\beta^{(n)} - \zeta \\
&\geq d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \mathbf{y}^{(n)} \rangle \\
&\quad - \|\mathbf{g}^{q(n)} + \boldsymbol{\nu}\|_2 \|\boldsymbol{\ell}^{q(n)} - \mathbf{y}^{(n)}\|_2 - 2G\beta^{(n)} - \zeta. \quad (40)
\end{aligned}
$$

Since $\|\boldsymbol{\nu}\|_2 \leq \tau$, $\|\mathbf{g}^{q(n)})\|_2 \leq G$, $\|\boldsymbol{\ell}^{q(n)} - \bar{\boldsymbol{\ell}}^{(n)}\|_2 \leq \beta^{(n)}$ and $\|\mathbf{d}^{(n)}\|_2 \leq \beta^{(n)}$, and using (35), it is possible to bound following quantities as

$$
\begin{aligned}
\|\mathbf{g}^{q(n)} + \boldsymbol{\nu}\|_2 &\leq G + \tau, \\
\|\boldsymbol{\ell}^{q(n)} - \mathbf{y}^{(n)}\|_2 &= \|\boldsymbol{\ell}^{q(n)} - \bar{\boldsymbol{\ell}}^q + \mathbf{d}^{(n)}\|_2 \leq 2\beta^{(n)} \text{ to obtain}
\end{aligned}
$$

$$
d^q(\mathbf{y}) \geq d^q(\mathbf{y}^{(n)}) + \langle \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \mathbf{y}^{(n)} \rangle - \underbrace{(\beta^{(n)}(4G + 2\tau) + \zeta)}_{\triangleq \epsilon^{(n)}/Q},
$$

which is (37). $\square$

(c) From (33), the inequality in (37) implies $(\mathbf{g}^{q(n)} + \boldsymbol{\nu}) \in \partial_{\epsilon^{(n)}/Q} q^q(\mathbf{y})$ with $\epsilon^{(n)}/Q = (\beta^{(n)}(4G + 2\tau) + \zeta)$. Summing over q yields

$$
d(\mathbf{y}) \geq d(\mathbf{y}^{(n)}) + \Big\langle \sum_q \mathbf{g}^{q(n)} + \boldsymbol{\nu}, \mathbf{y} - \mathbf{y}^{(n)} \Big\rangle - Q(\beta^{(n)}(4G + 2\tau) + \zeta),
$$

where the inequality holds for any $\boldsymbol{\nu}$ satisfying $\|\boldsymbol{\nu}\| \leq \tau$. Since τ can be selected such that $\beta^{(n)}/c^{(n)} \leq \tau$, then it is possible to choose $\boldsymbol{\nu} = \mathbf{d}^{(n)}/c^{(n)}$, which proves part (c). $\square$

(d) It is sufficient to write explicitly the update rule for $\mathbf{y}^{(n)}$. Using the definition of $\mathbf{y}^{(n+1)}$ in (35) and (35), $\mathbf{y}^{(n+1)}$ can be written as

$$
\begin{aligned}
\mathbf{y}^{(n+1)} &= \mathcal{P}_+\left[\frac{1}{Q} \sum_q \left(\boldsymbol{\ell}^{q(n)} - c^{(n)} \mathbf{g}^{q(n)} \right) \right] \\
&= \mathcal{P}_+\left[\bar{\boldsymbol{\ell}}^{(n)} - \frac{c^{(n)}}{Q} \sum_q \mathbf{g}^{q(n)} \right] \\
&= \mathcal{P}_+\left[\mathbf{y}^{(n)} + \mathbf{d}^{(n)} - \frac{c^{(n)}}{Q} \sum_q \mathbf{g}^{q(n)} \right] \\
&= \mathcal{P}_+\left[\mathbf{y}^{(n)} - \frac{c^{(n)}}{Q} \sum_q \left(\mathbf{g}^{q(n)} + \frac{\mathbf{d}^{(n)}}{c^{(n)}} \right) \right],
\end{aligned}
$$

which proves part (d). $\square$

Proof. (*Theorem 3*) Given Lemma 5, the sequence $\{\mathbf{y}^{(n)}\}$ is generated with an $\epsilon^{(n)}$-subgradient algorithm to maximize $d(\mathbf{y})$. For $n \geq 1$

$$
\mathbf{y}^{(n+1)} = \mathcal{P}_+\left[\mathbf{y}^{(n)} - (c^{(n)}/Q)\mathbf{h}^{q(n)} \right].
$$

Therefore, by using Proposition 4.2 in [45],

$$
\begin{aligned}
\lim_{n \to \infty} \sup d(\mathbf{y}^{(n)}) &= d^* + \lim_{n \to \infty} \sup Q(\beta^{(n)}(4G + 2\tau) + \zeta) \\
&= d^* + Q\zeta,
\end{aligned}
$$

holds, which concludes proof of the theorem. $\square$

APPENDIX E
PROOF OF THEOREM 4

Lemma 6. *Let $\mathbf{y}^{(n)}$ be defined as (35). Under the same assumptions as Theorem 3*

(a) For any $\mathbf{y} \in \mathbb{R}^+$,

$$
\sum_{t=1}^n \langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^{(t)} \rangle \geq -\frac{\|\mathbf{y}^{(1)} - \mathbf{y}\|_2^2}{2c^{(n)}/Q} - n\frac{c^{(n)}Q(G + \tau)^2}{2}; \quad (41)
$$

(b) For any $\mathbf{y} \in \mathbb{R}^+$,

$$
\sum_{t=1}^n \langle \mathbf{h}^{(n)}, \mathbf{y} - \mathbf{y}^* \rangle \geq -\frac{\|\mathbf{y}^{(1)} - \mathbf{y}\|_2^2}{2c^{(n)}/Q} - n\frac{c^{(t)}Q(G + \tau)^2}{2} - \epsilon^{(n)}, \quad (42)
$$

where $\epsilon^{(n)} = Q(\beta^{(n)}(4G + 2\tau) + \zeta)$.

Proof.

(a) Since $\|\mathbf{g}^{q(n)} + \boldsymbol{\nu}\|_2 \leq G + \tau$, using the update rule in (38), for any $\mathbf{y} \in \mathbb{R}^+$,

$$
\begin{aligned}
\|\mathbf{y}^{(t+1)} - \mathbf{y}\|_2^2 &= \left\| \mathbf{y}^{(t)} - \frac{c^{(t)}}{Q} \mathbf{h}^{(t)} - \mathbf{y} \right\|_2^2 \\
&= \left(\frac{c^{(t)}}{Q} \right)^2 \|\mathbf{h}(t)\|_2^2 - \frac{2c^{(t)}}{Q} \left\langle \mathbf{h}(t), \mathbf{y}^{(t)} - \mathbf{y} \right\rangle + \left\| \mathbf{y}^{(t)} - \mathbf{y} \right\|_2^2 \\
&\leq \|\mathbf{y}^{(t)} - \mathbf{y}\|_2^2 - \frac{2c^{(t)}}{Q} \left\langle \mathbf{h}^{(t)}, \mathbf{y}^{(t)} - \mathbf{y} \right\rangle + c^{(t)2}(G + \tau)^2,
\end{aligned}
$$

holds, where $\|\mathbf{h}^{(t)}\|_2 = \|\sum_q \mathbf{g}^{q(t)} + \mathbf{d}^{(t)}/c^{(t)})\|_2 \leq Q(G + \tau)$. Therefore, for any $\mathbf{y} \in \mathbb{R}^+$

$$
\langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^{(t)} \rangle \geq \frac{\|\mathbf{y}^{(t+1)} - \mathbf{y}\|_2^2 - \|\mathbf{y}^{(t)} - \mathbf{y}\|_2^2}{2c^{(t)}/Q} - \frac{c^{(t)}Q(G + \tau)^2}{2}. \quad (43)
$$

Summing across t leads to

$$\sum_{t=1}^{n} \langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^{(t)} \rangle \geq \frac{-\|\mathbf{y}^{(1)} - \mathbf{y}\|_2^2}{2c^{(n)}/Q} - n\frac{c^{(n)}Q(G+\tau)^2}{2},$$

which proves part (a). $\square$

(b) Since $\mathbf{h}^{(t)}$ is an $\epsilon^{(t)}$-subgradient of the dual function $d(\cdot)$ at $\mathbf{y}^{(t)}$, using the ϵ-subgradient inequality (34)

$$d(\mathbf{y}^*) \geq d(\mathbf{y}^{(t)}) + \left\langle \mathbf{h}^{(t)}, \mathbf{y}^* - \mathbf{y}^{(t)} \right\rangle - \epsilon^{(t)}$$

$$\Rightarrow \langle \mathbf{h}^{(t)}, \mathbf{y}^{(t)} - \mathbf{y}^* \rangle \geq d(\mathbf{y}^{(t)}) - d(\mathbf{y}^*) - \epsilon^{(t)} \geq -\epsilon^{(t)},$$

where the last inequality comes from the optimality condition $d(\mathbf{y}^{(t)}) \geq d(\mathbf{y}^*)$. Then, it follows

$$\begin{aligned}\langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^* \rangle &= \langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^{(t)} \rangle + \langle \mathbf{h}^{(t)}, \mathbf{y}^{(t)} - \mathbf{y}^* \rangle \\ &\geq \langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^{(t)} \rangle - \epsilon^{(t)}.\end{aligned} \quad (44)$$

Substituting the right-hand side of (43) into the right-hand side of (44), it follows

$$\begin{aligned}\langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^* \rangle \geq &\frac{\|\mathbf{y}^{(t+1)} - \mathbf{y}\|_2^2 - \|\mathbf{y}^{(t)} - \mathbf{y}\|_2^2}{2c^{(t)}/Q} \\ &- \frac{c^{(t)}Q(G+\tau)^2}{2} - \epsilon^{(t)}.\end{aligned}$$

Summing the above expression across t to get

$$\begin{aligned}\sum_{t=1}^{n} \langle \mathbf{h}^{(t)}, \mathbf{y} - \mathbf{y}^* \rangle \geq &- \frac{\|\mathbf{y}^{(1)} - \mathbf{y}\|^2}{2c^{(t)}/Q} - n\frac{c^{(t)}Q(G+\tau)^2}{2} \\ &- \sum_{t=1}^{n} \epsilon^{(t)},\end{aligned}$$

which proves part (b). $\square$

Proof. (Theorem 4)

(a) By concavity of the primal cost function $f^{(n)} \triangleq \sum_q f^{q(n)}$ and definition of $\mathbf{Q}^{q(n+1)}$ as a local Lagrangian maximizer of local Lagrangian function $\mathcal{L}^q(\mathbf{Q}^q, \boldsymbol{\ell}^q)$,

$$\begin{aligned}f^{(n)} &= \sum_q d^q\big(\boldsymbol{\ell}^{q(n)}\big) - \big\langle \mathbf{g}^{q(n)}, \boldsymbol{\ell}^{q(n)} \big\rangle \\ &\geq \frac{1}{n}\sum_{t=1}^{n}\sum_q d^q\big(\boldsymbol{\ell}^{q(t)}\big) - \big\langle \mathbf{g}^{q(t)}, \boldsymbol{\ell}^{q(t)} \big\rangle.\end{aligned} \quad (45)$$

is true. By (37) in Lemma 5 with $\mathbf{y} = \boldsymbol{\ell}^{q(t)}$

$$\begin{aligned}d^q(\boldsymbol{\ell}^{q(t)}) \geq &\ d^q(\mathbf{y}^{(t)}) + \left\langle \mathbf{g}^{q(t)}, \boldsymbol{\ell}^{q(t)} \right\rangle + \langle \boldsymbol{\nu}, \boldsymbol{\ell}^{q(t)} \rangle \\ &- \langle \mathbf{g}^{q(t)} + \boldsymbol{\nu}, \mathbf{y}^{(t)} \rangle - \epsilon^{(t)}/Q\end{aligned}$$

with $\epsilon^{(t)}/Q = \beta^{(t)}(4G + 2\tau) + \zeta$. Summing over q leads to

$$\begin{aligned}\sum_q d^q(\boldsymbol{\ell}^{q(t)}) \geq \sum_q \Big(&d^q(\mathbf{y}^{(t)}) + \langle \boldsymbol{\ell}^{q(t)}, \mathbf{g}^{q(t)} \rangle \\ &+ \langle \boldsymbol{\nu}, \boldsymbol{\ell}^{q(t)} \rangle - \langle \mathbf{g}^{q(t)} + \boldsymbol{\nu}, \mathbf{y}^{(t)} \rangle - \epsilon^{(t)}/Q\Big).\end{aligned}$$

Replacing $d^q(\boldsymbol{\ell}^{q(t)})$ in (45) with the lower bound in the preceding equation, (45) becomes

$$f^{(n)} \geq \frac{1}{n}\sum_{t=1}^{n}\sum_q d^q(\mathbf{y}^{(t)}) + \langle \boldsymbol{\nu}, \boldsymbol{\ell}^{q(t)} \rangle - \langle \mathbf{g}^{q(t)} + \boldsymbol{\nu}, \mathbf{y}^{(t)} \rangle - \frac{\epsilon^{(t)}}{Q}. \quad (46)$$

Since $\mathbf{h}^{(t)} = \sum_q \mathbf{g}^{q(t)} + \boldsymbol{\nu}$, the third term on the right-hand side in (46) can be replaced by the lower bound in

Lemma 6(a) by setting $\mathbf{y} = \mathbf{0}$. Furthermore, it is possible to bound $\langle \boldsymbol{\nu}, \boldsymbol{\ell}^{q(t)} \rangle \geq -\tau\Lambda$, which comes by construction with $\|\boldsymbol{\nu}\|_2 \leq \tau$ and $\|\boldsymbol{\ell}^{q(t)}\|_2 \leq \Lambda$. Hence, (46) can be rewritten as

$$f^{(n)} \geq d(\mathbf{y}^{(n)}) - Q\tau\Lambda - \frac{\|\mathbf{y}^{(1)}\|_2^2}{2nc^{(n)}/Q} - \frac{c^{(n)}Q(G+\tau)^2}{2} - \frac{1}{n}\sum_{t=1}^{n}\epsilon^{(t)}.$$

By strong duality $d^* = f^*$ and $d(\mathbf{y}) \geq d^*, \forall \mathbf{y} \in \mathbb{R}^+$, so that the equation above can be rewritten as

$$f^{(n)} \geq f^* - Q\tau\Lambda - \frac{\|\mathbf{y}^{(1)}\|_2^2}{2nc^{(n)}/Q} - \frac{c^{(n)}Q(G+\tau)^2}{2} - \frac{1}{n}\sum_{t=1}^{n}\epsilon^{(t)}.$$

In addition, from Assumption 5 $\frac{1}{n}\sum_{t=1}^{n}\epsilon^{(t)} \leq \epsilon^{(1)}$ and $\|\mathbf{y}^{(1)}\|_2^2 \leq \Lambda^2$ since $\epsilon^{(n)}$ is a nondecreasing sequence (because $\beta^{(n)}$ is an nondecreasing sequence) with $\epsilon^{(1)} \geq \cdots \geq \epsilon^{(n)}$, for $n > 1$. The expression above can be further rewritten as

$$\begin{aligned}f^{(n)} \geq &\ f^* - Q\tau\Lambda - \frac{\Lambda^2}{2nc^{(n)}/Q} - \frac{c^{(n)}Q(G+\tau)^2}{2} \\ &- Q(\beta^{(1)}(4G + 2\tau) + \zeta)\end{aligned}$$

and part (a) of the theorem is proven. $\square$

(b) Given any dual optimal solution $\mathbf{y}^*$,

$$f^{(n)} = f^{(n)} + \left\langle \sum_q \mathbf{h}^{q(n)}, \mathbf{y}^* \right\rangle - \left\langle \sum_q \mathbf{h}^{q(n)}, \mathbf{y}^* \right\rangle \quad (47)$$

is true. Also the first two terms on the right-hand side can be rewritten as

$$\begin{aligned}f^{(n)} + \left\langle \sum_q \mathbf{h}^{q(n)}, \mathbf{y}^* \right\rangle &= f^{(n)} + \left\langle \sum_q \mathbf{g}^{q(n)}, \mathbf{y}^* \right\rangle \\ &\quad + Q\langle \mathbf{d}^{(n)}/c^{(n)}, \mathbf{y}^* \rangle \\ &\leq f^{(n)} + \left\langle \sum_q \mathbf{g}^{q(t)}, \mathbf{y}^* \right\rangle + Q\Lambda\tau,\end{aligned}$$

where the Cauchy-Schwarz inequality was used to bound

$$\langle \mathbf{d}^{(n)}/c^{(n)}, \mathbf{y}^* \rangle \leq \|\mathbf{d}^{(n)}/c^{(n)}\|_2 \|\mathbf{y}^*\|_2 \leq \tau\Lambda.$$

Next, from the saddle point property, i.e., $\mathcal{L}(\mathbf{Q}^{(n)}, \mathbf{y}^*) \leq \mathcal{L}(\mathbf{Q}^*, \mathbf{y}^*) = f^*$, is used to further rewrite the bound as

$$\begin{aligned}f^{(n)} + \left\langle \sum_q \mathbf{g}^{q(n)}, \mathbf{y}^* \right\rangle + Q\tau\Lambda &= \mathcal{L}(\mathbf{Q}^{(n)}, \mathbf{y}^*) + Q\tau\Lambda \\ &\leq f^* + Q\tau\Lambda. \quad (48)\end{aligned}$$

Also, it is possible to upper bound $-\langle \mathbf{y}^*, \mathbf{h}^{(n)} \rangle$ by substituting $\mathbf{y} = 2\mathbf{y}^* \in \mathbb{R}^+$ inside part (b) of Lemma 6,

$$\begin{aligned}-\langle \mathbf{h}^{(n)}, \mathbf{y}^* \rangle &\leq \frac{\|\mathbf{y}^{(n)} - 2\mathbf{y}^*\|_2^2}{2c^{(n)}/Q} + \frac{c^{(n)}Q(G+\tau)^2}{2} + \epsilon^{(n)} \\ &\leq \frac{5\Lambda^2}{2c^{(n)}/Q} + \frac{c^{(n)}Q(G+\tau)^2}{2} + \epsilon^{(n)}, \quad (49)\end{aligned}$$

where the last inequality was obtained due to the fact that $\|\mathbf{y}^{(n+1)} - 2\mathbf{y}^*\|_2^2 \leq \|\mathbf{y}^{(n)}\|_2^2 + \|2\mathbf{y}^*\|_2^2 \leq 5\Lambda^2$ by Assumption 5. Therefore, by putting (48) and (49) into (47), an upper bound on $f^{(n)}$ can be obtained as

$$\begin{aligned}f^{(n)} \leq &\ f^* + Q\tau\Lambda + \frac{5\Lambda^2}{2nc^{(n)}/Q} + \frac{c^{(n)}Q(G+\tau)^2}{2} \\ &+ Q(\beta^{(1)}(4G + 2\tau) + \zeta),\end{aligned}$$

which concludes the proof. $\square$

REFERENCES

[1] 3GPP, "TR 36.819 coordinated multi-point operation for LTE physical layer aspects", 2013.

[2] G. Interdonato, E. Björnson, H.Q. Ngo, P. Frenger, and E.G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. and Netw.*, vol. 197, pp. 1–13, 2019.

[3] Qualcomm: How can comp extend 5g nr to high capacity and ultra-reliable communications? [Online]. Available: https://www.qualcomm.com/media/documents/files/how-comp-can-extend-5g-nr-to-high-capacity-and-ultra-reliable-communications.pdf.

[4] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, pp. 102–111, Feb. 2011.

[5] Y. Cheng, S. Drewes, A. Philipp, and M. Pesavento, "Joint network topology optimization and multicell beamforming using mixed integer programming," *IEEE ITG WSA*, pp. 187–192, Mar. 2012.

[6] J. Zhao, T.Q.S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 2762–2775, Jun. 2013.

[7] K.-G. Nguyen, Q.-D. Vu, M. Juntti, and L.-N. Tran, "Globally optimal beamforming design for downlink CoMP transmission with limited backhaul capacity," in Proc. IEEE ICASSP, Mar. 2017.

[8] P. Luong, L.-N. Tran, C. Despins and G. Gagnon, "Joint beamforming and remote radio head selection in limited fronthaul C-RAN," in Proc. IEEE VTC, Sep. 2016.

[9] Q. Shi, M. Razaviyayn, Z.-Q. Luo and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, pp. 4331–4340, Sep. 2011.

[10] S.S. Christensen, R. Agarwal, E. Carvalho, and J.M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[11] R. Agarwal and J.M. Cioffi, "Beamforming design for the MIMO downlink for maximizing weighted sum-rate," *ISITA*, pp. 1–6, Dec. 2008.

[12] P. Komulainen, A. Tolli, and M. Juntti, "Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2204–2218, May 2013.

[13] P.C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Multicell MISO downlink weighted sum-rate maximization: A distributed approach," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 556–570, Feb. 2013.

[14] A. Tölli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2 pp. 570–580, Feb. 2011.

[15] C. Shen, T.-H.Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Process.*, no.6, vol. 60, pp. 2988–3003, Jun. 2012.

[16] 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.*, no.2, vol. 31, pp. 226–240, Feb. 2013.

[17] J. Kaleva *et al.*, "Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[18] J. Kaleva *et al.*, "Decentralized joint precoding with pilot-aided beamformer estimation," *IEEE Trans. Signal Process.*, no.9, vol. 66, pp. 2330–2341, May 2018.

[19] I. Atzeni, I., B. Gouda, and A. Tölli, "Distributed precoding design via over-the-air signaling for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, no.2, vol. 20, pp. 1201–1216, Feb. 2021.

[20] R. Olfati-Saber and R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Automatic Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

[21] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," *J. Optim. Theory Appl.*, vol. 168, no. 1, pp. 172–197, May 2015.

[22] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, Oct. 2017.

[23] S. Liang, L.Y. Wang, and G. Yin, "Distributed smooth convex optimization with coupled constraints," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 347–353, Jan. 2020.

[24] A. Falsone, I. Notarnicola, G. Notarsefano, and M. Prandini, "Tracking-ADMM for distributed constraint-coupled optimization," *Automatica*, vol. 117, Jul. 2020.

[25] T. Sherson, R. Heusden, and W.B. Kleijn, "On the distributed method of multipliers for separable convex optimization problems," *IEEE Trans. Signal Inf. Proces. Over Netw.*, vol. 5, no. 3, pp. 495–510, Mar. 2019.

[26] B. Johansson, T. Keviczky, M. Johansson, and K.H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in Proc. IEEE CDC, Dec. 2008.

[27] J.C. Duchi, A. Agarwal, and M.J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Automatic Control*, no.3, vol. 57, pp. 592–606, Mar. 2012.

[28] C.-Y. Chang and C.C. Fung, "Sparsity enhanced mismatch model for robust spatial intercell interference cancelation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 125–139, Jan. 2015.

[29] C.-Y. Chang and C.C. Fung, "Sparsity enhanced mismatch model for robust spatial intercell interference management in heterogeneous networks with doubly-selective fading channels," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2671–2684, Jul. 2015.

[30] C.C. Fung and S.-T. Tai, "Orthogonalized" sparsity enhanced mismatch models for wireless heterogeneous networks," *Physical Commun.*, vol. 51, pp. 1–14, 2022.

[31] A. Nagate, S. Nabatame, D. Ogata, K. Hoshino and T. Fujii, "Field experiment of CoMP joint transmission over x2 interface for LTE-advanced," in Proc. IEEE VTC, Jun. 2013.

[32] X. Yu, Q. Cui, and M. Haenggi, "Coherent joint transmission in downlink heterogeneous cellular networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 274–277, Apr. 2018.

[33] J. Sun, W. Chen, P. Gaal, and J. Montojo, "Non-coherent joint transmission techniques," *World International Property Organization WO 2018/031872 A1*, 2018.

[34] 3GPP, *TR 38.913 v14.3.9 5g: Study on scenarios and requirements for next generation access technologies*, 2018.

[35] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM J. Optimization*, vol. 6, no 2, pp. 342–361, 1996.

[36] M. Servetnyk and C.C. Fung, "Distributed joint transmitter design and selection using augmented ADMM," in Proc. IEEE ICASSP, May 2019.

[37] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *J. of Optim. Theory Appl.*, vol. 129, no 3, pp. 469–488, Nov. 2006.

[38] S.M. Robinson, "Linear convergence of epsilon-subgradient descent methods for a class of convex functions," *Mathematical Programming*, vol. 86, no. 1, pp. 41–50, 1999.

[39] M. Servetnyk and C.C. Fung, "Precoding and selection for coordinated multipoint transmission in fronthaul-constrained cloud-RAN," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 51–55, Jan. 2020.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction of method of multipliers," *Foundations Trends Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[41] L. Vandenberghe, V.R. Balakrishnan, R. Wallin, A. Hansson, and T. Roh, "Interior-point algorithms for semidefinite programming problems derived from the KYP lemma," *Henrion D., Garulli A. (eds) Positive Polynomials in Control, Lecture Notes in Control and Information Science*, Springer, Berlin, Heidelberg, vol. 312, pp. 195–238, Aug. 2005.

[42] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 50–61, May 2010.

[43] *3GPP TS 38.101-1 v17.6.0*, "3GPP Technical Specification Group Radio Access Network; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 17)", 2017.

[44] J. Hansryd and J. Edstam, "Microwave capacity evolution," *Ericsson Review*, vol. 1, pp. 22–27, 2011.

[45] A. Nedic, "Subgradient methods for convex minimization," *Ph.D. Thesis, Massachusetts Institute of Technology*, 2002.

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

[47] M. Grant, and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[48] ApS, M. O. S. E. K., *MOSEK optimization suite*, 2017.

Mykola Servetnyk (S'17-M'20) received his B.Sc. degree in Electronics Engineering in 2015 and Ph.D. degree in Electronics Engineerin & Computer Science in 2020, both from National Chiao Tung University, Hsinchu, Taiwan. He also received a B.Sc. degree in 2017 in Applied Mathematics from Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. He was a Visiting Research Assistant in 2019 at the University of Houston, TX, USA. He is currently working as a Software Engineer in tests for wireless communications systems. His research interests include mobile communications, distributed algorithm design, and data science.

Carrson C. Fung received his B.S. degree from Carnegie Mellon University, Pittsburgh, PA, USA, in 1994, M.S. degree from Columbia University, New York City, NY, USA, in 1996 and Ph.D. degree at The Hong Kong University of Science and Technology in 2005, all in Electrical Engineering. He was the Recipient of the prestigious Sir Edward Youde Ph.D. Fellowship in 2001-2002. He was the Recipient of the National Chiao Tung University College of Electrical & Computer Engineering Excellent Teaching Award in 2011, 2017, and 2018. He received the National Chiao Tung University College of Electrical & Computer Engineering Outstanding Teaching Award in 2020. He also received the National Chiao Tung University Excellent Teaching Award in 2019. He became a Fellow of the Advance Higher Education Academy in 2020. He was a Member of Technical Staff at AT&T and Lucent Technologies Bell Laboratories, Holmdel, NJ, USA, from 1994-1999, where he worked on video and audio coding. He was also a Researcher at the Hong Kong Applied Science and Technology Research Institute (ASTRI) in 2005, where he worked on MIMO-OFDM systems and a Senior DSP Engineer at Sennheiser Research Lab in Palo Alto, CA, USA, in 2006, where he worked on microphone and microphone array technologies. He is currently an Associate Professor at the National Yang Ming Chiao Tung University in Hsinchu, Taiwan. His research interests include graph signal processing, data science, and optimization.