DISTRIBUTED JOINT TRANSMITTER DESIGN AND SELECTION USING AUGMENTED ADMM

Mykola Servetnyk and Carrson C. Fung

Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan 300

ABSTRACT

This work considers a design of network in which multiple transmission points (TPs) cooperatively serve users by jointly precoding shared data. Considered problem formulation jointly designs the beamformers and performs TP-UE link selection, which aims in improving overall system rate. Proposed distributed Augmented ADMM algorithm features parallelization among TPs, which has practical importance for computational load distribution and reducing signaling overhead in backhaul. This approach is different from others in literature because it solves a design problem that involves a coupling constraint which no existing algorithm is able to solve. Simulation results are also provided to show that the proposed distributed algorithm performance outperforms previously proposed distributed consensus optimization method and is comparable to its centralized counterpart.

***Index Terms*—** Coordinated multipoint transmission, TP selection, interference management, distributed optimization.

1. INTRODUCTION

The increasing need for spectral efficiency imposes the demand for effective interference management and transmission coordination. Coordinated multipoint transmission (CoMP) allows joint transmission from cooperating TPs, which gives ability to simultaneously transmit data by sharing same channel resources, such as time and frequency, to its users to boost reception performance. Several joint transmit and TP selection design strategies have been proposed in literature to reduce signaling overhead. [1] tackled the joint precoder and clustering problem by formulating the problem as a mixed-integer convex problem which minimizes the transmit energy subject to SINR QoS constraint, where convex formulation was obtained upon reformulation of the constraint. However, the approach only applies for case of a single receive antenna. Two algorithms were proposed in [2] with similar problem formulation, where one is based on iterative reweighed ℓ_1-norm minimization, the other is based on solving ℓ_2-norm relaxed problem and then iteratively removing the links that correspond to the smallest transmit power. [3] extended previous designs by including per TP backhaul constraint and solved the design problem by customized branch and bound algorithm applied to discrete monotonic optimization.

Unfortunately, none of these works is appropriate for large-sized networks due to computational complexity at centralized unit that increases proportionally to number of TPs in the network. Distributed optimization creates a valid alternative and overcomes privacy limitations by allowing each agent, or node, to keep its own data and not transmitting them across the network while distributing the computational effort, which has been extensively studied in [4–7]. [4], [5] and its extended version [6] proposed linear transceiver design algorithm for sum-rate maximization that is based on iterative minimization of weighted MSE, but solving the problem requires iterative design of transmitter and receiver. [7] proposed gradient based distributed joint beamforming, where pilot contamination in TDD is used for transceiver design.

The agents in the network coordinate as a swarm to perform a certain task and reach *consensus*. [8] provides a theoretical framework for wide range of consensus algorithms for multiagent networked systems. Amongst all of these algorithms, dual decomposition has been widely used. [9] and [10] have developed consensus-based proximal-dual decomposition for distributed processing, however, with certain limitation in the form of the constraint which limits its application and will be described later. In this work, a distributed consensus algorithm using augmented ADMM (AADMM) scheme is proposed, which includes an adaptive "resource" allocation scheme to overcome previously mentioned limitation.

Notations: Upper (lower) bold face letters indicate matrices (column vectors). Superscript ^{H} denotes Hermitian, ^{T} denotes transposition. $\mathbf{A} \succeq 0$ designate $\mathbf{A}$ as a symmetric positive semidefinite matrix. $\mathbf{1}_M$ denotes an $M \times 1$ vector, containing 1 in all of their entries. $tr(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$. $\|\cdot\|$ denotes Frobenius norm, unless specifically mentioned. $|\mathbf{A}|$ denotes the elementwise magnitude value of $\mathbf{A}$. $[\cdot]_+$ denotes the projection onto $\mathbb{R}_+$.

2. SYSTEM MODEL & PROBLEM FORMULATION

Assume the network consists of a set of TPs $\mathcal{Q} = \{1, \ldots, Q\}$, each having n_T transmit antennas. Set $\mathcal{I} = \{1, \ldots, I\}$ de-

This work has been partially supported by the Ministry of Science and Technology Grant 107-2221-E-009-071 and Ministry of Education project Trusted Intelligent Edge/Fog Computing Technology RSC 107RSA0021.

notes users that should be served by subset of $\mathcal{Q}$, known as the cooperating set, with each user equipped with n_R receive antennas. Channel between the qth TP and the ith user is denoted as $\mathbf{H}_i^q \in \mathbb{C}^{n_R \times n_T}$. The precoder matrix from the qth TP to the ith user is denoted as $\mathbf{F}_i^q \in \mathbb{C}^{n_T \times B}$, where B 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 intracell interference and $\mathbf{n}_i$ denotes AWGN with known variance σ_i^2 and $\mathbf{s}_i \in \mathbb{C}^B$ is data symbol for user i that satisfies $\mathbb{E}[\mathbf{s}_i^H \mathbf{s}_i] = 1$.

The joint TP selection and precoder design problem can be formulated by maximizing sum received signal power subject to instantaneous leakage interference power being constrained below the parameter I_{th} for all users in the cooperating set and the transmit power below P for all cooperating TPs. The number of active TPs can be controlled by including a sparsity inducing ℓ_0 regularization term in the objective. Therefore, the design problem can be formulated as

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

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

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

where $\alpha \in \mathbb{R}_+$ is regularization parameter that represents the associated cost for assigning TP to the users. Increases in α will undoubtely promote sparsity in the precoder vector, thus lowering operation cost. Since there is no available analytic rate equation for MU-MIMO, in order to improve system rate, maximization of sum of received signal power is considered.

It can be observed that (1a) is not convex. This can be easily circumvented by defining $\mathbf{Q}_i^q \triangleq \mathbf{F}_i^q \mathbf{F}_i^{qH}$ such that the sum 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 could be rewritten in similar way. Also, ℓ_0 can be relaxed and replaced with ℓ_1 norm, so that (1) becomes

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

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

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

where maximization with respect to $\mathbf{Q}$ denotes maximization with respect to $\mathbf{Q}_i^q$ for all $i \in \mathcal{I}$ and for all $q \in \mathcal{Q}$.

$\mathbf{F}_i^q$ can be retrieved from $\mathbf{Q}_i^q$ by using randomization procedure similar to the one in [11]. Specifically, $Nrand$ number of $\mathbf{F}_i^q = \frac{1}{B} \mathbf{U}_{\mathbf{Q}_i^q} \mathbf{\Lambda}_{\mathbf{Q}_i^q}^{1/2} \mathbf{E}$ 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 interference and transmit power constraints are kept. The remaining $\mathbf{F}_i^q$s are compared, and the one maximizing the objective is selected.

3. DISTRIBUTED DESIGN ALGORITHM

Proposed algorithm consists of three steps based on the ADMM algorithm [12]. Key idea is to define constrained variable, such that objective and constraints could be optimized independently. The fast iterative shrinkage-thresholding algorithm (FISTA) [13] is used in the first step to optimize objective. A consensus-based dual decomposition method is used in the second step to solve the constraint coupled problem in a distributive manner. Step 2 is the communications step in the entire algorithm, where neighboring TPs exchange information about the dual and slack variable associated with the instantaneous leakage interference power threshold. The threshold is regarded as a resource and a proposed adaptive resource method is proposed for its allocation. Therefore, private information, such as CSI/CQI, TP location nor precoder are not transmitted. This enhances security and reduces signaling overhead. Finally, third step updates dual variable of the ADMM constraint. All these steps can be done in parallel, which will be shown in the sequel.

3.1. Basic derivations

Define the constraint variable $\mathbf{Q}_{i\,c}^q$ and constraint $\mathbf{Q}_{i\,c}^q = \mathbf{Q}_i^q$. Also define the scaled dual variable $\mathbf{Q}_{i\,s}^q$ [12]. Next, rewrite (2) such that the constraints will only include variable $\mathbf{Q}_{i\,c}^q$ and objective only includes $\mathbf{Q}_i^q$. Lagrangian relaxation will then be applied to the constraints with an added proximity term to form the augmented Lagrangian

$$\mathcal{L}(\mathbf{Q}, \mathbf{Q}_c, \mathbf{Q}_s) = \sum_q \sum_i tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}) - \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T} - \frac{\rho}{2} \|\mathbf{Q}_i^q - \mathbf{Q}_{i\,c}^q + \mathbf{Q}_{i\,s}^q\|^2 + \frac{\rho}{2} \|\mathbf{Q}_{i\,s}^q\|^2. \tag{3}$$

3.2. Step 1: Objective optimization

In this step, (3) will be maximized with respect to $\mathbf{Q}_i^q$ for different q using the forward-backward algorithm [14] as it involves the 1-norm non-smooth function. This can be done by defining $f_1^q(\mathbf{Q}_i^q) \triangleq tr(\mathbf{H}_i^q \mathbf{Q}_i^q \mathbf{H}_i^{qH}) - \frac{\rho}{2} \|\mathbf{Q}_i^q - \mathbf{Q}_{i\,c}^q + \mathbf{Q}_{i\,s}^q\|^2$ and $f_2^q(\mathbf{Q}_i^q) \triangleq \alpha \mathbf{1}_{n_T}^T |\mathbf{Q}_i^q| \mathbf{1}_{n_T}$. Since $f_1^q(\mathbf{Q}_i^q)$ is a smooth function, it can be optimized via gradient method, where stepsize at the pth iteration, $t^{(p)}$, can be computed using backtracking search. The step direction equals the gradient of $f_1^q(\mathbf{Q}_i^q)$, i.e. $\nabla f_1^q(\mathbf{Q}_i^q) = \mathbf{H}_i^{qH} \mathbf{H}_i^q - \frac{\rho}{2}(\mathbf{Q}_i^q - \mathbf{Q}_{i\,c}^q + \mathbf{Q}_s^q)$. $f_2^q(\mathbf{Q}_i^q)$ is non-smooth and is optimized by applying threshold-shrinkage operator $\mathcal{T}_\gamma(\cdot)$ (proximal operator of 1-norm function) with input $\mathbf{Q}_i^q$ which optimized $f_1^q(\cdot)$. The threshold-shrinkage operator is given as $\mathcal{T}_\gamma[[\mathbf{X}]_{ij}] = \max(|[\mathbf{X}]_{ij}| - \gamma, 0)\,\mathrm{sign}([\mathbf{X}]_{ij})$ and γ is known as shrinkage threshold. $\mathbf{Q}_i^q$ at iteration p can be computed as

$$\mathbf{Q}_i^{q(p+1)} = \mathcal{T}_{\alpha t^{(p)}}[\mathbf{Q}_i^{q(p)} - t^{(p)} \nabla f(\mathbf{Q}_i^{q(p)})], \tag{4}$$

Step 1 will terminate when $\|\nabla f_1^q(\mathbf{Q}_i^{q(p)})\| \leq \epsilon_{fista}$ for some $\epsilon_{fista} \geq 0$, whose value is indicated in Table 2.

3.3. Step 2: Consensus Based Optimization

The second step involves maximizing (3) with respect to $\mathbf{Q}_{i\,c}^{q}$. 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 come to agreement. For convenience, define $L \triangleq I_{th}$. In the case of (2), L is regarded as a resource that is available to all agents but its exact value is unknown at each node. [9] and [10] considered a similar problem, however, the constraint is assumed to be of the form $\sum_q g^q(x^q) \leq 0$, with $g^q(x^q)$ being a convex function of local variable x^q. In other words, [9] and [10] assumed that L is evenly distributed amongst all agents as $L_{ij}^q = \frac{L}{Q} \in \mathbb{R}$ so that the leakage interference power constraint can be rewritten as $\sum_q g_{ij}^q(\mathbf{Q}_{i\,c}^q) \triangleq \sum_q tr\left(\mathbf{H}_j^q \mathbf{Q}_{i\,c}^q \mathbf{H}_j^{qH}\right) - L_{ij}^q \leq 0$ which has the same form as the constraint considered in [9,10]. This is however not optimal since L_{ij}^q is not necessarily equal $\forall q$.

An adaptive algorithm is proposed herein to resolve this problem. Notice that $\sum_q L_{ij}^q = I_{th}$. Denote objective function in step 2 as $f^q \triangleq \sum_i -\frac{\rho}{2}\|\mathbf{Q}_i^{q(n+1)} - \mathbf{Q}_{i\,c}^q + \mathbf{Q}_{i\,s}^{q(n)}\|^2$. Further, denote λ_{ij} as Lagrange multiplier associated with the interference leakage constraint for user pair (i,j) and define concatenated vectors $\boldsymbol{\lambda} \triangleq [\lambda_{12}, \lambda_{13}, ..., \lambda_{1I}, ..., \lambda_{ij}, ..., \lambda_{II-1}]^T$ and $\mathbf{g}^q \triangleq [g_{12}^q, g_{13}^q, ..., g_{1I}^q, ..., g_{ij}^q, ..., g_{II-1}^q]^T$ (the dependency on $\mathbf{Q}_{i\,c}^q$ has been omitted for brevity). It is important to emphasize that $\boldsymbol{\lambda}$ is a coupling variable, i.e. elements inside $\boldsymbol{\lambda}$ contains Lagrange multipliers that are applied to all nodes. Therefore, a local copy of $\boldsymbol{\lambda}$, called $\boldsymbol{\lambda}^q$, is created and the Lagrangian function that is optimized in step 2 at the qth node can be written in separable form as $\mathcal{L}(\mathbf{Q}_{i\,c}, \boldsymbol{\lambda}) = \sum_q f^q + \boldsymbol{\lambda}^{qT}\mathbf{g}^q = \sum_q \mathcal{L}^q(\mathbf{Q}_{i\,c}^q, \boldsymbol{\lambda}^q)$.

$\boldsymbol{\lambda}^q$ will be exchanged for each q node with its one-hop neighbors in set $\mathcal{N}^q$ during consensus optimization. Network over which TPs communicate is described by *consensus matrix* [15] $\mathbf{W} \in \mathbb{R}^{Q\times Q}$ with the properties that $[\mathbf{W}]_{rq} = 0$ if $r \notin \mathcal{N}^q$, $\mathbf{W} = \mathbf{W}^T$, $\mathbf{W}\mathbf{1}_Q = \mathbf{1}_Q$ and $\lim_{\varphi\to\infty} \mathbf{W}^{\varphi} = \mathbf{1}\mathbf{1}^T/Q$. Consensus shall be established using the dual variable $\boldsymbol{\lambda}^q$. To do so, define the consensus dual auxiliary variable

$$\boldsymbol{\ell}^q = \sum\nolimits_{r\in\mathcal{N}^q} [\mathbf{W}]_{rq}\boldsymbol{\lambda}^r. \tag{5}$$

Then the Lagrange dual problem to find $\boldsymbol{\lambda}^q$ becomes $\min_{\boldsymbol{\lambda}}^q \max_{\mathbf{Q}_{i\,c}^q} \left\{\mathcal{L}^q(\mathbf{Q}_{i\,c}^q, \boldsymbol{\lambda}^q) - \frac{\|\boldsymbol{\lambda}^q - \boldsymbol{\ell}^q\|^2}{2c}\right\}$ for each TP, where second term in the inner objective is a proximal term and acts as incentive for consensus. Scalar c balances emphasis on objective versus consensus on dual variable $\boldsymbol{\lambda}^q$. Maximization with respect to $\mathbf{Q}_{i\,c}^q$ at the nth iteration is

$$\mathbf{Q}_{i\,c}^{q(n+1)} = \arg\max_{\mathbf{Q}_{i\,c}^q \in \mathcal{C}^q} f^q + \boldsymbol{\ell}^{q(n)T}\mathbf{g}^{q(n)} \text{ s.t. (2c)} \tag{6}$$

at point $\boldsymbol{\ell}^{q(n)}$ instead of $\boldsymbol{\lambda}^{q(n)}$ [9, 10], which has to be initialized when $n = 0$ or obtained from (5) when $\boldsymbol{\lambda}^{q(n+1)} = \arg\min_{\boldsymbol{\lambda}^q \geq 0} \sum_i \boldsymbol{\lambda}^{qH}\widetilde{\mathbf{g}}^{q(n+1)} - \frac{\|\boldsymbol{\lambda}^q - \boldsymbol{\ell}^{q(n)}\|^2}{2c^{(n)}}$ becomes available. Note that $\widetilde{\mathbf{g}}^{q(n+1)}$ equals $\mathbf{g}^q$, but evaluated at $\mathbf{Q}_{i\,c}^{q(n+1)}$. This is a constrained maximization problem with a quadratic objective which can be solved explicitly as

$$\boldsymbol{\lambda}^{q(n+1)} = [\boldsymbol{\ell}^{q(n)} + c^{(n)}\widetilde{\mathbf{g}}^{q(n+1)}]_+. \tag{7}$$

$c^{(n)}$ is a nonincreasing sequence of positive reals and satisfies $\sum_{n=0}^{\infty} c^{(n)} = \infty$ and $\sum_{n=0}^{\infty} c^{(n)2} \leq \infty$.

3.4. Adaptive strategy for computing L_{ij}^q

Since the coupling constraint $\sum_q g_{ij}^q(\mathbf{Q}_{i\,c}^q)$ must have prior knowledge about L_{ij}^q and since even distribution of L is not optimal, an adaptive resource allocation algorithm is proposed to allocate L across all agents which achieves better performance than the even distribution scheme in [9, 10]. The scheme is illustrated as follows. Initialization is done with $L_{ij}^{q(0)} = \frac{L}{Q}$, i.e. equal distribution. Define the *slack leakage interference power* as $S_{ij}^{q(n)} \triangleq L_{ij}^{q(n)} - \widetilde{L}_{ij}^{q(n)}$ where $\widetilde{L}_{ij}^{q(n)} \triangleq tr(\mathbf{H}_j^{q(n)}\mathbf{Q}_i^{q(n)}\mathbf{H}_j^{q(n)H})$ is the *instantaneous leakage interference power*. $S_{ij}^{q(n)}$ can be viewed as the amount of excess leakage interference power (resource) that is not needed by the qth TP while serving the ith user. Then a new value for the instantaneous leakage interference power threshold can be updated as $L_{ij}^{q(n+1)} \triangleq \widetilde{L}_{ij}^{q(n)} + \sum_{r\in\mathcal{N}^q}[\mathbf{W}]_{rq}S_{ij}^{q(n)}$. After $L_{ij}^{q(n+1)}$ is computed at the nth iteration, it will be inserted into $\mathbf{g}^{q(n+1)}$ and $S_{ij}^{q(n+1)}$ will be computed for the next iteration. It should be noted that all nodes will be running the above steps in parallel. Convergence and optimality of this scheme is proven in [16].

3.5. Step 3: Dual Descent and ADMM parameter update

Step 3 involves the minimization of (3) with respect to the scaled dual variable $\mathbf{Q}_{i\,s}^q$, which can be done with simple gradient descent. The result becomes

$$\mathbf{Q}_{i\,s}^{q(m+1)} = \mathbf{Q}_{i\,s}^{q(m)} + \left(\mathbf{Q}_i^{q(m)} - \mathbf{Q}_{i\,c}^{q(m)}\right), \tag{8}$$

where the superscript (m) is the iteration index for step 3. $\mathbf{Q}_i^{q(m)}$ and $\mathbf{Q}_{i\,c}^{q(m)}$ equal to the most updated result from step 1 and 2, respectively. Note that ρ will vary during the iteration in order to improve convergence and make the performance to be less dependent on the initial choice of α [12]. Consequently, the scaled dual variable should be scaled as

$$(\rho^{(m+1)}, \mathbf{Q}_s^{(m+1)}) = \begin{cases} (\mu\rho^{(m)}, \frac{\mathbf{Q}_s^{(m+1)}}{\mu}), \text{ if } |r_p| \geq \tau|r_s| \\ (\frac{\rho^{(m)}}{\mu}, \mu\mathbf{Q}_s^{(m+1)}), \text{ if } |r_p| \leq \tau|r_s| \\ (\rho^{(m)}, \mathbf{Q}_s^{(m+1)}), \text{ otherwise} \end{cases}, \tag{9}$$

where $r_p = \|\mathbf{Q}_{i\,c}^{q(m)} - \mathbf{Q}_i^{q(m)}\|$ and $r_s = \|\mathbf{Q}_{i\,c}^{q(m)} - \mathbf{Q}_{i\,c}^{q(m-1)}\|$ denote the primal and dual residuals, respectively. μ and τ are parameters and their values are given in Table 2.

Overall algorithm is summarized in Algorithm 1 and convergence of step 2 is described in [16] where the Laplacian spectral gap plays an important role, while convergence of ADMM is well known [12].

Algorithm 1: Distributed consensus optimization using proposed AADMM.

Result: Precoder matrices $\mathbf{Q}_i^q\ \forall i \in \mathcal{I}, \forall q \in \mathcal{Q}$

0\. Initialize: $\mathbf{Q}_s^{q(0)}, \mathbf{Q}_c^{q(0)}, \boldsymbol{\lambda}^{q(0)}, \boldsymbol{\ell}^{q(0)}, \boldsymbol{L}^{q(0)}, m = 0$

while $|r_p^{(m)}| \geq \epsilon_{glo}$ & $|r_d^{(m)}| \geq \epsilon_{glo}$ **do**

- $m = m + 1$
- 1\. Local primal optimization: Set $p = 0$. For each TP
  - **while** $\|\nabla f_1^q(\mathbf{Q}_i^q)\| \leq \epsilon_{fista}$ **do**
    - $p = p + 1$; Compute $\mathbf{Q}_i^{q(p+1)}$ using (4)
- 2\. Consensus optimization step
  - Set $n = 0$. For each TP
  - **while** $\|\mathbf{Q}_c^{q(n+1)} - \mathbf{Q}_c^{q(n)}\|_F \leq \epsilon_{cons}$ **do**
    - $n = n + 1$; Update $c^{(n)}$;
    - Receive $\boldsymbol{\lambda}^{q(m+1)}$ from neighbors and update $\boldsymbol{\ell}^{q(m+1)}$ by (5)
    - Update $\mathbf{Q}_c^{q(n+1)}$ by (6) and update $S_{ij}^{q(n+1)}$
    - Update $\boldsymbol{\lambda}^{q(n+1)}$ by (7)
    - Receive $S_{ij}^{q(n+1)}$ from neighbors and update $L^{q(n+1)}$ according to Sec. 3.4
- 3\. Dual ascent step: compute $\mathbf{Q}_s^{q(m)}$ by (8)
- 4\. Update primal dual residuals and $\rho^{(m+1)}$ by (9).

4. NUMERICAL RESULTS

Example of simulated network is shown in Fig. 1. Each sector around a TP (green triangle) consists of three UEs (red squares) that are placed closer to sector edge to promote situation analogous to cell-edge users in traditional LTE networks. Pink arrow denotes TP serving a certain UE. Other network and algorithm parameters are summarized in Table 1 and 2.

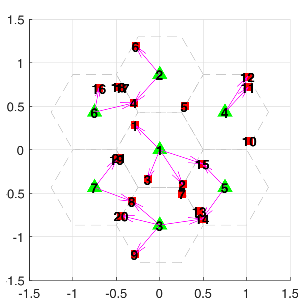

Fig. 1: Example of simulated network

# of TPs/UEs	$Q = 7, K = 21$
# of Antennas	$n_T = 4, n_R = 2$
I_{th}	$10^{-4}W$
P^q	$1W$
σ^2	-33 dB
Ref loss(dB)	60
PL exponent	3.76
Shadowing	10 dB
Tx antenna gain	10 dB

Table 1: Simulated network parameters

Fig. 2 shows convergence behavior when each TP is connected to two and three nearest neighbors, labeled as 2NN and 3NN, respectively. Results for a fully connected network (FC) and its centralized counterpart, i.e. solving (2) directly, are also shown as performance benchmark. Weights $[\mathbf{W}]_{rq}$ are selected to be $1/|\mathcal{N}^q|$ for each node q so messages from one node can be disseminated to all other nodes at the fastest rate for an undirected connected graph.

$\rho^{(0)}$	ϵ_{fista}	ϵ_{glo}	μ	τ	$Nrand$	$c^{(n)}$
2	10^{-6}	10^{-6}	1.1	5	10^4	$\frac{10^4}{n+1}$

Table 2: Simulated algorithm parameters

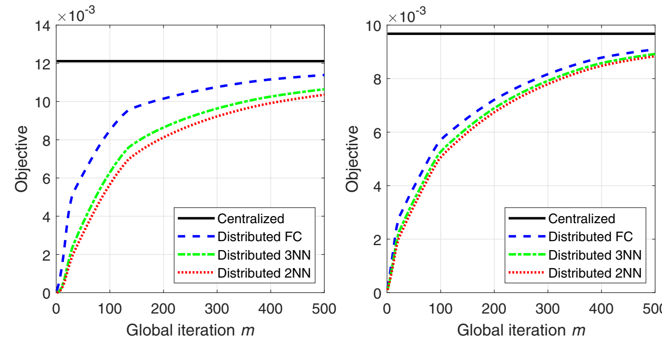

Fig. 2: Two instances of algorithm convergence for different network connectivity scenarios.

Performance of proposed distributed algorithm in terms of objective value and sum rate (kbps/Hz) is shown in Fig. 3. Results for different scenarios highlight the impact that connectivity in network has on performance. Fig. 3 also depicts the case of distributed algorithm with $L_{ij}^q = L/Q$ [9, 10] and is labeled as EqD in order to highlight the performance gain of the proposed AADMM method (with adaptive allocation).

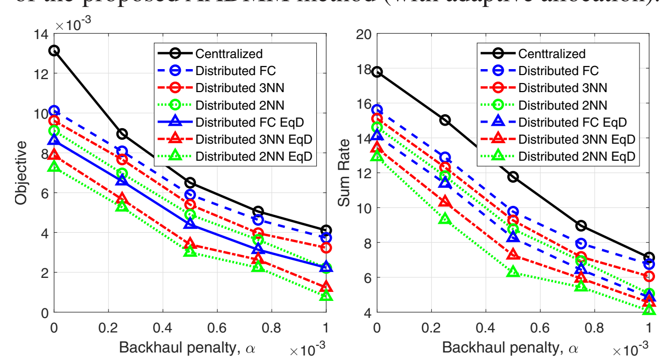

Fig. 3: Performance of the proposed problem formulation

5. CONCLUSION

A novel AADMM algorithm for distributed optimization is proposed for solving the joint transmitter design and selection problem. The algorithm includes an adaptive resource allocation method that circumvents the coupling constraint problem that previous algorithms failed to address, while maintaining convergence. Proposed scheme has large importance for next generation communication networks as it eliminates requirement for central unit, which reduces backhaul signaling overhead and latency caused by transmitting CSI to upper layers of the network. In addition, since communications step only involves the exchange of dual and slack variables, it enhances overall security in the network.

6. REFERENCES

[1] Y. Cheng, S. Drewes, A. Philipp, and M. Pesavento, "Joint network topology optimization and multicell beamforming using mixed integer programming," in *2012 International ITG Workshop on Smart Antennas (WSA)*, 2012, pp. 187–192.

[2] J. Zhao, T. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2762–2775, 2013.

[3] Kien-Giang Nguyen, Quang-Doanh Vu, Markku Juntti, and Le-Nam Tran, "Globally optimal beamforming design for downlink comp transmission with limited backhaul capacity," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3649–3653.

[4] P. Komulainen, A. Tölli, and M. Juntti, "Effective csi signaling and decentralized beam coordination in tdd multi-cell mimo systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2204–2218, 2013.

[5] 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," vol. 59(9), pp. 4331–4340, 2011.

[6] 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 Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 226–240, 2013.

[7] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with qos constraints for interfering broadcast channel via successive convex approximation," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2788–2802, 2016.

[8] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on automatic control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[9] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," *Journal of Optimization Theory and Applications*, vol. 168, no. 1, pp. 172–197, 2016.

[10] 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, 2017.

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

[12] S. Boyd and et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[14] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[15] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *47th IEEE Conference on Decision and Control, CDC 2008; Cancun; 9 December 2008 through 11 December 2008*. IEEE, 2008, pp. 4185–4190.

[16] M. Servetnyk and C. C. Fung, "Distributed joint tp selection and precoder design using augmented consensus based dual decomposition," unpublished.