D3M: DISTRIBUTED MULTI-CELL MULTIGROUP MULTICASTING

P. Song, G. Scutari, F. Facchinei, and L. Lampariello

ABSTRACT

The paper studies the max-min fair multicast multigroup beamforming problem in a multi-cell environment, with perfect (instantaneous or statistical) Channel State Information (CSI). We propose a new general *distributed* algorithmic framework based on INner Convex Approximations (INCA): the nonsmooth NP-hard problem is replaced by a sequence of smooth strongly convex subproblems, which can be solved in a *distributed* fashion across the cells, with limited communication overhead. Differently from renowned semidefinite-relaxation-based schemes, the INCA algorithm is proved to *always* converge to a d-stationary solution of the aforementioned class of problems. Numerical results show that it compares favorably with state-of-the-art algorithms.

Index Terms— Distributed optimization, inner convex approximation, multicell multigroup multicasting.

1. INTRODUCTION

Multicast beamforming is a part of the Evolved Multimedia Broadcast Multicast Service in the Long-Term Evolution standard [1] for efficient audio and video streaming. Multicast beamforming utilizes multiple transmit antennas and some form of CSI to steer transmitted power towards a group of subscribers while limiting interference to other users and systems [2]. Multicasting can be broadly classified into a) single-group multicasting (see [3] for state-of-the-art results), where all the subscribers request a common data stream from the transmitter; and b) multiple-group multicasting [4], where different groups of subscribers request different data streams from the transmitter. Two alternative criteria have been widely considered in the literature to design the beampatterns, namely: a) the maximization of the minimum received Signal to Interference plus Noise Ratio (SINR) subject to a transmit power constraint, which is commonly referred to as the Max-Min Fair (MMF) beamforming problem [4]; and b) the minimization of the transmit power subject to Quality-of-Service (QoS) guarantees at the receivers of all the users [2, 5].

This paper focuses on the MMF beamforming problem for *multiple*-group multicasting; both single-cell and multi-cell scenarios are considered. This problem is nonconvex, due to the nonconvexity of the SINR functions. Special instances of the general formulation exhibit ad-hoc structures that allow them to be solved efficiently, leveraging equivalent (quasi-)convex reformulations; see, e.g., [3, 6, 7]. In the case of general channel vectors, however, the (single-cell) MMF beamforming problem was proved to be NP-hard [4]. This has motivated a lot of interest to pursuit approximate solutions that approach optimal performance at moderate complexity. SemiDefinite Relaxations (SDR) followed by Gaussian randomization (SDR-G) have been extensively studied in the literature to obtain good suboptimal solutions [4, 8, 9, 10], with theoretical bound guarantees [11, 12]. For a large number of antennas or users, however, the quality of the approximation obtained by SDR-G methods deteriorates considerably. In fact, SDR-based approaches return feasible points that in general may not be even stationary for the original nonconvex problem. Moreover, in a multi-cell scenario, SDR-G is not suitable for a *distributed* implementation across the cells.

Two schemes based on heuristic convex approximations have been recently proposed in [13] and [14] (the latter based on earlier work [15]) for the *single-cell* multiple-group MMF beamforming problem. While extensive experiments show that these schemes achieve better solutions than SDR-G approaches, their theoretical convergence and guarantees remain an open question. Finally, we are not aware of any *distributed* scheme with provable convergence for the *multi-cell* MMF beamforming problem.

Building on our recent developments [16, 17], in this paper we fill this gap and propose the first *distributed* algorithm converging to d-stationary solutions of the aforementioned MMF beamforming problems. The algorithmic framework employs a novel convex approximation technique: the nonsmooth NP-hard problem is replaced by a sequence of smooth strongly convex subproblems, whereby the nonconvex SINR constraints are replaced by proper upper convex approximations; we term it "INner Convex Approximation" (INCA) algorithm. In a multi-cell scenario, it naturally leads to a *distributed* implementation with limited signaling across the base stations (BSs). Numerical results show that our INCA schemes reach better solutions than SDR-G approaches with high probability, while having comparable computational complexity.

The rest of the paper is organized as follows. Sec. 2 introduces the INCA algorithm in the simplified setting of a single-cell environment; the multi-cell case along with the distributed implementation of INCA is studied in Sec. 3. Preliminary numerical results are presented in Sec. 4, while Sec. 5 draws some conclusions.

2. SINGLE-CELL MULTIGROUP MULTICASTING

Consider a wireless multicast downlink network comprising a single Base BS, equipped with N_t transmit antennas, and M active users, which have a single receive antenna. There are K multicast groups, let $\mathcal{G}_k$ denote the k-th group, with $k \in \mathcal{K} \triangleq \{1, \cdots, K\}$. Each receiver listens to a single multicast, implying that $\mathcal{G}_1, \cdots, \mathcal{G}_K$ form a partition of the set of the M users. Denoting by $\mathbf{w}_k \in \mathbb{C}^{N_t}$ the beamforming weight vector for transmission to group k, the joint Max-Min Fair (MMF) beamforming problem reads [4]

$$\begin{array}{ll} \underset{\mathbf{w} \triangleq (\mathbf{w}_k)_{k=1}^K}{\text{maximize}} & U(\mathbf{w}) \triangleq \min_{k \in \mathcal{K}} \min_{i \in \mathcal{G}_k} \dfrac{1}{\mu_i} \dfrac{\mathbf{w}_k^H \mathbf{H}_i \mathbf{w}_k}{\sum_{\ell \neq k} \mathbf{w}_\ell^H \mathbf{H}_i \mathbf{w}_\ell + \sigma_i^2} \\ \text{subject to} & \sum_{k=1}^K \|\mathbf{w}_k\|_2^2 \leq P, \end{array} \quad (1)$$

where $\mathbf{H}_i$ is a positive semidefinite matrix modeling the channel between the BS and user i; specifically, $\mathbf{H}_i = \mathbf{h}_i \mathbf{h}_i^H$ if instantaneous CSI is assumed, with $\mathbf{h}_i \in \mathbb{C}^{N_t}$ denoting the frequency-flat quasi-static channel vector from the BS to user i; and $\mathbf{H}_i = \mathbb{E}(\mathbf{h}_i \mathbf{h}_i^H)$ represents the spatial correlation matrix if only long-term CSI is available (in the latter case, no special structure for $\mathbf{H}_i$ is assumed). The constant $1/\mu_i > 0$ is a predetermined factor accounting for possibly different grades of service; and σ_i^2 is the variance of the AWGN at receiver i. We denote by $\mathcal{W}$ the (convex) feasible set of (1). We remark that one can add further (convex) constraints in $\mathcal{W}$, such as per-antenna power constraints, null or interference constraints; the

Song is with the with the School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing, China; email: peiransong@bistu.edu.cn. Scutari is with the School of Industrial Engineering, Purdue University, West-Lafayette, IN, USA; email: gscutari@purdue.edu. Facchinei and Lapariello are with the Dept. of Computer, Control, and Management Engineering, University of Rome "La Sapienza", Rome, Italy; emails: <facchinei, lampariello>@diag.uniroma1.it. The work of Scutari was supported by the USA National Science Foundation under Grants CMS 1218717, CCF 1564044, and CAREER Award 1254739. The work of Facchinei was partially supported by the MIUR project PLATINO, Grant n. PON01_01007.

algorithmic framework we are going to introduce is still applicable.

Problem (1) has been proved to be NP-hard [4]. Therefore our focus is on computing efficiently (d-)stationary solutions of (1).

Definition 1 (d-stationarity) *Let $U'(\mathbf{w};\mathbf{d})$ denote the directional derivative of U at $\mathbf{w} \triangleq (\mathbf{w}_k)_{k=1}^K$ in the direction $\mathbf{d} \in \mathbb{C}^{N_t \cdot K}$, defined as $U'(\mathbf{w};\mathbf{d}) \triangleq \lim_{t\downarrow 0} (U(\mathbf{w}+t\mathbf{d}) - U(\mathbf{w}))/t$. A tuple $\mathbf{w}^\star \triangleq (\mathbf{w}_k^\star)_{k=1}^K \in \mathcal{W}$ is a d-stationary solution of (1) if*

$$U'(\mathbf{w}^\star;\mathbf{w}-\mathbf{w}^\star) \le 0, \quad \forall \mathbf{w} \in \mathcal{W}. \tag{2}$$

Of course, (local/global) optimal solutions of (1) satisfy (2).

Equivalent reformulation: To compute a stationary solution of the nonconvex and nonsmooth problem (1), we preliminarily rewrite (1) in the following equivalent *smooth* (still nonconvex) form: introducing the slack variables $t \ge 0$, and $\boldsymbol{\beta} \triangleq (\beta_i > 0)_{i=1}^M$, we have

$$\begin{aligned}
\underset{t\ge 0,\,\boldsymbol{\beta},\,\mathbf{w}}{\text{maximize}} \quad & t \\
\text{subject to} \quad & \underbrace{\mu_i \cdot t \cdot \beta_i - \mathbf{w}_k^H \mathbf{H}_i \mathbf{w}_k}_{\triangleq g_i(t,\beta_i,\mathbf{w}_k)} \le 0,\ \forall i,\ k \\
& \textstyle\sum_{\ell\neq k} \mathbf{w}_\ell^H \mathbf{H}_i \mathbf{w}_\ell + \sigma_i^2 \le \beta_i,\ \forall i,\ k \\
& \textstyle\sum_{k=1}^K \|\mathbf{w}_k\|_2^2 \le P.
\end{aligned} \tag{3}$$

We denote by $\mathcal{Z}$ the feasible set of (3).

Problems (1) and (3) are equivalent in the following sense.

Proposition 2 *Given problems (1) and (3), the following hold:*

(a) $\mathbf{w}^\star$ *is a d-stationary solution of (1) if and only if there exist $t^\star$ and $\boldsymbol{\beta}^\star \triangleq \{\beta_i^\star\}_{i=1}^M$ such that $(t^\star,\boldsymbol{\beta}^\star,\mathbf{w}^\star)$ is a stationary solution of (3);*

(b) *Every stationary solution $(t^\star,\boldsymbol{\beta}^\star,\mathbf{w}^\star)$ of (3) is regular, i.e, the Mangasarian-Fromovitz Constraint Qualification (see, e.g., [18, Sec. 3.2]) is satisfied at $(t^\star,\boldsymbol{\beta}^\star,\mathbf{w}^\star)$.*

While the equivalence between (1) and (3) in terms of global optimal solutions is a well-known fact in the literature, the same result in terms of stationary solutions is, to the best of our knowledge, new. It follows from Proposition 2 that one can focus w.l.o.g. on the smooth reformulation (3). Building on the framework developed in [16], the next section is devoted to the design of an efficient algorithm for the computation of the stationary solutions of (3).

2.1. Algorithmic design

Problem (3) is nonconvex due to the nonconvex constraint functions $g_i(t,\beta_i,\mathbf{w}_k)$. The proposed approach consists then in solving a sequence of strongly *convex approximations* of (3) wherein each g_i is replaced by a suitably chosen upper convex surrogate. More formally, denoting by $\mathbf{z}^\nu \triangleq (t^\nu,\boldsymbol{\beta}^\nu,\mathbf{w}^\nu)$ the current iterate, the convex subproblem at iteration $\nu \ge 1$ reads

$$\begin{aligned}
\underset{t\ge 0,\,\boldsymbol{\beta},\,\mathbf{w}}{\text{maximize}} \quad & t - \frac{\tau_t}{2}(t-t^\nu)^2 - \tau_{\mathbf{w}} \|\mathbf{w}-\mathbf{w}^\nu\|_2^2 - \tfrac{\tau_{\boldsymbol{\beta}}}{2}\|\boldsymbol{\beta}-\boldsymbol{\beta}^\nu\|^2 \\
\text{subject to} \quad & \tilde{g}_i(t,\beta_i,\mathbf{w}_k;t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu) \le 0,\ \forall i,k, \\
& \textstyle\sum_{\ell\neq k} \mathbf{w}_\ell^H \mathbf{H}_i \mathbf{w}_\ell + \sigma_i^2 \le \beta_i,\ \forall i,k, \\
& \textstyle\sum_{k=1}^K \|\mathbf{w}_k\|_2^2 \le P,
\end{aligned} \tag{4}$$

where each $\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu)$, function of $(t,\beta_i,\mathbf{w}_k)$, is an upper convex approximation of $g_i(t,\beta_i,\mathbf{w}_k)$ at the current iterate $(t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu)$. In the objective function of (4) we added a proximal regularization to make it strongly convex; therefore, problem (4) has a unique solution, which we denote by $(\hat{t}^\nu,\hat{\boldsymbol{\beta}}^\nu,\hat{\mathbf{w}}^\nu)$.

On the surrogate functions $\tilde{g}_i$: The surrogate functions $\tilde{g}_i$ need to be chosen to satisfy some technical conditions (cf. [16, Assumption 3]). Here we recall only the key assumptions (the others are readily satisfied in practice), namely: $\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu)$ is a convex global upper bound of $g_i(t,\beta_i,\mathbf{w}_k)$ such that $\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t,\beta_i,\mathbf{w}_k) = g_i(t,\beta_i,\mathbf{w}_k)$ and $\nabla_1\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t,\beta_i,\mathbf{w}_k) = \nabla g_i(t,\beta_i,\mathbf{w}_k)$, for all tuples $(t,\beta_i,\mathbf{w}_k)$ that are feasible for (3), where $\nabla_1\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t,\beta_i,\mathbf{w}_k)$ denotes the partial gradient of $\tilde{g}_i$ with respect to $(t,\beta_i,\mathbf{w}_k)$ evaluated at $(t,\beta_i,\mathbf{w}_k;t,\beta_i,\mathbf{w}_k)$. The upper bound conditions guarantee that every solution $(\hat{t}^\nu,\hat{\boldsymbol{\beta}}^\nu,\hat{\mathbf{w}}^\nu)$ is feasible for (3), whereas the gradient consistency condition ensures that the approximations have locally the same first order behavior of the original functions. Two examples of valid surrogates are given next.

Note that $g_i(t,\beta_i,\mathbf{w}_k)$ is the sum of a bilinear function and a concave one, namely: $g_i(t,\beta_i,\mathbf{w}_k) = g_{i,1}(t,\beta_i) + g_{i,2}(\mathbf{w}_k)$, with

$$g_{i,1}(t,\beta_i) \triangleq \mu_i \cdot t \cdot \beta_i, \text{ and } g_{i,2}(\mathbf{w}_k) \triangleq -\mathbf{w}_k^H \mathbf{H}_i \mathbf{w}_k. \tag{5}$$

A valid surrogate $\tilde{g}_i$ is then obtained as follows: i) one can linearize $g_{i,2}(\mathbf{w}_k)$ around $\mathbf{w}_k^\nu$, that is,

$$\begin{aligned}
\tilde{g}_{i,2}(\mathbf{w}_k;\mathbf{w}_k^\nu) \triangleq & -(\mathbf{w}_k^\nu)^H \mathbf{H}_i \mathbf{w}_k^\nu - \left\langle \nabla_{\mathbf{w}_k^*} g_{i,2}(\mathbf{w}_k^\nu),\, \mathbf{w}_k - \mathbf{w}_k^\nu \right\rangle \\
& \ge g_{i,2}(\mathbf{w}_k)
\end{aligned} \tag{6}$$

with $\nabla_{\mathbf{w}_k^*} g_{i,2}(\mathbf{w}_k^\nu) = \mathbf{H}_i\mathbf{w}_k^\nu$ and $\langle \mathbf{a},\mathbf{b}\rangle \triangleq 2\,\mathrm{Re}\{\mathbf{a}^H\mathbf{b}\}$; and ii) upper bound $g_{i,1}(t,\beta_i)$ around $(t^\nu,\beta_i^\nu) \neq (0,0)$ as

$$\tilde{g}_{i,1}(t,\beta_i;t^\nu,\beta_i^\nu) \triangleq \frac{\mu_i}{2}\left(\frac{\beta_i^\nu}{t^\nu}\,t^2 + \frac{t^\nu}{\beta_i^\nu}\,\beta_i^2\right) \ge g_{i,1}(t,\beta_i). \tag{7}$$

Overall, this results in the following surrogate function which satisfies all the aforementioned conditions (and [16, Assumption 3]):

$$\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu) \triangleq \tilde{g}_{i,1}(t,\beta_i;t^\nu,\beta_i^\nu) + \tilde{g}_{i,2}(\mathbf{w}_k;\mathbf{w}_k^\nu). \tag{8}$$

Another example of valid approximation can be readily obtained using a different bound for the bilinear term $g_{i,1}(t,\beta_i)$ in (5). Rewriting $g_{i,1}(t,\beta_i)$ as the difference of two convex functions, $g_{i,1}(t,\beta_i) = \frac{\mu_i}{2}((t+\beta_i)^2 - (t^2+\beta_i^2))$, the desired convex upper bound of $g_{i,1}(t,\beta_i)$ can be obtained by linearizing the concave part of $g_{i,1}(t,\beta_i)$ around (t^ν,β_i^ν) while retaining the convex part, which leads to

$$\begin{aligned}
\widehat{g}_{i,1}(t,\beta_i;t^\nu,\beta_i^\nu) \triangleq & \frac{\mu_i}{2}\left((t+\beta_i)^2 - (t^\nu)^2 - (\beta_i^\nu)^2\right) \\
& -\mu_i\,\left(t^\nu\,(t-t^\nu) + \beta_i^\nu\,(\beta_i-\beta_i^\nu)\right).
\end{aligned} \tag{9}$$

The resulting valid surrogate function is then

$$\tilde{g}_i(t,\beta_i,\mathbf{w}_k;t^\nu,\beta_i^\nu,\mathbf{w}_k^\nu) \triangleq \widehat{g}_{i,1}(t,\beta_i;t^\nu,\beta_i^\nu) + \tilde{g}_{i,2}(\mathbf{w}_k;\mathbf{w}_k^\nu). \tag{10}$$

The INCA Algorithm: The proposed method, described in Algorithm 1, consists in solving the sequence of convexified subproblems (4) [using as surrogate functions $\tilde{g}_i$ either (8) or (10)], followed by a step-size procedure. Convergence is established in Theorem 3, whose proof can be found in [19, Th 10].

Algorithm 1: INCA Algorithm for Problem (1)

Data: $\gamma^\nu \in (0,1]$, $\mathbf{z}^0 \triangleq (t^0,\boldsymbol{\beta}^0,\mathbf{w}^0) \in \mathcal{Z}$, with $t^0 > 0$, and $(\tau_t,\tau_{\mathbf{w}},\tau_{\boldsymbol{\beta}}) > \mathbf{0}$. Set $\nu = 0$;

(S.1): If $\mathbf{z}^\nu \triangleq (t^\nu,\boldsymbol{\beta}^\nu,\mathbf{w}^\nu)$ is a stationary solution of (1): STOP;

(S.2): Compute the unique solution $\hat{\mathbf{z}}^\nu \triangleq (\hat{t}^\nu,\hat{\boldsymbol{\beta}}^\nu,\hat{\mathbf{w}}^\nu)$ of (4);

(S.3): Update $\mathbf{z} \triangleq (t,\boldsymbol{\beta},\mathbf{w})$: set $\mathbf{z}^{\nu+1} = \mathbf{z}^\nu + \gamma^\nu(\hat{\mathbf{z}}^\nu - \mathbf{z}^\nu)$;

(S.4): $\nu \leftarrow \nu + 1$ and go to step (S.1)

Theorem 3 *Let $\{\mathbf{z}^\nu = (t^\nu, \boldsymbol{\beta}^\nu, \mathbf{w}^\nu)\}$ be the sequence generated by Algorithm 1. Choose any τ_t, $\tau_{\mathbf{w}}$, $\tau_{\boldsymbol{\beta}} > 0$ and step-size sequence $\{\gamma^\nu\}$ such that $\gamma^\nu \in (0,1]$, $\gamma^\nu \to 0$, and $\sum_\nu \gamma^\nu = +\infty$. Then, $\{\mathbf{z}^\nu\}$ is bounded (with $t^\nu > 0$, for all $\nu \geq 1$), and every of its limit points $(\bar{t}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{w}})$ is a stationary solution of (3), such that $\bar{t} > 0$. Therefore, $\bar{\mathbf{w}}$ is a d-stationary solution of problem (1). Furthermore, if the algorithm does not stop after a finite number of steps, none of the $\bar{\mathbf{w}}$ above is a local minimum of U.*

Theorem 3 offers some flexibility in the choice of free parameters $(\tau_t, \tau_{\mathbf{w}}, \tau_{\boldsymbol{\beta}})$ and $\{\gamma^\nu\}_\nu$, while guaranteeing convergence of Algorithm 1. For instance, many choices are possible for $\{\gamma^\nu\}_\nu$ satisfying condition i); a practical rule that we found effective in our experiments is [20]: $\gamma^{\nu+1} = \gamma^\nu(1 - \varepsilon\gamma^\nu)$, with $\gamma^0 \in (0,1]$ and $\varepsilon \in (0, 1/\gamma^0)$. Also, one can relax the computation of the exact solution $(\hat{t}^\nu, \hat{\boldsymbol{\beta}}^\nu, \hat{\mathbf{w}}^\nu)$ of (4) and allow for inexact solutions [20]; we omit further details because of the space limitation.

3. THE MULTI-CELL CASE

Consider now a multicell multicast system comprising K BSs (cells), each equipped with N_t transmit antennas. For notational simplicity, we assume w.l.o.g. that each BS serves a single multicast group of single antenna users; let $\mathcal{G}_k$ denote the group of users served by the k-th BS, with $k \in \mathcal{K} \triangleq \{1, \cdots, K\}$; $\mathcal{G}_1, \ldots, \mathcal{G}_K$ is a partition of $\mathcal{K}$. The extension of the proposed algorithm to the multi-group case is straightforward. Denoting by $\mathbf{w}_k \in \mathbb{C}^{N_t}$ the beamforming vector for transmission at BS k, the coordinated multicell MMF beamforming problem is as follow (see, e.g., [10])

$$\begin{array}{ll} \underset{\mathbf{w}\triangleq(\mathbf{w}_k)_{k=1}^K}{\text{maximize}} & \underset{i\in\mathcal{G}_k,\, k\in\mathcal{K},}{\min} \dfrac{\mathbf{w}_k^H \mathbf{H}_{k,i,k}\mathbf{w}_k}{\sum_{\ell\neq k}\mathbf{w}_\ell^H \mathbf{H}_{\ell,i,k}\mathbf{w}_\ell + \sigma_{i,k}^2} \\ \text{subject to} & \|\mathbf{w}_k\|_2^2 \leq P_k, \ \forall k \in \mathcal{K}, \end{array} \tag{11}$$

where $\mathbf{H}_{\ell,i,k} = \mathbf{h}_{\ell,i,k}\mathbf{h}_{\ell,i,k}^H$ or $\mathbb{E}(\mathbf{h}_{\ell,i,k}\mathbf{h}_{\ell,i,k}^H) \succeq \mathbf{0}$ represents the instantaneous or long-term CSI matrix, respectively, with $\mathbf{h}_{\ell,i,k} \in \mathbb{C}^{N_t}$ denoting the channel vector from the ℓ-th BS to the i-th user in cell k; $\sigma_{i,k}^2$ is the variance of the AWGN at the user i in the cell k. Different QoS among users can be readily accommodated by mulitplying each SINR in (11) by a predetermined positive factor, which we will tacitly assume to be absorbed in the channel matrices $\mathbf{H}_{k,i,k}$.

Distributed algorithms: In the above multi-cell setting, our goal is to design *distributed* algorithms wherein each BS locally computes its own beamforming vector, with limited signaling among the other BSs. Following similar steps as in the single-cell case, it is not difficult to check that one can cast the computation of d-stationary solutions of (11) into the iterative solution of the following sequence of strongly convex subproblems: given the current iterate $(t^\nu, \boldsymbol{\beta}^\nu, \mathbf{w}^\nu)$, with $\boldsymbol{\beta}^\nu \triangleq ((\beta_{i,k})_{i\in\mathcal{G}_k})_{k\in\mathcal{K}} > \mathbf{0}$, $\mathbf{w} \triangleq (\mathbf{w}_k)_{k\in\mathcal{K}}$, and $t^\nu \neq 0$,

$$\begin{array}{lll} \underset{t\geq 0,\,\boldsymbol{\beta},\,\{\mathbf{w}_k\}_{k=1}^K,}{\text{maximize}} & \multicolumn{2}{l}{t - \dfrac{\tau_t}{2}(t - t^\nu)^2 - \tau_{\mathbf{w}}\|\mathbf{w} - \mathbf{w}^\nu\|_2^2 - \frac{\tau_{\boldsymbol{\beta}}}{2}\|\boldsymbol{\beta} - \boldsymbol{\beta}^\nu\|^2} \\ \text{subject to} & \text{(a):} & \tilde{g}_{i,k}(t, \beta_{i,k}, \mathbf{w}_k; t^\nu, \beta_{i,k}^\nu, \mathbf{w}_k^\nu) \leq 0, \ \forall i, k, \\ & \text{(b):} & \sum_{\ell\neq k}\mathbf{w}_\ell^H \mathbf{H}_{\ell,i,k}\mathbf{w}_\ell + \sigma_{i,k}^2 \leq \beta_{i,k}, \ \forall i, k, \\ & \multicolumn{2}{l}{\|\mathbf{w}_k\|_2^2 \leq P_k, \ \forall k \in \mathcal{K},} \end{array} \tag{12}$$

where the surrogate functions $\tilde{g}_{i,k}$ are chosen as [cf. (8)]

$$\begin{aligned} \widetilde{g}_{i,k}\left(t, \beta_{i,k}, \mathbf{w}_i; t^\nu, \beta_{i,k}^\nu, \mathbf{w}_i^\nu\right) &\triangleq \frac{1}{2}\left(\frac{\beta_{i,k}^\nu}{t^\nu}t^2 + \frac{t^\nu}{\beta_{i,k}^\nu}\beta_{i,k}^2\right) \\ &- (\mathbf{w}_k^\nu)^H \mathbf{H}_{k,i,k}\mathbf{w}_k^\nu - \langle \mathbf{H}_{k,i,k}\mathbf{w}_k^\nu, \mathbf{w}_k - \mathbf{w}_k^\nu\rangle. \end{aligned}$$

We denote by $\hat{\mathbf{z}}^\nu \triangleq (\hat{t}^\nu, \hat{\boldsymbol{\beta}}^\nu, \hat{\mathbf{w}}^\nu)$ the unique solution of (12). Using (12), we can readily apply Algorithm 1, where $\hat{\mathbf{z}}^\nu$ in Step 3 is now interpreted as the unique solution of (12); convergence to d-stationary solutions of (11) is still guaranteed under Theorem 3.

The above algorithm is however centralized, because the subproblems (12) do not decouple across the BSs. A distributed solution method for (12) can be obtained, exploiting the additive separability in the BSs' variables of the objective function and constraints in (12), as outlined next. Denoting by $\boldsymbol{\lambda} \triangleq ((\lambda_{i,k})_{i\in\mathcal{G}_k})_{k\in\mathcal{K}}$ and $\boldsymbol{\mu} \triangleq ((\mu_{i,k})_{i\in\mathcal{G}_k})_{k\in\mathcal{K}}$ the multipliers associated to the constraints (a) and (b) in (12), respectively, $\boldsymbol{\sigma}^2 \triangleq ((\sigma_{i,k}^2)_{i\in\mathcal{G}_k})_{k\in\mathcal{K}}$, and given $(t^\nu, \boldsymbol{\beta}^\nu, \mathbf{w}^\nu)$, the (partial) Lagrangian of (12) can be shown to have the following structure: $\mathcal{L}^\nu(t, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}_{(1)}^\nu(t, \boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{k=1}^K \mathcal{L}_{(2)}^\nu(\mathbf{w}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{k=1}^K \mathcal{L}_{(3)}^\nu(\boldsymbol{\beta}_k, \boldsymbol{\lambda}, \boldsymbol{\mu})$, where

$$\mathcal{L}_{(1)}^\nu(t, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq -t + \tfrac{\tau_t}{2}(t - t^\nu)^2 + \tfrac{\boldsymbol{\lambda}^T\boldsymbol{\beta}^\nu}{2t^\nu}t^2 + \boldsymbol{\mu}^T\boldsymbol{\sigma}^2;$$

$$\begin{aligned} \mathcal{L}_{(2)}^\nu(\mathbf{w}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) &\triangleq \tau_{\mathbf{w}}\|\mathbf{w}_k - \mathbf{w}_k^\nu\|^2 - \textstyle\sum_{i\in\mathcal{G}_k}\lambda_{i,k}(\mathbf{w}_k^\nu)^H \mathbf{H}_{k,i,k}\mathbf{w}_k^\nu \\ &+ \textstyle\sum_{i\in\mathcal{G}_k}\left[\sum_{\ell\neq k}\mu_{i,\ell}\mathbf{w}_k^H\mathbf{H}_{k,i,\ell}\mathbf{w}_k - \lambda_{i,k}\langle\mathbf{H}_{k,i,k}\mathbf{w}_k^\nu, \mathbf{w}_k - \mathbf{w}_k^\nu\rangle\right]; \end{aligned}$$

$$\mathcal{L}_{(3)}^\nu(\boldsymbol{\beta}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \tfrac{\tau_{\boldsymbol{\beta}}}{2}\|\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^\nu\|^2 - \boldsymbol{\mu}_k^T\boldsymbol{\beta}_k + \textstyle\sum_{i\in\mathcal{G}_k}\frac{\lambda_{i,k}t^\nu}{2\beta_{i,k}^\nu}\beta_{i,k}^2.$$

The above structure of the Lagrangian leads naturally to the following decomposition of the dual function

$$\begin{aligned} D^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \min_{t\geq 0}\mathcal{L}_{(1)}^\nu(t, \boldsymbol{\lambda}, \boldsymbol{\mu}; t^\nu) \\ &+ \sum_{k\in\mathcal{K}}\min_{\|\mathbf{w}_k\|_2^2\leq P_k}\mathcal{L}_{(2)}^\nu(\mathbf{w}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{k\in\mathcal{K}}\min_{\boldsymbol{\beta}_k\geq\mathbf{0}}\mathcal{L}_{(3)}^\nu(\boldsymbol{\beta}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}). \end{aligned}$$

The unique solutions of the above optimization problems can be computed in closed form:

$$\begin{aligned} \hat{t}^\nu(\boldsymbol{\lambda}) &\triangleq \underset{t\geq 0}{\text{argmin}}\,\mathcal{L}_{(1)}^\nu(t, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \left[\frac{1 + \tau_t \cdot t^\nu}{\tau_t + \boldsymbol{\lambda}^T\boldsymbol{\beta}^\nu/t^\nu}\right]_+, \\ \hat{\boldsymbol{\beta}}_k^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}) &\triangleq \underset{\boldsymbol{\beta}_k\geq\mathbf{0}}{\text{argmin}}\,\mathcal{L}_{(3)}^\nu(\boldsymbol{\beta}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \left(\left[\frac{\tau_{\boldsymbol{\beta}}\cdot\beta_{i,k}^\nu + \mu_{i,k}}{\tau_{\boldsymbol{\beta}} + \lambda_{i,k}\cdot t^\nu/\beta_{i,k}^\nu}\right]_+\right)_{i\in\mathcal{G}_k}, \\ \hat{\mathbf{w}}_k^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k^\star) &\triangleq \underset{\|\mathbf{w}_k\|_2^2\leq P_k}{\text{argmin}}\,\mathcal{L}_{(2)}^\nu(\mathbf{w}_k, \boldsymbol{\lambda}, \boldsymbol{\mu}) = (\xi_k^\star\mathbf{I} + \mathbf{A}_k)^{-1}\mathbf{b}_k^\nu, \end{aligned} \tag{13}$$

where $[x]_+ \triangleq \max(0, x)$; $\mathbf{A}_k \triangleq \tau_{\mathbf{w}}\mathbf{I} + \sum_{i\in\mathcal{G}_k}\sum_{\ell\neq k}\mu_{i,\ell}\mathbf{H}_{k,i,\ell}$; $\mathbf{b}_k^\nu \triangleq \left(\tau_{\mathbf{w}}\mathbf{I} + \sum_{i\in\mathcal{G}_k}\lambda_{i,k}\mathbf{H}_{k,i,k}\right)\mathbf{w}_k^\nu$; and $\xi_k^\star$, which is such that $0 \leq \xi_k^\star \perp \|\hat{\mathbf{w}}_k^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k^\star)\|^2 - P_k \leq 0$, can be computed as follows. Denoting by $\mathbf{U}_k\mathbf{D}_k\mathbf{U}_k^H$ the eigendecomposition of $\mathbf{A}_k$, we have $f_k(\xi_k) \triangleq \|\hat{\mathbf{w}}_k^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k)\|^2 - P_k = \sum_{j=1}^{N_t}\frac{[\mathbf{U}_k^H\mathbf{b}_k^\nu\mathbf{b}_k^{\nu H}\mathbf{U}_k]_{jj}}{(\xi_k + [\mathbf{D}_k]_{jj})^2} - P_k$. Therefore, $\xi_k^\star = 0$ if $f_k(0) < 0$; otherwise $\xi_k^\star$ is such that $f_k(\xi_k^\star) = 0$, which can be computed using bisection on $[0, \sqrt{\mathrm{tr}(\mathbf{U}_k^H\mathbf{b}_k^\nu\mathbf{b}_k^{\nu H}\mathbf{U}_k)/P_k} - \min_j[\mathbf{D}_k]_{jj}]$.

Finally, note that the dual function $D^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is differentiable with gradient given by: denoting $\mathbf{z}^\nu \triangleq (t^\nu, \boldsymbol{\beta}^\nu, \mathbf{w}^\nu)$,

$$\begin{aligned} \nabla_{\lambda_{i,k}}D^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \widetilde{g}_{i,k}\left(\hat{t}^\nu(\boldsymbol{\lambda}), \hat{\beta}^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}), \hat{\mathbf{w}}_k^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k^\star); \mathbf{z}^\nu\right), \\ \nabla_{\mu_{i,k}}D^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \sum_{\ell\neq k}\hat{\mathbf{w}}_\ell^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k^\star)^H\mathbf{H}_{\ell,i,k}\hat{\mathbf{w}}_\ell^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}, \xi_k^\star) \\ &\quad + \sigma_{i,k}^2 - \hat{\beta}_{i,k}^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu}), \end{aligned} \tag{14}$$

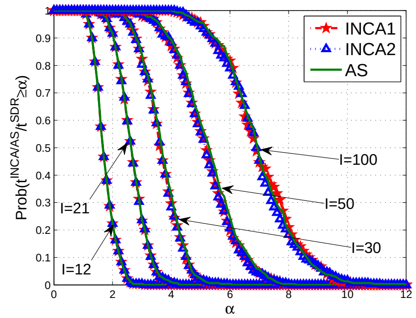
for all $i \in \mathcal{G}_k$ and $k \in \mathcal{K}$. Using (14), the dual problem $\max_{\boldsymbol{\lambda},\boldsymbol{\mu}\geq\mathbf{0}}D^\nu(\boldsymbol{\lambda}, \boldsymbol{\mu})$ can be then solved in a distributed way with convergence guarantee using, e.g., the gradient (or Newton) scheme with diminishing step-size; we omit further details.

The distributed algorithm is a double loop scheme: the inner loop deals with the update of the multipliers $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, given $\mathbf{z}^\nu \triangleq (t^\nu, \boldsymbol{\beta}^\nu, \mathbf{w}^\nu)$; whereas the outer loop consists in updating $\mathbf{z}^\nu$ via $\mathbf{z}^{\nu+1} = \mathbf{z}^\nu + \gamma^\nu (\hat{\mathbf{z}}^\nu - \mathbf{z}^\nu)$ wherein $\hat{\mathbf{z}}^\nu \triangleq \left(\hat{t^\nu}(\boldsymbol{\lambda}^\infty), \hat{\boldsymbol{\beta}^\nu}(\boldsymbol{\lambda}^\infty, \boldsymbol{\mu}^\infty),\right.$ $\left.\hat{\mathbf{w}}^\nu(\boldsymbol{\lambda}^\infty, \boldsymbol{\mu}^\infty)\right)$, with $(\boldsymbol{\lambda}^\infty, \boldsymbol{\mu}^\infty)$ denoting the solution from the inner loop. The inner and outer updates can be performed in a fairly distributed way among the cells. More specifically, let $\{(\boldsymbol{\lambda}^n, \boldsymbol{\mu}^n)\}$ denote the sequence generated by solving the dual problem; given $(\boldsymbol{\lambda}^n, \boldsymbol{\mu}^n)$, the BSs can compute $\hat{\mathbf{w}}_k^\nu(\boldsymbol{\lambda}^n, \boldsymbol{\mu}^n)$ and $\hat{\boldsymbol{\beta}}_k^\nu(\boldsymbol{\lambda}^n, \boldsymbol{\mu}^n)$ in parallel [cf. (13)]; to do so, they need only local information (within the cell). The update of $\hat{t^\nu}(\boldsymbol{\lambda}^n)$ and the multipliers $\left(\boldsymbol{\lambda}^{n+1}, \boldsymbol{\mu}^{n+1}\right)$ require some coordination among the BSs: it can be either carried out by a BS header or locally by all the BSs if a consensus-like scheme is employed.

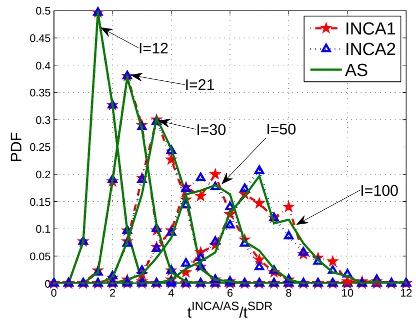
4. NUMERICAL RESULTS

In this section, we present some numerical results validating the proposed approach and algorithmic framework. We compare our INCA algorithm with the renowned SDR-G scheme in [4]. For INCA, we considered two instances, corresponding to the two alternative approximation strategies (7) and (9); we will term them "INCA1" and "INCA2", respectively. The setup of our experiment is the following. We simulated a single BS system; the transmitter is equipped with $N_t = 8$ transmit antennas and serves $K = 2$ multicast groups, each with $I_k \triangleq I$ single-antenna users, $\forall k \in \mathcal{K}$. Different numbers of users per group are considered, namely: $I = 12, 24, 30, 50, 100$. The proposed INCA algorithms are simulated using the step-size rule $\gamma^\nu = \gamma^{\nu-1}\left(1 - \epsilon\gamma^{\nu-1}\right)$ [16], with $\gamma^0 = 1$, where $\epsilon = 10^{-2}$; the proximal gain is set to $\tau = 10^{-5}$. The iterate is terminated when the absolute value of the difference of the objective function in two consecutive iterations is less than 10^{-3}. For the SDR-G in [4], 300 Gaussian samples are taken during the randomization phase, where the principal component of the relaxed SDP solution is also included as a candidate; the best value of the resulting objective function is denoted by t^{SDR}. To be fair, for the proposed INCA schemes, we considered 300 random feasible starting points and kept the best value of the objective function at convergence, denoted by t^{INCA}. We then compared the performance of the two algorithms in terms of the ratio $t^{\text{INCA}}/t^{\text{SDR}}$. As benchmark, we also report the results achieved using the standard nonlinear programming solver in Matlab, specifically the active-set algorithm in 'fmincon'; we refer to it as "AS" algorithm and denote by t^{AS} the best value of the objective function at convergence (obtained over the same random initializations of the INCA schemes). In Fig. 1 a) we plot the probability that $t^{\text{INCA/AS}}/t^{\text{SDR}} \geq \alpha$ versus α, for different values of I (number of users per group), and $\text{SNR} \triangleq P/\sigma^2 = 3\text{dB}$; this probability is estimated taking 300 independent channel realizations. The figures show a significant gain of the proposed INCA methods. For instance, when $I = 30$, the minimum achieved SINR of all INCA methods is about *at least* three times and *at most* 5 times the one achieved by SDR-G, with probability one. It seems that the gap tends to grow in favor of the INCA methods, as the number of users increases. In Fig. 1 b) we plot the distribution of $t^{\text{INCA/AS}}/t^{\text{SDR}}$. For instance, when $I = 30$, the minimum achieved SINR of all INCA methods is on average about four times the one achieved by SDR-G; the variance is about 1.

We observe that, while the proposed schemes compare favorably with the commercial off-the-shelf software (in terms of quality of the solution and convergence speed), they allow for a distributed implementation in a multi-cell scenario with convergence guarantees; off-the-shelf softwares instead lacks this important feature. In



(a)



(b)

Fig. 1. (a):Prob$(t^{\text{INCA/AS}}/t^{\text{SDR}} \geq \alpha)$ versus α, for $I = 12, 21, 30, 50, 100$; (b): Estimated p.d.f. of $t^{\text{INCA/AS}}/t^{\text{SDR}}$.

Fig. 2, we compare the (second-order) distributed implementation of INCA (cf. Sec. 3) with the centralized one. We simulated a system composed of $K = 4$ BSs, each of them equipped with $N_t = 4$ transmit antennas and serving one multicast group. Each group has $I = 3$ single-antenna users. In both loops (inner and outer), the iterate is terminated when the absolute value of the difference of the objective function in two consecutive iterations is less than 10^{-2}. Fig. 2 shows the evolution of the objective function t versus the iterations. For the distributed algorithm, the number of iterations counts both the inner and outer iterations. The algorithms converge quite fast to the same stationary point of (11).

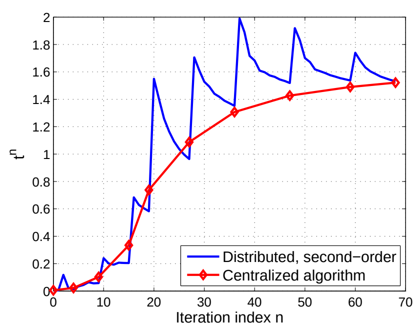


Fig. 2. Minimum rate vs. iterations.

5. CONCLUSIONS

We proposed the INCA algorithm, a new algorithmic framework for the *distributed* computation of d-stationary solutions of the multicell MMF multicast beamforming, with convergence guarantees. To the best of our knowledge, this is the first *provable distributed* algorithm for this class of problems. Numerical results show that our INCA schemes reach better solutions than renowned SDP relaxation-based approaches with high probability, while having similar computational complexity.

6. REFERENCES

[1] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (embms) in lte-advanced: Overview and rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.

[2] Sidiropoulos N. D., Davidson T. N., and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. on Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[3] B. Gopalakrishnan and N. D. Sidiropoulos, "High performance adaptive algorithms for single-group multicast beamforming," *IEEE Trans. on Signal Process.*, vol. 63, no. 16, pp. 4373–4384, Aug. 2015.

[4] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. on Signal Process.*, vol. 56, no. 3, pp. 1268 – 1279, Mar. 2008.

[5] A. B. Gershman, N. D. Sidiropoulos, S. ShahbazPanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming: From receive to transmit and network designs," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.

[6] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. on Signal Process.*, vol. 55, no. 10, pp. 4916–4927, Oct. 2007.

[7] G. Dartmann and G. Ascheid, "Equivalent quasi-convex form of the multicast max-min beamforming problem," *IEEE Trans. on Vehicular Technology*, vol. 62, no. 9, pp. 4643–4648, Nov. 2013.

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

[9] G.-W Hsu, H.-H Wang, H.-J Su, and P. Lin, "Joint beamforming for multicell multigroup multicast with per-cell power constraints," in *2014 IEEE 25th Annual Int. Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, Washington, DC, Sept. 2-5 2014, pp. 527 – 532.

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

[11] Z.-Q. Luo, W.-K. Ma, A. Man-Cho So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[12] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi, "Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem," *IEEE Trans. on Signal Process.*, vol. 56, no. 8, pp. 3932–3943, Aug. 2008.

[13] A. Schad and M. Pesavento, "Max-min fair transmit beamforming for multi-group multicasting," in *2012 International ITG Workshop on Smart Antennas (WSA)*, Dresden, Germany, March 7-8 2012, pp. 115 – 118.

[14] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup beamforming for per-antenna power constrained large-scale arrays," *arXiv:1503.06966*, Mar. 2015.

[15] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex qcqps," *IEEE Signal Process. Let.*, vol. 22, no. 7, pp. 804–808, Jul. 2015.

[16] G. Scutari, F. Facchinei, L. Lampariello, and P. Song, "Parallel and distributed methods for nonconvex optimization - Part I: Theory," *arXiv:1410.4754*, Oct. 2014.

[17] G. Scutari, F. Facchinei, L. Lampariello, and P. Song, "Parallel and distributed methods for nonconvex optimization," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 4-9 2014, pp. 840–844.

[18] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problem*, Springer-Verlag, New York, 2003.

[19] G. Scutari, F. Facchinei, L. Lampariello, P. Song, and S. Sardellitti, "Parallel and distributed methods for nonconvex optimization - Part II: Applications," *arXiv:1601.04059*, Jan. 2016.

[20] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multiuser systems," *IEEE Trans. on Signal Process.*, vol. 63, no. 3, pp. 641–656, Feb. 2014.