Multicell MISO Downlink Weighted Sum-Rate Maximization: A Distributed Approach

Pradeep Chathuranga Weeraddana, *Student Member, IEEE*, Marian Codreanu, *Member, IEEE*, Matti Latva-aho, *Senior Member, IEEE*, and Anthony Ephremides, *Life Fellow, IEEE*

***Abstract*—We develop an easy to implement distributed method for weighted sum-rate maximization (WSRMax) problem in a multicell multiple antenna downlink system. Unlike the recently proposed minimum weighted mean-squared error based algorithms, where at each iteration all mobile terminals needs to estimate the covariance matrices of their received signals, compute and feedback over the air certain parameters to the base stations (BS), our algorithm operates without any user terminal assistance. It requires only BS to BS signalling via reliable backhaul links (e.g., fiber, microwave links) and all required computation is performed at the BSs. The algorithm is based on primal decomposition and subgradient methods, where the original nonconvex problem is split into a master problem and a number of subproblems (one for each BS). A novel sequential convex approximation strategy is proposed to address the nonconvex master problem. In the case of subproblems, we adopt an existing iterative approach based on second-order cone programming and geometric programming. The subproblems are coordinated to find a (possibly suboptimal) solution to the master problem. Subproblems can be solved by BSs in a fully asynchronous manner, though the coordination between subproblems should be synchronous. Numerical results are provided to see the behavior of the algorithm under different degrees of BS coordination. They show that the proposed algorithm yields a good tradeoff between the implementation-level simplicity and the performance.**

***Index Terms*—Distributed optimization, geometric programming, primal decomposition, second-order cone programming, subgradient method, successive convex approximations, wireless networks.**

I. Introduction

The weighted sum-rate maximization (WSRMax) problem plays a central role in many network control and optimization methods, e.g., in [1]–[9] it is the basis for physical layer resource allocation. Unfortunately, in the case of wireless networks, the WSRMax problem is NP-hard [10]. Therefore, we have to rely on centralized and exponentially complex global optimization approaches [11], [12] for computing an exact solution. As a result, many optimal network design methods developed so far require a centralized implementation. However, finding even suboptimal but distributed methods for WSRMax is crucial for practical use.

Distributed implementation of WSRMax problem has been investigated in [13]–[17] in the context of digital subscriber loops (DSL) networks. Those systems are inherently consisting of single-input and single-output (SISO) links. Related algorithms for SISO wireless ad hoc networks and SISO orthogonal frequency division multiple access cellular systems are found in [18]–[21]. However, in the case of multi antenna cellular systems, the decision variables space is, of course, larger, e.g., joint optimization of transmit beamforming patterns, transmit powers, and link activations is required. Therefore, designing efficient distributed methods for WSRMax is a more challenging task due to the extensive amount of message passing required to resolve the coupling between variables.

Several distributed methods for WSRMax in multiple-input and single-output (MISO) cellular networks have been proposed in [22]–[28]. Specifically, in [22] a two-user MISO interference channel (IC)[1] is considered and a distributed algorithm is derived by using the commonly used high signal-to-interference-plus-noise ratio (SINR) approximation [29]. Moreover, another approximation, which relies on zero forcing (ZF) beamforming is introduced in [22] to address the problem in the case of multiuser MISO IC. Authors in [23] proposed a method based on a distributed pricing mechanism to address the problem. Both methods in [22], [23] are restricted to MISO IC (i.e., one user per cell) and are not applicable in the more general interfering broadcast channels, where there are many users per cell. The methods proposed in [24]–[26] derived the necessary (but not sufficient) optimality conditions for the WSRMax problem and used it as the basis for their distributed solution. However, many parameters must be selected heuristically to construct a potential distributed solution and there is in general no systematic method to find those parameters. In particular, the algorithms in [24], [25] are designed for systems with very limited backhaul signaling resources and do not consider any iterative base station (BS) coordination mechanism to resolve the out-of-cell interference coupling. Even though, the method proposed in [26] relies on stringent requirements on the message passing between BSs during each iteration of the algorithm, their results show that

[1] K-user MISO IC means that there are K transmitter-receiver pairs, where the transmitters have multiple antennas and the receivers have single antennas.

Manuscript received July 28, 2012; accepted September 25, 2012. Date of publication October 16, 2012; date of current version January 08, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wolfgang Utschick. This work was supported by EU projects Hycon2, Hydrobionets, and the VR project In network Optimization, the Finnish Funding Agency for Technology and Innovation (Tekes), Academy of Finland, Nokia, Nokia Siemens Networks, Elektrobit, Graduate School in Electronics, Telecommunications, and Automation (GETA) Foundations, Nokia Foundation, NSF Grant CCF-0728966, CCF-0905209, and US Army Research Office Grant W911NF-08-1-0238.

P. C. Weeraddana is with Automatic Control Laboratory, KTH Royal Institute of Technology, Stockholm 100 44, Sweden (e-mail: chatw@kth.se).

M. Codreanu and M. Latva-aho are with Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland (e-mail: codreanu@ee.oulu.fi; matti.latva-aho@ee.oulu.fi).

A. Ephremides is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: etony@ece.umd.edu).

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

Digital Object Identifier 10.1109/TSP.2012.2225060

BS coordination can provide considerable gains as compared to uncoordinated methods. An inexact cooperate descent algorithm for the case where each BS is serving only one cell edge user has been proposed in [27]. The method proposed in [28] considers a per data stream power constraint for simplicity, and thus their method does not apply in case of the more realistic power constraints at the BS, e.g., sum power constraint at the BS transmitter, per antenna power constraints. Centralized methods for WSRMax in multi antenna cellular networks are derived in [30]–[34].

Many optimization criteria other than the weighted sum-rate have been considered in references [35]–[43] to distributively optimize the system resources (e.g., beamforming patterns, transmit powers, etc.) in multi antenna cellular networks. In particular, the references [35]–[38] used the characterization of the Pareto boundary of the MISO interference channel [44] as the basis for their distributed methods. Their proposed methods do not employ any BS coordination mechanism to resolve the out-of-cell interference coupling. These algorithm can perform poorly, especially if the degrees of freedom available at BS transmitter is insufficient to avoid interference. The method proposed in [39] is designed for sum-rate maximization and uses high SINR approximation. A cooperative beamforming algorithm is proposed in [40] for MISO IC, where each BS can transmit only to a single user. Their proposed method employs an iterative BS coordination mechanism to resolve the out-of-cell interference coupling. However, the convexity properties exploited for distribution of the problem are destroyed when there are more than one user is served by any BS. In [41]–[43] distributed algorithms have been derived to minimize a total (weighted) transmitted power or the maximum per antenna power across the BSs subject to SINR constraints at the user terminals.

Recently, an interesting distributed algorithm for WSRMax is proposed by Shi *et al.* [45], which exploits a nontrivial equivalence between the WSRMax problem and a weighted sum mean squared error minimization problem. In the rest of the paper, we refer to this method as *WMMSE algorithm* as suggested in [45]. Each iteration of WMMSE algorithm essentially consists of the following three steps: 1) received signal covariance estimation at each user terminal, 2) computation and feedback of certain parameters from user terminals to BSs over *the air interface*, and 3) transmit beamformer adjustment at each BS. In practice, performing *perfect* covariance estimation and *perfect* feedback during each iteration can be very challenging. In the presence of user terminal imperfections, such as estimation and feedback errors, the algorithm's performance can degrade and its convergence can be less predictable.

In this paper we provide an alternative distributed method for WSRMax problem in a multicell MISO downlink system. Unlike the WMMSE algorithm [45], our method does not rely on user terminals' assistance such as estimations, computations, and feedback information to BSs over the air interface during iterations. The proposed method require only the BS to BS synchronized communication, where all the signalling overhead is exchanged through reliable backhaul links (e.g., fiber and microwave links). All the necessary computation can be carried out *asynchronously* at each BS without any involvement of the user terminals. Thus, our algorithm is well suited for systems where the user terminal support is not allowed or not desirable. Our algorithm is based on primal decomposition methods and subgradient methods [46]. Specifically, we first apply primal decomposition techniques to split the problem into a master problem and many subproblems. For master problem, we develop a novel sequential convex approximation strategy [47] together with a subgradient method that relies on BS coordinations. The master problem resolves the out-of-cell interference power, which is also known as the interference temperature in the context of cognitive radio networks [40]. In the case of subproblems, we adopt an existing algorithm originally proposed in [31, Sec. 4.3], which is based on second-order cone programming (SOCP) [48] and geometric programming (GP) [49]. These subproblems (or BS optimizations) can be carried out in a fully asynchronous manner. We show the monotonic convergence properties of the algorithm, with appropriate choice of the stopping criterion for the subgradient method. We also provide practical stopping criteria, which are favorable for implementing the algorithm, but at the expense of a sacrificing the monotone convergence. Numerical results are provided to compare our method with WMMSE algorithm [45], the GP/SOCP based algorithm proposed in [31, Sec. 4.3], and the distributed algorithm proposed in [24], [25]. The behavior of the algorithm under different degrees of BS coordination is also discussed and numerically illustrated. Preliminary results of this paper can be found in [50].

The rest of the paper is organized as follows. The system model and problem formulation are presented in Section II. In Section III we present the problem decomposition, where we develop a novel sequential convex approximation strategy for addressing the nonconvex master problem. Our proposed distributed algorithm is presented in Section IV. The numerical results are presented in Section V and Section VI concludes our paper.

Notations: All boldface lower case and upper case letters represent vectors and matrices respectively and calligraphy letters represent sets. We use $\mathbb{R}_+$ to denote the set of nonnegative real numbers. The set of complex numbers is denoted by $\mathbb{C}$, the set of complex n-vectors is denoted $\mathbb{C}^n$. $|x|$ denotes the absolute value of the complex number x, $\|\mathbf{x}\|_2$ denotes the ℓ_2-norm of the complex vector $\mathbf{x}$, and $\mathbf{vec}(\mathbf{X})$ denotes the vector obtained by stacking the columns of matrix $\mathbf{X}$. The identity matrix is denoted by $\mathbf{I}$. The superscript $(\cdot)^{\mathrm{H}}$ stands for Hermitian transpose, the superscript $(\cdot)^\star$ is used to denote a solution of an optimization problem, and $\mathrm{E}\{\cdot\}$ denotes statistical expectation. The notation $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \Sigma_{\mathbf{x}})$ indicates that $\mathbf{x}$ is complex Gaussian distributed with mean $\bar{\mathbf{x}}$ and covariance $\Sigma_{\mathbf{x}}$.

II. System Model and Problem Formulation

A multicell MISO downlink system, with N BSs each equipped with T transmit antennas is considered. The set of all BSs is denoted by $\mathcal{N}$ and we label them with the integer values $n = 1, \ldots, N$. The *transmission region* of each BS is modeled as a disc with radius R_{BS} centered at the location of the BS. Single data stream is transmitted for each user. We denote the set of all data streams in the system by $\mathcal{L}$ and label them with the integer values $l = 1, \ldots, L$. The transmitter node (i.e., the BS) of lth data stream is denoted by $tran(l)$ and the receiver node of lth data stream is denoted by $rec(l)$. We have $\mathcal{L} = \cup_{n \in \mathcal{N}} \mathcal{L}(n)$, where $\mathcal{L}(n)$ denotes the set of data

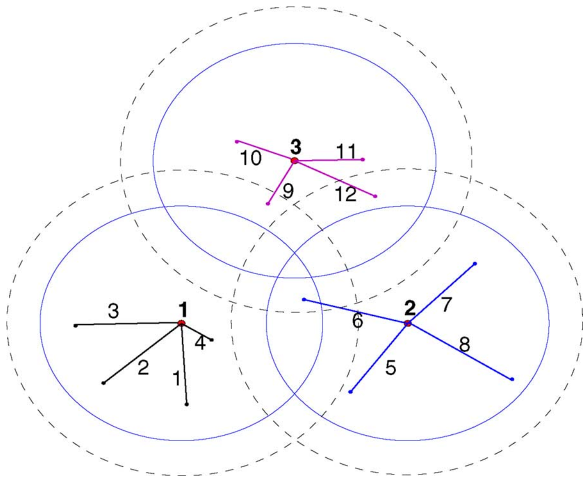

Fig. 1. Multicell network, $\mathcal{N} = \{1,2,3\}$, $\mathcal{L} = \{1,\ldots,12\}$, $\mathcal{L}(1) = \{1,\ldots,4\}$, $\mathcal{L}(2) = \{5,\ldots,8\}$, $\mathcal{L}(3) = \{9,\ldots,12\}$. The area inside solid-lined circles around BS 1, 2, and 3 represent the associated transmission regions of each BS and the area inside dash-lined circles around BSs represent the associated interference regions of each BS.

streams transmitted by nth BS. Note that the users of the data streams transmitted by each BS are necessarily located inside the transmission region of the BS (see Fig. 1).

The antenna signal vector transmitted by nth BS is given by

$$\mathbf{x}_n = \sum_{l\in\mathcal{L}(n)} \sqrt{p_l}\, d_l \mathbf{v}_l, \tag{1}$$

where $p_l \in \mathbb{R}_+$ denotes the power, $d_l \in \mathbb{C}$ represents the information symbol, and $\mathbf{v}_l \in \mathbb{C}^T$ is the beamformer, all associated to lth data stream. We assume that d_l and $\mathbf{v}_l$ are normalized such that $\mathrm{E}|d_l|^2 = 1$ and $\|\mathbf{v}_l\|_2 = 1$. Moreover, we assume independent data streams, i.e., $\mathrm{E}\{d_l d_j^*\} = 0$ for all $l,j \in \mathcal{L}$, where $l \neq j$.

The signal received at $rec(l)$ is given by

$$\begin{aligned} y_l =& \mathbf{h}_{ll}^{\mathrm{H}}\sqrt{p_l}d_l\mathbf{v}_l + \sum_{\substack{j\in\mathcal{L}(tran(l))\\ j\neq l}} \mathbf{h}_{jl}^{\mathrm{H}}\sqrt{p_j}d_j\mathbf{v}_j \\ &+ \sum_{j\in\mathcal{L}\setminus\mathcal{L}(tran(l))} \mathbf{h}_{jl}^{\mathrm{H}}\sqrt{p_j}d_j\mathbf{v}_j + z_l \qquad (2)\\ =& \mathbf{h}_{ll}^{\mathrm{H}}\sqrt{p_l}d_l\mathbf{v}_l + \sum_{\substack{j\in\mathcal{L}(tran(l))\\ j\neq l}} \mathbf{h}_{jl}^{\mathrm{H}}\sqrt{p_j}d_j\mathbf{v}_j \\ &+ \sum_{i\in\mathcal{N}\setminus\{tran(l)\}}\sum_{j\in\mathcal{L}(i)} \mathbf{h}_{jl}^{\mathrm{H}}\sqrt{p_j}d_j\mathbf{v}_j + z_l, \qquad (3)\end{aligned}$$

where $\mathbf{h}_{jl}^{\mathrm{H}} \in \mathbb{C}^{1\times T}$ is the channel matrix between $tran(j)$ and $rec(l)$, and z_l is circular symmetric complex Gaussian noise with variance σ_l^2. Note that the second term in (3) represents the intra-cell interference and the third term represents the out-of-cell interference. The received SINR of lth data stream is given by

$$\gamma_l = \frac{p_l\left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(tran(l)),j\neq l} p_j\left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_j\right|^2 + \sum_{i\in\mathcal{N}\setminus\{tran(l)\}} z_{il}}, \tag{4}$$

where $z_{il} = \sum_{j\in\mathcal{L}(i)} p_j|\mathbf{h}_{jl}^{\mathrm{H}}\mathbf{v}_j|^2$ represents the out-of-cell interference power from ith BS to $rec(l)$, which is typically known as *interference temperature* in the context of cognitive radio networks [40].

The out-of-cell interference term in (4) (i.e., $\sum_{i\in\mathcal{N}\setminus\{tran(l)\}} z_{il}$) prevents resource allocation (RA) on an intra-cell basis and demands centralized RA methods. To facilitate potential distributed algorithms for RA, we make the following assumption: transmissions from ith BS *do interfere* the lth data stream transmitted by BS $n \neq i$, if the distance between ith BS and $rec(l)$ is smaller than a threshold R_{int}.[2] The disc with radius R_{int} centered at the location of any BS is referred to as the *interference region* of the BS, see Fig. 1. Thus, if ith BS is located at a distance larger than R_{int} to $rec(l)$, then the associated z_{il} components are set to zero.[3] Based on the assumption above, we can express γ_l as

$$\gamma_l = \frac{p_l\left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(tran(l)),j\neq l} p_j\left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_j\right|^2 + \sum_{i\in\mathcal{N}_{\mathrm{int}}(l)} z_{il}}, \tag{5}$$

where $\mathcal{N}_{\mathrm{int}}(l) \subseteq \mathcal{N}\setminus\{tran(l)\}$ is the set of out-of-cell interfering BSs that are located at a distance less than R_{int} to $rec(l)$. For example, in Fig. 1 we have $\mathcal{N}_{\mathrm{int}}(9) = \{1\}, \mathcal{N}_{\mathrm{int}}(12) = \{2\}$, $\mathcal{N}_{\mathrm{int}}(6) = \{1,3\}$, and $\mathcal{N}_{\mathrm{int}}(l) = \emptyset$ for all $l \in \mathcal{L}\setminus\{6,9,12\}$. It is worth noting that the shape of the transmission and interference regions can be arbitrary closed contours around the BSs instead of the circles. This can mean arbitrary associations of users to BSs. However, without loss of generality, we can use disc model, which simplifies the presentation. Finally, it is useful to define the set $\mathcal{L}_{\mathrm{int}}$ of data streams that are subject to out-of-cell interference, i.e., $\mathcal{L}_{\mathrm{int}} = \{l \mid l \in \mathcal{L}, \mathcal{N}_{\mathrm{int}}(l) \neq \emptyset\}$. For example, in Fig. 1 we have $\mathcal{L}_{\mathrm{int}} = \{6,9,12\}$.

Let β_l be an arbitrary positive weight associated with lth data stream. We consider the case where all receivers are using *single-user detection* (i.e., a receiver decodes its intended signal by treating all other interfering signals as noise). Assuming that the power allocation is subject to a maximum power constraint $\sum_{l\in\mathcal{L}(n)} p_l\|\mathbf{v}_l\|_2 \leq p_n^{\max}$ for each BS $n \in \mathcal{N}$, the problem of WSRMax can be expressed as, see (6) at the bottom of the next page, where the variables are $\{p_l, \mathbf{v}_l\}_{l\in\mathcal{L}}$ and $\{z_{il}\}_{l\in\mathcal{L}_{\mathrm{int}}, i\in\mathcal{N}_{\mathrm{int}}(l)}$ and $\ln(\cdot)$ is the natural logarithm. The weights β_l, $l = 1,\ldots,L$ assign different priorities to different users. For example, in the context of physical layer resource allocation in optimal cross-layer control policies, β_l represents queue backlog associated with data stream l [2]. Note that we can simply replace the constraint $\sum_{l\in\mathcal{L}(n)} p_l\|\mathbf{v}_l\|_2^2 \leq p_n^{\max}$ with $\sum_{l\in\mathcal{L}(n)} p_l \leq p_n^{\max}$, because $\|\mathbf{v}_l\|_2 = 1$.

III. Problem Decomposition, Master Problem, and Subproblems

In this section, we develop the main building blocks required to derive the distributed algorithm for problem (6),

[2]Similar assumptions are made in [51] in the context of arbitrary wireless networks.

[3]The value of R_{int} is chosen such that the power of the interference term is below the noise level and this commonly used approximation is made to avoid unnecessary coordinations between distant BSs. The effect of nonzero z_{il} terms can be accurately modeled by changing the statistical characteristics of noise z_l at $rec(l)$. However, those issues are extraneous to the main focus of the paper.

namely, the master problem and the subproblems. To do this, we first break problem (6) into a master problem and N subproblems (one for each BS), by treating out-of-cell interference powers $\{z_{il}\}_{l\in\mathcal{L}_{\rm int}, i\in\mathcal{N}_{\rm int}(l)}$ as complicating variables. In the case of the master problem, we develop a novel sequential convex approximation strategy to circumvent the difficulties due to the inherent nonconvexity of problem (6). In the case of the subproblem, we adopt the method originally proposed in [31, Sec. 4.3], which is essentially based on SOCP and GP techniques.

A. Primal Decomposition

We start by first reformulating problem (6) as, see (7) at the bottom of the page, where the variables are $\{p_l, \mathbf{v}_l\}_{l\in\mathcal{L}}$ and $\{z_{il}\}_{l\in\mathcal{L}_{\rm int}, i\in\mathcal{N}_{\rm int}(l)}$. Problem (6) and (7) are equivalent, since 1) function $\ln(\cdot)$ is increasing and 2) the objective function of problem (7) is increasing in z_{il}, and therefore the first set of constraints holds with equality at the optimal point.

Let $\mathcal{L}_{\rm int}(n)$ denote the set of links for which base station n acts as an out-of-cell interferer. In particular, $\mathcal{L}_{\rm int}(n) = \{l | l \in \mathcal{L}_{\rm int}, n \in \mathcal{N}_{\rm int}(l)\}$. By noting that the sets $\{(l,i) | l \in \mathcal{L}_{\rm int}, i \in \mathcal{N}_{\rm int}(l)\}$ and $\{(l,n) | n \in \mathcal{N}, l \in \mathcal{L}_{\rm int}(n)\}$ are identical, we can rewrite the first inequality constraint of problem (7) as

$$z_{nl} \geq \sum_{j\in\mathcal{L}(n)} p_j \left|\mathbf{h}_{jl}^{\rm H}\mathbf{v}_j\right|^2, \quad n \in \mathcal{N}, \quad l \in \mathcal{L}_{\rm int}(n). \tag{8}$$

Now we treat z_{nl} as complicating variables and use primal decomposition techniques to split problem (7) into a master problem and N subproblems (one for each BS). The master problem updates the complicating variables $\{z_{nl}\}_{n\in\mathcal{N}, l\in\mathcal{L}_{\rm int}(n)}$ to maximize the overall weighed sum rate (i.e., to maximize the objective of original problem (6)). To express the master problem compactly, let us denote the vector $\{z_{nl}\}_{n\in\mathcal{N}, l\in\mathcal{L}_{\rm int}(n)}$ of out-of-cell interference components by $\mathbf{z}$. The master problem is given by

$$\begin{array}{ll} \text{minimize} & \sum_{n\in\mathcal{N}} f_n(\mathbf{z}) \\ \text{subject to} & \mathbf{z} \succeq \mathbf{0}, \end{array} \tag{9}$$

where the variable is $\mathbf{z}$ and $f_n(\mathbf{z})$ is the optimal value of the nth subproblem given by, see (10) at the bottom of the page, with variables $\{p_l, \mathbf{v}_l\}_{l\in\mathcal{L}(n)}$. To simplify the presentation, it is

$$\begin{array}{ll} \text{maximize} \sum_{n\in\mathcal{N}}\sum_{l\in\mathcal{L}(n)}\beta_l \ln\left(1 + \dfrac{p_l\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum\limits_{\substack{j\in\mathcal{L}(n)\\ j\neq l}} p_j\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_j\right|^2 + \sum\limits_{i\in\mathcal{N}_{\rm int}(l)} z_{il}}\right) & \\ \text{subject to} \quad z_{il} = \sum_{j\in\mathcal{L}(i)} p_j\left|\mathbf{h}_{jl}^{\rm H}\mathbf{v}_j\right|^2, \quad l \in \mathcal{L}_{\rm int}, \quad i \in \mathcal{N}_{\rm int}(l) & \\ \qquad \sum_{l\in\mathcal{L}(n)} p_l \|\mathbf{v}_l\|_2^2 \leq p_n^{\max}, \quad n \in \mathcal{N} & \\ \qquad \|\mathbf{v}_l\|_2 = 1, \quad p_l \geq 0, \quad l \in \mathcal{L}. & \end{array} \tag{6}$$

$$\begin{array}{ll} \text{minimize} - \sum_{n\in\mathcal{N}}\sum_{l\in\mathcal{L}(n)}\beta_l \ln\left(1 + \dfrac{p_l\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum\limits_{\substack{j\in\mathcal{L}(n)\\ j\neq l}} p_j\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_j\right|^2 + \sum\limits_{i\in\mathcal{N}_{\rm int}(l)} z_{il}}\right) & \\ \text{subject to} \quad z_{il} \geq \sum_{j\in\mathcal{L}(i)} p_j\left|\mathbf{h}_{jl}^{\rm H}\mathbf{v}_j\right|^2, \quad l \in \mathcal{L}_{\rm int}, \quad i \in \mathcal{N}_{\rm int}(l) & \\ \qquad \sum_{l\in\mathcal{L}(n)} p_l \leq p_n^{\max}, \ n \in \mathcal{N} & \\ \qquad \|\mathbf{v}_l\|_2 = 1, \quad p_l \geq 0, \ l \in \mathcal{L}. & \end{array} \tag{7}$$

$$\begin{array}{ll} \text{minimize} - \sum_{l\in\mathcal{L}(n)}\beta_l \ln\left(1 + \dfrac{p_l\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum\limits_{\substack{j\in\mathcal{L}(n)\\ j\neq l}} p_j\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_j\right|^2 + \sum\limits_{i\in\mathcal{N}_{\rm int}(l)} z_{il}}\right) & \\ \text{subject to} \quad z_{nl} \geq \sum_{j\in\mathcal{L}(n)} p_j\left|\mathbf{h}_{jl}^{\rm H}\mathbf{v}_j\right|^2, \quad l \in \mathcal{L}_{\rm int}(n) & \\ \qquad \sum_{l\in\mathcal{L}(n)} p_l \leq p_n^{\max} & \\ \qquad \|\mathbf{v}_l\|_2 = 1, \quad p_l \geq 0, \quad l \in \mathcal{L}(n). & \end{array} \tag{10}$$

also useful to introduce the following equivalent reformulation of problem (10):

$$\begin{aligned}
\text{minimize} \quad & -\sum_{l\in\mathcal{L}(n)} \beta_l \ln(1+\gamma_l) \\
\text{subject to} \quad & \gamma_l \le \frac{p_l \left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} p_j \left|\mathbf{h}_{ll}^{\mathrm{H}}\mathbf{v}_j\right|^2 + \sum_{i\in\mathcal{N}_{\mathrm{int}}(l)} z_{il}}, \\
& l\in\mathcal{L}(n) \\
& z_{nl} \ge \textstyle\sum_{j\in\mathcal{L}(n)} p_j \left|\mathbf{h}_{jl}^{\mathrm{H}}\mathbf{v}_j\right|^2, \quad l\in\mathcal{L}_{\mathrm{int}}(n) \\
& \textstyle\sum_{l\in\mathcal{L}(n)} p_l \le p_n^{\max} \\
& \|\mathbf{v}_l\|_2 = 1, \quad p_l\ge 0, \quad l\in\mathcal{L}(n),
\end{aligned} \tag{11}$$

where the variable is $\{p_l,\gamma_l,\mathbf{v}_l\}_{l\in\mathcal{L}(n)}$. The equivalence of problem (10) and (11) follows since the objective function of problem (11) is decreasing in γ_l, and therefore the first set of constraints holds with equality at the optimal point.

B. Master Problem

Computing the objective value of the master problem (9) requires the solution of each subproblem (10), which is NP-hard [10]. Moreover, even if we would be able to *solve* the subproblems, we *cannot* apply standard subgradient methods to solve the master problem (9) since it is *not convex*. To address these difficulties, we develop a novel method that solves successive approximated variants of the original master problem (9). Each approximated problem can be transformed into a convex problem by a change of variables. To solve the resulting convex problems, we propose a subgradient method. It is important to note that, the approximations and variable transformations mentioned above are such that we can always rely on subproblems (10) (i.e., BS optimizations) to compute a subgradient. Details of the subproblem solution method are deferred to Section III-C.

We start by approximating the objective function of problem (9) with an upper bound function, which in turn is used to obtain the approximation of the master problem. We refer to the resulting approximation as *the approximated master* problem. Next, we derive an equivalent convex form of the approximated master problem, followed by the subgradient methods to solve it.

1) Derivation of an Upper Bound Function for the Master Problem: The key idea is as follows: we first carry out partial minimization of problem (11) to yield an initial upper bound on $f_n(\mathbf{z})$.[4] Then the initial upper bound is further modified by using a well known monomial approximation, so that convex optimization techniques can be readily employed.

To simplify the presentation, let $\mathcal{H}$ denote the feasible set of problem (11). For some fixed normalized $\check{\mathbf{v}}_l$, let $\check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}) = \{(p_l,\gamma_l)_{l\in\mathcal{L}(n)} | (p_l,\gamma_l,\check{\mathbf{v}}_l)_{l\in\mathcal{L}(n)} \in \mathcal{H}\}$. Now we can write the following relations: see (12)–(17) shown at the bottom of the page. The first equality follows from the definition of $f_n(\mathbf{z})$ and from the equivalence of problems (10) and (11), (13) follows from partial minimization of the function over $\{p_l,\gamma_l\}_{l\in\mathcal{L}(n)}$ while $\{\mathbf{v}_l\}_{l\in\mathcal{L}(n)}$ being fixed such that $\{\mathbf{v}_l=\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}$, (14) follows trivially from the properties of $\ln(\cdot)$ function, (15) follows from the monomial lower bound on $1+\gamma_l$, i.e., $1+\gamma_l \ge \check{\gamma}_l^{\frac{-\check{\gamma}_l}{1+\check{\gamma}_l}}(1+\check{\gamma}_l)\gamma_l^{\frac{\check{\gamma}_l}{1+\check{\gamma}_l}}$, where $\check{\gamma}_l$ is an arbitrary positive number[5] [52, Lem. 1], (16) follows from the

[4]The minimum value of a function with respect to the all set of variables is always better than the minimum value of the function with respect to a subset of variables while others being fixed.

[5]This bound is typically used in conjunction with an iterative method, which uses local approximations. The parameter $\check{\gamma}_l$ is usually the point at which the approximation is made.

$$\begin{aligned}
f_n(\mathbf{z}) &= \inf_{(p_l,\gamma_l,\mathbf{v}_l)_{l\in\mathcal{L}(n)}\in\mathcal{H}} -\sum_{l\in\mathcal{L}(n)} \beta_l \ln(1+\gamma_l) && (12)\\
&\le \inf_{(p_l,\gamma_l)_{l\in\mathcal{L}(n)}\in\check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)})} -\sum_{l\in\mathcal{L}(n)} \beta_l \ln(1+\gamma_l) && (13)\\
&= \inf_{(p_l,\gamma_l)_{l\in\mathcal{L}(n)}\in\check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)})} \ln\left(\prod_{l\in\mathcal{L}(n)} (1+\gamma_l)^{-\beta_l}\right) && (14)\\
&\le \inf_{(p_l,\gamma_l)_{l\in\mathcal{L}(n)}\in\check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)})} \ln \prod_{l\in\mathcal{L}(n)} \left(\check{\gamma}_l^{\frac{-\check{\gamma}_l}{1+\check{\gamma}_l}}(1+\check{\gamma}_l)\gamma_l^{\frac{\check{\gamma}_l}{1+\check{\gamma}_l}}\right)^{-\beta_l} && (15)\\
&= \ln \underbrace{\inf_{(p_l,\gamma_l)_{l\in\mathcal{L}(n)}\in\check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)})} \left(\prod_{l\in\mathcal{L}(n)} \left(\check{\gamma}_l^{-\frac{\check{\gamma}_l}{1+\check{\gamma}_l}}(1+\check{\gamma}_l)\gamma_l^{\frac{\check{\gamma}_l}{1+\check{\gamma}_l}}\right)^{-\beta_l}\right)}_{\check{f}_n(\mathbf{z})} && (16)\\
&= \ln\left(\check{f}_n(\mathbf{z})\right). && (17)
\end{aligned}$$

monotonic properties of $\ln(\cdot)$, and $\check{f}_n(\mathbf{z})$ is the optimal value of the following problem:[6]

$$\begin{aligned}
&\text{minimize} \prod_{l\in\mathcal{L}(n)} \left(\check{\gamma}_l^{-\check{\gamma}_l/(1+\check{\gamma}_l)}(1+\check{\gamma}_l)\right)^{-\beta_l} \prod_{l\in\mathcal{L}(n)} \gamma_l^{-\beta_l \frac{\check{\gamma}_l}{1+\check{\gamma}_l}} \\
&\text{subject to } \gamma_l \le \frac{p_l \left|\mathbf{h}_{ll}^{\mathrm{H}}\check{\mathbf{v}}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} p_j \left|\mathbf{h}_{ll}^{\mathrm{H}}\check{\mathbf{v}}_j\right|^2 + \sum_{i\in\mathcal{N}_{\mathrm{int}}(l)} z_{il}}, \\
&\qquad l \in \mathcal{L}(n)\setminus\mathcal{L}_{\mathrm{local}}(n) \\
&\qquad \gamma_l \le \frac{p_l \left|\mathbf{h}_{ll}^{\mathrm{H}}\check{\mathbf{v}}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} p_j \left|\mathbf{h}_{ll}^{\mathrm{H}}\check{\mathbf{v}}_j\right|^2}, \ l\in\mathcal{L}_{\mathrm{local}}(n) \\
&\qquad z_{nl} \ge \textstyle\sum_{j\in\mathcal{L}(n)} p_j \left|\mathbf{h}_{jl}^{\mathrm{H}}\check{\mathbf{v}}_j\right|^2, \ l\in\mathcal{L}_{\mathrm{int}}(n) \\
&\qquad \textstyle\sum_{l\in\mathcal{L}(n)} p_l \le p_n^{\max} \\
&\qquad p_l \ge 0, \ l\in\mathcal{L}(n),
\end{aligned} \tag{18}$$

where the variable is $\{p_l, \gamma_l\}_{l\in\mathcal{L}(n)}$ and $\mathcal{L}_{\mathrm{local}}(n)$ is the subset of data streams transmitted by nth BS, which are not interfered by any out-of-cell interference, i.e., $\mathcal{L}_{\mathrm{local}}(n) = \{l | \ l \in \mathcal{L}(n), \mathcal{N}_{\mathrm{int}}(l) = \emptyset\}$. Note that, the inequality (13) holds with equality if the *optimal* normalized beamforming directions of problem (11) is identical to $\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}$ and the inequality (15) holds with equality if $\{\gamma_l = \check{\gamma}_l\}_{l\in\mathcal{L}(n)}$.

From (12)—(17) we have $f_n(\mathbf{z}) \le \ln(\check{f}_n(\mathbf{z}))$, which holds for all $n\in\mathcal{N}$. Thus we have

$$\sum_{n\in\mathcal{N}} f_n(\mathbf{z}) \le \sum_{n\in\mathcal{N}} \ln\left(\check{f}_n(\mathbf{z})\right), \tag{19}$$

which gives an upper bound on the objective function of (9). The approximated master problem is obtained by replacing the objective function of the original master problem (9) by the upper bound function given in (19), i.e.,

$$\begin{aligned}
&\text{minimize} \quad \sum_{n\in\mathcal{N}} \ln\left(\check{f}_n(\mathbf{z})\right) \\
&\text{subject to} \quad \mathbf{z}\succeq \mathbf{0},
\end{aligned} \tag{20}$$

where the variables is $\mathbf{z}$. Though Problem (20) is not convex in its current form, it can be equivalently reformulated into a convex problem via a variable transformation as shown in the next section.

2) Convex Reformulation of the Approximated Master Problem: Let us first transform problem (20) by the logarithmic change of variables $\bar{z}_{il} = \ln z_{il}$ (so $z_{il} = e^{\bar{z}_{il}}$). This yields the problem

$$\text{minimize} \quad \sum_{n\in\mathcal{N}} \ln\left(\check{f}_n(e^{\bar{\mathbf{z}}})\right), \tag{21}$$

where the variable is $\bar{\mathbf{z}} = \{\bar{z}_{il}\}_{l\in\mathcal{L}_{\mathrm{int}}, i\in\mathcal{N}_{\mathrm{int}}(l)}$. Here we use the notation $e^{\mathbf{y}}$, where $\mathbf{y}$ is a vector, to mean componentwise exponentiation: $[e^{\mathbf{y}}]_k = e^{y_k}$.

Next we show that problem (21) is convex. To see this, we capitalize on perturbation and sensitivity analysis results for convex optimization problems [53]–[55].[7] In particular, we apply perturbation results to the convex form of GP (18). To do this, let us first perform the logarithmic change of variables $\bar{p}_l = \ln p_l$, $\bar{\gamma}_l = \ln\gamma_l$, logarithmic change of parameters $\bar{z}_{il} = \ln z_{il}$, and a logarithmic transformation of the objective and constraint functions of GP (18) to get its convex form: see (22) at the bottom of the page, where the variable is $\{\bar{p}_l, \bar{\gamma}_l\}_{l\in\mathcal{L}(n)}$ and $g_{jl} = |\mathbf{h}_{jl}^{\mathrm{H}}\check{\mathbf{v}}_j|^2$. Problem (22) possesses the following key features:

a. Since the optimal value of GP (18) is $\check{f}_n(\mathbf{z})$, the optimal value of problem (22) is given by $\ln(\check{f}_n(e^{\bar{\mathbf{z}}}))$.

[6]Here we have explicitly characterized the constraint $(p_l, \gamma_l)_{l\in\mathcal{L}(n)} \in \check{\mathcal{H}}(\{\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)})$.

[7]Basic sensitivity results are documented in [53, Sec. 5.6] and more general results can be found in [54, Chap. 2] and [55, Sec. 5.6].

$$\begin{aligned}
&\text{minimize} \sum_{l\in\mathcal{L}(n)} \frac{\beta_l\check{\gamma}_l}{1+\check{\gamma}_l}\bar{\gamma}_l + \ln\prod_{l\in\mathcal{L}(n)}\left(\check{\gamma}_l^{-\frac{\check{\gamma}_l}{1+\check{\gamma}_l}}(1+\check{\gamma}_l)\right)^{-\beta_l} \\
&\text{subject to } \ln\left(g_{ll}^{-1}e^{\bar{\gamma}_l-\bar{p}_l}\left(\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} g_{jl}e^{\bar{p}_j} + \sum_{i\in\mathcal{N}_{\mathrm{int}}(l)} e^{\bar{z}_{il}}\right)\right) \le 0, \quad l\in\mathcal{L}(n)\setminus\mathcal{L}_{\mathrm{local}}(n) \\
&\qquad \ln\left(g_{ll}^{-1}e^{\bar{\gamma}_l-\bar{p}_l}\left(\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} g_{jl}e^{\bar{p}_j}\right)\right) \le 0, \quad l\in\mathcal{L}_{\mathrm{local}}(n) \\
&\qquad \ln\left(\sum_{j\in\mathcal{L}(n)} g_{jl}e^{-\bar{z}_{nl}}e^{\bar{p}_j}\right) \le 0, \quad l\in\mathcal{L}_{\mathrm{int}}(n) \\
&\qquad \ln\left(\sum_{l\in\mathcal{L}(n)} (p_n^{\max})^{-1}e^{\bar{p}_l}\right) \le 0.
\end{aligned} \tag{22}$$

b. Objective function of problem (22) is jointly convex in $\{\bar{p}_l, \bar{\gamma}_l\}_{l\in\mathcal{L}(n)}$ *and* $\bar{\mathbf{z}}$.
c. The constraint functions of problem (22) become jointly convex in $\{\bar{p}_l, \bar{\gamma}_l\}_{l\in\mathcal{L}(n)}$ *and* $\bar{\mathbf{z}}$.

By using the perturbation and sensitivity result given in [55, Lem. 1] it follows that $\ln(\check{f}_n(e^{\bar{\mathbf{z}}}))$ is convex in $\bar{\mathbf{z}}$. Consequently, problem (21) is convex.

3) Subgradient Method to Solve the Convex Form of the Approximated Master Problem: In this subsection, we derive the subgradient method for solving problem (21). By invoking [55, Lem. 1], we can compute a subgradient of $\sum_{n\in\mathcal{N}} \ln(\check{f}_n(e^{\bar{\mathbf{z}}}))$ at $\bar{\mathbf{z}}$. Specifically, a subgradient is given by $\sum_{n\in\mathcal{N}} \{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\rm int}, i\in\mathcal{N}_{\rm int}(l)}$ and see (23) at the bottom of the page, where $\{\lambda_l^\star(e^{\bar{\mathbf{z}}})\}_{l\in\mathcal{L}(n)\setminus\mathcal{L}_{\rm local}(n)}$ denotes the optimal Lagrange multipliers associated with the first set of constraints of problem (22), $\{\mu_l^\star(e^{\bar{\mathbf{z}}})\}_{l\in\mathcal{L}_{\rm int}(n)}$ denotes the optimal Lagrange multipliers associated with the third set of constraints of (22), and $\{\bar{p}_l^\star(e^{\bar{\mathbf{z}}}), \bar{\gamma}_l^\star(e^{\bar{\mathbf{z}}})\}_{l\in\mathcal{L}(n)}$ denotes the optimal solution of problem (22). Each BS n can compute $\{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\rm int}, i\in\mathcal{N}_{\rm int}(l)}$ independently, which in turn are used to construct the subgradient of $\sum_{n\in\mathcal{N}} \ln(\check{f}_n(e^{\bar{\mathbf{z}}}))$ at $\bar{\mathbf{z}}$ via BS-BS coordination. Note that the *zero* in (23) are used to simplify the presentation. In practice, these zeros need *not* be exchanged between BSs during their coordinations.

The subgradient method for problem (21) is given by [46]

$$\bar{z}_{il}^{(j+1)} = \bar{z}_{il}^{(j)} - \theta^{(j)} \sum_{n\in\mathcal{N}} d_{il}^n(\bar{\mathbf{z}}^{(j)}),\ l\in\mathcal{L}_{\rm int},\ i\in\mathcal{N}_{\rm int}(l) \quad (24)$$

$$= \bar{z}_{il}^{(j)} - \theta^{(j)} \left(d_{il}^{i}\left(\bar{\mathbf{z}}^{(j)}\right) + d_{il}^{tran(l)}\left(\bar{\mathbf{z}}^{(j)}\right)\right), \quad l\in\mathcal{L}_{\rm int},\ i\in\mathcal{N}_{\rm int}(l), \quad (25)$$

where j is the current iteration index of the subgradient method and $\theta^{(j)} \in \mathbb{R}_+$ is a step size.[8] The second equality (25) follows from (23) after ignoring the zero elements. This suggest that, for computing the (i,l)th component of the subgradient, only *two* BSs (i.e., i and $tran(l)$) need to coordinate.

C. Subproblem: BS Optimization

Note that subproblem (11) is NP-hard [10], and therefore any practical solution method is reliant on approximations. The subproblem solution method presented in this section is essentially based on the Algorithm 4.3.1 originally proposed in [31, Sec. 4.3]. Here we briefly discuss the key idea of this algorithm for the sake of completeness.

The key idea of the algorithm is to carry out the optimization with respect to different subsets of variables by considering others fixed [31, Sec. 4.3]. First, by fixing the beamformers $\{\mathbf{v}_l\}_{l\in\mathcal{L}(n)}$, a GP of the form (18) is solved which locally approximates the original subproblem (11). This is a decent step. Then, for fixed $\{\gamma_l\}_{l\in\mathcal{L}(n)}$ values, a maximum power reduction factor $t^\star$ is found such that the SINR values are preserved. The maximum power reduction factor is given by the optimum $t^\star$ that solves the following problem:

$$\begin{aligned} &\text{minimize} \quad t \\ &\text{subject to}\ \gamma_l \le \frac{p_l\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_l\right|^2}{\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} p_j\left|\mathbf{h}_{ll}^{\rm H}\mathbf{v}_j\right|^2 + \sum_{i\in\mathcal{N}_{\rm int}(l)} z_{il}}, \\ &\qquad l\in\mathcal{L}(n) \\ &\qquad t^2 z_{nl} \ge \textstyle\sum_{j\in\mathcal{L}(n)} p_j\left|\mathbf{h}_{jl}^{\rm H}\mathbf{v}_j\right|^2, \quad l\in\mathcal{L}_{\rm int}(n) \\ &\qquad \textstyle\sum_{l\in\mathcal{L}(n)} p_l\|\mathbf{v}_l\|_2^2 \le t^2 p_n^{\max} \\ &\qquad \|\mathbf{v}_l\|_2 = 1, \quad p_l\ge 0, \quad l\in\mathcal{L}(n), \end{aligned} \quad (26)$$

where the variables are t and $\{p_l, \mathbf{v}_l\}_{l\in\mathcal{L}(n)}$.[9] Note that, we always have $t^\star \le 1$, and, hence, the saved power can be used to decrease the objective of original problem (11) by 1) setting $\{\mathbf{v}_l = \mathbf{v}_l^\star\}_{l\in\mathcal{L}(n)}$ and $\{p_l = p_l^\star/t^\star\}_{l\in\mathcal{L}(n)}$ and 2) increasing $\{\gamma_l\}_{l\in\mathcal{L}(n)}$ until the SINR constraints become tight. The result is again a descent step. The discussion above leads to the following descent algorithm which can be asynchronously solved by nth BS:

Algorithm 1: Finding a suboptimal solution for BS optimization problem (11) [31, Sec. 4.3]

1. Initialization; given a feasible beamformer configuration $\{\mathbf{v}_l^{(0)}\}_{l\in\mathcal{L}(n)}$, a feasible power allocation $\{p_l^{(0)}\}_{l\in\mathcal{L}(n)}$, and $\mathbf{z}$. Set iteration index $i=0$.
2. By setting $p_l = p_l^{(i)}$ and $\mathbf{v}_l = \mathbf{v}_l^{(i)}$, compute $\check{\gamma}_l$ for all $l\in\mathcal{L}(n)$ from (5).
3. By setting $\check{\mathbf{v}}_l = \mathbf{v}_l^{(i)}$ for all $l\in\mathcal{L}(n)$, solve problem (18). Denote the solution by $\{p_l^\star, \gamma_l^\star\}_{l\in\mathcal{L}(n)}$ and the optimal Lagrange multipliers by $\{\lambda_l^\star\}_{l\in\mathcal{L}(n)\setminus\mathcal{L}_{\rm local}(n)}$ and $\{\mu_l^\star\}_{l\in\mathcal{L}_{\rm int}(n)}$.
4. Stopping criterion; if the stopping criterion is satisfied STOP by returning $d_{il}^n(\cdot)$ by using (23) and the suboptimal solution $\{\check{p}_l, \check{\gamma}_l, \check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}$, where $\check{p}_l = p_l^\star$. Otherwise, update achieved SINR values $\gamma_l^{\rm tmp} = \gamma_l^\star$ for all $l\in\mathcal{L}(n)$.
5. By setting $\gamma_l = \gamma_l^{\rm tmp}$ for all $l\in\mathcal{L}(n)$, solve problem (26). Denote the solution by $t^\star$ and $\{p_l^\star, \mathbf{v}_l^\star\}_{l\in\mathcal{L}(n)}$. Update $p_l^{(i+1)} = p_l^\star/(t^\star)^2$ and $\mathbf{v}_l^{(i+1)} = \mathbf{v}_l^\star$ for all $l\in\mathcal{L}(n)$. Set $i=i+1$ and go to step 2.

[8]We chose diminishing nonsummable step lengths (i.e., $\theta^{(j)} = 1/j$), that guarantees the asymptotic convergence of the subgradient method [46].

[9]It is well known that problem (26) is equivalently formulated as a SOCP, see [31, Sec. 4.3].

$$d_{il}^n(\bar{\mathbf{z}}) = \begin{cases} \dfrac{\lambda_l^\star(e^{\bar{\mathbf{z}}}) e^{\bar{z}_{il}}}{\sigma_l^2 + \sum_{j\in\mathcal{L}(n), j\neq l} g_{jl} e^{\bar{p}_j^\star(e^{\bar{\mathbf{z}}})} + \sum_{m\in\mathcal{N}_{\rm int}(l)} e^{\bar{z}_{ml}}} & l\in\mathcal{L}(n)\setminus\mathcal{L}_{\rm local}(n), \quad i\in\mathcal{N}_{\rm int}(l) \\ -\mu_l^\star(e^{\bar{\mathbf{z}}}) & l\in\mathcal{L}_{\rm int}(n), \quad i=n \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

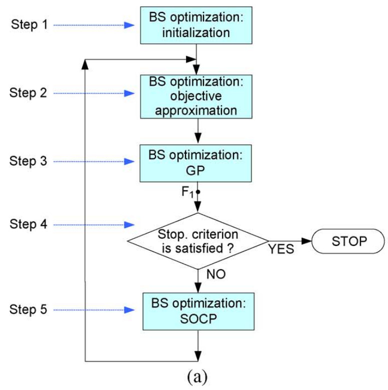

Fig. 2. Block diagrams of proposed algorithms. (a) Algorithm 1; (b) Algorithm 2.

The block diagram shown in Fig. 2(a) summarizes *Algorithm 1*. It is a descent algorithm and we refer the reader to [31] for more details.

Note that, step 3 of *Algorithm 1* solves problem (18) for some normalized $\check{\mathbf{v}}_l$. This is *the problem* that should be solved to find $d_{il}^n(\bar{\mathbf{z}})$ given in (23), which is then used to compute a subgradient $\sum_{n\in\mathcal{N}}\{d_{il}^n(\bar{\mathbf{v}})\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ for the objective of the approximated master problem (21). The observations above suggest that the local BS optimizations (i.e., *Algorithm 1*) can be employed to compute the subgradient in a distributed fashion. Specifically, the dual variables and the optimal solutions required to compute the subgradient elements $d_{il}^n(\bar{\mathbf{z}})$ are obtained as a by-product of the BS optimization process. These are, of course, desirable and favorable features that are exploited when developing our distributed WSRMax algorithm in Section IV.

IV. Distributed Algorithm

In this section we blend 1) the subgradient method, which solves an approximation of the master problem (9) (see Section III-B) and 2) *Algorithm 1*, which finds a suboptimal solution to subproblem (10) (see Section III-C). The result is an algorithm, which solves a series of approximated variants of the original master problem (9) via a subgradient method. Subgradients for the subgradient method are computed by coordinating the subproblems or the BS optimizations.

The main skeleton of the proposed distributed algorithm is depicted in Fig. 2(b), which is a smooth integration of the subgradient method (24) and *Algorithm 1* in an iterative manner. This results in Algorithm 2, see Fig. 2(b) for a concise block diagram.

The first step initializes *Algorithm 2*. Steps 2 represents the BS optimizations that are performed asynchronously in a decentralized fashion by each BS for fixed out-of-cell interference $\mathbf{z}$. BS optimizations terminate after the per BS stopping criterion is satisfied; see step 4 of *Algorithm 1*. At this stage each BS has its own solution and the subgradient part $\{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$. BS coordination is initiated at step 3. For example, each BSs coordinate to construct a subgradient $\sum_{n\in\mathcal{N}}\{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ and perform subgradient method (24), which must be *synchronous*. This updates global out-of-cell interference variable $\mathbf{z}$. At step 4, each BS performs their own GP to compute $\{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ for the next subgradient iteration. Step 5, is the stopping criterion for the

---

Algorithm 2: Distributed algorithm for WSRMax

1 Initialization; given the globally agreed initial out-of-cell interference $\mathbf{z}$, a feasible beamformer configuration $\{\mathbf{v}_l^{(0)}\}_{l\in\mathcal{L}(n)}$, and a feasible power allocation $\{p_l^{(0)}\}_{l\in\mathcal{L}(n)}$. Set subgradient iteration index $j=0$.

2 for $n=1$ to N
— performs *Algorithm 1* and return the subgradient contribution $\{d_{il}^n(\bar{\mathbf{z}})\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ and the suboptimal solution $\{\check{p}_l,\check{\gamma}_l,\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}$.

3 Set $\{\bar{z}_{il}^{(j)}=\ln z_{il}\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ and perform (24) to yield $\{\bar{z}_{il}^{(j+1)}\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$ and set $\mathbf{z}=\{e^{\bar{z}_{il}^{(j+1)}}\}_{l\in\mathcal{L}_{\text{int}},i\in\mathcal{N}_{\text{int}}(l)}$.

4 for $n=1$ to N
— solve problem (18). Denote the solution by $\{p_l^\star(\mathbf{z}),\gamma_l^\star(\mathbf{z})\}_{l\in\mathcal{L}(n)}$ and the optimal Lagrange multipliers by $\{\lambda_l^\star(e^{\bar{\mathbf{z}}})\}_{l\in\mathcal{L}(n)\setminus\mathcal{L}_{\text{local}}(n)}$ and $\{\mu_l^\star(e^{\bar{\mathbf{z}}})\}_{l\in\mathcal{L}_{\text{int}}(n)}$.
— Compute $d_{il}^n(\bar{\mathbf{z}})$ by using (23).

5 Stopping criterion; if the stopping criterion is satisfied, reset subgradient iteration index j, i.e., $j=0$, set $\{\mathbf{v}_l^{(0)}=\check{\mathbf{v}}_l\}_{l\in\mathcal{L}(n)}$, $\{p_l^{(0)}=p_l^\star(\mathbf{z})\}_{l\in\mathcal{L}(n)}$, and go to step 2. Otherwise increment subgradient iteration index j, i.e., $j=j+1$ and go to step 3.

---

subgradient method. If the stopping criterion is satisfied, Algorithm switches back to BS optimizations, i.e., step 2. Otherwise, the subgradient method is performed until the stopping criterion is satisfied. The algorithm continues in an iterative manner.

Fig. 3(a) depicts graphically the behavior of *Algorithm 2*. The nonconvex curve is the objective function of the master problem (9) after the logarithmic change of variables $\bar{z}_{il}=\ln z_{il}$. The convex curves are the objective functions of approximated master problems of the form (21), which are essentially parameterized by the current beamforming directions. The vertical arrows correspond to asynchronous per BS optimizations, i.e., step 2 depicted in Fig. 2(b). The horizontal arrows correspond to the subgradient method, i.e., step 3–5 depicted in Fig. 2(b). Fig. 3(a) shows that the algorithm switches between

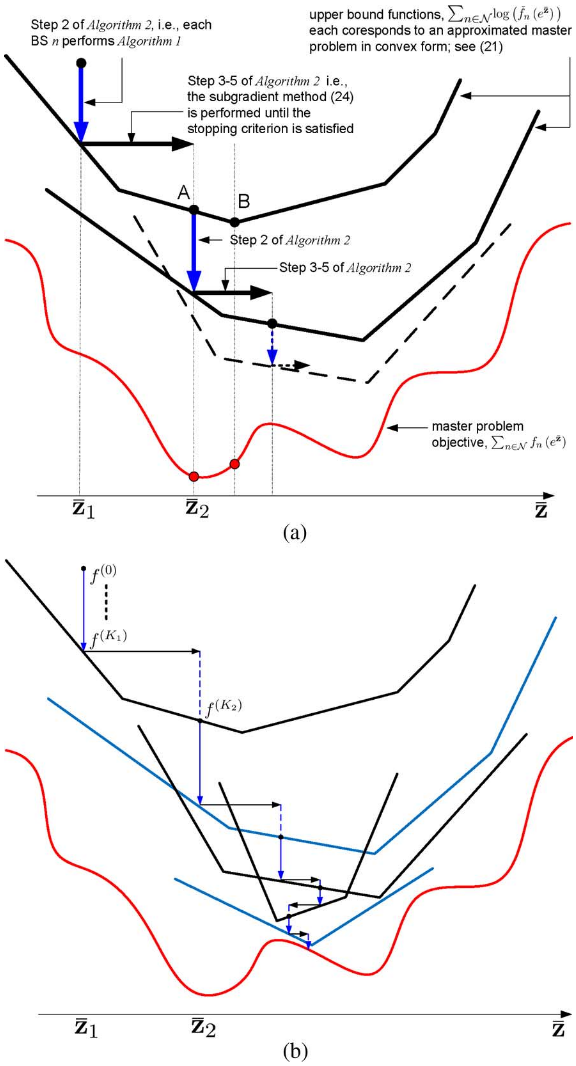

Fig. 3. The behavior of *Algorithm 2*; the objective function of problem (9) and (21) are shown in the domain of $\bar{\mathbf{z}}$. (a) Upper bounds; (b) monotonic convergence.

the per BS optimizations and the subgradient method. For example, by fixing out-of-cell interference at $\bar{\mathbf{z}}_1$, the algorithm performs per BS optimizations. Once a specified stopping criterion is satisfied, the algorithm stops BS optimizations and performs the subgradient method until a specified stopping criterion is satisfied. As a result, the out-of-cell interference values are changed from $\bar{\mathbf{z}}_1$ to $\bar{\mathbf{z}}_2$. The algorithm continues in an iterative manner.

The algorithm proposed in this section has following features, which simplify its practical implementation:

a. *Local channel state information (CSI)*: The nth BS requires to know only the channels to receiver nodes located inside its interference region. Specifically, nth BS should know channel matrices $\mathbf{h}_{jl}^{\mathrm{H}}$, where $l \in \mathcal{L}(n) \cup \mathcal{L}_{\mathrm{int}}(n)$ and $tran(j) = n$. This is similar to the CSI requirement in WMMSE algorithm (see [45, Sec. IV]).
b. *Asynchronism*: All the subproblems or BS optimizations can be carried out in a fully asynchronous fashion until a stopping criterion is satisfied.
c. *Fast Local optimization*: Each subsystem need to solve convex problems, which can be performed very fast provided the significant computing power available at each BS.
d. *Thin protocol*: Each BS does *not* need to reveal the entirety of its own subproblem during the BS coordination; only a little communication is needed, and therefore the protocol between BSs can be very light.
e. *Reliability*: To carry out the algorithm, only BS to BS synchronized signalling is required. This signalling can be carried out via reliable backhaul communication links such as microwave and fibre links.
f. *No user terminal involvement*: The user terminals do not require performing any processing associated with algorithm iterations and user to BS signalling is not required.

A. Monotonic Convergence of Algorithm 2

In this section we first show that *Algorithm 2* can generate a monotonically nonincreasing sequence of objective values, with appropriate choice of stopping criteria. In particular, we measure the objective value given by the algorithm just after each GP; see point 'F_1' of Fig. 2(a) and point 'F_2' of Fig. 2(b). Then we show the monotonic convergence of *Algorithm 2*.

Algorithm 2 starts with *Algorithm 1* (see step 2). Let $f^{(0)}, f^{(1)}, \ldots, f^{(K_1)}$ denote the sequence of objective values obtained during *Algorithm 1* iterations. Here K_1 is the number of *Algorithm 1* iterations until its stopping criterion is satisfied. Natural stopping criteria includes 1) running *Algorithm 1* for a fixed number of iterations or 2) running *Algorithm 1* until the objective value decrement between two successive iterations is below a certain predefined threshold. Since *Algorithm 1* contains nonascent steps (see Section III-C) we have

$$f^{(0)} \geq f^{(1)} \geq \ldots \geq f^{(K_1)}, \tag{27}$$

as depicted in Fig. 3(b).

Next, *Algorithm 2* switches to the subgradient method (24) (see step 3). Note that, the subgradient method is not a descent algorithm. Therefore, in order to obtain a monotonically nonincreasing sequence of objective values, we consider the following stopping criterion: running subgradient method until an objective value $f^{(K_2)}$ is achieved, such that $f^{(K_1)} \geq f^{(K_2)}$ (see Fig. 3(b)), where $K_2 = K_1 + J$ and $J > 1$ is the number of subgradient iterations.[10] Thus, we have

$$f^{(0)} \geq f^{(1)} \geq \ldots \geq f^{(K_1)} \geq f^{(K_2)}. \tag{28}$$

The switching between *Algorithm 1* and the subgradient method is done in an iterative manner. The result is a monotonically nonincreasing sequence of objective values $f^{(0)}, f^{(1)}, f^{(2)}, \ldots$ such that $f^{(i)} \geq f^{(i+1)}$, $i = 0, 1, 2, \ldots$. Moreover, note that the optimal objective value of problem (7) is *bounded*. This guarantees the monotonic convergence of *Algorithm 2* [56, Th. 3.14].

[10]In fact, the subgradient method, with diminishing nonsummable step lengths, ensures asymptotic convergence [46]. However, the requirement here is to iterate until a better objective value (compared to the initial objective value $f^{(K_1)}$) is obtained.

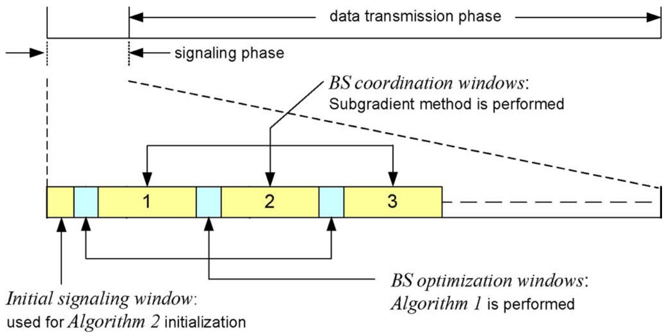

Fig. 4. An example signalling frame structure.

Note that the development of *Algorithm 2* is not based on Karush-Kuhn-Tucker (KKT) optimality conditions for the nonconvex problem (7). As a result, characterizing completely the solution structure of the proposed algorithm is a difficult task. For example, the (suboptimal) solution after the convergence of *Algorithm 2* may not necessarily be a locally optimal point of problem (7).

B. A Practical Stopping Criterion/Signalling Strategy

The stopping criteria discussed in Section IV-A are, of course, important to ensure the monotonic convergence of the algorithm. However, it is desirable to seek for stopping criteria, which are favorable for practical implementations of the algorithm, but with a violation of the monotonic convergence. In the sequel, we explain such an example strategy.

The key idea is to define time barriers; i.e., system checkpoints at which all BS must start their local optimizations (i.e., *Algorithm 1*) and system checkpoints at which all BS start coordination (i.e., the subgradient method). In particular, each BS transmissions are synchronized and the data transmission phase of each BS is preceded by a signalling phase, in which the rate/power allocation of each BS is determined via WSRMax; see Fig. 4. The signalling phase consists of three types of time slots called *initial signalling window*, *BS optimization window*, and *BS coordination window*. The initial signalling window is used for step 1 of *Algorithm 2*, i.e., the initialization step. The latter two types of windows (i.e., BS optimization window and BS coordination window) are repeated until the data transmission phase is reached as shown in Fig. 4. We define the BS optimization windows to be the the time periods where *Algorithm 1* is performed asynchronously. Therefore, during BS optimization windows, step 2 of *Algorithm 2* is carried out. The width of the window is determined by the maximum number of *Algorithm 1* iterations. The BS coordination windows are defined to be the time periods where the subgradient method is performed. Therefore, during any BS coordination window, step 3, step 4, and step 5 of *Algorithm 2* are carried out repeatedly. The width of the BS coordination window is determined by the maximum number of subgradient iterations. Typically, we may assume that the time period of any BS optimization window is significantly *smaller* compared to the time period of any BS coordination window because of the following reasons: 1) significant computing power available at BSs so that the BS optimization can be performed very fast, 2) BS coordination require backhaul message exchanges between BSs, which in turn demand stringent time requirements.

V. Numerical Examples

In this section we run our proposed *Algorithm 2* (Section IV) in multiuser multicell environments and the benefits due to different degrees of BS coordination are numerically evaluated. As benchmarks, we consider three algorithms:[11] 1) distributed WMMSE algorithm [45], 2) GP-SOCP based centralized algorithm proposed in [31, Sec. 4.3], and 3) the distributed algorithm proposed in [24], [25], which is based on a virtual SINR beamforming strategy. To emphasize the practical relevance of the proposed algorithm, we consider only the stopping criterion discussed in Section IV-B, which is based on time barriers or system checkpoints as shown in Fig. 4.

We consider an exponential path loss model, where the channel gains between BSs and users are given by

$$\mathbf{h}_{ij} = \sqrt{\left(\frac{d_{ij}}{d_0}\right)^{-\eta}}\mathbf{c}_{ij}, \tag{29}$$

where d_{ij} is the distance from the transmitter of ith data stream to the receiver of jth data stream, d_0 is the *far field reference distance* [57], η is the path loss exponent, and $\mathbf{c}_{ij} \in \mathbb{C}^T$ such that $\mathbf{c}_{ij} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ (i.e., frequency-flat fading with uncorrelated antennas). The first term of (29) represents the path loss factor and the second term models the Rayleigh small-scale fading. An arbitrarily generated set $\mathcal{C}$ of fading coefficients where $\mathcal{C} = \{\mathbf{c}_{ij} \mid i, j \in \mathcal{L}\}$ is referred to as a *single fading realization*. The variance of the noise is considered equal for all data streams, i.e., $\sigma_l^2 = N_0$ for all $l \in \mathcal{L}$ and the maximum power constraint is assumed the same for all nodes, i.e., $p_n^{\max} = p_0^{\max}$ for all $n \in \mathcal{N}$. We define the SNR operating point at a distance d [distance units] as

$$\mathrm{SNR}(d) = \begin{cases} p_0^{\max}/N_0 & d \le d_0 \\ p_0^{\max}/N_0\,(d/d_0)^{-\eta} & \text{otherwise.} \end{cases} \tag{30}$$

In all our simulations, we set $d_0 = 1, \eta = 4, p_0^{\max}/N_0 = 45\,\mathrm{dB}$, $\mathrm{SNR}(R_{\mathrm{int}}) = 0\,\mathrm{dB}$, where R_{int} is the radius of the interference regions of each BS,[12] and $\mathrm{SNR}(R_{\mathrm{BS}}) = 8$ dB, where R_{BS} is the radius of the transmission regions of each BS.

In our simulations two multicell multiuser wireless cellular networks as shown in Fig. 5 are considered. In the case of first network (i.e., Fig. 5(a)), there are $N = 2$ BSs with $T = 4$ antennas at each one. The BSs are located such that the distance between the two BSs is $D_{\mathrm{BS}} = 1.5 \times R_{\mathrm{BS}}$. In the case of second network (i.e., Fig. 5(b)), there are $N = 3$ BSs with $T = 4$ antennas at each one. Moreover, the BSs are located such that they form an equilateral triangle and the distance between any two BSs is $D_{\mathrm{BS}} = 1.5 \times R_{\mathrm{BS}}$. There are 4 users per each BS

[11]These three algorithms are not restricted to MISO IC. They can handle more general MISO interfering broadcast channel.

[12]Signal strength of BS's transmitted signal at a distance R_{int} is at most on the order of noise, Therefore, as we modeled in Section II, it is reasonable to consider that the interference cased by the BS outside the interference region is negligible.

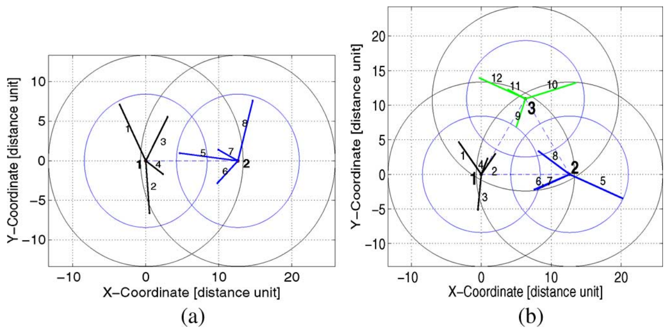

Fig. 5. (a) Multicell network 1, $\mathcal{N} = \{1,2\}$, $\mathcal{L} = \{1,\ldots,8\}$, $\mathcal{L}(1) = \{1,\ldots,4\}$, $\mathcal{L}(2) = \{5,\ldots,8\}$, $\mathcal{L}_{\rm int} = \{3,\ldots,7\}$; (b) Multicell network 2, $\mathcal{N} = \{1,2,3\}$, $\mathcal{L} = \{1,\ldots,12\}$, $\mathcal{L}(1) = \{1,\ldots,4\}$, $\mathcal{L}(2) = \{5,\ldots,8\}$, $\mathcal{L}(3) = \{9,\ldots,12\}$, $\mathcal{L}_{\rm int} = \{1,2,4,6,\ldots,11\}$.

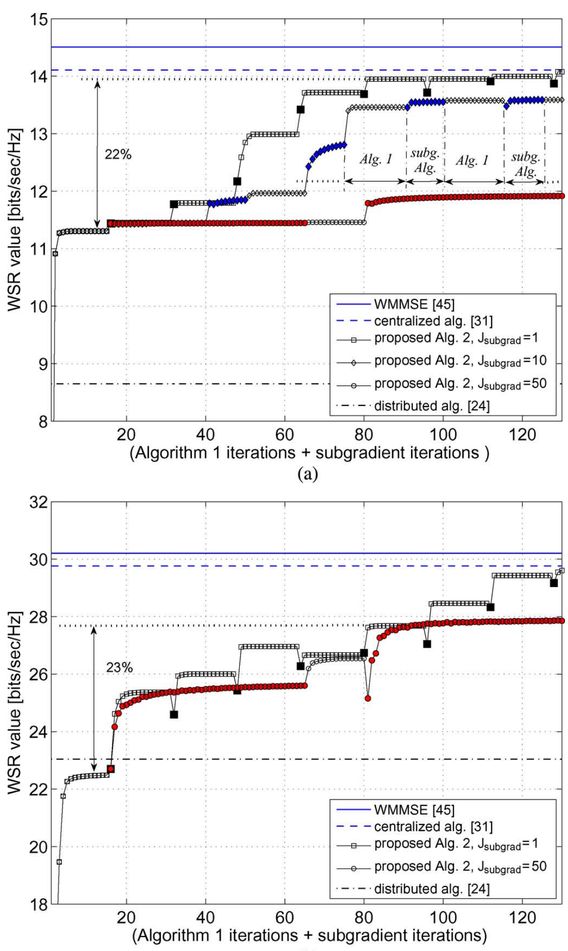

Fig. 6. Objective value versus GP iteration: (a) Multicell network 1; (b) Multicell network 2.

located inside the transmission region of the BS. The locations of users associated with BSs are arbitrarily chosen as shown in Fig. 5. A single data stream is transmitted for each user.

To see the behavior of *Algorithm 2*, we first consider a non-fading case where for each network (see Fig. 5), an arbitrary generated single fading realization is considered. We run the algorithm in both networks shown in Fig. 5. Fig. 6 shows the objective value of problem (6) computed at points 'F_1' and 'F_2' (see Fig. 2(a) and Fig. 2(b)). Here the X-axis of Fig. 6 represents combined *Algorithm 1* iterations and subgradient iterations. For simplicity, we denote the number of *Algorithm 1* iterations carried out during the BS optimization window by $J_{\rm BS-opt}$ and denote the number of subgradient iterations performed during the BS coordination window by $J_{\rm subgrad}$. Plots are drawn for the cases of $J_{\rm BS-opt} = 15$ and $J_{\rm subgrad} = 1,\ 10,\ 50$. Note that $J_{\rm subgrad}$ is a measure of the degree of BS coordination. For example, $J_{\rm subgrad} = 1$ means that the subgradient method is performed only once during any BS coordination window and $J_{\rm subgrad} = 50$ means that the subgradient method is carried out 50 consecutive times during any BS coordination window. Weights β_l of each data stream is arbitrarily chosen from the interval (0, 1]. In step 1 of *Algorithm 2*, the components of initial out-of-cell interference vector $\mathbf{z}$ are chosen on the order of noise variance N_0 (e.g., $0.5N_0$). Moreover, the normalized initial beamformers $\{\mathbf{v}_l^{(0)}\}_{l\in\mathcal{L}(n)}$ are randomly generated and a feasible uniform initial beamformer power allocation is chosen, i.e., $\left\{p_l^{(0)} = \alpha p_0^{\max}/T\right\}_{l\in\mathcal{L}(n)}$, where $\alpha \in (0,1]$ is chosen to ensure the feasibility of problem (18).

In order to describe the algorithm's behavior, let us first focus to Fig. 6(a), the case of $J_{\rm subgrad} = 1$. To distinguish *Algorithm 1* iterations from the subgradient iterations, we use two types of squares; transparent squares and solid squares. Specifically, the transparent squares correspond to the *Algorithm 1* iterations and the solid squares correspond to the subgradient iterations. Since $J_{\rm subgrad} = 1$, only a *single* subgradient iteration is performed during any BS coordination window. Furthermore, each BS perform 15 *Algorithm 1* iterations during any BS optimization window, since we have $J_{\rm BS-opt} = 15$. Note that the BS optimizations (*Algorithm 1*) are always nondecreasing steps.[13] The flattening of these line segments means that BS optimizations cannot further improve the system objective. Violation of overall monotonic behavior is inevitable since the subgradient method is not a descent algorithm in general [46]. Results show that BS coordination can gracefully resolve the out-of-cell interference (i.e., $\mathbf{z}$) via subgradient method. For example, the plot in the case of $J_{\rm subgrad} = 1$, shows a 22% increase in the weighted sum-rate (WSR), after having 5 subgradient iterations.

Fig. 6(a) further shows that the value of $J_{\rm subgrad}$, which parameterizes the degree of BS coordination has a significant effect on the overall WSR value. It is interesting to note that, a smaller number of *consecutive* subgradient iterations (e.g., $J_{\rm subgrad} = 1,\ 10$) can perform better compared to a larger number of consecutive subgradient iterations (e.g., $J_{\rm subgrad} = 50$). Such a behavior is very important in practice to reduce significantly the backhaul message exchanges during any BS coordination window. We can intuitively explain the behavior by considering the two points 'A' and 'B' in Fig. 3(a). In particular, point 'A' corresponds to a smaller $J_{\rm subgrad}$, where the (convex form) approximated master problem (21) is solved to a low accuracy. Point 'B' corresponds to a larger $J_{\rm subgrad}$, where the (convex form) approximated master problem is solved to a high accuracy. Of course, point 'B' is better than

[13]Nondecreasing because we have plotted the positive weighted sum-rate value instead of the negative value of it.

point 'A' for the *approximated* master problem, but not necessarily for the original master problem (9); see the master objective depicted in Fig. 3(a). This suggests that one need not solve each approximation to a high accuracy. Refining the approximation more often (which corresponds to a smaller $J_{\rm subgrad}$), rather than solving some approximated master problem to a high accuracy (which corresponds to a larger $J_{\rm subgrad}$) is more beneficial.

Fig. 6(b) shows the proposed algorithm behavior in the case of network setup 2 in Fig. 5(b). The behavior is very similar to the previous plots in Fig. 6(a). The network can yield substantial gains by performing just one subgradient iterations during any BS coordination window, i.e., less backhaul message exchanges between BSs. For example, the plot in the case of $J_{\rm subgrad} = 1$, shows a 23% increase in the WSR, after having 5 subgradient iterations; see Fig. 6(a). Fig. 6 also shows the performance of the considered benchmark algorithms *after* their convergence. In both networks, for the considered channel realizations, the performance of the distributed algorithm in [24] is significantly low. Note that, algorithm in [24] is well suited for lightly loaded scenarios (see [35, Fig. 4]), and therefore, it is intuitively expected this performance drop due to the lack of degrees of freedom available at BS transmissions to avoid interference. Results further show that the distributed WMMSE algorithm outperforms the proposed algorithm in both scenarios. Such results are intuitively expected because WMMSE algorithm *do* rely on user terminal assistance during algorithm's iterations compared to our proposed *Algorithm 2*. The good performance of the centralized algorithm compared to *Algorithm 2* agrees with the intuition that methods with a centralized controller can always outperform decentralized methods.

It is important to note, however, that all the considered algorithms are suboptimal methods to problem (6), and therefore their optimality is not guaranteed. As a result, they may experience different performance ranking for different channel realizations. One such case is illustrated in Fig. 7. The algorithms' parameters are same as in Fig. 6 except the fading realizations. Results show that *Algorithm 2* can outperforms WMMSE and the centralized algorithms.

In order to see the average behavior of the proposed algorithm, we consider a fading case. Here, we run *Algorithm 2* for 500 fading realization with $J_{\rm subgrad} = 1$ and $J_{\rm BS-opt} = 15$. Recall that the algorithm parameter $J_{\rm subgrad} = 1$ means that during any BS coordination window, only one subgradient iteration is performed. These are the only operations that require message exchanging between BSs via backhaul links. Moreover, subgradient iterations are the main implementation-level bottleneck, provided significant computing power at BSs, where *Algorithm 1* iterations can be performed fast and efficiently. Thus, it is interesting to see the average WSR value of problem (6) achieved at point 'F_3' of *Algorithm 2* (see Fig. 2(b)) after $m(=0,1,\ldots)$ subgradient iteration. In other words, we examine the evolution of average WSR versus the number of BS coordinations.

Fig. 8 shows the dependence of the average WSR value on the number of subgradient iterations in the case of considered network 1 and network 2. Note that, we have used the same figure to plot the dependence of the average objective value of

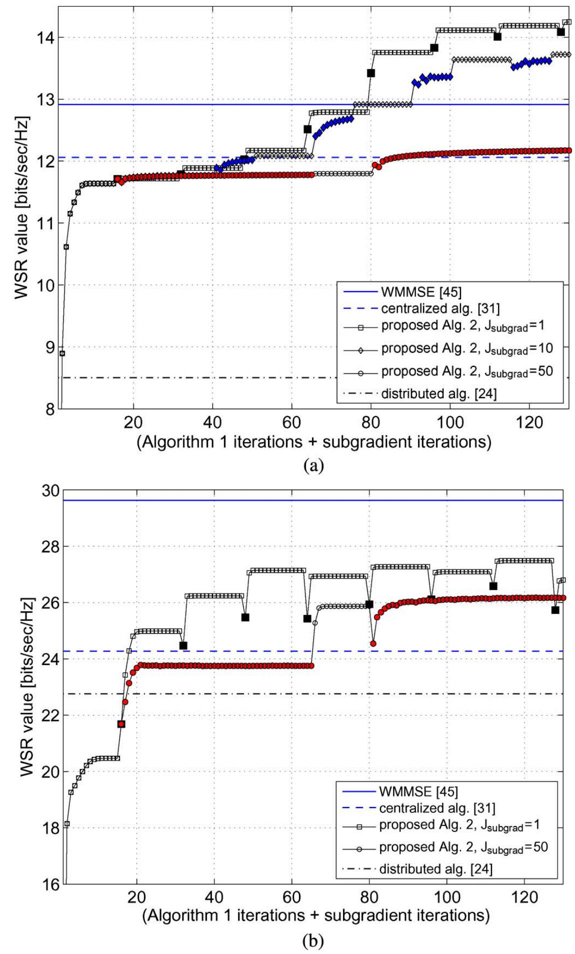

Fig. 7. Objective value versus GP iteration: (a) Multicell network 1; (b) Multicell network 2.

WMMSE algorithm on the number of iterations.[14] Results show that the BS coordination plays a critical role in the performance of *Algorithm 2*. For clarity, we denote the situation where the subgradient iterations $J_{\rm subgrad} = 0$ as *noncoordinating* case. In the case of network 1 (see Fig. 8(a)), more than 12% improvement in the average objective value is achieved within five BS coordinations compared to the noncoordinating case. For network 2 (see Fig. 8(b)), within five BS coordinations, more than 24% improvement in the average objective value is achieved as compared to the noncoordinating case.

Fig. 8 also shows that the average performance of WMMSE algorithm is better compared to that of *Algorithm 2*. This behavior is intuitively expected since, unlike the proposed *Algorithm 2*, the WMMSE algorithm benefits from user terminal assistance. Recall that, during each iterations, WMMSE algorithm requires user terminals assistance such as signal covariance estimations, computations, and feedback information to BSs over the air interface. In contrast, our proposed method require only BS-level synchronized communication and all the necessary computation is concentrated at the BSs. The result

[14]The subgradient iterations are analogous to WMMSE iterations in the following sense: both the subgradient iterations and the WMMSE algorithm iterations require message exchanges between nodes. Specifically, the subgradient method requires BS-BS message exchanges and WMMSE requires BS-user terminal as well as user terminal-BS message exchanges.

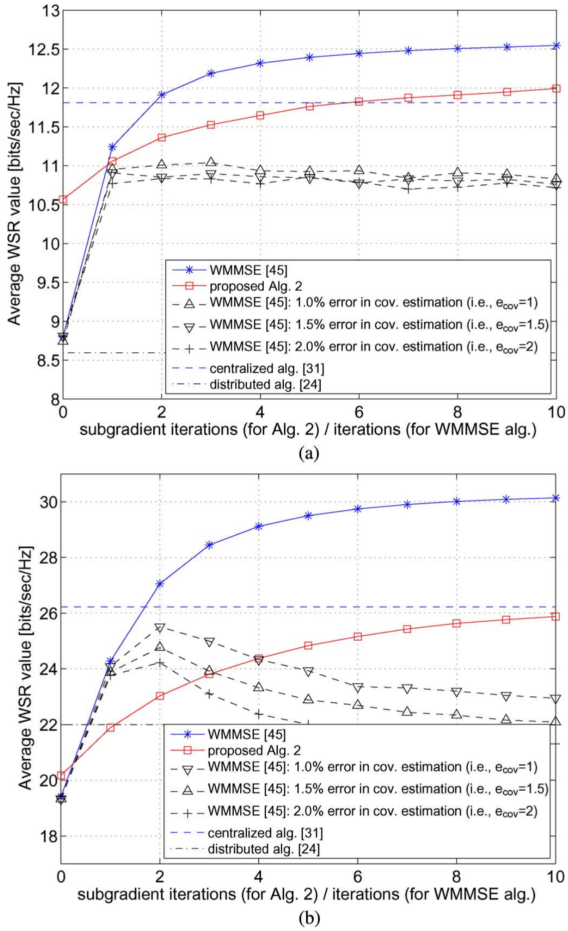

Fig. 8. Average objective value versus number of BS coordinations: (a) Multicell network 1; (b) Multicell network 2.

is naturally a trade off between performance gains and the implementation-level simplicity. For a fair comparison of *Algorithm 2* and WMMSE algorithm, we examine the sensitivity of WMMSE algorithm to imperfections on the signal covariance estimations at user terminals. Specifically, during each WMMSE iteration, we randomly perturb the error free signal covariance matrix J_l (which is a scalar in the case of MISO) at each user terminal l as follows: $J_l := J_l + J_l\,(xe_{\mathrm{cov}}/100)$, where x is a random variable with 2 equiprobable outcomes -1, 1 and e_{cov} is the amount of covariance perturbation. Results show that even small estimation errors have a significant effect on the performance of WMMSE algorithm. Moreover, in such situations, the convergence of the WMMSE method becomes less predictable. Thus, our algorithm is well suited for systems where the user terminal assistance is not desirable due to potential errors such as estimation errors and feedback errors.

Fig. 8 further shows that, the performance of *Algorithm 2* within several BS coordinations is comparable with that of the centralized algorithm [31, Sec. 4.3]. For example, in the case of network 1, *Algorithm 2* achieves around 99% of the average WSR value given by the centralized algorithm [31, Sec. 4.3]. Moreover, in the case of network 2, *Algorithm 2* yields around 94% of the average WSR value given by the considered centralized algorithm. Finally, we see that there is a substantial performance gap between *Algorithm 2* and the distributed algorithm in [24]. The main reason for such a performance drop of algorithm in [24] is the insufficient degree of freedom available at BS transmissions to cancel the interference it causes to the user terminals.

VI. Conclusions

We considered the weighted sum-rate maximization problem in a multicell multiple-input and single-output downlink system. The problem is nonconvex; in fact it is NP-hard. A distributed solution method for the problem is proposed. The main advantage of the proposed algorithm is its implementation-level simplicity. Unlike the minimum weighted mean-squared error based algorithms, our method does not demand user terminal assistance during each iteration. Our algorithm essentially require base station to base station (BS) communication, which are reasonably realizable, provided reliable backhaul links (e.g., fibre and microwave links) and significant computing power at BSs. As a result, a good trade-off between the performance gains and the implementation-level simplicity was achieved. The proposed algorithm was based on primal decomposition and subgradient methods. In particular, the main problem was split into a master problem and many subproblems (one for each base station). A novel sequential convex approximation strategy together with a subgradient method were blent to address the nonconvex master problem. Master problem solution relies on synchronous BS coordinations. A descent algorithm based on second-order cone programming and a geometric programming were adopted in the case of subproblems. The subproblems can be performed in a fully *asynchronous* manner. The monotonic convergence of the algorithm was established, with appropriate choice of stopping criteria at intermediate steps. Practical stopping criteria have also been proposed. Numerical experiments were performed to compare our method with existing state-of-the-art algorithms. Results suggest that our algorithm is well suited for systems where the user terminal assistance is not allowed or not desirable. Results further showed that the proposed algorithm could significantly improve the overall system performance with a small amount of BS coordinations. These observations are indeed important for deriving simple signalling protocols in the context of large-scale practical cellular communication systems.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments which have improved the presentation and the quality of the paper.

References

[1] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Trans. Autom. Contr.*, vol. 37, no. 12, pp. 1936–1949, Dec. 1992.

[2] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. San Rafael, CA: Morgan & Claypool, 2010, vol. 7, Synthesis Lectures on Communication Networks.

[3] M. J. Neely, E. Modiano, and C. E. Rohrs, “Power allocation and routing in multibeam satellites with time-varying channels,” *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[4] X. Lin and N. B. Shroff, “Joint rate control and scheduling in multihop wireless networks,” in *Proc. IEEE Int. Conf. Dec. and Cont.*, Atlantis, Paradise Island, Bahamas, Dec. 14–17, 2004, vol. 5, pp. 1484–1489.

[5] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[6] A. L. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Que. Syst.*, vol. 50, no. 4, pp. 401–457, Aug. 2005.

[7] A. Eryilmaz, R. Srikant, and J. R. Perkins, "Stable scheduling policies for fading wireless channels," *IEEE/ACM Trans. Netw.*, vol. 13, no. 2, pp. 411–424, Apr. 2005.

[8] X. Wu and R. Srikant, "Regulated maximal matching: A distributed scheduling algorithm for multi-hop wireless networkswith node-exclusive spectrum sharing," in *Proc. IEEE Conf. Decision Cont., Eur. Cont. Conf.*, Seville, Spain, Dec. 12–15, 2005, pp. 5342–5347.

[9] G. Sharma, R. R. Mazumdar, and N. B. Shroff, "On the complexity of scheduling in wireless networks," in *Proc. ACM Int. Conf. Mobile Comp. Net.*, Los Angeles, CA, Sep. 24–29, 2006, pp. 227–238.

[10] Z. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Select. Areas Commun.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[11] R. Horst, P. Pardolos, and N. Thoai*, Introduction to Global Optimiz.*, 2nd ed. Dordrecht, Boston, London: Kluwer Academic, 2000, vol. 48.

[12] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Weighted sum-rate maximization for a set of interfering links via branch and bound," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3977–3996, Aug. 2011.

[13] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.

[14] P. Tsiaflakis, M. Diehl, and M. Moonen, "Distributed spectrum management algorithms for multiuser DSL networks," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4825–4843, Oct. 2008.

[15] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.

[16] J. Papandriopoulos and S. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, Jun. 11–15, 2006, pp. 3270–3275.

[17] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Proc. IEEE Inform. Theory App. Workshop*, San Diego, CA, Jan. 29–Feb. 2 2007, pp. 414–420.

[18] J. Yuan and W. Yu, "Distributed cross-layer optimization of wireless sensor networks: A game theoretic approach," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, CA, Nov. 27–Dec. 1 2006.

[19] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[20] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *Proc. IEEE INFOCOM*, Rio De Janeiro, Brazil, Apr. 19–25 2009, pp. 1287–1295.

[21] W. Yu, T. Kwon, and C. Shin, "Joint scheduling and dynamic power spectrum optimization for wireless multicell networks," in *Proc. Conf. Inform. Sci. Syst. (CISS)*, Princeton, NJ, Mar. 21–23 2010, pp. 1–6.

[22] S.-H. Park, H. Park, and I. Lee, "Distributed beamforming techniques for weighted sum-rate maximization in MISO interference channels," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1131–1133, Dec. 2010.

[23] M. Rossi, A. M. Tulino, O. Simeone, and A. M. Haimovich, "Nonconvex utility maximization in Gaussian MISO broadcast and interference channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 22–27 2011, pp. 2960–2963.

[24] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimality properties and low-complexity solutions to coordinated multicell transmission," in *Proc. IEEE Global Telecommun. Conf.*, Miami, FL, Dec. 6–10, 2010, pp. 1–6.

[25] E. Björnson, N. Jaldén, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, measurementbased evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011.

[26] L. Venturino, N. Prasad, and X. Wang, "Coordinated linear beamforming in downlink multi-cell wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1451–1461, Apr. 2010.

[27] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1142–1157, Mar. 2011.

[28] W. Yu, T. Kwon, and C. Shin, "Multicell coordination via joint scheduling, beamforming and power spectrum adaptation," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 10–15, 2011, pp. 2570–2578.

[29] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[30] M. Codreanu, A. Tölli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4639–4655, Sep. 2007.

[31] M. Codreanu, "Multidimensional Adaptive Radio Links for Broadband Communications" Ph.D. thesis, Centre for Wireless Communications, University of Oulu. Acta Universitatis Ouluensis, Oulu, Finland, Nov. 2007 [Online]. Available: http://herkules.oulu.fi/isbn9789514286223

[32] G. Zheng, K.-K. Wong, and T.-S. Ng, "Throughput maximization in linear multiuser MIMO-OFDM downlink systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1993–1998, May 2008.

[33] S. Joshi, P. C. Weeraddana, M. Codreanu, and M. Latva-aho, "Weighted sum-rate maximization for MISO downlink cellular networks via branch and bound," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2090–2095, Apr. 2012.

[34] S. S. Christensen, R. Agarwal, E. Carvalho, and J. 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.

[35] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4298–4310, Aug. 2010.

[36] R. Zakhour and D. Gesbert, "Coordination on the MISO interference channel using the virtual SINR framework," in *Proc. ITG Workshop Smart Antennas*, Berlin, Germany, Feb. 16–18, 2009.

[37] R. Zakhour and D. Gesbert, "Distributed multicell-MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2444–2448, Aug. 2010.

[38] R. Zakhour, Z. Ho, and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels," in *Proc. IEEE Veh. Technol. Conf.*, Barcelona, Spain, Apr. 26–29, 2009, pp. 1–5.

[39] B. O. Lee, H. W. Je, I. Sohn, O.-S. Shin, and K. B. Lee, "Interference-aware decentralized precoding for multicell MIMO TDD systems," in *Proc. IEEE Global Telecommun. Conf.*, New Orleans, LA, Nov.30–Dec. 4 2008, pp. 1–5.

[40] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[41] 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.

[42] B. Song, R. L. Cruz, and B. D. Rao, "A simple joint beamforming and power control algorithm for multi-user MIMO wireless networks," in *Proc. IEEE Veh. Technol. Conf.*, Los Angeles, CA, Sep. 26–29, 2004, pp. 247–251.

[43] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[44] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[45] 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, no. 9, pp. 4331–4340, Sep. 2011.

[46] D. P. Bertsekas*, Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[47] S. Boyd, Sequential Convex Programming [Online]. Available: http://www.stanford.edu/class/ee364b/lectures/seq slides. pdf 2007

[48] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193–228, Nov. 1998.

[49] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimiz. Eng.*, vol. 8, no. 1, pp. 67–127, 2007.

[50] P. C. Weeraddana, M. Codreanu, and M. Latva-aho, "Multicell downlink weighted sum-rate maximization: A distributed approach," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, Pacific Grove, CA, Nov. 6–9, 2011, pp. 1368–1375.

[51] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[52] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, “Resource allocation for cross-layer utility maximization in wireless networks,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, Jul. 2011.

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

[54] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. New York: Academic., 1983, vol. 165.

[55] D. P. Palomar and M. Chiang, “Alternative distributed algorithms for network utility maximization: Framework and applications,” *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[56] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, Inc., 1976.

[57] A. Kumar, D. Manjunath, and J. Kuri, *Wireless Networking*. Burlington, MA, USA: Elsevier, 2008.

Pradeep Chathuranga Weeraddana (S’08) received the M.Eng. degree telecommunication from School of Engineering and Technology, Asian Institute of Technology, Thailand, in 2007, and the Ph.D. degree from University of Oulu, Oulu, Finland, in 2011.

He is currently working as a Postdoctoral Researcher in Automatic Control Labortatory, Electrical Engineering Department and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include application of optimization techniques in various application domains, such as signal processing, wireless communications, smart grids, privacy, and security.

Marian Codreanu (S’02–M’07) received the M.S. degree from the University Politehnica of Bucharest, Romania, in 1998, and the Ph.D. degree from University of Oulu, Oulu, Finland, in 2007.

From 1998 to 2002, he was a Teaching Assistant at the Telecommunications Department in University Politehnica of Bucharest. In 2002, he joined the Centre for Wireless Communications (CWC) at University of Oulu where he is currently holding a Senior Research Fellow position. In 2008, he was a Visiting Postdoctoral Researcher the University of Maryland, College Park.. His research interests include optimization, information theory, and signal processing for wireless communication systems and networks.

Dr. Codreanu received the Best Doctoral Thesis Prize within the area of all technical sciences in Finland in 2007. He was a Cochair of the Technical Program Committee (TPC) of The First Nordic Workshop on Cross-Layer Optimization in Wireless Networks in 2010, and a Cochair of the TPC of The Second Nordic Workshop on System and Network Optimization for Wireless (SNOW 2011).

Matti Latva-aho (S’96–M’98–SM’06) was born in Kuivaniemi, Finland, in 1968. He received the M.Sc., Lic.Tech., and Dr. Tech (Hons.) degrees in electrical engineering from the University of Oulu, Oulu, Finland in 1992, 1996, and 1998, respectively.

From 1992 to 1993, he was a Research Engineer at Nokia Mobile Phones, Oulu, Finland. During the years 1994–1998, he was a Research Scientist at Telecommunication Laboratory and Centre for Wireless Communications at the University of Oulu. Currently, he is the Head of Telecommunication Laboratory and Professor of Digital Transmission Techniques at the University of Oulu. He was Director of Centre for Wireless Communications at the University of Oulu during the years 1998–2006. His research interests are related to mobile broadband wireless communication systems. He has published 200 conference or journal papers in the field of wireless communications.

Anthony Ephremides (S’68–M’71–SM’77–F’84) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 1971.

He holds the Cynthia Kim Professorship of Information Technology at the Electrical and Computer Engineering Department of the University of Maryland, College Park, where he holds a joint appointment at the Institute for Systems Research, of which he was among the founding members in 1986. He has held various visiting positions at other Institutions (including MIT, UC Berkeley, ETH urich, INRIA, etc) and cofounded and codirected a NASA-funded Center on Satellite and Hybrid Communication Networks in 1991. He has been the President of Pontos, Inc, since 1980, and has served as President of the IEEE Information Theory Society in 1987, and as a Member of the IEEE Board of Directors in 1989 and 1990. He has been the General Chair and/or the Technical Program Chair of several technical conferences (including the IEEE Information Theory Symposium in 1991, 2000, and 2011, the IEEE Conference on Decision and Control in 1986, the ACM Mobihoc in 2003, and the IEEE Infocom in 1999). He has served on the Editorial Board of numerous journals and was the Founding Director of the Fairchild Scholars and Doctoral Fellows Program, a University-Industry Partnership from 1981 to 1985. He is the author of several hundred papers, conference presentations, and patents, and his research interests lie in the areas of communication systems and networks and all related disciplines, such as information theory, control and optimization, satellite systems, queueing models, signal processing, etc. He is especially interested in wireless networks and energy efficient systems

Dr. Ephremides has received the IEEE Donald E. Fink Prize Paper Award in 1991 and the first ACM Achievement Award for Contributions to Wireless Networking in 1996, as well as the 2000 Fred W. Ellersick MILCOM Best Paper Award, the IEEE Third Millennium Medal, the 2000 Outstanding Systems Engineering Faculty Award from the Institute for Systems Research, and the Kirwan Faculty Research and Scholarship Prize from the University of Maryland in 2001, and a few other official recognitions of his work. He also received the 2006 Aaron Wyner Award for Exceptional Service and Leadership to the IEEE Information Theory Society.