Asynchronous Distributed Downlink Beamforming and Power Control in Multi-cell Networks

Siduo Shen, Student Member, IEEE, and Tat-Ming Lok, Senior Member, IEEE,

Abstract—In this paper, we consider a multi-cell network where every base station (BS) serves multiple users with an antenna array. Each user is associated with only one BS and has a single antenna. Assume that only long-term channel state information (CSI) is available in the system. The objective is to minimize the network downlink transmission power needed to meet the users’ signal-to-interference-plus-noise ratio (SINR) requirements. For this objective, we propose an asynchronous distributed beamforming and power control algorithm which provides the same optimal solution as given by centralized algorithms. To design the algorithm, the power minimization problem is formulated mathematically as a non-convex problem. For distributed implementation, the non-convex problem is cast into the dual decomposition framework. Resorting to the theory about matrix pencil, a novel asynchronous iterative method is proposed for solving the dual of the non-convex problem. The methods for beamforming and power control are obtained by investigating the primal problem. At last, simulation results are provided to demonstrate the convergence and performance of the algorithm.

Index Terms—Beamforming, multi-cell networks, dual decomposition, matrix pencil, asynchronous distributed algorithm

I. INTRODUCTION

Downlink beamforming within a cell is an efficient technique for balancing the intra-cell interference, which has been studied since 1998 [1]–[6]. However, cell edge users may still suffer inter-cell interference in a multi-cell network. By coordinating beamforming and power control among BSs, inter-cell interference can be mitigated [2]. This kind of beamforming is called multi-cell beamforming. Multi-cell beamforming has its applications, like interference management in femto-cell/small-cell networks, or spatial spectrum sharing in cognitive radio networks.

One of the basic beamforming problems is to minimize the total transmission power subject to each user’s SINR constraint which is named the power minimization beamforming problem (PMBP). The problem differs for different kinds of CSI [7]. The PMBP is a convex problem [5], [6] with the instantaneous CSI which appears as a vector. However, the PMBP is non-convex [3] with the long-term CSI. The long-term CSI appears as a spatial correlation matrix whose rank is greater than one.

Centralized algorithms in [1] and [3] are derived for single-cell PMBP with either instantaneous CSI or long-term CSI. Their generalizations are applied to the multi-cell PMBP and provide the optimal solution. However, their distributed implementations are unavailable. The main reason is that they need strict coordination and synchronization among BSs which is impractical with nowadays backhaul network.

In this work, we consider the long-term CSI. Our main contribution is an algorithm that can be implemented asynchronously in a distributed fashion. The algorithm achieves the same optimal solution as given by the centralized algorithms. In the algorithm, a BS updates the beamforming vectors of its users. A BS only communicates with its neighboring BSs for coordination via backhaul networks. Motivated by the concept of “power on demand” [8], we let each user compute its required power and send it to its BS. The BSs allocate the users the required power. The term “asynchronous” has two folds of meanings. First, neither the BSs nor the users need to update simultaneously, or with the same frequency. This property makes the algorithm suitable for distributed implementation. Second, the algorithm converges with outdated coordination information. This implies that the algorithm is robust to the transmission delay and the packet loss of the backhaul network.

To design the algorithm, the PMBP is cast into the dual decomposition framework [9]. A novel asynchronous iterative method is proposed for solving the dual of the PMBP, which relies on the recent progress about the semidefiniteness of a matrix pencil [10]–[12]. With the optimal dual variables, a user’s beamforming vector is the null space of a linear combination of matrices which is computed by its BS. The power control algorithm falls in the classical iterative power control framework in [13]. Thus, the users can update the power asynchronously with local information. Besides, we also investigate the feasibility of the multi-cell PMBP with long-term CSI. The algorithm is derived assuming the PMBP being feasible.

A. Related Work

In the literature about PMBP, uplink-downlink duality is a major aspect [1], [2], [4]. The dual of the single-cell PMBP has a similar form to the uplink beamforming problem. Based on the dual of the PMBP, the virtual uplink problem is defined. The virtual uplink problem can be solved with a beamformer-power update method. The method is extended to the multi-cell case in [14] by viewing multiple BSs as an antenna array. The communication requirement may be difficult to meet in practice. Besides, method with limited coordination among BSs is proposed in [15].

The duality of the single-cell PMBP with instantaneous CSI is also studied with the Lagrangian theory in [5], [6].
An algorithm that is different from the algorithm in [1] is proposed. The work of [16] is an extension of [6] to the multi-cell case. An efficient distributed multi-cell beamforming algorithm is proposed. In [16], the system is time division duplex, i.e., reciprocal channel. Besides, instantaneous CSI is assumed, and the noise variances at each receiver are the same. In this case, the virtual uplink is the real uplink [4], which is convenient for distributed implementation. However, the algorithm is not amenable to the multi-cell PMBP with long-term CSI. Our work complements the work of [16] by assuming long-term CSI, non-reciprocal channel, and different noise variances at each user.

Another class of methods is based on the convex optimization techniques [3, 17]. The PMBP is relaxed to a convex semidefinite program (SDP). In [3], the relaxation gap is proved to be zero if the PMBP is feasible, which implies strong duality. This conclusion is generalized to the multi-cell case in [17]. With this approach, additional constraints can be considered in problem formulation [18, 19]. However, this method is hard to be implemented distributively.

Other related works are summarized as follows. In [20], the surrogate duality approach is used to maximize the minimum SINR under power constraint. In [21], power minimization of multi-cell multi-antenna beamforming system is studied. In [22], max-min weighted SINR problem subject to weighted sum power constraint is studied, where the problem is decoupled into multiple sub-problems by introducing a set of slack variables. In [23, 24], beamforming techniques are applied to hierarchical systems for interference reduction. Besides, there are works with other design objectives, such as transmission rate [25], energy-efficiency [26], and robustness [27].

B. Organization

The rest of this paper is organized as follows. In Section II, the system model and the problem formulation are given. In Section III, the feasibility of the PMBP is studied. Then, in Section IV, we introduce the computation methods for the solution of the PMBP. In Section V, we show how to implement the computation methods in an asynchronous distributed manner. The performance is further demonstrated in Section VI via simulation.

Notation: Let $A$ be a matrix. $A^{-1}$, $A^H$ and $A^T$ represent the inverse of $A$, the Hermitian transpose of $A$ and the transpose of $A$, respectively. Besides, $\det(A)$, $\text{rank}(A)$, $\text{null}(A)$, and $\rho(A)$ stand for the determinant of $A$, the rank of $A$, the null space of $A$, and the radius of $A$, respectively. If $A \succeq 0$, $A$ is a positive semidefinite/positive definite matrix. $I$ denotes an identity matrix with adaptive size. $E[\cdot]$ denotes the statistical expectation. When describing vectors’ relation, “$\leq$”, “$\geq$”, “$<$”, “$>$” are component-wise.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a multi-cell network with $M$ cells indexed by $m \in \mathcal{M} = \{1, \ldots, M\}$. Each cell has a BS with $N$ antennas and $K$ users. Within a cell, a user is indexed by $i \in \mathcal{I} = \{1, \ldots, K\}$. In the system scope, a user is identified by its cell index and its intra-cell index, i.e., $(m, i)$. User $(m, i)$ indicates the user $i$ in cell $m$. For later use, “$\forall (m, i)$” represents “for all users”.

The cells work in the same spectrum. The channel gain from BS $m$ to user $(n, i)$ is denoted with $h_{m,n,i} \in \mathbb{C}^{N \times 1}$, $m, n \in \mathcal{M}$, $i \in \mathcal{I}$. We assume flat-fading channels, i.e., no inter-symbol-interference. The long-term CSI is given as the spatial correlation matrices,

$$R_{m,n,i} = E[h_{m,n,i}h_{m,n,i}^H], \forall m, n \in \mathcal{M}, \forall i \in \mathcal{I},$$

which satisfy $\text{rank}(R_{m,n,i}) > 1$ and $R_{m,n,i} \succeq 0$. The long-term CSI assumption implies that the mean SINR is considered.

Let $s_{m,i} \in \mathbb{C}$ denotes the information signal transmitted from BS $m$, $m \in \mathcal{M}$, to its user $i$, $i \in \mathcal{I}$. Let $w_{m,i} \in \mathbb{C}^{N \times 1}$ be the beamforming vector for $s_{m,i}$. In the remaining of this work, we have $\|w_{m,i}\| = 1$ as default. The received signal of user $(m, i)$ is

$$y_{m,i} = h_{m,m,i}^H w_{m,i} s_{m,i} + \sum_{j \neq i} h_{m,j}^H w_{m,j} s_{m,j} + z_{m,i},$$

where $z_{m,i}$ is the additive white circularly symmetric Gaussian noise with variance $\sigma_{m,i}^2$. To simplify the notations, as [16] does, we have $\sum_{j \neq i} \sum_{k \neq j} = \sum_{k \neq j} \sum_{j \neq k}$ and $\sum_{n \neq m, j} = \sum_{m,n \neq m} \sum_{j \in \mathcal{I}}$. Similar notation schemes are used in the rest of this work.

The SINR of the user $(m, i)$ is defined as follows,

$$\sinr_{m,i} = \frac{p_{m,i} \|w_{m,i}\|^2 R_{m,m,i}}{\sum_{(n,j) \neq (m,i)} p_{n,j} \|w_{n,j}\|^2 R_{n,n,j} + \sigma_{m,i}^2},$$

where $p_{m,i} = E[\|s_{m,i}\|^2]$ is the allocated power of BS $m$ for user $(m, i)$ and $w_{m,i}^H R_{m,m,i} w_{m,i} = E[\|h_{m,m,i}\|^2]$. Besides, the notation $\sum_{(n,j) \neq (m,i)}$ represents summation over all the users except user $(m, i)$. In addition, we denote by $\gamma_{m,i}$ the SINR target of user $(m, i)$.

The objective is to minimize the total transmission power subject to each user’s SINR constraint. Mathematically, the objective is to find the solution of the following master problem,

$$\min_{\mathbf{p} \geq 0, \mathbf{y}} \sum_{m,i} p_{m,i} \text{ s.t. } \sinr_{m,i} \geq \gamma_{m,i}, \forall (m, i),$$

where we define $\mathbf{p} = [p_{1,1}, p_{1,2}, \ldots, p_{M,K}]^T$ and $\mathbf{y} = [w_{1,1}, w_{1,2}, \ldots, w_{M,K}]$. Since the constraints, $\|w_{m,i}\| = 1$, $\forall (m, i)$, are default, we omit them in this and the following optimization problems.

III. PROBLEM FEASIBILITY

The master problem (1) is a mathematical formulation of the multi-cell PMBP. If the master problem is feasible, all users’ SINR targets can be achieved. However, the master problem is not feasible for all sets of SINR targets. In this section, we study the feasibility of the master problem which can be viewed as a multi-cell extension of the work in [28].
The master problem is feasible if its feasible region is non-empty. The feasible region is characterized by \( \gamma_{m,i} \geq \gamma_{m,i}, \forall (m,i) \), which can be written as,

\[
P_{m,i} - \gamma_{m,i} \sum_{(n,j) \neq (m,i)} p_{n,j} w_{n,j}^H R_{m,m,i} w_{m,i} \geq \frac{\gamma_{m,i} \sigma^2_{m,i}}{w_{m,i}^H R_{m,m,i} w_{m,i}}, \forall (m,i).
\]  

(2)

The norm constraints of the beamforming vectors can be guaranteed by normalization, so they do not impact the feasibility of the master problem.

We rewrite the above inequalities in the following compact form,

\[
(I - \Gamma G)p \geq \eta,
\]

(3)

where \( \Gamma \) is a diagonal matrix, \( \eta \) is a vector, and \( G \) is a \( MK \times MK \) matrix. In particular, we have

\[
\Gamma = \text{diag}(\gamma_{1,1}, \gamma_{1,2}, \ldots, \gamma_{M,K}),
\]

the component of \( \eta \) as

\[
\eta_{m,i} = \frac{\gamma_{m,i} \sigma^2_{m,i}}{w_{m,i}^H R_{m,m,i} w_{m,i}},
\]

(4)

and the component of matrix \( G \) as

\[
G_{(m,i),(n,j)} = \frac{w_{n,j}^H R_{m,m,i} w_{m,i}}{w_{m,i}^H R_{m,m,i} w_{m,i}}.
\]

(5)

The feasibility of the master problem now depends on the inequality (3). Mathematically, according to the Perron-Frobenius theorem [29], we have the following lemma.

**Lemma 1.** The master problem is feasible if and only if there exists a set of beamforming vectors such that \( \rho(\Gamma G) < 1 \).

Similar conclusion can be found in [4], [7]. From system perspective, given the users’ SINR targets, \( \Gamma G \) captures the interference relationship among the users which mainly depends on the physical environment. If \( \rho(\Gamma G) < 1 \), the lower bound of \( p \) would be \( (I - \Gamma G)^{-1} \eta \). According to [4], \( p \) is in connection with the thermal noise. Since the feasibility of the master problem depends on \( \Gamma G \), to further reveal the physical meaning of \( \rho(\Gamma G) < 1 \), motivated by [4], [30], we have the following proposition.

**Proposition 1.** The master problem is feasible if and only if there exists a set of beamforming vectors such that the following inequalities hold for any \( p \geq 0 \),

\[
\frac{p_{m,i} w_{m,i}^H R_{m,m,i} w_{m,i}}{\sum_{(n,j) \neq (m,i)} p_{n,j} w_{n,j}^H R_{m,m,i} w_{n,j}} > \gamma_{m,i}, \forall (m,i).
\]

(6)

**Proof:** See Appendix [A].

The ratio on the left hand side of inequality (6) implies that the signal to interference ratio should be strictly larger than the target SINR. It shows mathematically that the feasibility mainly depends on the existence of the beamforming vectors that can coordinate the interference.

In summary, Lemma 1 and Proposition 1 can be viewed as further characterizations of the feasible region of the master problem. They can provide practical design insights. The above results remind us that feasibility can be achieved by removing the strongly interfered users, or decreasing some users’ SINR targets. As shown in [31]–[33], user scheduling and network-wide congestion control can be employed for feasibility. To focus on the algorithm design, we assume the master problem is feasible in the derivation of the algorithm.

**IV. COMPUTATIONAL SOLUTIONS**

In this section, we derive a series of computation methods for solving the mathematical problem (1). They are the underlying principles of the asynchronous distributed beamforming and power control algorithm. We derive these methods with their asynchronous distributed implementations in mind.

For distributed implementation, we start by employing the Lagrangian dual decomposition method [5], [6], [9]. Furthermore, since strong duality holds for the master problem, which is proved in [7], the optimality of the dual decomposition method is guaranteed. The Lagrangian of the master problem is written as follows,

\[
L(p, \Upsilon, \lambda) = \sum_{m,i} \lambda_{m,i} \sigma^2_{m,i} + \sum_{m,i} p_{m,i} w_{m,i}^H R_{m,m,i} w_{m,i}
\]

\[+ \sum_{m,i} \lambda_{m,i} \left( 1 + \frac{1}{\gamma_{m,i}} \right) p_{m,i} w_{m,i}^H R_{m,m,i} w_{m,i}
\]

\[+ \sum_{m,i} \lambda_{m,i} \sum_{n,j} p_{n,j} w_{n,j}^H R_{m,m,i} w_{n,j},
\]

(7)

where \( \lambda = [\lambda_{1,1}, \lambda_{1,2}, \ldots, \lambda_{M,K}]^T \) and \( \lambda \geq 0 \). About the last term of the above equation, we have

\[
\sum_{m,i} \lambda_{m,i} \sum_{n,j} p_{n,j} w_{n,j}^H R_{m,m,i} w_{n,j}
\]

\[= \sum_{m,i} \sum_{n,j} p_{m,i} w_{m,i}^H \lambda_{n,j} R_{n,m,i} w_{n,j},
\]

Then, equation (7) can be rewritten as,

\[
L(p, \Upsilon, \lambda) = \sum_{m} \sum_{i} \lambda_{m,i} \sigma^2_{m,i}
\]

\[+ \sum_{m,i} \left( \sum_{i} p_{m,i} w_{m,i}^H \right) \left( 1 - \left( 1 + \frac{1}{\gamma_{m,i}} \right) \lambda_{m,i} R_{m,m,i} w_{m,i} \right).
\]

(8)

Furthermore, we define

\[
L_m(p_m, \Upsilon_m, \lambda_m) = \sum_{m} \lambda_{m,i} \sigma^2_{m,i}
\]

\[+ \sum_{m,i} \lambda_{m,i} \left( 1 + \frac{1}{\gamma_{m,i}} \right) p_{m,i} R_{m,m,i}
\]

\[+ \sum_{n,j} \left( \lambda_{n,j} R_{n,m,j} \right) w_{n,j},
\]

(9)

where \( p_m = [p_{m,1}, \ldots, p_{m,K}]^T \) and \( \Upsilon_m = [w_{m,1}, \ldots, w_{m,K}] \). Then, the Lagrangian is

\[
L(p, \Upsilon, \lambda) = \sum_{m} L_m(p_m, \Upsilon_m, \lambda).
\]

(10)
According to the Lagrangian theory, the primal problem is

$$\min_{p \geq 0} \ L(p, \mathbf{Y}, \lambda).$$

(11)

Based on equation (11), the primal problem can be decomposed into $M$ sub-primal problems,

$$\min_{p_m \geq 0} \ L_m(p_m, \mathbf{Y}_m, \lambda).$$

(12)

With the optimal dual variables, the beamforming and power control can be performed within each BS by solving the corresponding sub-primal problem.

To avoid being trivial, we require, $\forall (m, i)$,

$$I - \left(1 + \frac{1}{\gamma_{m,i}}\right) \lambda_{m,i} \mathbf{R}_{m,m,i} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \succeq 0.$$ \hspace{1cm} (13)

Otherwise, the optimum of the primal problem would be negative infinity.

From (11) and (13), the dual problem is

$$\max_{\lambda \geq 0} \sum_{m,i} \lambda_{m,i} \sigma_{m,i}^2$$

s.t. $I - \left(1 + \frac{1}{\gamma_{m,i}}\right) \lambda_{m,i} \mathbf{R}_{m,m,i} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \succeq 0$, \hspace{1cm} $\forall (m, i).$ \hspace{1cm} (14)

The dual problem can be centrally solved by CVX [34]. However, a computation method designed for distributed implementation is still unavailable.

If instantaneous CSI is used, inequalities (13) are linear inequalities in essence. When the corresponding dual problem achieves optimality, they are all active, i.e., establishing with equality. Based on this property, an efficient iterative method is derived in [5], [6], [16].

However, since the ranks of the spatial correlation matrices are greater than one, the inequalities in (13) are linear matrix inequalities. This difference makes it challenging to derive a similar iterative algorithm.

A. Methods for Computing Dual Variables

One of our major contributions is an asynchronous iterative method for computing the dual variables. To design the method, we first propose the following lemma.

**Lemma 2.** Given $N \times N$ Hermitian matrices, $\mathbf{A} \succeq 0$, $\mathbf{B} \succeq 0$, the minimum non-negative eigenvalue of matrix pencil $\mathbf{A} - \mu \mathbf{B}$, denoted by $\mu_+(\mathbf{A}, \mathbf{B})$, exists and is finite. Assuming variable $\mu \geq 0$, we have

$$\mathbf{A} - \mu \mathbf{B} \succeq 0 \iff \mu \leq \mu_+(\mathbf{A}, \mathbf{B}).$$

**Proof:** See Appendix B

**Remark.** The related definitions about matrix pencil are provided in Appendix B. The general method to compute the eigenvalues of a matrix pencil is based on the generalized Schur decomposition (15) (see Definition 5 in Appendix B). When matrix $\mathbf{B}$ is non-singular, the set of eigenvalues of matrix pencil $\mathbf{A} - \mu \mathbf{B}$ is the same as the set of eigenvalues of matrix $\mathbf{AB}^{-1}$.

Next, we rewrite the constraints of the dual problem (14) as

$$
(1 + \frac{1}{\gamma_{m,i}})^{-1} \left( \mathbf{I} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \right) - \lambda_{m,i} \mathbf{R}_{m,m,i} \succeq 0,
\forall (m, i). \hspace{1cm} (15)
$$

Since correlation matrices $\mathbf{R}_{m,m,i} \succeq 0, \forall n, m \in \mathcal{M}, \forall i \in \mathcal{I}$, we have $\left( \mathbf{I} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \right) \succeq 0$ and $\mathbf{R}_{m,m,i} \succeq 0$. According to Lemma 2, inequalities in (15) are equivalent to

$$\lambda_{m,i} \leq
\mu_+ \left( (1 + \frac{1}{\gamma_{m,i}})^{-1} \left( \mathbf{I} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \right), \mathbf{R}_{m,m,i} \right),
\forall (m, i). \hspace{1cm} (16)
$$

Furthermore, since the right hand side of (16) also contains the dual variables, we define the following functions,

$$J_{m,i}(\lambda) = \mu_+ \left( (1 + \frac{1}{\gamma_{m,i}})^{-1} \left( \mathbf{I} + \sum_{n,j} \lambda_{n,j} \mathbf{R}_{m,n,j} \right), \mathbf{R}_{m,m,i} \right),
\forall (m, i). \hspace{1cm} (17)
$$

Combining (16) and (17), we have $\lambda \leq J(\lambda)$, where $J(\lambda) = [J_{1,1}(\lambda), J_{1,2}(\lambda), \ldots, J_{M,K}(\lambda)]^T$. Thus far, the dual problem (14) is equivalent to

$$\max_{\lambda \geq 0} \sigma^T \lambda \quad \text{s.t.} \quad \lambda \leq J(\lambda),$$

where $\sigma = [\sigma_{1,1}^2, \sigma_{1,2}^2, \ldots, \sigma_{M,K}^2]^T$.

To solve the above problem, we propose an iterative dual computation method which converges under asynchronous implementation. Briefly, the method is to update the dual variables with $J(\lambda)$. We first propose the following theorem about its synchronous implementation.

**Theorem 1.** From any feasible initial point $\lambda(0)$, the sequence generated by the following iterations,

$$\lambda(t + 1) = J(\lambda(t)), \quad t = 0, 1, 2, \ldots,$$

converges to a unique point $\lambda^*$ which satisfies $\lambda^* = J(\lambda^*)$, and is the optimal solution of the dual problem.

**Proof:** See Appendix C

**Remark.** The key point of proving the above theorem is to prove that the specific function $J(\lambda)$ is a standard function as defined in [13]. Besides, the work in [36] proved the convergence of a similar fixed point iterations for instantaneous CSI. Moreover, the key feature of the synchronous implementation is that when updating at the $t$th iteration, all the dual variables are updated with the same input dual variable $\lambda(t)$. When the dual variables are updated with the different input dual variables, the iterative dual computation method is implemented asynchronously.
We employ the asynchronous algorithm model in [9]. Define \( \lambda(f_{m,i}(t)) \) as the input dual variables used for generating \( \lambda_{m,i}(t+1) \), where

\[
\lambda(f_{m,i}(t)) = [\lambda_{1,1}(f_{m,i}^{1,1}(t)), \lambda_{1,2}(f_{m,i}^{1,2}(t)), \ldots, \lambda_{M,K}(f_{m,i}^{M,K}(t))]^T.
\]

(18)

\[
f_{m,i}(t) = [f_{m,i}^{1,1}(t), f_{m,i}^{1,2}(t), \ldots, f_{m,i}^{M,K}(t)]^T.
\]

(19)

Function \( f_{m,i}^{n,j}(t) \in \{0,1,\ldots,t-1, t\} \) is the time index of dual variable \( \lambda_{n,j} \) used for updating \( \lambda_{m,i}(t+1) \). So, \( \lambda_{n,j}(f_{m,i}^{n,j}(t)) \) is the value of the dual variable \( \lambda_{n,j} \) used for updating \( \lambda_{m,i}(t+1) \).

The asynchronous model also captures the case that the dual variables are not updated in every iteration. Denote by \( T_{m,i} \) the set of time indexes at which dual variable \( \lambda_{m,i} \) is updated. For example, \( T_{m,i} = \{1,3,7, \ldots\} \). We assume that \( T_{m,i} \) is infinite, and, provided any time \( t_0 \), there is \( t_1 \) such that \( f_{m,i}^{n,j}(t) \geq t_0 \), for all \( t_1 \). The physical meaning of this assumption is that the transmission of the dual variables does not fail in all the iterations.

The convergence of the iterative dual computation method under asynchronous implementation is formally given as the following theorem.

**Theorem 2.** From any feasible initial point \( \lambda(0) \), the sequence generated by the following iterations, \( \forall (m,i) \),

\[ \lambda_{m,i}(t+1) = \left\{ \begin{array}{ll} J_{m,i}(\lambda(f_{m,i}(t))), & t \in T_{m,i}, \\ \lambda_{m,i}(t), & \text{otherwise,} \end{array} \right. \]

converges to a unique point \( \lambda^* \) which satisfies \( \lambda^* = J(\lambda^*) \), and is the optimal solution of the dual problem.

**Proof:** We have proved that \( J(\lambda) \) is a standard function in Appendix C. The proof of this theorem is also based on this property. With the reasoning similar to [13], this theorem can be proved based on Theorem 1 and the conclusions from [9] pp. 431-433.

In summary, the iterative dual computation method converges to the same optimal point \( \lambda^* \) under both synchronous and asynchronous implementations.

### B. Methods for Computing Primal Variables

We derive the methods for computing the primal variables in this section.

1) **Beamforming:** Thanks to the dual decomposition, each cell’s beamforming problem relates to one sub-primal problem. The optimality conditions for the \( m \)th cell’s sub-primal problem are

\[
\frac{\partial L_m(p_{m,i}^T, \lambda_{m,i})}{\partial w_{m,i}} = \left( I - \left( 1 + \frac{1}{\gamma_{m,i}} \right) \lambda_{m,i} R_{m,m,i} \right) w_{m,i} + \sum_{n,j} \lambda_{n,j} R_{m,n,j} w_{m,i} = 0, \quad \forall i \in I,
\]

(20)

which are equivalent to

\[
w_{m,i} = \text{null} \left( I - \left( 1 + \frac{1}{\gamma_{m,i}} \right) \lambda_{m,i} R_{m,m,i} + \sum_{n,j} \lambda_{n,j} R_{m,n,j}, \forall i \in I. \right.
\]

(21)

Furthermore, define

\[
E_{m,i}(\lambda) = I - \left( 1 + \frac{1}{\gamma_{m,i}} \right) \lambda_{m,i} R_{m,m,i} + \sum_{n,j} \lambda_{n,j} R_{m,n,j}
\]

which is Hermitian. If the matrix \( E_{m,i}(\lambda^*) \) is full rank, the null space will be a zero vector. From the computation of the dual variables, we can see that the matrix \( E_{m,i}(\lambda^*) \) has an eigenvalue to be 0. Thus, \( E_{m,i}(\lambda^*) \) is not full rank and has a non-zero null space.

It requires that the beamforming of each cell can only be performed after the dual variables are optimal. The beamforming could be delayed for both numerical reasons, and practical reasons, like transmission delay. Our method enables beamforming to be performed at any time.

When \( E_{m,i}(\lambda) \) is full rank, we have the eigendecomposition of it as \( V_{m,i} \Lambda_{m,i} V_{m,i}^{-1} \), where \( \Lambda_{m,i} \) is a diagonal matrix and \( V_{m,i} \) is a unitary matrix. Let

\[
k' = \arg \min_{k=1, \ldots, N} \| V_{m,i} \Lambda_{m,i} V_{m,i}^{-1} \|.
\]

(22)

Define a diagonal matrix \( \tilde{\Lambda}_{m,i} \) whose components are the same as \( \Lambda_{m,i} \) except \( [\tilde{\Lambda}_{m,i}]_{k',k'} = 0 \). The beamforming vector can be computed as,

\[
w_{m,i} = \left( V_{m,i} \tilde{\Lambda}_{m,i} V_{m,i}^{-1} \right) .
\]

(22)

**Remark.** This method provides a set of suboptimal beamforming vectors when \( \lambda \) is not optimal. Most importantly, this method does not affect the result when the dual variables are optimal.

2) **Power Control:** Going back to the master problem with the optimal beamforming vectors, the master problem reduces to a typical power control problem [13].

\[
\min_{p \geq 0} \sum_{m,i} p_{m,i} \quad \text{s.t.} \quad p_{m,i} \geq I_{m,i}(p), \forall (m,i), \quad (23)
\]

where

\[
I_{m,i}(p) = \left( \sum_{n,j} p_{n,j} g(n,j),(m,i) + \sigma_{m,i}^2 \right) \left( 1 + \frac{1}{\gamma_{m,i}} \right) g_{(m,i),(m,i)}^{-1} \forall (m,i).
\]

(24)

with power gains \( g(n,j),(m,i) = w_n^H R_{n,m,i} w_n, \forall m,n \in \mathcal{M}, \forall i \in I \).

Furthermore, let \( I(p) = [I_{1,1}(p), I_{1,2}(p), \ldots, I_{N,K}(p)]^T \).

Since \( p \geq I(p) \), the solution of problem (23) can be computed with the following iterative method,

\[
p(t+1) = I(p(t)).
\]

(25)
Moreover, this iterative power exchange method can be implemented asynchronously in a distributed fashion as given in \[13\].

V. ASYNCHRONOUS DISTRIBUTED BEAMFORMING AND POWER CONTROL

In this section, we introduce the asynchronous distributed beamforming and power control algorithm which is based on the above computation methods. For ease of demonstrating the features of the algorithm, we describe the algorithm in the form of threads which are given in Table I. The threads can work in parallel without strict coordination. Besides, with the proposed algorithm, the information exchange among BSs is illustrated in Fig. 1.

The CSI thread is used to collect the long-term CSI. Each user can estimate the long-term CSI, from its associated BS to it, and, from the neighboring BSs to it. This can be done through pilot channels. The users send the long-term CSI to their BSs. When a BS interferes with a user of another cell, or its user is interfered with by other BSs, it exchanges the long-term CSI with the related neighboring BSs. Comparing to the algorithms using instantaneous CSI, the communication overheads here are relatively low. The larger $T_{\text{CSI}}$ is, the smaller the time average communication overhead is. Moreover, we assume $T_{\text{CSI}} \gg \max \{T_{\text{DC}}, T_{\text{BF}}, T_{\text{PC}}\}$, i.e., the long-term CSI can be viewed as fixed for the other threads.

The DC thread is the key to the asynchronous implementation of the algorithm. Each BS runs a DC thread which updates the dual variables of its users. Mathematically, the update of a dual variable needs all the dual variables. Indeed, in addition to its own dual variables, a BS only needs to know the dual variables of the users interfered by it. The BSs communicate via the backhaul network. In practice, transmission delay and packet loss could happen, which means that some BSs use outdated information. According to Theorem 2 the BSs can still compute the optimal dual variables in this situation. Furthermore, the BSs do not need to update the dual variables simultaneously, nor with the same frequency. As to the computational complexity, each BS needs to compute the generalized eigenvalues. If using the QZ method \[35\], the computational complexity of a BS per iteration is about $O(K N^3)$.

Each BS runs a BF thread to compute the beamforming vectors of its users. With the beamforming computation method in \[22\], the update frequency of the BF thread does not need to coordinate with the DC thread. The needed dual variables in the computation are fetched from the DC thread. The main computation load of the BF thread is to compute the eigenvalues and the null spaces of $K$ matrices with size $N \times N$.

The PC thread runs at each user. Each user computes the required power with equation \[24\] based on the local information: the total received signal power and the power gain between it and its BS. Each user sends the demanded power to its BS and the BS allocates the user the required power. This kind of power control can be understood as "power on demand" \[8\]. Besides, this iterative power control method \[24\] falls in the iterative power control framework in \[13\]. Thus, the users can update the power asynchronously and distributively. Besides, the computational complexity for each user is low.

In summary, the DC thread is the key to the asynchronous implementation of the algorithm. As long as the optimal dual variables are achieved, the optimal beamforming and power control can be realized by the BSs and the users asynchronously. In other words, if the iterative dual computation method could not converge under asynchronous implementation, the beamforming and power control algorithm could
not converge asynchronously either. With Theorem 2, we conclude that the asynchronous distributed beamforming and power control algorithm can provide the optimal solution to the master problem.

VI. SIMULATION

In this section, we demonstrate the performance of the proposed beamforming and power control algorithm via simulation. The downlink channel model captures the path loss and the spatial correlation. Let $R_{n,m,i} = (d_{n,m,i})^\delta R_{n,m,i}$, $\forall m,n \in M$, $\forall i \in I$, where $d_{n,m,i}$ is the distance from BS $n$ to user $(m,i)$ and $\chi = -3$. According to [3], the spatial correlation matrix is approximated by, $\forall m,n \in M$, $\forall i \in I$, 

$$
\begin{bmatrix}
[R_{n,m,i} (\theta_{n,m,i}, \sigma_\theta )]_{k,l}
\end{bmatrix}_{k,l} = e^{j\pi (k-l) \sin \theta_{m,i}} e^{-\frac{\sigma_\theta (k-l) \sin \theta_{m,i}}{10}}, \quad k,l = 1,\ldots,N,
$$

where $\theta_{n,m,i}$ means that user $(m,i)$ locates at $\theta_{n,m,i}$ relative to the array broadside of BS $n$. Each user is surrounded by local scatters corresponding to a spread angle of $\sigma_\theta = 2^\circ$. Without loss of generality, the SINR targets of the users are assumed to be the same. The other simulation settings are given as follows, if not specified. Let $N = 4$, $\sigma_\theta^2 = 10^{-12}$, $\gamma_{m,i} = 0.1$, $\forall (m,i)$. The distance between two BSs is 2000. The users are randomly located within their cells.

As pointed above, the iterative dual computation method is the key to the asynchronous implementation of the whole algorithm. The asynchronous implementation here is to let each dual variable has a positive probability being not updated. The probability is denoted by $P$. This setup simulates the cases that the BSs do not update simultaneously and that packet loss or delay happens.

Fig. 2(a) shows the convergence of the iterative dual computation method under asynchronous implementation with $P = 0.5$. In case 1, we set $\lambda(0) \leq J(\lambda(0))$ resulting in increasing sequences. In case 2, we set $\lambda(0) \geq J(\lambda(0))$ resulting in decreasing sequences. To show the effect of the asynchronous implementation, the convergence of the iterative dual computation method under synchronous implementation is illustrated in Fig. 2(b). Comparing Fig. 2(a) and Fig. 2(b), we notice that the method under asynchronous implementation has a slower convergence speed. Besides, Fig. 3 shows the convergence of the iterative computation method with $M = 7$, $K = 2$, $P = 0.5$, where only the dual variables of each cell’s first user are shown.

To further confirm this observation, we study the influence of the asynchronous implementation with $M = 4$, $K = 4$. The four BSs are put at the four corners of a square. Table I shows the average iteration numbers needed for the iterative dual computation method to converge under different values of $P$. The convergence criteria is $\|\lambda(t+1) - \lambda(t)\| \leq 10^{-5}$. We can see that, as $P$ increases, the average number of iterations also increases.

Next, we show the convergence of the asynchronous distributed beamforming and power control algorithm with $M = 4$, $K = 4$. In [7], the authors prove that the optimum of the SDP-relaxed master problem (1) is the same as the master problem as long as the master problem is feasible. Based on this conclusion, we use CVX to compute the optimum of the SDP-relaxed master problem as a benchmark. It is plotted as the black dot curve in Fig. 4. Furthermore, we assume $T_{\text{CSI}} \gg T_{\text{DC}} = T_{\text{BF}} = T_{\text{DC}}$. Specifically, the CSI is assumed to be fixed. In each iteration, the BSs update the beamforming vectors once and the users update the power allocations once. The dual variables are updated according to the probability $P$ in each iteration. Fig. 4 shows that the beamforming and power control algorithm converges to the benchmark under different $P$. Besides, Fig. 4 shows that when $P$ is large, the total transmission power could have large fluctuations. When $P = 0$, the algorithm converges quickly. This manifestation reminds us that in practice reliable backhaul networks are
Fig. 3. The convergence of the iterative dual computation method under asynchronous implementation. $P = 0.5$, $M = 7$, $K = 2$. The seven cells are located according to the hexagonal model similar to [22] Fig. 2.

Fig. 4. Convergence of the asynchronous distributed beamforming and power control algorithm with different update failure probability. $M = 4$, $K = 4$.

important for the performance of the algorithm.

If a BS computes the beamforming vectors and power allocations treating the inter-cell interference as noise, we call this kind of beamforming the single-cell beamforming. The performance gain of multi-cell beamforming over single-cell beamforming is illustrated in Fig. 5. Each curve in Fig. 5 is plotted via 1000-time averaging. Only the setup, with which the master problem is feasible, is used.

In both the cases $M = 4$, $K = 2$ and $M = 4$, $K = 3$, the averaged power consumption of the multi-cell beamforming is less than that of the single-cell beamforming. Furthermore, the power consumption of the single-cell beamforming increases more quickly than the multi-cell beamforming. It is because the single-cell beamforming only coordinates intra-cell interference. When the SINR target increases, a BS needs to increase power to increase the SINR. When all BSs do so, the inter-cell interference received by the users will increase. As a result, the BSs need to further increase power. However, when using multi-cell beamforming, inter-cell interference is coordinated. The BSs increase power mainly for increasing the users’ SINRs.

Besides, the increase of the total power consumption implies that certain BS’s power consumption must increase. In practice, the maximum transmission power of a BS is constrained. Considering each BS’s power constraint, the method used in [6], [22] can be used to find the minimum power consumption. Briefly, the power minimization problem can be decomposed into two levels by decomposing the BSs’ power constraints. Our method here can be used to solve the lower level problem.

VII. CONCLUSION

In this paper, our main contribution is an asynchronous distributed beamforming and power control algorithm for the multi-cell networks. The algorithm is derived via investigating a multi-cell PMBP which assumes long-term CSI, non-reciprocal channel, and different noise variances at each user. Via exploring the special problem structure, we are able to propose a novel asynchronous iterative dual computation method to compute the dual of the multi-cell PMBP. Besides, beamforming and power control could be done within each cell by the BSs and the users asynchronously in a distributed fashion. At last, the convergence and performance of the algorithm are also demonstrated via simulation.

APPENDIX A

PROOF OF PROPOSITION [1]

We first introduce the definitions of Z-matrix, P-matrix, and K-matrix for the proof of this proposition.

**Definition 1.** (Definition 2) i) A matrix $M \in \mathbb{R}^{n \times n}$ is called Z-matrix if its off-diagonal entries are all non-positive; ii) A matrix $M \in \mathbb{R}^{n \times n}$ is called P-matrix if it reverses the sign of no nonzero vector, i.e., $x_i[Mx] \leq 0 \Rightarrow x = 0$; iii) If a matrix is both Z-matrix and P-matrix, it is called K-matrix.

The K-matrix is related to the matrix radius according to the following lemma.
Lemma 3. (Lemma 1) Let \( M \in \mathbb{R}^{n \times n} \) be a K-matrix, \( N \in \mathbb{R}^{n \times n} \) be a non-negative matrix, then \( \rho(M^{-1}N) < 1 \) if and only if \( M - N \) is a K-matrix.

According to Lemma 1 we need to prove that the inequalities (6) is the necessary and sufficient condition for \( \rho(\Gamma G) < 1 \). Since \( \rho(\Gamma G) = \rho(I^{-1}\Gamma G) < 1 \) where \( I \) is a K-matrix, we conclude that \( I - \Gamma G \) is a K-matrix according to Lemma 3.

Furthermore, \( I - \Gamma G \) is a P-matrix whose components are

\[
[I - \Gamma G]_{(m,i),(n,j)} = \begin{cases} -\frac{\gamma_{m,i}}{w_{m,i} R_{m,m} + w_{m,i} R_{m,n} - w_{n,j} R_{n,m} + w_{n,j} R_{n,n}}, & (m,i) \neq (n,j) \\ 1, & (m,i) = (n,j). \end{cases}
\]

According to the definition of P-matrix, we have

\[
p_{m,i}[(I - \Gamma G)p]_{m,i} > 0, \forall (m,i),
\]

where \([I - \Gamma G]p\) is the component of vector \((I - \Gamma G)p\) corresponding to user \((m,i)\).

If inequalities (27) hold for all \( p_{m,i} > 0 \), we obtain \([I - \Gamma G]p_{m,i} > 0, \forall (m,i)\), which are equivalent to (6).

APPENDIX B
Proof of Lemma 2

To prove this lemma, we resort to the theoretical results about the semidefiniteness of a matrix pencil [11], [12], [35]. We first introduce some necessary terminologies about matrix pencils, i.e., the following two definitions, which can be found in [35] pp. 375-377. The eigenvalues of a matrix pencil are defined as follows.

Definition 2. (Eigenvalue and Eigenvector of a Matrix Pencil)
The set of all the form \( A - \mu B \) with \( \mu \in \mathbb{C} \) is said to be a pencil. The eigenvalues of the pencil are elements of the set \( \mu(A, B) \) defined by

\[
\mu(A, B) = \{ z \in \mathbb{C} : \det(A - zB) = 0 \}.
\]

If \( \mu \in \mu(A, B) \) and

\[
A x = \mu B x, \ x \neq 0,
\]

then \( x \) is referred to as an eigenvector of \( A - \mu B \).

Definition 3. (Generalized Schur Decomposition) If \( A \) and \( B \) are in \( \mathbb{C}^{n \times n} \), then there exist unitary \( Q \) and \( Z \) such that \( Q^H A Z = T \) and \( Q^H B Z = S \) are upper triangular. If for some \( k, \sigma_{kk} \) and \( s_{kk} \) are both zero, then \( \mu \in \mu(A, B) = \mathbb{C} \). Otherwise

\[
\mu \in \mu(A, B) = \{ \frac{t_{ii}}{s_{ii}} : s_{ii} \neq 0 \}.
\]

With the above definitions, we start the formal proof of Lemma 2. More strict proof of similar results in the general case can be found in the study about semidefiniteness intervals of matrix pencil [11 Corollary 3.7], [12 Theorem 2.1]. Our proof is given as follows.

A. Proof of Existence and Finiteness of \( \mu_+(A, B) \):
According to [10], we have

\[
\mu_+(A, B) = \min_{x^H B x = 1} x^H A x.
\]

Since \( A \succ 0 \), \( B \succeq 0 \) by assumption, the minimum of \( \min_{x^H B x = 1} x^H A x \) exists and is finite due to convexity.

B. Proof of \( A - \mu B \succeq 0 \Rightarrow \mu \leq \mu_+(A, B) \):

Since \( A - \mu B \) is positive semidefinite, we have

\[
x^H A x - \mu x^H B x \geq 0, \forall x \neq 0
\]

(28)

Since \( A \succ 0 \), \( B \succeq 0 \), we have \( x^H A x > 0 \) and \( x^H B x \geq 0 \).

If \( x \in \{ z : z^H B z = 0, z \neq 0 \} \), then \( \mu \geq 0 \) by assumption.

If \( x \in \{ z : z^H B z > 0, z \neq 0 \} \), then \( \mu \leq \frac{x^H A x}{x^H B x} \) based on (28).

Furthermore, \( \mu \leq \frac{x^H A x}{x^H B x} \) is equivalent to

\[
\mu \leq \min_{x : x^H B x = 1} \frac{x^H A x}{x^H B x}.
\]

The value of \( \min_{x : x^H B x = 1} \frac{x^H A x}{x^H B x} \) is equivalent to the optimal value of the following problem

\[
\min_{x : x^H B x = 1} x^H A x \quad \text{s.t.} \quad x^H B x = 1,
\]

because we can always scale \( x \) such that \( x^H B x = 1 \). According to [10], the minimizing vector is the eigenvector of matrix pencil \( A - \mu B \), whereas the minimum is the corresponding eigenvalue. Thus, we can conclude that \( \mu \leq \mu_+(A, B) \), i.e., being smaller or equal to the minimum non-negative eigenvalue of matrix pencil \( A - \mu B \).

C. Proof of \( A - \mu B \succeq 0 \IFF \mu \leq \mu_+(A, B) \):

This proof is based on the determinant of the matrix pencil \( A - \mu B \).

Let \( \beta_1(\mu) \geq \beta_2(\mu) \geq \ldots \geq \beta_N(\mu) \) be the eigenvalues of matrix \( A - \mu B \) given \( \mu \). We have

\[
\det(A - \mu B) = \prod_{k=1}^{N} \beta_k(\mu).
\]

In addition, we have the following lemma.

Lemma 4. Eigenvalues \( \beta_k(\mu), k = 1, \ldots, N \), are continuous functions of \( \mu \).

Proof: Let \( \mu_0 \in (0, \mu_+(A, B)) \). To prove the continuity is to prove

\[
\lim_{\epsilon \to 0} \beta_k(\mu_0 + \epsilon) = \lim_{\mu \to \mu_0} \beta_k(\mu) = \beta_k(\mu_0), \forall k.
\]

Let \( \lambda_k(A) \) represents the kth eigenvalue of matrix \( A \), assuming \( \lambda_k(A) \geq \lambda_{k+1}(A) \). We have

\[
\beta_k(\mu_0 + \epsilon) = \lambda_k(A - (\mu_0 + \epsilon)B) = \lambda_k(A - \mu_0 B - \epsilon B), \forall k.
\]

The matrix \( \epsilon B \) can be viewed as a perturbation to matrix \( A - \mu_0 B \).
Next, we resort to the matrix perturbation theory. Since \( A, B \) are symmetric matrices, according to the Hoffman-Wielandt Theorem [17] pp. 189), we have
\[
\sum_k \left( \lambda_k (A - \mu_0 B - \epsilon B) - \lambda_k (A - \mu_0 B) \right)^2 \leq \epsilon^2 ||B||_F^2,
\]
where \( || \cdot ||_F \) represents Frobenius matrix norm. We can see that
\[
(\lambda_k (A - \mu_0 B - \epsilon B) - \lambda_k (A - \mu_0 B))^2 \geq 0, \forall k.
\]
In addition, since \( ||B||_F^2 \) is finite, \( \lim_{\epsilon \to 0} ||B||_F^2 = 0 \), Therefore, we can conclude that
\[
\lim_{\epsilon \to 0} (\lambda_k (A - \mu_0 B - \epsilon B) - \lambda_k (A - \mu_0 B))^2 = 0, \forall k.
\]
That is, \( \lim_{\epsilon \to 0} \beta_k (\mu_0 + \epsilon) = \beta_k (\mu_0), \forall k \). This completes the proof.

Moreover, based on Definition [3] we have
\[
\det(A - \mu B) = \det(T - \mu S) \det(Q) \det(Z) = \prod_{k=1, \ldots, N} (t_{kk} - \mu s_{kk}).
\]
With (30), we have
\[
\det(A - \mu B) = \prod_{k=1, \ldots, N} (t_{kk} - \mu s_{kk}) = \prod_{k=1, \ldots, N} \beta_k (\mu).
\]
(31)

It is easy to see that if \( \mu = 0 \), \( \det(A - \mu B) > 0 \) and all the eigenvalues are positive. Let \( \mu \) increase continuously from 0. From the polynomial in (31), we can see that the determinant keeps to be positive until \( \mu = \mu_+(A, B) \) where \( \det(A - \mu B) = 0 \). With Lemma [4] it implies that no \( \beta_k (\mu) \) changes sign during the process. So, we can conclude that if \( 0 \leq \mu \leq \mu_+(A, B) \), then \( A - \mu B \succeq 0 \).

**APPENDIX C**

**PROOF OF THEOREM [1]**

The proof of this theorem is motivated by the work of [13]. We need to prove \( J(\lambda) \) is a standard function as defined in [13]. That is, for all \( \lambda \geq 0 \), the following properties are satisfied:

- Positivity: If \( \lambda \geq 0 \), \( J(\lambda) \geq 0 \).
- Monotonicity: If \( \lambda \geq \lambda' \), \( J(\lambda) \geq J(\lambda') \).
- Scalability: \( \alpha J(\lambda) \geq J(\alpha \lambda), \forall \alpha > 1 \).

If \( J(\lambda) \) is a standard function, then the remaining proof is identical to the proof in [13]. Our contribution here is to prove that the specific function \( J(\lambda) \) is standard.

**A. Proof of Positivity:**

The dual variables \( \lambda \) are greater than or equal to zero trivially. Since \( J_{m,i}(\lambda), \forall (m, i) \), are the non-negative eigenvalues, we have \( J(\lambda) \geq 0 \).

**B. Proof of Monotonicity:**

To prove the monotonicity, we first propose the following lemma.

**Lemma 5.** Let \( N \times N \) Hermitian matrices \( B, C, D \succeq 0 \). If \( \beta' \geq \beta \geq 0 \), then
\[
\mu_+(C + \beta'D, B) \geq \mu_+(C + \beta D, B),
\]
(32)

**Proof:** Let \( \mu_+ = \mu_+(C + \beta D, B) \) and \( v \) be the corresponding generalized eigenvector. Similarly, let \( \mu'_+ = \mu_+(C + \beta'D, B) \) and \( u \) be the corresponding generalized eigenvector.

From equation (29), we have
\[
\mu_+ = u^H (C + \beta D) v,
\]
while
\[
\mu'_+ = u'^H (C + \beta'D) u = u'^H (C + \beta D) u + (\beta' - \beta) u'^H D u.
\]
Since \( v \) minimize \( u'^H (C + \beta D) x \) subject to \( u'^H B x = 1 \), we have
\[
u'^H (C + \beta D) u \geq u'^H (C + \beta D) v.
\]
Furthermore, since \( \beta' \geq \beta \) and \( D \succeq 0 \), we have
\[
(\beta' - \beta) u'^H D u \geq 0.
\]
Then, we can conclude that \( \mu'_+ \geq \mu_+ \).

Suppose that, from \( \lambda \) to \( \lambda' \), only \( \lambda_{n,j} \) changes to \( \lambda'_{n,j} \) while the other entries remain unchanged.

Let
\[
B = R_{m,m,i},
\]
\[
C = \left(1 + \frac{1}{\gamma_{m,i}} \right)^{-1} \left(I + \sum_{(i,j) \neq (m,i)} \lambda_{i,j} R_{m,j,i} \right),
\]
\[
D = \left(1 + \frac{1}{\gamma_{m,i}} \right)^{-1} R_{m,j,i},
\]
(33)

According to Lemma [5] if \( \lambda'_{n,j} \geq \lambda_{n,j} \), then
\[
\mu_+(C + \lambda'_{n,j} D, B) \geq \mu_+(C + \lambda_{n,j} D, B),
\]
which means \( J_{m,i}(\lambda') \geq J_{m,i}(\lambda), \forall (m, i) \).

If multiple entries increase from \( \lambda \) to \( \lambda' \), it is equivalent to let them increase one by one. Consequently, if \( \lambda \geq \lambda' \), \( J(\lambda) \geq J(\lambda') \) according to Lemma [5].

**C. Proof of Scalability:**

Below, we prove that \( \alpha J_{m,i}(\lambda) > J_{m,i}(\alpha \lambda), \forall (m, i) \), with \( \alpha > 1 \).

Let
\[
B = R_{m,m,i},
\]
\[
C = \left(1 + \frac{1}{\gamma_{m,i}} \right)^{-1} \sum_{n,j} \alpha \lambda_{n,j} R_{m,n,j},
\]
\[
D = \left(1 + \frac{1}{\gamma_{m,i}} \right)^{-1} I.
\]
(34)

Then, we have
\[
J_{m,i}(\alpha \lambda) = \mu_+(C + D, B),
\]
and
\[ \alpha J_{\lambda}(\lambda) = \mu_+(C + \alpha D, B) \]
because of
\[ \mu_+(C + \alpha D, B) = \min_{x \in \mathbb{R}_{m,m}, x=1} x^H \left( 1 + \frac{1}{\rho_{m,t}} \right)^{-1} \left( \alpha I + \sum_{n,j} \alpha \lambda_{n,j} R_{m,n,j} \right) x \]
\[ = \alpha J_{\lambda}(\lambda). \]
Recall that, in the proof of Lemma 5 if \( D > 0 \), \((\beta')^H D^H D u > 0\), then
\[ \mu_+(C + \beta' D, B) > \mu_+(C + \beta D, B). \]
Since \( \alpha > 1 \), according to Lemma 5 we can conclude that
\[ \alpha J(\lambda) > J(\lambda), \forall \alpha > 1. \]
With the above results, the theorem can be proved by a line of reasoning similar to that in [13]. The sketch of the remaining proof is as follows. First prove that the uniqueness of the fixed point. If \( \lambda(0) \geq J(\lambda(0)) \), the sequences are decreasing. If \( \lambda(0) < J(\lambda(0)) \), the sequences are increasing. The sequences converge to the unique fixed point.

REFERENCES
[1] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, “Transmit beamforming and power control for cellular wireless systems,” IEEE J. Sel. Areas Commun., vol. 16, no. 8, pp. 1437–1450, Oct. 1998.
[2] E. Visotsky and U. Madhow, “Optimum beamforming using transmit antenna arrays,” in Proc. Veh. Technol. Conf. (VTC), Houston, TX, 1999, pp. 851–856.
[3] M. Bengtsson and B. Ottersten, “Optimal downlink beamforming using semidefinite optimization,” in Proc. 37th Annu. Allerton Conf. Commun., Control and Computing, 1999, pp. 987–996.
[4] M. Schubert and H. Boche, “Solution of the multiuser downlink beamforming problem with individual SINR constraints,” IEEE Trans. Veh. Technol., vol. 53, no. 1, pp. 18–28, Jan. 2004.
[5] A. Wiesel, Y. Eldar, and S. Shamai, “Linear precoding via conic optimization for fixed MIMO receivers,” IEEE Trans. Signal Process., vol. 54, no. 1, pp. 161–176, Jan. 2006.
[6] W. Yu and T. Lan, “Transmit optimization for the multi-antenna downlink with per-antenna power constraints,” IEEE Trans. Signal Process., vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
[7] M. Bengtsson and B. Ottersten, “Optimal and suboptimal transmit beamforming,” in Handbook of Antennas in Wireless Communications, 1st ed. CRC Press, 2001, pp. 18–1–18–33.
[8] X. Lin and T. M. Lok, “Distributed power control for one-to-many transmissions in gaussian interference channels,” IEEE Trans. Commun., vol. 60, no. 8, pp. 2363–2375, Aug. 2012.
[9] D. P. Bertsekas and I. N. Tsitsiklis, Parallel and distributed computation. Prentice Hall Inc., 1989.
[10] G. Stewart, “Perturbation bounds for the definite generalized eigenvalue problem,” Linear algebra and its applications, vol. 23, pp. 69–85, Feb. 1979.
[11] J. Kovač-Striko and K. Veselić, “Trace minimization and definiteness of symmetric pencils,” Linear algebra and its applications, vol. 216, pp. 139–158, Feb. 1995.
[12] X. Liang, R.-C. Li, and Z. Bai, “Trace minimization principles for positive semi-definite pencils,” Linear Algebra and Its Applications, vol. 438, no. 7, pp. 3058–3106, Apr. 2013.
[13] R. D. Yates, “A framework for uplink power control in cellular radio systems,” IEEE J. Sel. Areas Commun., vol. 13, no. 7, pp. 1341–1347, Sep. 1995.
[14] C. Botella, G. Pinto, A. Gonzalez, and M. Dieg, “Coordination in a multi-cell multi-antenna multi-user w-cdma system: A beamforming approach,” IEEE Trans. Wireless Commun., vol. 7, no. 11, pp. 4479–4485, Nov. 2008.
[15] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, “Distributed multiecell beamforming with limited intercell coordination,” IEEE Trans. Signal Process., vol. 59, no. 2, pp. 728–738, Feb. 2011.
[16] 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.
[17] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, “Convex optimization-based beamforming,” IEEE Signal Process. Mag., vol. 27, no. 3, pp. 62–75, May 2010.
[18] N. Vucic and H. Boche, “Robust QoS-Constrained optimization of downlink multiuser MISO systems,” IEEE Trans. Signal Process., vol. 57, no. 2, pp. 714–725, Feb. 2009.
[19] Y. Huang and D. Palomar, “A dual perspective on separable semidefinite programming with applications to optimal downlink beamforming,” IEEE Trans. Signal Process., vol. 58, no. 8, pp. 4271–4277, Aug. 2010.
[20] G. Marteau, X. Gou, W. Arief, and G. Ascheid, “On the duality of the max-min beamforming problem with per-antenna and per-antenna-array power constraints,” IEEE Trans. Veh. Technol., vol. 62, no. 2, pp. 606–619, Feb. 2013.
[21] Z. Xiang, M. Tao, and X. Wang, “Coordinated multicast beamforming in multiecell networks,” IEEE Trans. Wireless Commun., vol. 12, no. 1, pp. 12–21, Jan. 2013.
[22] D. Cai, T. Quek, C.-W. Tan, and S. Low, “Max-min sir coordinated multipoint downlink transmission–duality and algorithms,” IEEE Trans. Signal Process., vol. 60, no. 10, pp. 5384–5395, Oct. 2012.
[23] Y. Yeong, T. Q. Quek, and H. Shin, “Beamforming optimization for multiuser two-tier networks,” IEEE J. Commun. Networks, vol. 13, no. 4, pp. 327–338, Aug. 2011.
[24] J.-H. Noh and S.-J. Oh, “Beamforming in a multi-user cognitive radio system with partial channel state information,” IEEE Trans. Wireless Commun., vol. 12, no. 2, pp. 616–625, Feb. 2013.
[25] R. Zhang and S. Cui, “Cooperative interference management with MISO beamforming,” IEEE Trans. Signal Process., vol. 58, no. 10, pp. 5450–5458, Oct. 2010.
[26] G. Lim and T. Ciminini, L.J., “Energy-efficient cooperative beamforming in clustered wireless networks,” IEEE Trans. Wireless Commun., vol. 12, no. 3, pp. 1376–1385, Mar. 2013.
[27] K.-Y. Wang, N. Jacklin, Z. Ding, and C.-Y. Chi, “Robust MISO transmit optimization under outage-based QoS constraints in two-tier heterogeneous networks,” IEEE Trans. Wireless Commun., vol. 12, no. 4, pp. 1883–1897, 2013.
[28] S. Chen, H. Fang, and T. M. Lok, “Design of downlink beamformer for real-time and non-real-time services,” in Proc. IEEE Wireless Commun. and Networking Conf. (WCNC), Shanghai, 2013, pp. 3529–3534.
[29] R. A. Horn and C. R. Johnson, Matrix analysis. Cambridge Univ. press, 1990.
[30] J.-S. Pang, G. Scutari, D. Palomar, and F. Facchinei, “Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach,” IEEE Trans. Signal Process., vol. 58, no. 6, pp. 3251–3271, Jun. 2010.
[31] O. Somekh, O. Simeone, Y. Bar-Ness, and A. Haimovich, “CTH11-2: Distributed multi-cell zero-forcing beamforming in cellular downlink channels,” in Proc. Global Telecommun. Conf. (GLOBECOM), San Francisco, CA, 2006, pp. 1–6.
[32] M. Vemula, A. D幼儿, J. Ling, and C. Papadias, “Inter-cell coordination, opportunistic beamforming and scheduling,” in Proc. IEEE Inter. Conf. Communications (ICC), Istanbul, 2006, pp. 5319–5324.
[33] R. Stridh, M. Bengtsson, and B. Ottersten, “System evaluation of optimal downlink beamforming with congestion control in wireless communication,” IEEE Trans. Wireless Commun., vol. 5, no. 4, pp. 743–751, Apr. 2006.
[34] M. Grant, S. Boyd, and Y. Ye, “CVX: Matlab software for disciplined convex programming,” 2008.
[35] G. H. Golub and C. F. Van Loan, Matrix computations. Johns Hopkins Univ. Press, 1996.
[36] D. Cai, T. Quek, and C.-W. Tan, “A unified analysis of max-min weighted sir for mimo downlink system,” IEEE Trans. Signal Process., vol. 59, no. 8, pp. 3850–3862, Aug. 2011.
[37] G. W. Stewart and J.-g. Sun, Matrix perturbation theory. Academic press, 1990.