This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

SINR Balancing via Base Station Association, Beamforming and Power Control in Downlink Multicell MISO Systems

Mengjie Xie, *Student Member, IEEE*, and Tat-Ming Lok, *Senior Member, IEEE*

***Abstract*—This paper considers the downlink channel of multicell multiuser multiple-input single-output (MISO) systems with arbitrary architecture. We aim to maximize the minimum weighted signal-to-interference-plus-noise ratio (SINR) through user-base station (BS) association, coordinated beamforming among BSs and power control subject to per BS power constraints. The problem is known to be NP-hard. In the high-SNR regime where the performance is interference-limited, we globally optimize the upper bound, which is achieved asymptotically, of the balanced SINR. In the low-SNR regime where the performance is restricted by the tightest per BS power constraint, we balance the transmit power using relaxed integer programming (RIP) and group sparse optimization (GSO) techniques. The two goals can be integrated to achieve a universally better performance. From the perspective of optimization methods, the corresponding algorithms are based on Lagrangian relaxation and are guaranteed to converge. Simulation results show that the proposed algorithms outperform the existing algorithms designed for the downlink single-input single-output (SISO) systems.**

***Index Terms*—downlink multicell MISO; SINR balancing; max-min SINR; base station association; relaxed integer programming; group sparse optimization.**

I. Introduction

The concept of coordinated multipoint (CoMP) has been introduced to deal with inter-cell interference in LTE-advanced and WiMAX. In CoMP networks, interference can be exploited by joint processing of the data or mitigated by coordinated scheduling and beamforming [1]. As data traffic grows and spectral efficiency approaches its limit, node density is urged to be increased. A strategy is to deploy low-power nodes together with high-power nodes to form a heterogeneous network (HetNet) [2]. Considering the imbalanced BS power budgets and density of deployments, a more flexible user-BS association scheme is required to balance the load and power in HetNet [3]. Meanwhile, the emergence of cloud radio access network (C-RAN) offers a practical platform for joint processing. In C-RAN, BSs can be clustered, either fully or partially, statically or dynamically, and share data within the clusters [4]. BS clustering yields a many-to-many relationship for user-BS association. This paper addresses the SINR balancing problem via joint design of the following techniques: user-BS association, coordinated beamforming and power control in a general downlink multicell MISO system, and extensively, dynamic BS clustering when the architecture supports data sharing in backhaul links.

SINR is a commonly used metric for quality-of-service (QoS). The SINR balancing problem, also known as the max-min weighted SINR problem, has been widely studied for decades. For single-antenna systems, the concept of SINR balancing is first introduced in [5] and later applied to the cellular system in [6]. [7] solves the problem over power assignment in single-cell downlink MISO systems. [8]–[10] study the SINR balancing problem using nonlinear Perron-Frobenius theory. [7]–[10] focus on single-cell systems subject to a single (weighted) sum power constraint. However, in downlink systems with multiple power constraints, the active power constraint at optimum and the noise variance in the dual uplink problem are uncertain [11]–[13]. Concerning per BS power constraints, Cai et al. [11] relax the multiple power constraints to be a single weighted sum power constraint followed by subgradient-based update of the weights; Huang et al. [13] use the subgradient projection method to update the uncertain noise variance in the dual problem.

Beside multiple per BS power constraints, user-BS association introduces extra challenge to the SINR balancing problem because of its interdependence with channel matrices and power constraints. It has been stressed early in [14] that there is no Pareto optimal solution for joint power control and BS association in downlink multicell SISO systems, and later in [15] and [16] that the problem is NP-hard in general. [16] studies joint power control and BS association under per BS power constraints in downlink multicell SISO systems, which is linked with the uplink problem in [15] by the same sum power. The authors also propose an improved algorithm to manage the imbalanced power budget in HetNet [16]. To the best of our knowledge, the problem investigated by [16] is closest to what this paper does, except that beamforming is not involved in downlink multicell SISO systems in [16].

In this paper, we target at SINR balancing via joint BS association, beamforming and power control in downlink multicell MISO systems subject to per BS power constraints. We deconstruct the problem from a new perspective by extracting two factors that limit the performance, namely, interference and the tightest power constraint. In the interference-limited case, we optimize the asymptotic upper bound of the balanced SINR. Concerning the tightest power constraint, we balance the transmit power at BSs using relaxed integer programming (RIP) [4] and group sparse optimization (GSO) [17]

This work was supported in part by the General Research Fund from the Research Grants Council, Hong Kong Special Administrative Region, under Project CUHK 14203616.

Mengjie Xie and Tat M. Lok are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: {xm012, tmlok}@ie.cuhk.edu.hk.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

techniques. The two goals are combined ultimately to resolve the SINR balancing problem. To achieve these various goals, we propose a novel Lagrangian relaxation framework that accommodates various algorithms.

The rest of the paper is structured as follows. Section II introduces the system model and formulates the problem. Section III establishes some preliminary theoretical results to guide the design of algorithms. In section IV, we propose algorithms converging to the global optimum of the asymptotic upper bound of the balanced SINR in the high SNR regime. Section V starts with RIP-based and GSO-based approaches to balance the transmit power and enhance the performance at low SNR, followed by an integrated algorithm that solves the SINR balancing problem in universal SNR regime. Section VI demonstrates simulation results and Section VII concludes the paper.

II. System Model

We consider a downlink multicell MISO network with K multi-antenna BSs and N single-antenna users sharing the channel without frequency or time extensions. Denote by $\mathcal{K} = \{1, 2, \ldots, K\}$ and $\mathcal{N} = \{1, 2, \ldots, N\}$ the index sets for BSs and users, respectively. Each user $n \in \mathcal{N}$ is associated with exactly one BS with the index $\kappa_n \in \mathcal{K}$. Each BS $k \in \mathcal{K}$ may serve multiple users or not associate with any user. BS k has $M^{[k]}$ antennas and a power budget $P^{[k]}$.

BS k transmits a data stream $x_n(t)$ to user n with $\kappa_n = k$ using power p_n and a beamforming vector $\mathbf{v}_n^{[k]}$ of dimension $M^{[k]} \times 1$, where the data stream and the beamformer are normalized such that $\mathbb{E}[x_n(t)^\star x_n(t)] = 1$ and $\mathbf{v}_n^{[k]H}\mathbf{v}_n^{[k]} = 1$. We assume $N > M^{[k]}$ for all $k \in \mathcal{K}$ such that zero-forcing beamforming is not applicable and that each user receives interference from at least one other user. The transmitted signal to user n can be written as $\sqrt{p_n}\mathbf{v}_n^{[\kappa_n]}x_n(t)$. The received signal at user n is

$$\begin{aligned} y_n(t) =& \sqrt{p_n}\mathbf{h}_n^{[\kappa_n]H}\mathbf{v}_n^{[\kappa_n]}x_n(t) \\ &+ \sum_{\substack{m\neq n\\ m\in\mathcal{N}}} \sqrt{p_m}\mathbf{h}_n^{[\kappa_m]H}\mathbf{v}_m^{[\kappa_m]}x_m(t) + z_n \end{aligned} \tag{II.1}$$

where $\mathbf{h}_n^{[k]}$ is the $M^{[k]} \times 1$ channel state information (CSI) vector from BS k to user n. CSI is assumed to be known at the BS side. In simulation, the channel is modeled as Rayleigh flat fading and the entries of all $\mathbf{h}_n^{[k]}$ are drawn from i.i.d. zero-mean unit-variance circularly symmetric complex Gaussian distribution. The second term in (II.1) captures all intra-cell and inter-cell interfering signals. z_n is the additive white Gaussian noise (AWGN) with noise power η_n.

The received SINR of user n is given by

$$\Gamma_n = \frac{p_n g_{nn}}{\sum_{\substack{m\neq n\\ m\in\mathcal{N}}} p_m g_{nm} + \eta_n} \tag{II.2}$$

where $g_{nm} = ||\mathbf{h}_n^{[\kappa_m]H}\mathbf{v}_m^{[\kappa_m]}||^2$ is the effective channel gain or interference imposed by user m on user n.

The goal is to determine κ_n, $\mathbf{v}_n^{[\kappa_n]}$, and p_n for all $n \in \mathcal{N}$, in order to balance the weighted SINR $\frac{\Gamma_n}{\gamma_n}$ for max-min fairness. γ_n may reflect some long-term priority of user n.

The problem is mixed-integer programming and is NP-hard [15], [16]. In Section V, RIP and GSO techniques are adopted to relax the problem, where a many-to-many relationship is allowed for user-BS association in intermediate steps. Therefore, before formulating the problem, we first introduce consistent descriptions of the association strategy $\mathbb{W}_n$ and the beamforming strategy $\mathbb{V}_n$ of user n in place of the integer mapping function κ_n and the beamforming vector $\mathbf{v}_n^{[\kappa_n]}$.

The association strategy of user n is represented by

$$\mathbb{W}_n = \left(w_n^{[1]}, w_n^{[2]}, \ldots, w_n^{[K]}\right) \tag{II.3}$$

where

$$w_n^{[k]} = \begin{cases} 1, & k = \kappa_n \\ 0, & k \neq \kappa_n \end{cases} \tag{II.4}$$

The feasible set of $\mathbb{W}_n$ is denoted by

$$\mathcal{W}_n = \left\{ \mathbb{W}_n : w_n^{[k]} \in \{0,1\}, \forall k \in \mathcal{K}; \sum_{k\in\mathcal{K}} w_n^{[k]} = 1 \right\} \tag{II.5}$$

Later in Section V, the feasible set is relaxed so that $w_n^{[k]}$ is not restricted to be binary.

Let the beamforming strategy of user n consist of the beamformers at all K BSs, given by

$$\mathbb{V}_n = \left(\mathbf{v}_n^{[1]}, \mathbf{v}_n^{[2]}, \ldots, \mathbf{v}_n^{[K]}\right) \tag{II.6}$$

The feasible set of $\mathbb{V}_n$ is denoted by

$$\mathcal{V}_n = \left\{ \mathbb{V}_n : \mathbf{v}_n^{[k]H}\mathbf{v}_n^{[k]} = 1, \forall k \in \mathcal{K} \right\} \tag{II.7}$$

Note that $\mathbf{v}_n^{[k]}$ for $k \neq \kappa_n$ are redundant arguments which will not take effect in actual transmission. However, we reserve these arguments which become effective in relaxed-integer programming in Section V.

Concerning per BS power constraints, we further construct a vector $\mathbf{w}^{[k]} = [w_1^{[k]}, w_2^{[k]}, \ldots, w_N^{[k]}]^T$ for each BS k and a power vector $\mathbf{p} = [p_1, p_2, \ldots, p_N]^T$. Now the SINR balancing problem can be formulated as

$$\begin{aligned} \max_{\mathbf{p},\ \mathbb{W}_n, \mathbb{V}_n, \forall n\in\mathcal{N}} \quad & \min_{n\in\mathcal{N}} \ \frac{\Gamma_n}{\gamma_n} \\ \text{s.t.} \quad & \mathbf{w}^{[k]T}\mathbf{p} \leq P^{[k]},\ \forall k \in \mathcal{K} \\ & \mathbb{V}_n \in \mathcal{V}_n,\ \forall n \in \mathcal{N} \\ & \mathbb{W}_n \in \mathcal{W}_n,\ \forall n \in \mathcal{N} \end{aligned} \tag{P}$$

In the first constraint, $\mathbf{w}^{[k]T}\mathbf{p} = \sum_{n:\kappa_n=k} p_n$ is the total transmit power at BS k constrained by the power budget $P^{[k]}$. In the second and third constraints, the feasible sets of the beamforming strategies and association strategies are given by (II.7) and (II.5), where the beamformers $\mathbf{v}_n^{[k]}$ are orthonormal and $w_n^{[k]}$ are binary.

In terms of feasibility of problem (P), the power vector $\mathbf{p}$ can be viewed as BS-centric, constrained by the power budgets of BSs. The beamforming strategy $\mathbb{V}_n$ can be considered as user-centric, which means each user can make a feasible decision without being constrained by others. The association strategy $\mathbb{W}_n$ is user-centric with respect to the feasible set

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

$\mathcal{W}_n$, but is meanwhile restricted by the BS-centric power constraints.

For briefness of notation, we consider $\mathbb{W}_n$ as user-centric and further define $\mathbb{S}_n = (\mathbb{W}_n, \mathbb{V}_n)$ with the feasible set $\mathcal{S}_n = \mathcal{W}_n \times \mathcal{V}_n$ to combine the association strategy and beamforming strategy of user n. The strategy profile of all users can be denoted by $\mathbb{S} = (\mathbb{S}_1, \mathbb{S}_2, \dots, \mathbb{S}_N)$ with the feasible set $\mathcal{S} = \prod_{n\in\mathcal{N}} \mathcal{S}_n$. Problem (P) is restated as the following problem (P′) in terms of the strategy profile $\mathbb{S}$.

$$\begin{aligned} \max_{\mathbf{p},\,\mathbb{S}} \quad & \min_{n\in\mathcal{N}} \frac{\Gamma_n}{\gamma_n} \\ \text{s.t.} \quad & \mathbf{w}^{[k]T}\mathbf{p} \le P^{[k]},\ \forall k \in \mathcal{K} \\ & \mathbb{S} \in \mathcal{S} \end{aligned} \tag{P′}$$

The NP-hardness of the joint design problem has been established in [15] and [16]. Therefore, we design algorithms to solve the relaxed problem in the rest of the paper.

III. Preliminary Discussion

In this section, we discuss the SINR balancing problem (P′) with respect to the power allocation vector $\mathbf{p}$ and the strategy profile $\mathbb{S}$ separately.

A. Revisiting Downlink Power Control

First, for a fixed strategy profile $\mathbb{S}$, problem (P′) is reduced to the classic downlink power control problem.

$$\begin{aligned} \max_{\mathbf{p}} \quad & \min_{n\in\mathcal{N}} \frac{\Gamma_n}{\gamma_n} \\ \text{s.t.} \quad & \mathbf{w}^{[k]T}\mathbf{p} \le P^{[k]},\ \forall k \in \mathcal{K} \end{aligned} \tag{$\mathrm{P}_{\mathbb{S}}$}$$

We have assumed that each user will receive interference from at least one other user, which implies that the weighted SINR of all users are equal at optimum. Denote the optimal value by

$$\frac{1}{\rho^*} = \frac{\Gamma_n^*}{\gamma_n},\ \forall n \in \mathcal{N} \tag{III.1}$$

A closed-form expression of ρ^* and the corresponding power allocation vector $\mathbf{p}^*$ are given in [11]. For consistency with [11], we define the cross channel matrix $\mathbf{F}$ of dimension $N\times N$ and a vector $\hat{\boldsymbol{\gamma}}$ of length N as follows.

$$[\mathbf{F}]_{nm} = \begin{cases} 0, & n = m \\ g_{nm}, & n \neq m \end{cases} \tag{III.2}$$

$$\hat{\boldsymbol{\gamma}} = \left[\frac{\gamma_1}{g_{11}}, \frac{\gamma_2}{g_{22}}, \dots, \frac{\gamma_N}{g_{NN}}\right]^T \tag{III.3}$$

The dependence of the effective channel gain g_{nm} upon the strategy profile $\mathbb{S}$ is omitted in the above notation. Let $\mathbf{D}(\hat{\boldsymbol{\gamma}})$ represent the diagonal matrix with the elements of $\hat{\boldsymbol{\gamma}}$ on the diagonal. Define $\mathbf{C}^{[k]} = \mathbf{D}(\hat{\boldsymbol{\gamma}})(\mathbf{F} + \frac{1}{P^{[k]}}\boldsymbol{\eta}\mathbf{w}^{[k]T})$ and $j = \arg\max_k \rho(\mathbf{C}^{[k]})$, then the optimal value and solution are given by

$$\rho^* = \rho(\mathbf{C}^{[j]}) \tag{III.4}$$

$$\mathbf{p}^* = \frac{P^{[j]}}{\mathbf{w}^{[j]T}\mathbf{x}(\mathbf{C}^{[j]})}\mathbf{x}(\mathbf{C}^{[j]}) \tag{III.5}$$

where $\rho(\cdot)$ and $\mathbf{x}(\cdot)$ denote the Perron-Frobenius (PF) eigenvalue and the right PF eigenvector.

The closed-form expression of ρ^* reveals that the optimal value is restricted by the tightest power constraint which corresponds to the largest $\rho(\mathbf{C}^{[k]})$.

It can also be noticed that the power constraints vanish in the interference-limited case. When the noise power $\boldsymbol{\eta} = \mathbf{0}$, the optimal ρ^* and the corresponding optimal power allocation vector $\mathbf{p}^*$ become

$$\rho^* = \rho(\mathbf{C}) \tag{III.6}$$

$$\mathbf{p}^* = a \cdot \mathbf{x}(\mathbf{C}) \tag{III.7}$$

where $\mathbf{C} = \mathbf{D}(\hat{\boldsymbol{\gamma}})\mathbf{F}$ and a can be any positive scaling factor.

From the above discussion on the closed-form solution to ($\mathrm{P}_{\mathbb{S}}$), we observe two limiting factors in the the original problem (P′): one is the tightest per BS power constraint, which depends on the association strategy and power allocation; the other is the effective channel gain or interference, which depends on the association strategy and beamforming strategy.

In addition, the uplink-downlink duality also has a Perron-Frobenius characterization as follows.

$$\rho(\mathbf{C}^{[j]}) = \rho\left(\mathbf{D}(\hat{\boldsymbol{\gamma}})(\mathbf{F}^T + \frac{1}{P^{[j]}}\mathbf{w}^{[j]}\boldsymbol{\eta}^T)\right) \tag{III.8}$$

In the dual uplink problem, $\boldsymbol{\eta}$ corresponds to the weight on power and $\mathbf{w}^{[j]}$ corresponds to the noise vector. Now we have a clear vision of the difficulty in applying uplink-downlink duality to solve the problem ($\mathrm{P}_{\mathbb{S}}$): the power constraint $P^{[j]}$ and the noise $\mathbf{w}^{[j]}$ are uncertain in the dual uplink problem. Moreover, $\mathbf{w}^{[k]}$ are also variables in the original problem (P′).

B. Feasible SINR Balancing Level

In the rest of this paper, we refer to $\frac{1}{\rho} = \min_n \frac{\Gamma_n}{\gamma_n}$ as the balanced SINR and $\rho = \max_n \frac{\gamma_n}{\Gamma_n}$ as the SINR balancing level.

In this subsection, we study the feasible SINR balancing level ρ with respect to the strategy profile $\mathbb{S}$ in the interference-limited case where the noise power approaches zero or the power budget approaches infinity. To be specific, we relax the explicit power constraints to be the finite-power requirement and allow for any noise power $\boldsymbol{\eta} \ge \mathbf{0}$. In the interference-limited case, $\rho(\mathbf{C})$ is the asymptotic lower bound of all feasible ρ. Any $\rho > \rho(\mathbf{C})$ is referred to as feasible in the sense that it can be achieved with finite power.

Theorem 1 interprets the feasibility of ρ for $\boldsymbol{\eta} > \mathbf{0}$ more rigorously without recourse to the closed-form solution. Before that, we define the following matrices

$$\mathbf{G}(\rho) = \rho\mathbf{D}^{-1}(\hat{\boldsymbol{\gamma}}) - \mathbf{F} \tag{III.9}$$

$$\mathbf{A}(\rho) = \mathbf{G}(\rho)\mathbf{D}(\hat{\boldsymbol{\gamma}}) \tag{III.10}$$

$$\mathbf{B} = \rho\mathbf{I} - \mathbf{A}(\rho) \tag{III.11}$$

It is assumed that each user will receive interference from or impose interference to another user, which implies that $\mathbf{F}$ is irreducible. Following the definitions (III.9) to (III.11), $\mathbf{G}(\rho)$, $\mathbf{A}(\rho)$ and $\mathbf{B}$ are all irreducible matrices.

Theorem 1. *ρ is feasible in the interference-limited case, namely, there exists a positive and finite power allocation* *vector such that $\frac{\Gamma_n}{\gamma_n} \geq \frac{1}{\rho}, \forall n \in \mathcal{N}$ in the presence of noise, if and only if any of the following statements holds:*

1) *There exists $\mathbf{0} < \mathbf{p} < \infty$ such that $\mathbf{G}(\rho)\mathbf{p} \geq \boldsymbol{\eta}$ for any $\boldsymbol{\eta} > 0$;*
2) *$\mathbf{G}(\rho)$ is a non-singular M-matrix;*
3) *$\mathbf{A}(\rho)$ is a non-singular M-matrix;*
4) *$\rho(\mathbf{B}) < \rho$.*

Proof. The equivalence between the main statement and statement 1) can be obtained directly by expanding $\mathbf{G}(\rho)\mathbf{p} \geq \boldsymbol{\eta}$ entry-wise.

According to (III.9) and (III.10), $\mathbf{G}(\rho) \in \mathbf{Z}^{n,n}$ and $\mathbf{A}(\rho) \in \mathbf{Z}^{n,n}$ where $\mathbf{Z}^{n,n}$ denotes the set of all $n \times n$ real matrices whose off-diagonal elements are less than or equal to zero. Next, we prove the equivalence between statements 1) and 2), 2) and 3), and 3) and 4) sequentially.

- Condition K_{33} [18, Theorem 1] states that $\mathbf{G}(\rho) \in \mathbf{Z}^{n,n}$ is a non-singular M-matrix if and only if there exists $\mathbf{x} > \mathbf{0}$ with $\mathbf{G}(\rho)\mathbf{x} > \mathbf{0}$. Therefore, statement 2) holds given statement 1). On the other direction, let $\mathbf{b} = \mathbf{G}(\rho)\mathbf{x}$ for such $\mathbf{x}$, then we can find a $\mathbf{p} = \max_n \frac{\eta_n}{b_n}\mathbf{x}$ such that $\mathbf{G}(\rho)\mathbf{p} \geq \boldsymbol{\eta}$. Therefore, statement 1) holds given statement 2).
- The condition that there exists $\mathbf{x} > \mathbf{0}$ with $\mathbf{G}(\rho)\mathbf{x} > \mathbf{0}$, is equivalent to that there exists $\mathbf{y} = \mathbf{D}^{-1}(\hat{\gamma})\mathbf{x} > \mathbf{0}$ with $\mathbf{A}(\rho)\mathbf{y} = \mathbf{G}(\rho)\mathbf{x} > \mathbf{0}$. Therefore, statements 2) and 3) are equivalent.
- According to [18, Definition], $\mathbf{A}(\rho)$ is expressed in the form $\mathbf{A}(\rho) = \rho\mathbf{I} - \mathbf{B}$ where $\mathbf{B}$ has non-negative elements and $\rho(\mathbf{B}) \leq \rho$, then $\mathbf{A}(\rho)$ is an M-martix. Furthermore, $\mathbf{A}(\rho)$ is a non-singular M-matrix if and only if $\rho(\mathbf{B}) < \rho$.

□

Extensively, when $\rho = \rho(\mathbf{B})$, $\mathbf{A}(\rho)$ and $\mathbf{G}(\rho)$ are singular. The following lemma can be established in the absence of noise, i.e., for $\boldsymbol{\eta} = \mathbf{0}$. The proof is similar to that of Theorem 1. The corresponding properties of irreducible M-matrices, either singular or non-singular, can be found in [19, Theorem 3.4].

Lemma 1. *ρ is feasible in the absence of noise if and only if any of the following statements holds:*

1) *There exists $\mathbf{0} < \mathbf{p} < \infty$ such that $\mathbf{G}(\rho)\mathbf{p} \geq \mathbf{0}$;*
2) *$\mathbf{G}(\rho)$ is an M-matrix;*
3) *$\mathbf{A}(\rho)$ is an M-matrix;*
4) *$\rho(\mathbf{B}) \leq \rho$.*

Noticing that $\mathbf{B} = \mathbf{F}\mathbf{D}(\hat{\gamma})$ and therefore $\rho(\mathbf{B}) = \rho(\mathbf{C})$, we can unify the intuitive results based on the closed-form solution with Theorem 1 and Lemma 1. In addition, if $\mathbf{G}(\rho)$ is an irreducible non-singular M-matrix, then $\mathbf{G}(\rho)$ is strictly inverse-positive [20, Theorem A]. That is, $\mathbf{G}^{-1}(\rho)$ exits and $\mathbf{G}^{-1}(\rho) > 0$. Then for any feasible ρ, the optimal power allocation vector $\mathbf{p}^*(\rho)$ that achieves the same weighted SINR for all users can be found directly by $\mathbf{p}^*(\rho) = \mathbf{G}^{-1}(\rho)\boldsymbol{\eta}$.

To conclude the section, the SINR balancing problem (P′) can be relaxed and solved by the following two main steps: (1) find the strategy profile $\mathbb{S}$ that yields a minimum feasible ρ satisfying the statements in Theorem 1 and Lemma 1; (2) for fixed $\mathbb{S}$, find the optimal power allocation vector satisfying the power constraints given by (III.5). In later sections, we will design algorithms based on these two main steps. Obviously, the difficulty lies mostly in the first step. We will propose a Lagrangian relaxation framework to approach the minimum value of ρ. Recall that in Subsection III.A, we intuitively summarized two factors that limits the minimum value of ρ: (1) interference; (1) the tightest power constraint. The two corresponding goals are: (1) to minimize the asymptotic lower bound of ρ; (2) to balance the transmit power. The statements in Theorem 1 and Lemma 1 serve as the feasibility conditions of ρ in the various problems in the Lagrangian relaxation framework.

IV. Asymptotic Lower Bound of the SINR Balancing Level

In this section, we propose algorithms that minimize the asymptotic lower bound of the SINR balancing level ρ. We adopt statement 3) in Lemma 1 to form the feasible region of the auxiliary variable ρ. That is, $\mathbf{A}(\rho)$ is an M-matrix, or equivalently, there exists $\mathbf{y} > \mathbf{0}$ such that $\mathbf{A}(\rho)\mathbf{y} \geq \mathbf{0}$. Because $\mathbf{B} = \rho\mathbf{I} - \mathbf{A}(\rho)$ and $\mathbf{B}$ is dependent on $\mathbb{S}$, $\mathbf{A}(\rho)\mathbf{y} \geq \mathbf{0}$ can be written as $\mathbf{B}(\mathbb{S})\mathbf{y} \leq \rho\mathbf{y}$. The problem is formed as follows.

$$\min_{\rho,\ \mathbb{S}\in\mathcal{S},\ \mathbf{y}>\mathbf{0}} \quad \rho \qquad (\mathrm{P}_1)$$
$$\text{s.t.} \quad \mathbf{B}(\mathbb{S})\mathbf{y} \leq \rho\mathbf{y}$$

We have discussed the condition for $\mathbf{A}(\rho)$ to be a singular M-matrix in last section. It can be easily proved that the above constraint is active if and only if $\rho = \rho(\mathbf{B}(\mathbb{S}))$ and $\mathbf{y} = \mathbf{x}(\mathbf{B}(\mathbb{S}))$.

A. Lagrangian Relaxation Method

To solve problem (P_1), we adopt the Lagrangian relaxation method [21] by penalizing the constraint with the Lagrangian multipliers $\boldsymbol{\lambda} \geq \mathbf{0}$. The Lagrangian function is $\mathcal{L}(\rho, \mathbb{S}, \mathbf{y}, \boldsymbol{\lambda}) = \rho + \boldsymbol{\lambda}^T\mathbf{B}(\mathbb{S})\mathbf{y} - \rho\boldsymbol{\lambda}^T\mathbf{y}$. By letting $\boldsymbol{\lambda}^T\mathbf{y} = 1$, $\mathcal{L}(\rho, \mathbb{S}, \mathbf{y}, \boldsymbol{\lambda})$ is equivalent to

$$\mathcal{L}'(\mathbb{S}, \mathbf{y}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^T\mathbf{B}(\mathbb{S})\mathbf{y} \qquad (\mathrm{IV}.1)$$

Due to the non-negative penalty term, $\rho \geq \mathcal{L}'(\mathbb{S}, \mathbf{y}, \boldsymbol{\lambda})$ for all jointly feasible ρ, $\mathbb{S}$, $\mathbf{y}$ and $\boldsymbol{\lambda}$. The partially dualized problem is

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0},\boldsymbol{\lambda}^T\mathring{\mathbf{y}}=1} \quad g(\boldsymbol{\lambda}) \qquad (\mathrm{P}_1^{\mathrm{Dual}})$$

where

$$g(\boldsymbol{\lambda}) = \min_{\mathbb{S}\in\mathcal{S},\mathbf{y}>\mathbf{0}} \mathcal{L}'(\mathbb{S}, \mathbf{y}, \boldsymbol{\lambda}) \qquad (\mathrm{P}_1^{\mathrm{in}})$$

is referred to as the inner problem, and $\mathring{\mathbf{y}}$ is the solution to $\mathbf{y}$ in the inner problem.

Lagrangian relaxation algorithms are typically based on the following principles: $g(\boldsymbol{\lambda})$ is a lower bound of the optimal result of (P_1) because $\rho^* \geq \mathcal{L}'(\mathbb{S}^*, \mathbf{y}^*, \boldsymbol{\lambda}) \geq g(\boldsymbol{\lambda})$ for any feasible $\boldsymbol{\lambda}$; solving ($\mathrm{P}_1^{\mathrm{Dual}}$) over $\boldsymbol{\lambda}$ gives a tighter bound; a sequence of $\boldsymbol{\lambda}$ can be found to guide $g(\boldsymbol{\lambda})$ towards ρ^*. For problem (P_1), the inner problem ($\mathrm{P}_1^{\mathrm{in}}$) can be solved efficiently. The problem is that when the dualized constraint has a complex structure as in (P_1), updating the dual variable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

via solving the dual problem does not guarantee convergence towards the optimal value.

For (P_1), the ideal choice of $\boldsymbol{\lambda}$ is what yields strong duality, in which case the penalty term is zero. We can impose the zero-penalty condition to get an updated $\hat{\boldsymbol{\lambda}}$ for the solution to (P_1^{in}), denoted by $\hat{\mathbb{S}}$ and $\mathring{\mathbf{y}}$, and a feasible $\hat{\rho}$ despite their optimality. The zero-penalty condition means that the relaxed constraint is active in (P_1), which happens only when

$$\hat{\rho} = \rho(\mathbf{B}(\hat{\mathbb{S}})) \tag{IV.2}$$

Therefore, $\hat{\boldsymbol{\lambda}}$ satisfying $\hat{\rho} = \hat{\boldsymbol{\lambda}}^T \mathbf{B}(\hat{\mathbb{S}})\mathring{\mathbf{y}}$ is given by the right PF eigenvector of $\mathbf{B}^T(\hat{\mathbb{S}})$ normalized such that $\hat{\boldsymbol{\lambda}}^T \mathring{\mathbf{y}} = 1$, written as

$$\hat{\boldsymbol{\lambda}} = \frac{\mathbf{x}(\mathbf{B}^T(\hat{\mathbb{S}}))}{\mathring{\mathbf{y}}^T \mathbf{x}(\mathbf{B}^T(\hat{\mathbb{S}}))} \tag{IV.3}$$

$\hat{\boldsymbol{\lambda}}$ given above satisfies the zero-penalty condition for any random $\mathring{\mathbf{y}} > \mathbf{0}$, although such $\hat{\rho}$ and $\hat{\mathbb{S}}$ are only feasible with active constraints for a particular $\mathring{\mathbf{y}} = \mathbf{x}(\mathbf{B}(\hat{\mathbb{S}}))$ in problem (P_1). We stress here that the zero-penalty condition is simply an updating rule for the dual variable $\boldsymbol{\lambda}$. The convergent performance is analyzed in detail after we present the algorithm.

As a brief summary of this subsection, the Lagrangian relaxation algorithm to solve (P_1) is implemented as follows: (1) for a given $\hat{\boldsymbol{\lambda}}$, solve ($\mathrm{P}_1^{\text{in}}$) for $\hat{\mathbb{S}}$ and $\mathring{\mathbf{y}}$; (2) update $\hat{\boldsymbol{\lambda}}$ according to (IV.3); (3) repeat (1) and (2) till convergence.

B. Algorithm Design

First, we solve the inner problem (P_1^{in}), which is not elaborated in last subsection. Notice that the nth column of $\mathbf{B}(\mathbb{S})$ depends only on $\mathbb{S}_n$ and that $\mathbf{y}$ is always positive. (P_1^{in}) can be decomposed into N parallel sub-problems with the nth given by

$$\min_{\mathbb{S}_n \in \mathcal{S}_n} \quad f_n(\mathbb{S}_n, \boldsymbol{\lambda}) \tag{$\mathrm{P}_n^{\text{Sub}}$}$$

where $f_n(\mathbb{S}_n, \boldsymbol{\lambda})$ is the nth element of the vector

$$\mathbf{f}(\mathbb{S}, \boldsymbol{\lambda}) = \mathbf{B}^T(\mathbb{S})\boldsymbol{\lambda} \tag{IV.4}$$

Denote by $\hat{\mathbb{S}}_n$ the solutions to the sub-problems ($\mathrm{P}_n^{\text{Sub}}$). Accordingly, the strategy profile $\hat{\mathbb{S}} = (\hat{\mathbb{S}}_1, \hat{\mathbb{S}}_2, \ldots, \hat{\mathbb{S}}_N)$ is the solution to $\mathbb{S}$ in the inner problem (P_1^{in}). We see that these solutions are regardless of the value of $\mathbf{y}$, so any random positive vector $\mathring{\mathbf{y}}$ can be a solution to $\mathbf{y}$ in (P_1^{in}).

($\mathrm{P}_n^{\text{Sub}}$) is a mixed-integer programming problem. For given $\boldsymbol{\lambda}$ and fixed $\kappa_n = k$, $f_n(\cdot)$ is a function of the beamformer $\mathbf{v}_n^{[k]}$

$$f_n^{[k]}(\mathbf{v}_n^{[k]}) = \gamma_n \frac{\mathbf{v}_n^{[k]H} \mathbf{R}_n^{[k]} \mathbf{v}_n^{[k]}}{\mathbf{v}_n^{[k]H} \mathbf{Q}_n^{[k]} \mathbf{v}_n^{[k]}} \tag{IV.5}$$

where $\mathbf{Q}_n^{[k]} = \mathbf{h}_n^{[k]} \mathbf{h}_n^{[k]H}$ and $\mathbf{R}_n^{[k]} = \mathbf{H}_{-n}^{[k]} \mathbf{D}(\boldsymbol{\lambda}_{-n}) \mathbf{H}_{-n}^{[k]H}$ for $\mathbf{H}_{-n}^{[k]} = \left[\mathbf{h}_1^{[k]}, \ldots, \mathbf{h}_{n-1}^{[k]}, \mathbf{h}_{n+1}^{[k]}, \ldots, \mathbf{h}_N^{[k]}\right]$ and $\boldsymbol{\lambda}_{-n} = [\lambda_1, \ldots, \lambda_{n-1}, \lambda_{n+1}, \ldots, \lambda_N]^T$. We have assumed $N > M^{[k]}$ for all $k \in \mathcal{K}$, so $\mathbf{R}_n^{[k]}$ has full rank almost surely.

It turns out that ($\mathrm{P}_n^{\text{Sub}}$) can be further decomposed into K sub-problems with respect to $\mathbf{v}_n^{[k]}$, written as

$$\min_{\mathbf{v}_n^{[k]H} \mathbf{v}_n^{[k]} = 1} \quad f_n^{[k]}(\mathbf{v}_n^{[k]}) \tag{$\mathrm{P}_{n[k]}^{\text{Sub}}$}$$

The problem is equivalent to maximizing the Rayleigh quotient $\mathbf{v}_n^{[k]H} \mathbf{Q}_n^{[k]} \mathbf{v}_n^{[k]} / \mathbf{v}_n^{[k]H} \mathbf{R}_n^{[k]} \mathbf{v}_n^{[k]}$. The solution is known to be

$$\hat{\mathbf{v}}_n^{[k]} = \mathbf{v}_1\left((\mathbf{R}_n^{[k]})^{-1} \mathbf{Q}_n^{[k]}\right) \tag{IV.6}$$

and the minimized value is

$$f_n^{[k]}(\hat{\mathbf{v}}_n^{[k]}) = \gamma_n / \sigma_1\left((\mathbf{R}_n^{[k]})^{-1} \mathbf{Q}_n^{[k]}\right) \tag{IV.7}$$

where $\sigma_1(\cdot)$ is the largest eigenvalue and $\mathbf{v}_1(\cdot)$ is the corresponding eigenvector.

The beamforming strategy $\hat{\mathbb{V}}_n$ in ($\mathrm{P}_n^{\text{Sub}}$) consists of $\hat{\mathbf{v}}_n^{[k]}$ given by (IV.6) for all $k \in \mathcal{K}$. The association strategy $\hat{\mathbb{W}}_n$ is obtained by comparing $f_n^{[k]}(\hat{\mathbf{v}}_n^{[k]})$ over all $k \in \mathcal{K}$. Specifically, $\hat{\mathbb{W}}_n$ consists of $\hat{w}_n^{[k]}$ as follows

$$\hat{w}_n^{[k]} = \begin{cases} 1, & k = \arg\min_j f_n^{[j]}(\hat{\mathbf{v}}_n^{[j]}) \\ 0, & \text{otherwise} \end{cases} \tag{IV.8}$$

Intuitively, the optimal beamforming strategy $\hat{\mathbb{V}}_n$ is to find beamformers that maximize SIR of user n at all K BSs in the dual uplink channel, where the power allocation vector is given by $\boldsymbol{\lambda}$. The optimal association strategy $\hat{\mathbb{W}}_n$ is to find the BS that maximizes the SIR weighted by $\frac{1}{\lambda_n \gamma_n}$ for user n in the dual uplink channel.

It follows that the solution to ($\mathrm{P}_n^{\text{Sub}}$) is $\hat{\mathbb{S}}_n = (\hat{\mathbb{W}}_n, \hat{\mathbb{V}}_n)$ and the solution to (P_1^{in}) is $\hat{\mathbb{S}} = (\hat{\mathbb{S}}_1, \hat{\mathbb{S}}_2, \ldots, \hat{\mathbb{S}}_N)$.

Now the algorithm that solves (P_1) iteratively is listed below. Since $\mathring{\mathbf{y}} > \mathbf{0}$ can be chosen randomly in the inner problem (P_1^{in}), we let $\mathring{\mathbf{y}} = \mathbf{1}$.

Algorithm 1

```
Initialize λ(0) > 0, 1^T λ(0) = 1;
while |ρ(t) − ρ(t − 1)| > ϵ do
    Update t = t + 1;
    for n = 1 to N do
        Solve (P_n^Sub) for S_n(t);
    end
    Update λ(t) = x(B^T(S(t))), 1^T λ(t) = 1, and
      ρ(t) = ρ(B(S(t)));
end
Calculate p* by (III.5) for S* = S(t);
Output: S* and p*.
```

ϵ is a small positive value to deal with rounding error. Theoretically, ϵ can be set to zero. The convergence of Algorithm 1 is asserted by the following Proposition.

Proposition 1. *$\rho(t)$ decreases monotonically and converges to the global optimum of (P_1) in Algorithm 1.*

Proof. Let $m = \arg\max_n \frac{f_n(\mathbb{S}_n(t), \boldsymbol{\lambda}(t-1))}{\lambda_n(t-1)}$, we first prove the monotonic decrement of $\rho(t)$ by establishing

$$\begin{aligned} \rho(t) &\leq \frac{f_m(\mathbb{S}_m(t), \boldsymbol{\lambda}(t-1))}{\lambda_m(t-1)} \\ &\leq \frac{f_m(\mathbb{S}_m(t-1), \boldsymbol{\lambda}(t-1))}{\lambda_m(t-1)} = \rho(t-1) \end{aligned} \tag{IV.9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

The first inequality is obtained from the min-max version of the Collatz-Wielandt formula [22, Theorem 2.7], namely $\min_{\mathbf{x}>\mathbf{0}} \max_{n} (\mathbf{A}\mathbf{x})_n/x_n = \rho(\mathbf{A})$ for an irreducible non-negative matrix $\mathbf{A}$. Equality holds if and only if $\boldsymbol{\lambda}(t-1) = \mathbf{x}(\mathbf{B}^T(\mathbb{S}(t))) = \boldsymbol{\lambda}(t)$, in which case $\frac{f_n(\mathbb{S}_n(t), \boldsymbol{\lambda}(t-1))}{\lambda_n(t-1)}$ is equal to $\rho(t)$ for all $n \in \mathcal{N}$. The second inequality is because $\mathbb{S}_m(t) = \arg\min_{\mathbb{S}_m} f_m(\mathbb{S}_m, \boldsymbol{\lambda}(t-1))$ according to line 5 in Algorithm 1. Equality holds if and only if $\mathbb{S}_m(t-1) = \mathbb{S}_m(t)$. For equalities to hold concurrently in the first and second inequalities, the condition $\mathbb{S}_m(t-1) = \mathbb{S}_m(t)$ should be satisfied for all $m \in \mathcal{N}$. The third equality is due to the way we set $\boldsymbol{\lambda}(t-1)$ and $\rho(t-1)$ in line 7 in Algorithm 1. Therefore, $\rho(t)$ decreases monotonically.

In addition, Algorithm 1 stops at $\rho(t) = \rho(t-1)$ when the strategy profile $\mathbb{S}(t) = \mathbb{S}(t-1)$. Because the beamforming strategy has closed-form solutions and the association strategy has limited number of combinations, Algorithm 1 does converge to a particular value $\hat{\rho} = \rho(\mathbf{B}(\hat{\mathbb{S}}))$. Next we prove the global optimality of such $\hat{\rho}$ and $\hat{\mathbb{S}}$.

The convergence of Algorithm 1 implies that $\hat{\mathbb{S}}_n = \arg\min_{\mathbb{S}_n} f_n(\mathbb{S}_n, \hat{\boldsymbol{\lambda}})$ for all $n \in \mathcal{N}$ with $\hat{\boldsymbol{\lambda}} = \mathbf{x}(\mathbf{B}^T(\hat{\mathbb{S}}_n))$. Denote the global optimal solution and value of (P$_1$) by $\mathbb{S}^*$ and ρ^*. Because $\rho \geq \rho(\mathbf{B}(\mathbb{S}))$, the minimal ρ^* must be equal to $\rho(\mathbf{B}(\mathbb{S}^*))$ with $\mathbf{y}^* = \mathbf{x}(\mathbf{B}(\mathbb{S}^*))$. Assume to the contrary that $\rho(\mathbf{B}(\mathbb{S}^*)) < \rho(\mathbf{B}(\hat{\mathbb{S}}))$. The max-min version of the Collatz-Wielandt formula [22, Theorem 2.7] says that $\max_{\mathbf{x}>\mathbf{0}} \min_{n} (\mathbf{A}\mathbf{x})_n/x_n = \rho(\mathbf{A})$. Therefore, for any $\boldsymbol{\lambda} > \mathbf{0}$, there exists a user m with $\frac{f_m(\mathbb{S}_m^*, \boldsymbol{\lambda})}{\lambda_m} \leq \rho(\mathbf{B}(\mathbb{S}^*)) < \rho(\mathbf{B}(\hat{\mathbb{S}}))$. Taking $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}$, we have $f_m(\mathbb{S}_m^*, \hat{\boldsymbol{\lambda}}) < \rho(\mathbf{B}(\hat{\mathbb{S}}))\hat{\lambda}_m = f_m(\hat{\mathbb{S}}_m, \hat{\boldsymbol{\lambda}})$, which contradicts with the optimality of $\hat{\mathbb{S}}_m$. Therefore, $\hat{\mathbb{S}}$ is the global optimizer and $\hat{\rho}$ is the global optimum of (P$_1$). □

As we have discussed in Section III.B, the optimal value $\rho^* = \rho(\mathbf{B}(\mathbb{S}^*))$ returned by Algorithm 1 is an asymptotic lower bound. Therefore, $\mathbb{S}^*$ and the corresponding $\mathbf{p}^*$ based on $\mathbb{S}^*$ can be approximated solutions to the original problem (P$'$) at high SNR.

C. A Variation of Algorithm 1

In Algorithm 1, the strategy $\mathbb{S}_n$ is determined by solving the sub-problem $(\text{P}_n^{\text{Sub}})$, which can be explained as maximization of the weighted SIR in the dual uplink channel for user n. It is noticed that the objective function $f_n(\cdot)$ in $(\text{P}_n^{\text{Sub}})$ depends only on the beamforming pattern but not the norm of the beamformers. In this sense, Algorithm 1 lacks the ability of power control. To compensate for this deficiency, we introduce a different metric and propose a variation of Algorithm 1, which helps incorporating power control into the design of algorithms in next Section. Define

$$\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) = \mathbf{v}_n^{[k]H}\left((f_n/\gamma_n)\mathbf{Q}_n^{[k]} - \mathbf{R}_n^{[k]}\right)\mathbf{v}_n^{[k]} \tag{IV.10}$$

where $\mathbf{Q}_n^{[k]}$ and $\mathbf{R}_n^{[k]}$ are the same as in Subsection IV.B, and

$$\phi_n(\mathbb{S}_n, \boldsymbol{\lambda}, f_n) = \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) \tag{IV.11}$$

In place of $(\text{P}_n^{\text{Sub}})$ and $(\text{P}_{n[k]}^{\text{Sub}})$, consider the following problems

$$\max_{\mathbb{S}_n\in\mathcal{S}_n} \quad \phi_n(\mathbb{S}_n, \boldsymbol{\lambda}, f_n) \qquad (\text{P}_n^{\text{Sub2}})$$

$$\max_{\mathbf{v}_n^{[k]H}\mathbf{v}_n^{[k]}=1} \quad \phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) \qquad (\text{P}_{n[k]}^{\text{Sub2}})$$

$(\text{P}_{n[k]}^{\text{Sub2}})$ has closed-form solution and optimal value given by

$$\hat{\mathbf{v}}_n^{[k]} = \mathbf{v}_1\left((f_n/\gamma_n)\mathbf{Q}_n^{[k]} - \mathbf{R}_n^{[k]}\right) \tag{IV.12}$$

$$\hat{\phi}_n^{[k]} = \sigma_1\left((f_n/\gamma_n)\mathbf{Q}_n^{[k]} - \mathbf{R}_n^{[k]}\right) \tag{IV.13}$$

Accordingly, the beamforming strategy $\hat{\mathbb{V}}_n$ consists of $\hat{\mathbf{v}}_n^{[k]}$ for all $k \in \mathcal{K}$. The association strategy $\hat{\mathbb{W}}_n$ consists of $\hat{w}_n^{[k]}$ given by

$$\hat{w}_n^{[k]} = \begin{cases} 1, & k = \arg\max_j \hat{\phi}_n^{[j]} \\ 0, & \text{otherwise} \end{cases} \tag{IV.14}$$

$\hat{\mathbb{W}}_n$ and $\hat{\mathbb{V}}_n$ constitute $\hat{\mathbb{S}}_n$ as the solution to $(\text{P}_n^{\text{Sub2}})$.

Intuitively, we are maximizing the weighted signal-minus-interference for user n in the dual uplink channel in $(\text{P}_n^{\text{Sub2}})$, instead of the weighted SIR in $(\text{P}_n^{\text{Sub}})$. Precisely, $(\text{P}_{n[k]}^{\text{Sub2}})$ is related to $(\text{P}_{n[k]}^{\text{Sub}})$ as follows. For fixed $\mathbf{v}_n^{[k]}$ and $\boldsymbol{\lambda}$, $\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n)$ is an increasing function of f_n. When $f_n = f_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda})$, $\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) = 0$. Consider the following steps: 1) start with some $\mathring{\mathbf{v}}_n^{[k]}$ and calculate $\mathring{f}_n = f_n^{[k]}(\mathring{\mathbf{v}}_n^{[k]}, \boldsymbol{\lambda})$ which yields $\phi_n^{[k]}(\mathring{\mathbf{v}}_n^{[k]}, \boldsymbol{\lambda}, \mathring{f}_n) = 0$; 2) solve $(\text{P}_{n[k]}^{\text{Sub2}})$ for $\hat{\mathbf{v}}_n^{[k]}$ which yields $\phi_n^{[k]}(\hat{\mathbf{v}}_n^{[k]}, \boldsymbol{\lambda}, \mathring{f}_n) \geq 0$; 3) update $\hat{f}_n = f_n^{[k]}(\hat{\mathbf{v}}_n^{[k]}, \boldsymbol{\lambda})$ which yields $\phi_n^{[k]}(\hat{\mathbf{v}}_n^{[k]}, \boldsymbol{\lambda}, \hat{f}_n) = 0$. It follows that $\hat{f}_n \leq \mathring{f}_n$. In this sense, $\hat{\mathbf{v}}_n^{[k]}$ is a better solution to $(\text{P}_{n[k]}^{\text{Sub}})$, compared with the original solution $\mathring{\mathbf{v}}_n^{[k]}$.

The variation of Algorithm 1 is listed below. The procedure is mostly the same as Algorithm 1, except that we initialize $\mathbf{f}(0) = \mathbf{f}(\mathbb{S}(0), \boldsymbol{\lambda}(0))$ for some random $\mathbb{S}(0)$ in line 1, solve $(\text{P}_n^{\text{Sub2}})$ instead of $(\text{P}_n^{\text{Sub}})$ in line 5, and update $\mathbf{f}(t) = \rho(t)\boldsymbol{\lambda}(t)$ in line 7.

Algorithm 1$'$

1 Initialize $\boldsymbol{\lambda}(0) > \mathbf{0}$, $\mathbf{1}^T\boldsymbol{\lambda}(0) = 1$, and $\mathbf{f}(0)$;
2 **while** $|\rho(t) - \rho(t-1)| > \epsilon$ **do**
3 Update $t = t + 1$;
4 **for** $n = 1$ *to* N **do**
5 Solve $(\text{P}_n^{\text{Sub2}})$ for $\mathbb{S}_n(t)$;
6 **end**
7 Update $\boldsymbol{\lambda}(t) = \mathbf{x}(\mathbf{B}^T(\mathbb{S}(t)))$, $\mathbf{1}^T\boldsymbol{\lambda}(t) = 1$, $\rho(t) = \rho(\mathbf{B}(\mathbb{S}(t)))$, and $\mathbf{f}(t) = \rho(t)\boldsymbol{\lambda}(t)$;
8 **end**
9 Calculate $\mathbf{p}^*$ by (III.5) for $\mathbb{S}^* = \mathbb{S}(t)$;
Output: $\mathbb{S}^*$ and $\mathbf{p}^*$.

In fact, Algorithm 1$'$ also falls into the category of the Lagrangian relaxation algorithms. Consider the following problem

$$\min_{\rho,\ \mathbb{S}\in\mathcal{S},\ \mathbf{y}>\mathbf{0}} \quad \rho \qquad (\text{P}_1')$$
$$\text{s.t.} \quad \mathbf{G}(\rho)\mathbf{y} \geq \mathbf{0}$$

Statement 2) of Lemma 1 is adopted as the feasibility condition of ρ in (P$'_1$), instead of statement 3) in (P$_1$). The corresponding inner problem of (P$'_1$) is to minimize the penalty of violating the feasibility condition of ρ, given by

$$\min_{\mathbb{S}\in\mathcal{S},\mathbf{y}>\mathbf{0}} \quad -\boldsymbol{\lambda}^T\mathbf{G}(\rho)\mathbf{y} \qquad (\text{P}^{\text{in}}_{1'})$$

Since $\phi_n(\mathbb{S}_n,\boldsymbol{\lambda},\rho\lambda_n) = (\mathbf{G}^T(\rho)\boldsymbol{\lambda})_n$, (P$^{\text{Sub2}}_n$) can be explained as sub-problems of the inner problem (P$^{\text{in}}_{1'}$), if f_n is updated to be $\rho\lambda_n$ as in line 7 in Algorithm 1$'$.

Algorithm1$'$ achieves the same goal as Algorithm 1. $\rho(t)$ decreases monotonically and converges to the global optimum of (P$'_1$) as well as (P$_1$). The proof is similar to that of Proposition 1. (IV.9) still holds with a different explanation of the second inequality. The second inequality can be written as $f'_m(t)/\lambda_m(t-1) \le f_m(t-1)/\lambda_m(t-1)$ where $f'_m(t) = f_m(\mathbb{S}_m(t),\boldsymbol{\lambda}(t-1))$. Because $\phi_m(\mathbb{S}_m(t),\boldsymbol{\lambda}(t-1),f_m(t-1)) \ge \phi_m(\mathbb{S}_m(t-1),\boldsymbol{\lambda}(t-1),f_m(t-1)) = 0 = \phi_m(\mathbb{S}_m(t),\boldsymbol{\lambda}(t-1),f'_m(t))$, we have $f'_m(t) \le f_m(t-1)$. In a word, maximizing $\phi_n(\mathbb{S}_n,\boldsymbol{\lambda},f_n)$ followed by the way of updating f_n decreases $f_n(\mathbb{S}_n,\boldsymbol{\lambda})$ indirectly.

From the above convergence analysis, it can be seen that as long as $\mathbb{S}_n(t)$ satisfies $\phi_n(\mathbb{S}_n(t),\boldsymbol{\lambda}(t-1),f_n(t-1)) \ge 0$ for all $n\in\mathcal{N}$ and $\boldsymbol{\lambda}(t)$ and $\mathbf{f}(t)$ are updated as in line 7, $\rho(t)$ is guaranteed to be non-incremental. To be specific, the non-incremental condition of ρ in the Lagrangian relaxation framework can be written as

$$\sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n) \ge 0, \quad \forall n\in\mathcal{N} \qquad \text{(IV.15)}$$

This is the key contribution of Algorithm 1$'$ to this paper and intrigues different design of the inner problem in the Lagrangian relaxation framework in next section.

V. Power Balancing for SINR Balancing

In this section, we include power control in the SINR balancing problem, which faces two major challenges: multiple per BS power constraints and dependence on the integer-valued association strategy. Referring to the closed-form expressions of ρ^* and $\mathbf{p}^*$ given by (III.4) and (III.5), the influence of power constraints is more notable in the noise-limited low SNR regime. The performance is limited by the tightest power constraint and thus can be improved via power balancing among BSs. This limitation may be crucial in HetNet where power budgets of BSs tend to variate considerably. Therefore, the power balancing level, denoted by $\alpha = \max_k \frac{1}{P^{[k]}}\mathbf{w}^{[k]T}\mathbf{p}$, is an important factor concerned in this section. To tackle the second challenge, we introduce relaxed integer programming in the following.

A. Power Balancing based on Relaxed Integer Programming

In previous sections, the association strategy $\mathbb{W}_n$ is composed of $w_n^{[k]} \in \{0,1\}$. In this section, the restriction is relaxed to be $0 \le w_n^{[k]} \le 1$, giving the following relaxed feasible set of $\mathbb{W}_n$.

$$\tilde{\mathcal{W}}_n = \left\{ \mathbb{W}_n : 0 \le w_n^{[k]} \le 1, \forall k\in\mathcal{K}; \sum_{k\in\mathcal{K}} w_n^{[k]} = 1 \right\} \qquad \text{(V.1)}$$

The feasible set of the beamforming strategy $\mathbb{V}_n$ is unchanged. Accordingly, we can construct the association strategy profile $\mathbb{W} = (\mathbb{W}_1,\mathbb{W}_2,\ldots,\mathbb{W}_N)$ and the beamforming strategy profile $\mathbb{V} = (\mathbb{V}_1,\mathbb{V}_2,\ldots,\mathbb{V}_N)$ with the feasible sets $\tilde{\mathcal{W}} = \prod_{n\in\mathcal{N}}\tilde{\mathcal{W}}_n$ and $\mathcal{V} = \prod_{n\in\mathcal{N}}\mathcal{V}_n$, respectively. The feasible set of the full strategy profile $\mathbb{S} = (\mathbb{W},\mathbb{V})$ is denoted by $\tilde{\mathcal{S}} = \tilde{\mathcal{W}}\times\mathcal{V}$. All previous notations, such as Γ_n, $\mathbf{G}(\rho)$, $\mathbf{A}(\rho)$ and $\mathbf{B}$, remain the same by rewriting the effective channel gain as

$$g_{nm} = \sum_{k\in\mathcal{K}} w_m^{[k]}||\mathbf{h}_n^{[k]H}\mathbf{v}_m^{[k]}||^2 \qquad \text{(V.2)}$$

The physical meaning of the relaxation is that each user n can be simultaneously associated with all K BSs providing the transmit power p_n with the proportion $w_n^{[k]}$. In other words, BS k transmits the data stream $x_n(t)$ with the beamformer $\mathbf{v}_n^{[k]}$ and power $p_n w_n^{[k]}$. Therefore, the effective channel gain is given by (V.2).

Consider the following RIP-based power balancing problem.

$$\begin{aligned} \min_{\alpha,\ \mathbb{S}\in\tilde{\mathcal{S}}} \quad & \alpha & (\text{P}_{\text{RIP}}) \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} w_n^{[k]}\hat{p}_n \le \alpha P^{[k]},\ \forall k\in\mathcal{K} \\ & \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n) \ge 0,\ \forall n\in\mathcal{N} \end{aligned}$$

where α is the power balancing level. $\hat{p}_n$ is the nth element of the estimated power allocation vector $\hat{\mathbf{p}}$. $\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n)$ follows the definition (IV.10) in previous section. The second constraint is exactly the non-incremental condition of the SINR balancing level ρ in the Lagrangian relaxation framework given by (IV.15).

The beamformer $\mathbf{v}_n^{[k]}$ only affects $\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n)$. We can maximize $\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n)$ over $\mathbf{v}_n^{[k]}$ to obtain the beamforming strategy profile $\hat{\mathbb{V}}$. Then for fixed $\hat{\mathbb{V}}$, (P$_{\text{RIP}}$) becomes a linear programming (LP) problem with respect to the variables $w_n^{[k]}, \forall n\in\mathcal{N}, \forall k\in\mathcal{K}$, and can be solved efficiently to obtain an optimal solution $\hat{\mathbb{W}}$. It can be shown that $\hat{\mathbb{S}} = (\hat{\mathbb{W}},\hat{\mathbb{V}})$ is indeed an optimal, although not necessarily unique, solution to (P$_{\text{RIP}}$). Denote by $\tilde{\mathcal{W}}(\hat{\mathbb{V}})$ the conditional feasible set of $\mathbb{W}$ given $\hat{\mathbb{V}}$. Needless to mention, $\tilde{\mathcal{W}}(\hat{\mathbb{V}})$ is a subset of $\tilde{\mathcal{W}}$. For any $\mathbb{W}\in\tilde{\mathcal{W}}$, if $(\mathbb{W},\mathbb{V})$ is a feasible strategy profile in (P$_{\text{RIP}}$), then $(\mathbb{W},\hat{\mathbb{V}})$ is also feasible. The first constraint depends only on $\mathbb{W}$. The second constraint is satisfied because $\hat{\mathbf{v}}_n^{[k]}$ maximizes $\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n)$ and $w_n^{[k]}$ is always positive. Therefore, any $\mathbb{W}\in\tilde{\mathcal{W}}$ is also an element in $\tilde{\mathcal{W}}(\hat{\mathbb{V}})$, which implies $\tilde{\mathcal{W}} \subseteq \tilde{\mathcal{W}}(\hat{\mathbb{V}})$. Together with $\tilde{\mathcal{W}}(\hat{\mathbb{V}}) \subseteq \tilde{\mathcal{W}}$, we obtain $\tilde{\mathcal{W}} = \tilde{\mathcal{W}}(\hat{\mathbb{V}})$. This means solving $\mathbb{V}$ first will not change the feasible region of $\mathbb{W}$ nor the optimal value which depends only on $\mathbb{W}$.

Maximizing $\phi_n^{[k]}(\mathbf{v}_n^{[k]},\boldsymbol{\lambda},f_n)$ over $\mathbf{v}_n^{[k]}$ is exactly problem (P$^{\text{Sub}}_{n[k]}$) in previous section. The optimal solution $\hat{\mathbf{v}}_n^{[k]}$ and optimal value $\hat{\phi}_n^{[k]}$ have been given by (IV.12) and (IV.13), respectively. The beamforming strategy profile $\hat{\mathbb{V}}$ is constructed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

by $\hat{\mathbf{v}}_n^{[k]}$. The association strategy profile $\hat{\mathbb{W}}$ is obtained by solving the following LP problem.

$$\begin{aligned} \min_{\alpha,\ \mathbb{W}\in\tilde{\mathcal{W}}} \quad & \alpha & (\mathrm{P}'_{\mathrm{RIP}}) \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} w_n^{[k]}\hat{p}_n \le \alpha P^{[k]},\ \forall k\in\mathcal{K} \\ & \sum_{k\in\mathcal{K}} w_n^{[k]}\hat{\phi}_n^{[k]} \ge 0,\ \forall n\in\mathcal{N} \end{aligned}$$

$\hat{\mathbb{S}} = (\hat{\mathbb{W}}, \hat{\mathbb{V}})$ is then an optimal solution to ($\mathrm{P}_{\mathrm{RIP}}$). ($\mathrm{P}_{\mathrm{RIP}}$) is a candidate inner problem in the Lagrangian relaxation framework. Although it does not optimize ρ directly, it improves the performance implicitly, especially in the noise-limited case, by balancing the transmit power among BSs.

B. Power Balancing based on Group Sparse Optimization

Group sparse optimization is a popular technique for power control in C-RAN. GSO treats beamformers of the same BS as a group and aims to sparsify as many groups as possible in the aggregated beamforming matrix. It is suited to address BS-centric metrics, such as power budget, backhaul traffic capacity in [4] and static power consumption in [17].

The GSO-based power balancing problem is similar to ($\mathrm{P}_{\mathrm{RIP}}$), formulated as

$$\begin{aligned} \min_{\alpha,\ \mathbb{S}\in\tilde{\mathcal{S}}} \quad & \alpha & (\mathrm{P}_{\mathrm{GSO}}) \\ \text{s.t.} \quad & \hat{\beta}^{[k]} \sum_{n\in\mathcal{N}} w_n^{[k]}\hat{p}_n \le \alpha P^{[k]},\ \forall k\in\mathcal{K} \\ & \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) \ge 0,\ \forall n\in\mathcal{N} \end{aligned}$$

In the first constraint, there is a weight

$$\hat{\beta}^{[k]} = \frac{1}{\sum_{n\in\mathcal{N}} \hat{w}_n^{[k]}\hat{p}_n + \tau} \tag{V.3}$$

where $\hat{w}_n^{[k]}$ is the estimated solution obtained from last iteration and τ is a small positive regularization constant. Let $\hat{p}^{[k]} = \sum_{n\in\mathcal{N}} \hat{w}_n^{[k]}\hat{p}_n$ represent the estimated transmit power of BS k and rewrite the first constraint as $\frac{1}{P^{[k]}}\hat{\beta}^{[k]}\hat{p}^{[k]} \le \alpha$. In the limiting situation, the left side has two possible outcomes

$$\frac{1}{P^{[k]}}\hat{\beta}^{[k]}\hat{p}^{[k]} \to \begin{cases} \dfrac{1}{P^{[k]}}, & \hat{p}^{[k]} > 0 \\ 0, & \hat{p}^{[k]} = 0 \end{cases} \tag{V.4}$$

BSs with large $P^{[k]}$ allow a smaller α while BSs with small $P^{[k]}$ tend to be shut down during the minimization of α.

($\mathrm{P}_{\mathrm{GSO}}$) is solved in the same way as ($\mathrm{P}_{\mathrm{RIP}}$). First we solve ($\mathrm{P}^{\mathrm{Sub}}_{n[k]}$) for the beamforming strategy profile $\hat{\mathbb{V}}$. Then for fixed $\hat{\mathbb{V}}$, solve the simplified LP version of ($\mathrm{P}_{\mathrm{GSO}}$), which is not elaborated here, to obtain $\hat{\mathbb{W}}$.

C. Combination of SINR Balancing and Power Balancing

In Section IV, we optimize the asymptotic lower bound of the SINR Balancing level in the interference-limited case by ignoring the power constraints. In previous two subsections, we optimize the power balancing level which has certain advantage in the noise-limited case. The two objectives can be combined to achieve a universally better performance.

Consider the following conceptual problem

$$\begin{aligned} \min_{\rho,\ \mathbb{S}\in\tilde{\mathcal{S}}} \quad & \rho & (\mathrm{P}_2) \\ \text{s.t.} \quad & \mathbf{G}(\rho)\hat{\mathbf{p}} \ge \alpha\boldsymbol{\eta} \end{aligned}$$

The constraint is the feasibility condition of ρ based on statement 1) in Theorem 1 taking $\mathbf{p} = \hat{\mathbf{p}}/\alpha$. As before, $\hat{\mathbf{p}}$ is the estimated power allocation vector and α is the power balancing level such that $\hat{\mathbf{p}}/\alpha$ satisfies all power constraints. The penalty term is $\boldsymbol{\lambda}^T(\alpha\boldsymbol{\eta} - \mathbf{G}(\rho)\hat{\mathbf{p}})$ with the Lagrangian multiplies $\boldsymbol{\lambda}$. Recall that $(\mathbf{G}^T(\rho)\boldsymbol{\lambda})_n = \phi_n(\mathbb{S}_n, \boldsymbol{\lambda}, f_n) = \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n)$, if f_n is updated to be $\rho\lambda_n$. The inner problem that minimizes the penalty term subject to additional power constraints and non-incremental condition of ρ is formulated as follows.

$$\begin{aligned} \min_{\alpha,\ \mathbb{S}\in\tilde{\mathcal{S}}} \quad & (\boldsymbol{\lambda}^T\boldsymbol{\eta})\alpha - \sum_{n\in\mathcal{N}} \hat{p}_n \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} w_n^{[k]}\hat{p}_n \le \alpha P^{[k]},\ \forall k\in\mathcal{K} & (\mathrm{P}_2^{\mathrm{in}}) \\ & \sum_{k\in\mathcal{K}} w_n^{[k]}\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n) \ge 0,\ \forall n\in\mathcal{N} \end{aligned}$$

($\mathrm{P}_2^{\mathrm{in}}$) can be considered as a combination of the two inner problems ($\mathrm{P}_{\mathrm{RIP}}$) and ($\mathrm{P}_{1'}^{\mathrm{in}}$), which minimizes the power balancing level α and maximizes a weighted sum of signal-minus-interference in the dual uplink channel $\phi_n(\cdot)$, respectively. Again, ($\mathrm{P}_2^{\mathrm{in}}$) can be solved in two steps. The beamforming strategy profile $\hat{\mathbb{V}}$ consists of $\hat{\mathbf{v}}_n^{[k]}$ given by (IV.12). Then by replacing $\phi_n^{[k]}(\mathbf{v}_n^{[k]}, \boldsymbol{\lambda}, f_n)$ with the value $\hat{\phi}_n^{[k]}$ given by (IV.13), ($\mathrm{P}_2^{\mathrm{in}}$) is simplified to an LP problem and solved for $\hat{\mathbb{W}}$.

D. Algorithm Design

Before proposing the iterative Lagrangian relaxation algorithms, we shall address some non-trivial problems, such as the estimation of the power allocation vector $\hat{\mathbf{p}}$. $\hat{\mathbf{p}}$ can be updated by (III.5) in each iteration. In fact, $\hat{\mathbf{p}}$ can be scaled arbitrarily because the actual power allocation vector is adjusted by the power balancing level α. Moreover, $\hat{\mathbf{p}}$ is merely an estimation and need not be precise during intermediate steps. Tested by simulation, (III.7) is also an eligible candidate. To make $\hat{\mathbf{p}}$ concordant with its role as an estimated power vector, we formally adopt (III.5) to calculate $\hat{\mathbf{p}}$ in the algorithm.

Another problem is that the returned solution is in the relaxed feasible set $\tilde{\mathcal{S}} = \tilde{\mathcal{W}} \times \mathcal{V}$. Each user may be associated with multiple BSs as a result. When the system architecture, such as C-RAN, supports data sharing in the backhaul, the result leads to a dynamic BS clustering scheme. Otherwise, the returned association strategy $\mathbb{W} \in \tilde{\mathcal{W}}$ should be cast into a feasible strategy $\mathbb{W}^\dagger \in \mathcal{W}$, which will degrade the performance inevitably. The casting procedure is simply to select the BS that provides the largest power, namely, let

$$w_n^{[k]\dagger} = \begin{cases} 1, & k = \arg\max_j w_n^{[j]} \\ 0, & \text{otherwise} \end{cases} \tag{V.5}$$

If a user is allowed to associate with multiple BSs, a partial casting procedure is also practicable. $w_n^{[k]}$ is cast into $w_n^{[k]\dagger}$ only when $\max_k w_n^{[k]}$ is larger than some threshold θ indicating that user n has a dominant resource provider. Otherwise $w_n^{[k]}$ remains unchanged. Such partially cast association strategy is denoted by $\mathbb{W}^{\ddagger}$ with

$$w_n^{[k]\ddagger} = \begin{cases} w_n^{[k]\dagger}, & \max_j w_n^{[j]} \geq \theta \\ w_n^{[k]}, & \text{otherwise} \end{cases} \tag{V.6}$$

Both the fully cast $\mathbb{W}^{\dagger}$ and partially cast $\mathbb{W}^{\ddagger}$ are adopted to evaluate the performance of the algorithms.

Now the algorithms can be consolidated as follows. By solving the different inner problems ($\mathrm{P_{RIP}}$), ($\mathrm{P_{GSO}}$) and ($\mathrm{P_2^{in}}$) and fully casting $\mathbb{W}(t)$ to $\mathbb{W}^{\dagger}(t)$, the algorithms are named as Algorithm RIP, Algorithm GSO and Algorithm 2, respectively. By partially casting $\mathbb{W}(t)$ to $\mathbb{W}^{\ddagger}(t)$, the algorithms are named as Algorithm RIP(JP), Algorithm GSO(JP) and Algorithm 2(JP), respectively, where JP stands for joint processing of the data by multiple BSs.

Algorithm RIP / GSO / 2 (JP)

Initialize $\boldsymbol{\lambda}(0) > \mathbf{0}$, $\mathbf{1}^T\boldsymbol{\lambda}(0) = 1$, and $\mathbf{f}(0)$;
while $|\rho(t) - \rho(t-1)| > \epsilon$ **do**
  Update $t = t + 1$;
  Solve ($\mathrm{P_{RIP}}$) or ($\mathrm{P_{GSO}}$) or ($\mathrm{P_2^{in}}$) for $\mathbb{S}(t)$;
  Update $\boldsymbol{\lambda}(t) = \mathbf{x}(\mathbf{B}^T(\mathbb{S}(t)))$, $\mathbf{1}^T\boldsymbol{\lambda}(t) = 1$, $\rho(t) = \rho(\mathbf{B}(\mathbb{S}(t)))$, $\mathbf{f}(t) = \rho(t)\boldsymbol{\lambda}(t)$, and $\hat{\mathbf{p}}(t)$ given by (III.5);
end
Cast $\mathbb{W}(t)$ to $\mathbb{W}^{\dagger}(t)$ or $\mathbb{W}^{\ddagger}(t)$;
Calculate $\mathbf{p}^*$ by (III.5) for $\mathbb{S}^* = (\mathbb{W}^{\dagger}(t), \mathbb{V}(t))$ or $(\mathbb{W}^{\ddagger}(t), \mathbb{V}(t))$;
Output: $\mathbb{S}^*$ and $\mathbf{p}^*$.

The algorithms proposed in this section inherit the framework of Algorithm $1'$ in last section with different inner problems ($\mathrm{P_{RIP}}$), ($\mathrm{P_{GSO}}$) and ($\mathrm{P_2^{in}}$). These problems do not optimize ρ directly but improves ρ under an analytical guidance. ρ is guaranteed to converge. In addition, the stopping criterion ϵ can be adjusted to tradeoff the running time and precision of the algorithms.

VI. Simulation Results

We conduct simulations for Algorithm 1, Algorithm RIP, Algorithm GSO and Algorithm 2 in downlink MISO systems with different power budgets. Algorithm 1 and Algorithm 2 are also tested in downlink SISO system for comparison with the algorithms in the literature [16]. Because we study a general system architecture in this paper, the channel between any transmit antenna and any user is generated from i.i.d. zero-mean unit-variance circularly-symmetric Gaussian distribution. The noise power and weight of SINR are assumed to be equal for all users. Specifically, we let $\gamma_n = 1$ and $\eta_n = \eta$ for all $n \in \mathcal{N}$. The power budgets of BSs are heterogeneous. We let the maximum and minimum power budgets be 20dB and 10dB, respectively. We monitor the performance for $\mathrm{SNR} = \min_k P^{[k]}/\eta$ ranging from 0 to 30dB. The demonstrated results are averaged over 1000 Monte-Carlo samples.

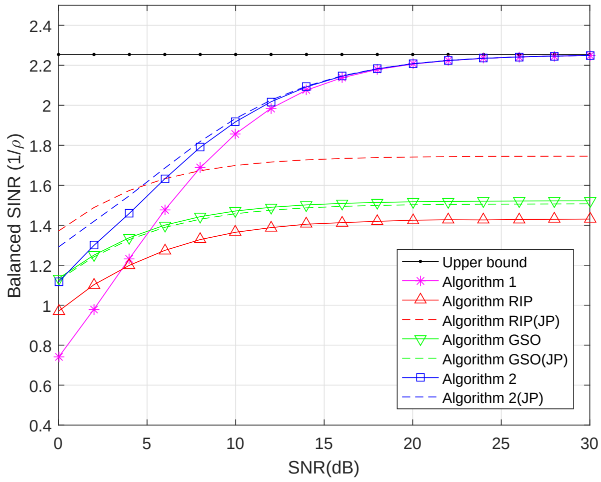

Fig. 1. Balanced SINR for $N = 15$, $K = 6$, $M^{[k]} = 8$ and power budget tuple being $(100, 100, 10, 10, 10, 10)$w.

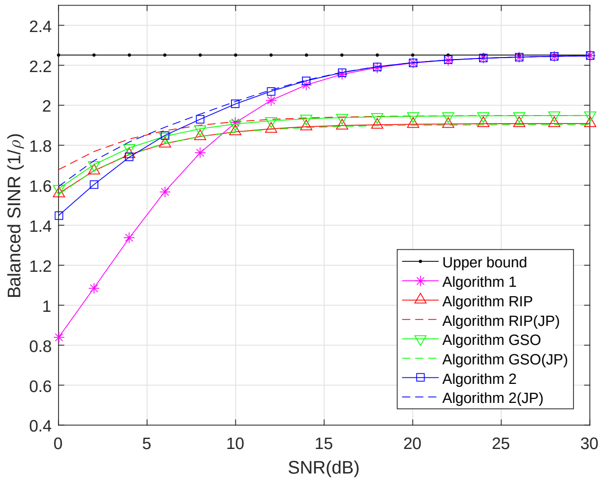

Fig. 2. Balanced SINR for $N = 15$, $K = 6$, $M^{[k]} = 8$ and power budget tuple being $(100, 100, 100, 100, 10, 10)$w.

In the first three figures, we run the proposed algorithms and plot the balanced SINR $\gamma_n/\rho = 1/\rho$ for the downlink MISO systems with $N = 15$, $K = 6$ and $M^{[k]} = 8$ for all $k \in \mathcal{K}$. The power budget tuples in Fig. 1 to Fig. 3 are $(100, 100, 10, 10, 10, 10)$w, $(100, 100, 100, 100, 10, 10)$w and $(100, 80, 60, 40, 20, 10)$w, respectively, which are assigned randomly to the BSs. The upper bound is the balanced SINR in the absence of noise and power constrains. It can be seen that Algorithm 1 approaches the upper bound asymptotically at high SNR. On the other hand, RIP-based and GSO-based algorithms which focus on power balancing has better performance at low SNR. Algorithm 2 exploits the advantageous features of these algorithms and achieves a comprehensively

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

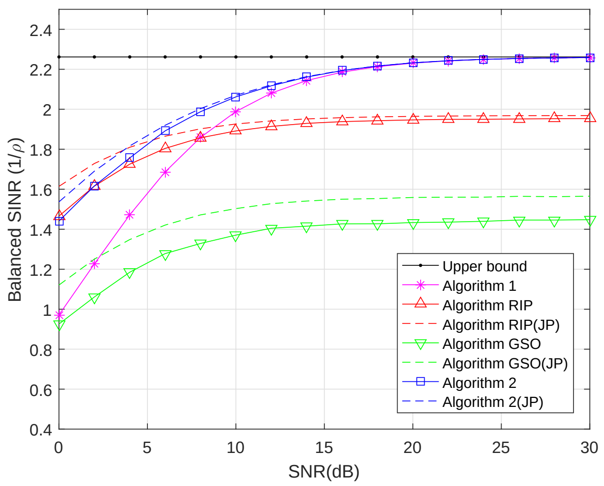

Fig. 3. Balanced SINR for $N = 15$, $K = 6$, $M^{[k]} = 8$ and power budget tuple being $(100, 80, 60, 40, 20, 10)$w.

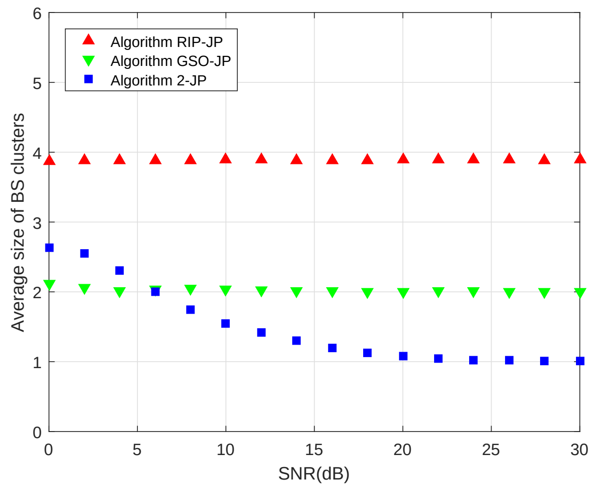

Fig. 4. Averaged size of BS clusters for casting threshold $\theta = 0.9$ with the same system setting as in Fig. 1.

better performance.

In general, partial casting together with joint processing achieves better performance than full casting. The difference is evident for Algorithm RIP and Algorithm RIP(JP) in Fig. 1. Fig. 4 shows the average BS cluster size, which is the average number of associated BS per user, for partial casting algorithms with a casting threshold $\theta = 0.9$. It is clear that Algorithm RIP(JP) ends up with larger BS cluster size, compared with Algorithm GSO(JP) and Algorithm 2(JP). Consequently, performance degradation is severer for Algorithm RIP(JP) under inappropriate casting. By inappropriate casting we mean that a considerable number of users turn out to associate with the low-power BSs. The problem is alleviated when there are more high-power BSs in Fig. 2, or when the power budgets are more balanced in Fig. 3.

The BS cluster size resulted by Algorithm GSO(JP) is smaller, because GSO-based algorithms tend to shut down as many low-power BSs as possible. As a result, the performance is affected less by inappropriate casting. Compared with RIP-based algorithms, this feature of GSO is advantageous in some systems as in Fig. 1 and disadvantageous in some systems as in Fig. 3. In Fig. 2, RIP-based algorithms and GSO-based algorithms are approximately equally good. Nevertheless, GSO-based algorithms are useful in dealing with other BS-centric metrics, which is not elaborated in this paper.

Overall, Algorithm 2 is a fairer way to manage imbalanced power budgets. It has acceptable performance at both low and high SNR. When joint processing is allowed, Algorithm 2(JP) also yields a more flexible and smaller BS cluster size. Therefore, we choose Algorithm 1 and Algorithm 2 to compare with the algorithms DLSum and DLSumA proposed in [16] for downlink SISO systems. The simulated system has $N = 15$ users and $K = 12$ single-antenna BSs. There are 4 macro BSs with power budget 20dB and 8 pico BSs with power budget 10dB.

The comparison algorithms DLSum and DLSumA proposed in [16] are based on uplink-downlink duality. In DLSum, the assocation strategy is obtained by solving the dual uplink problem subject to sum power constraint via an iterative method. Then for fixed association strategy, the power allocation vector in the downlink is calculated iteratively. The steps in DLSum are similar to Algorithm 1. The multiple per BS power constraints are ignored in Algorithm 1 but replaced by the sum power constraint in DLSum. The power allocation vector is found by a closed-form solution in Algorithm 1 but iteratively in DLSum. Therefore, in terms of complexity of the algorithms, DLSum and Algorithm 1 are essentially the same, if we had also chosen to find the power allocation vector iteratively. However, Algorithm 1 outperforms DLSum, as is shown in Fig. 5. This is because connecting the uplink and downlink by the same sum power is not adequate in the presence of multiple per BS power constraints. Rather than sum power, the performance is more limited by interference and the tightest power constraint. In some cases where interference is more vital, ignoring the multiple power constraints yields better performance than relaxing them to be a single sum power constraint.

The other comparison algorithm DLSumA is an advanced version of DLSum. The authors of [16] propose DLSumA in consideration of the imbalanced power budgets of BSs in HetNet. DLSumA transfers the imbalanced power budgets to amplification on CSI and calculates the effective sum power returned by DLSum. Then DLSum is run again based on the effective sum power. Although DLSumA remedies the deficiency due to sum power relaxation to some extent, the underlying principle is the same as DLSum, namely connecting uplink and downlink by the same sum power. As can be seen in Fig. 5, DLSum and DLSumA approach the same asymptotic value at high SNR, which is inferior to the upper bond achieved asymptotically by Algorithm 1 and Algorithm 2. In terms of complexity, DLSumA is twice as complex as DLSum. The complexity of Algorithm 2 is difficult to quantify due to the adopted RIP technique. However, in each iteration of Algorithm 2, the complexity is bounded by that of an LP problem. Also, the inner problem in each iteration is well-grounded to decrease ρ.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2785400, IEEE Transactions on Wireless Communications

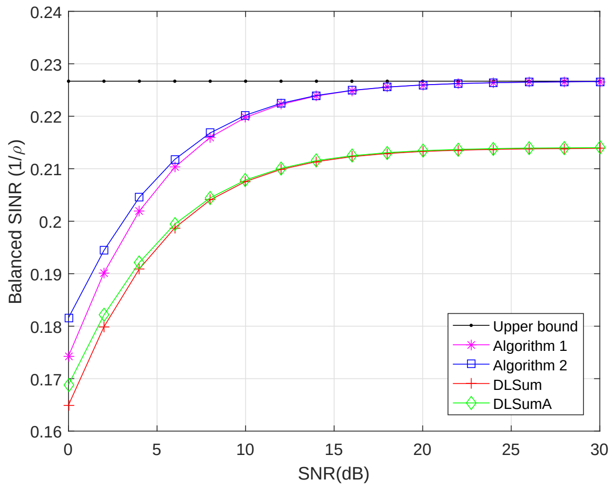

Fig. 5. Balanced SINR for $N = 15$, $K = 12$, $M^{[k]} = 1$ and power budget tuple being $(100, 100, 100, 100, 10, 10, 10, 10, 10, 10, 10, 10)$w.

VII. Conclusion

In this paper, we design user-BS association strategy, beamforming strategy and power allocation vector to balance SINR in downlink multicell MISO systems. The problem is divided into two main steps which solve for the strategies and power allocation vector separately. We propose algorithms based on Lagrangian relaxation with various inner problems. In the interference-limited case, the inner problems minimize the asymptotic lower bound of the SINR balancing level. We also design inner problems to minimize the power balancing level, which improves the balanced SINR implicitly, based on RIP and GSO. The two objectives are then combined to achieve a comprehensively better performance. Intuitively, the problems are designed in consideration of two limiting factors, namely interference and the tightest power constraint. Theoretically, the problems are well-founded by statements in Theorem 1 and Lemma 1. Simulation results show that the proposed algorithms outperform the existing algorithms designed for the downlink SISO system. In addition, a dynamic BS clustering scheme is produced as a byproduct. The various designs of the inner problem exhibit the extensibility of the framework. For instance, GSO-based algorithms can be extended to allow for other practical BS-centric metrics in future works.

References

[1] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, pp. 102–111, February 2011.

[2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *IEEE Wireless Communications*, vol. 18, pp. 10–21, June 2011.

[3] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2706–2716, June 2013.

[4] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[5] J. M. Aein, "Power balancing in systems employing frequency reuse," *COMSAT Technical Review*, vol. 3, no. 2, pp. 277–299, 1973.

[6] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 305–311, Aug 1992.

[7] W. Yang and G. Xu, "Optimal downlink power assignment for smart antenna systems," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 6, pp. 3337–3340 vol.6, May 1998.

[8] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing mse on multiuser downlink: Optimality, fast algorithms and equivalence via max-min sinr," *IEEE Transactions on Signal Processing*, vol. 59, pp. 6127–6143, Dec 2011.

[9] D. W. H. Cai, T. Q. S. Quek, and C. W. Tan, "A unified analysis of max-min weighted sinr for mimo downlink system," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3850–3862, Aug 2011.

[10] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, and M. S. Alouini, "Max-min sinr in large-scale single-cell mu-mimo: Asymptotic analysis and low-complexity transceivers," *IEEE Transactions on Signal Processing*, vol. 65, pp. 1841–1854, April 2017.

[11] D. W. H. Cai, T. Q. S. Quek, C. W. Tan, and S. H. Low, "Max-min sinr coordinated multipoint downlink transmission - duality and algorithms," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5384–5395, Oct 2012.

[12] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2646–2660, June 2007.

[13] Y. Huang, G. Zheng, M. Bengtsson, K. K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Transactions on Signal Processing*, vol. 59, pp. 728–738, Feb 2011.

[14] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Downlink power control and base station assignment," *IEEE Communications Letters*, vol. 1, pp. 102–104, July 1997.

[15] W. Liu, R. Sun, Z. Luo, and J. Li, "Globally optimal joint uplink base station association and beamforming," *CoRR*, vol. abs/1512.04927, 2015.

[16] R. Sun, M. Hong, and Z. Q. Luo, "Joint downlink base station association and power control for max-min fairness: Computation and complexity," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 1040–1054, June 2015.

[17] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in c-ran," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 494–508, Jan 2015.

[18] R. Plemmons, "M-matrix characterizations.inonsingular m-matrices," *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 175 – 188, 1977.

[19] G. Poole and T. Boullion, "A survey on m-matrices," *SIAM Review*, vol. 16, no. 4, pp. 419–427, 1974.

[20] C. M. Jr. and M. Stadelmaier, "Singular m-matrices and inverse positivity," *Linear Algebra and its Applications*, vol. 22, pp. 139 – 156, 1978.

[21] C. Lemaréchal, *Lagrangian Relaxation*, pp. 112–156. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001.

[22] K. Chang, L. Qi, and T. Zhang, "A survey on the spectral theory of nonnegative tensors," *Numerical Linear Algebra with Applications*, vol. 20, no. 6, pp. 891–912, 2013.