Multicell Coordinated Beamforming With Rate Outage Constraint—Part II: Efficient Approximation Algorithms

Wei-Chiang Li, Tsung-Hui Chang, *Member, IEEE*, and Chong-Yung Chi, *Senior Member, IEEE*

Abstract—This paper studies the coordinated beamforming (CoBF) design for the multiple-input single-output interference channel, provided that only channel distribution information is known to the transmitters. The problem under consideration is a probabilistically constrained optimization problem which maximizes a predefined system utility subject to constraints on rate outage probability and power budget of each transmitter. Our recent analysis has shown that the outage-constrained CoBF problem is intricately difficult, e.g., NP-hard. Therefore, the focus of this paper is on suboptimal but computationally efficient algorithms. Specifically, by leveraging on the block successive upper bound minimization (BSUM) method in optimization, we propose a Gauss-Seidel type algorithm, called distributed BSUM algorithm, which can handle differentiable, monotone and concave system utilities. By exploiting a weighted minimum mean-square error (WMMSE) reformulation, we further propose a Jocobi-type algorithm, called distributed WMMSE algorithm, which can optimize the weighted sum rate utility in a fully parallel manner. Both algorithms are shown to converge to the stationary points of the original NP-hard problems. To further provide a performance benchmark, a relaxed approximation method based on polyblock outer approximation is also proposed. Simulation results show that the proposed algorithms are significantly superior to the existing successive convex approximation method in both performance and computational efficiency, and can yield promising approximation performance by comparing with the performance benchmark.

Index Terms—Convex optimization, coordinated beamforming (CoBF), interference channel, outage probability.

I. INTRODUCTION

COORDINATED MULTIPOINT (CoMP) has been recognized as an effective approach for interference management in wireless cellular networks [2]. There are two main types of cooperation, namely *MIMO cooperation* and *interference coordination*, which offer a trade-off between the performance gain and induced communication overhead in the backhaul network [3]. Via high-capacity delay-free backhaul, the coordinated base stations (BSs) for the MIMO cooperation share all the channel state information (CSI) and users' data, so they perform as a virtual multiple-antenna BS and high spectrum efficiency can be achieved. For interference coordination, the BSs only share CSI in order to jointly design, e.g., power allocation and beamforming strategies, to mitigate the inter-cell interference. Compared with MIMO cooperation, the interference coodination requires a relatively modest amount of backhaul communication [4], and therefore is still viable when the backhaul capacity is limited. To study the interference coordination scheme, we consider the commonly used interference channel (IFC) model [5], where multiple transmitters simultaneously communicate with their respective receivers over a common frequency band, and hence interfere with each other.

This paper focuses on the multiple-input single-output (MISO) IFC, wherein the transmitters are equipped with multiple antennas while the receivers are equipped with single antenna. Our interest lies in the *coordinated beamforming* (CoBF) design where the transmitters cooperate to optimize their beamforming vectors in order to maximize a network-wide utility function, e.g., the sum rate, proportional fairness rate, harmonic mean rate, or the max-min-fairness (MMF) rate. Most of the existing works have assumed that the transmitters have the perfect CSI. Under this assumption, the MMF CoBF problem has been shown to be polynomial-time solvable [6] and efficient algorithms have been proposed [6]–[9]. However, for the sum rate, proportional fairness rate and harmonic mean rate, the utility maximization CoBF problem is difficult and has been shown NP-hard in general [6]. As a result, most of the research efforts have been devoted to developing suboptimal but efficient approximation algorithms; see, e.g., [6], [10]–[16] and also [17]–[19] for game theoretic approaches. Global optimization algorithms are also available in [20]–[22], but they are efficient only when the number of users is small.

In practical wireless environments, acquiring accurate users' CSI is difficult, especially in a mobile network. By contrast, the channel distribution information (CDI) remains unchanged for a relatively long period of time, and thus is easier to obtain. Moreover, the CDI based CoBF design need not perform beamforming optimization frequently, so the required backhaul bandwidth for implementing the CDI based CoBF design can be significantly smaller when compared with the CSI based

CoBF design. However, given only CDI at the transmitters, the data transmission would suffer from outage with a nonzero probability, i.e., reliable data transmission cannot be guaranteed all the time, due to channel fading. In view of this, the outage-aware CoBF design, which concerns the probability of rate outage, has attracted extensive attention recently. For example, the outage balancing CoBF problem was studied in [23]–[25], the outage-constrained power minimization problem was considered in [23], [26], and the outage-constrained utility maximization problem was studied in [27]–[29]. It turns out that the outage probability constrained CoBF problem is a very difficult optimization problem. Specifically, it has been shown in [30] that the outage balancing problem in [25] is in fact NP-hard. Besides, the outage-constrained CoBF problem [27]–[29] is NP-hard in general with not only the sum rate but also the MMF rate (under the MISO setting) [30]. This implies that efficient algorithms for high-quality approximate solutions are indispensable. In [29], a successive convex approximation (SCA) algorithm and a distributed SCA (DSCA) algorithm were proposed to handle the outage-constrained CoBF problem. However, the computational complexity of the two algorithms is high, hence preventing them from being used in practical scenarios with a moderate to large number of users.

In this paper, we propose two efficient distributed CoBF algorithms for the outage-constrained utility maximization problem, one referred to as the distributed *block successive upper bound minimization* (DBSUM) algorithm and the other referred to as the distributed *weighted minimum mean-square error* (DWMMSE) algorithm. The DBSUM algorithm is a Gauss-Seidel type algorithm, derived based on a judicious reformulation of the outage-constrained problem and application of the BSUM method in [31]. The DBSUM algorithm can handle a general class of monotonic, differentiable concave utilities. On the other hand, the DWMMSE algorithm is custom-devised for the weighted sum rate utility, and is a Jocobi-type algorithm so that all the transmitters can update their respective beamformers in a fully parallel manner. A common merit of the two algorithms is that the subproblems to be solved at each iteration are easily implementable, with problem dimension independent of the number of users. So, the two algorithms are computationally efficient and scalable with the size of the network. Moreover, the two algorithms are guaranteed to converge to the stationary points, i.e., the points satisfying the first-order optimality condition, of the original NP-hard problems. To further provide a benchmark for performance evaluation of the proposed DBSUM and DWMMSE algorithms, we also present a constraint relaxation technique for the outage-constrained CoBF problem. The constraint-relaxed problem is solved by a *polyblock outer approximation* (POA) algorithm [32] to obtain an upper bound for the optimal utility value of the original outage-constrained CoBF problem, in spite of tremendous computation time. We show by computer simulations that the proposed algorithms significantly outperform the DSCA algorithm [29] in both performance and computational efficiency, and exhibit better scalability with respect to (w.r.t.) the number of users. Moreover, by comparing with the performance upper bound obtained by the POA algorithm, it can be corroborated that the proposed algorithms achieve high approximation accuracy in general.

Synopsis: In Section II, we present the system model and problem formulations. The proposed DBSUM algorithm and DWMMSE algorithm are presented in Section III and Section IV, respectively. In Section V, we present the POA algorithm which serves as a benchmark performance upper bound for the two proposed algorithms. Simulation results are then provided in Section VI to demonstrate the efficacy of the proposed algorithms. Finally, the conclusions are drawn in Section VII.

Notations: The set of n-dimensional real vectors and complex vectors are denoted by $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. The set of non-negative real vectors is denoted by $\mathbb{R}^n_+$. The superscripts 'T' and 'H' represent the matrix transpose and conjugate transpose, respectively. We denote $\|\cdot\|$ as the vector Euclidean norm. $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \succ \mathbf{0}$) and $\mathbf{a} \succeq \mathbf{0}$ ($\mathbf{a} \succ \mathbf{0}$) mean that the matrix $\mathbf{A}$ is positive semidefinite (definite) and the vector $\mathbf{a}$ is componentwise nonnegative (positive). We use the expression $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{Q})$ if $\mathbf{x}$ is circularly symmetric complex Gaussian distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{Q}$. We denote $\ln(\cdot)$ and $\Pr\{\cdot\}$ as the natural log function and the probability function, respectively. The principal eigenvalue of a matrix $\mathbf{A}$ is denoted by $\lambda_{\max}(\mathbf{A})$. $\{a_{ik}\}$ denotes the set of all a_{ik} with subscripts i, k covering all the admissible integers that are defined in the context, and $\{a_{ik}\}_k$ denotes the set of all a_{ik} with the first subscript equal to i. The set $\{a_{ik}\}_{k \neq j}$ is defined by the set $\{a_{ik}\}_k$ excluding a_{ij}.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider a K-user MISO IFC where K transmitter-receiver pairs share a common spectral band. Each transmitter is equipped with N_t antennas, and all the receivers have single antenna. Assume that transmit beamforming is used for data transmission. Specifically, let $\boldsymbol{x}_i = \boldsymbol{w}_i s_i$ denote the signal intended for user i, where $\boldsymbol{w}_i \in \mathbb{C}^{N_t}$ and $s_i \in \mathbb{C}$ are the beamforming vector and the information signal, respectively. The received signal at receiver i is thus given by

$$x_i = \boldsymbol{h}_{ii}^H \boldsymbol{x}_i + \sum_{k=1, k \neq i}^{K} \boldsymbol{h}_{ki}^H \boldsymbol{x}_k + n_i, i = 1, \ldots, K, \quad (1)$$

where $\boldsymbol{h}_{ki} \in \mathbb{C}^{N_t}$ denotes the MISO channel from transmitter k to receiver i and $n_i \in \mathbb{C}$ is the additive Gaussian noise at receiver i which has zero mean and variance $\sigma_i^2 > 0$. The channels $\boldsymbol{h}_{ki}$ are assumed to be complex Gaussian distributed with zero mean and covariance matrix $\mathbf{Q}_{ki} \succeq \mathbf{0}$, i.e., $\boldsymbol{h}_{ki} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{ki})$, for all $i, k = 1, \ldots, K$. Assume Gaussian signaling, e.g., $s_i \sim \mathcal{CN}(0, 1)$, and assume that each receiver i decodes the information s_i from the received signal with other users' interference treated as noise (i.e., single user detection). Then, the instantaneous achievable rate (in bits/sec/Hz) of the ith user is given by

$$r_i(\{\boldsymbol{h}_{ki}\}_k, \{\boldsymbol{w}_k\}) = \log_2\left(\frac{|h_{ii}^H w_i|^2}{\sum_{k \neq i} |h_{ki}^H w_k|^2 + \sigma_i^2}\right). \quad (2)$$

We assume that only CDI is available at the transmitters; that is, the transmitters know only the channel covariance matrices $\mathbf{Q}_{ik}, i, k = 1, \ldots, K$. Under such circumstances, users might suffer from transmission outage. Specifically, let $R_i > 0$ be the transmission rate of the ith user. The outage event that

$r_i(\{\boldsymbol{h}_{ki}\}_{k=1}^{K}, \{\boldsymbol{w}_k\}_{k=1}^{K}) < R_i$ will occur with a nonzero probability due to channel fading. Our goal is to optimize the transmit beamformers $\{\boldsymbol{w}_i\}_{i=1}^{K}$ so that a predefined system utility, which concerns the system throughput or user fairness, or considers a proper tradeoff between the two, is maximized under both transmission outage probability and transmit power constraints. Mathematically, this can be formulated as the following outage-constrained CoBF problem:

$$\max_{\substack{\boldsymbol{w}_i \in \mathbb{C}^{N_t}, R_i \geq 0, \\ i=1,\ldots,K}} \quad U(R_1, \ldots, R_K) \tag{3a}$$

$$\text{s.t.} \quad \Pr\{r_i(\{\boldsymbol{h}_{ki}\}_k, \{\boldsymbol{w}_k\}) < R_i\} \leq \epsilon_i, \tag{3b}$$

$$\|\boldsymbol{w}_i\|^2 \leq P_i, i = 1, \ldots, K, \tag{3c}$$

where $U(R_1, \ldots, R_K)$ denotes the system utility of interest, $P_i > 0$ is the power constraint of user i, and $\epsilon_i \in (0, 1)$ is the maximal tolerable rate outage probability for $i = 1, \ldots, K$. The outage probability constraint (3b) guarantees that the rate outage probability is no larger than a specified threshold ϵ_i, which is usually small, e.g., $\epsilon_i = 0.1$. According to [23], [29], under $\boldsymbol{h}_{ki} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{Q}_{ki}) \ \forall k, i$, the outage probability in (3b) has a closed-form expression, and constraint (3b) can be explicitly expressed as

$$\ln \rho_i + \frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \boldsymbol{Q}_{ii} \boldsymbol{w}_i} + \sum_{k \neq i} \ln\left(1 + \frac{(2^{R_i}-1)\boldsymbol{w}_k^H \boldsymbol{Q}_{ki} \boldsymbol{w}_k}{\boldsymbol{w}_i^H \boldsymbol{Q}_{ii} \boldsymbol{w}_i}\right) \leq 0, \tag{4}$$

where $\rho_i \triangleq 1 - \epsilon_i$ for $i = 1, \ldots, K$.

As seen from (4), the outage-constrained CoBF problem (3) is in general nonconvex and appears difficult to deal with. In fact, our recent complexity analyses in [30] have shown that problem (3) can be computationally intractable. In particular, it has been shown in [30] that problem (3) is NP-hard in general for the weighted sum-rate utility $U(R_1, \ldots, R_K) = \sum_{i=1}^{K} \alpha_i R_i$, where $\alpha_i > 0$ for $i = 1, \ldots, K$, are the priority weights of users. Moreover, for the weighted min-rate (also known as the max-min-fairness (MMF) rate) utility $U(R_1, \ldots, R_K) = \min_{i \in \{1, \ldots, K\}} R_i/\alpha_i$, problem (3) is also NP-hard in general if $N_t \geq 2$. Since maximizing the MMF rate is known polynomial-time solvable under perfect CSI [6], this implies that the outage-constrained CoBF problem (3) is indeed more challenging. In view of the computational intractability of (3), in the subsequent Section III and Section IV, we propose two algorithms that can efficiently achieve high-quality approximate solutions to problem (3).

III. OUTAGE-CONSTRAINED COBF BY DISTRIBUTED BSUM ALGORITHM

Let us make the following assumptions on the system utility $U(\cdot)$. Firstly, $U(\cdot)$ is nondecreasing w.r.t. $R_1, \ldots, R_K$, respectively, as users always desire to increase the transmission rate as long as it is possible. Secondly, $U(\cdot)$ is jointly concave w.r.t. $R_1, \ldots, R_K$, as concavity enforces user fairness [33]. These assumptions are general enough to include some commonly adopted system utilities such as the weighted sum-rate utility, proportional fairness utility, harmonic mean utility, and the min-rate (i.e., MMF rate) utility [6]. Under these assumptions, we show in this section how the outage-constrained problem (3) can be efficiently handled in a distributed manner by the

block successive upper bound minimization (BSUM) method reported in [31].

A. Equivalent Reformulation

The key ingredient of the proposed method lies in the following equivalent reformulation of (3):

Proposition 1: *Problem (3) is equivalent to the following problem*

$$\max_{\boldsymbol{w}_i \in \mathbb{C}^{N_t}, i=1,\ldots,K} \quad U(R_1(\{\boldsymbol{w}_i\}), \ldots, R_K(\{\boldsymbol{w}_i\})) \tag{5a}$$

$$\text{s.t.} \quad \|\boldsymbol{w}_i\|^2 \leq P_i, i = 1, \ldots, K, \tag{5b}$$

where

$$R_i(\{\boldsymbol{w}_k\}) \triangleq \log_2(1 + \xi_i(\{\boldsymbol{w}_k\}_{k \neq i})\boldsymbol{w}_i^H \boldsymbol{Q}_{ii} \boldsymbol{w}_i), \tag{6}$$

and $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i}) > 0$ is a continuously differentiable function of $\{\boldsymbol{w}_k\}_{k \neq i}$ and is a unique solution to the equation

$$\Phi_i(\xi_i, \{\boldsymbol{w}_k\}_{k \neq i}) \triangleq$$
$$\ln \rho_i + \sigma_i^2 \xi_i + \sum_{k \neq i} \ln(1 + (\boldsymbol{w}_k^H \boldsymbol{Q}_{ki} \boldsymbol{w}_k) \cdot \xi_i) = 0, \tag{7}$$

for $i = 1, \ldots, K$.

Proposition 1 can be proved by exploiting the fact that the left-hand side function in (4) is monotonic[1] in $\frac{(2^{R_i}-1)}{\boldsymbol{w}_i^H \boldsymbol{Q}_{ii} \boldsymbol{w}_i}$. The idea is the same as the one reported in [30, Lemma 1] and interested readers may refer to [30, Appendix A] for the detailed proof.

By comparing problem (5) with problem (3), one can observe that the rate outage constraints in (3) [and (4)] have been judiciously incorporated with the objective function and it is the function $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i})$ that implicitly characterizes the impact of cross-link interference plus noise on receiver i. Indeed, as seen from (6), $R_i(\{\boldsymbol{w}_k\})$ is analogous to the achievable rate of a channel with channel matrix $\boldsymbol{Q}_{ii}$ and interference-plus-noise power $1/\xi_i(\{\boldsymbol{w}_k\}_{k \neq i})$. The key advantage of formulation (5) is that the constraint set is separable w.r.t. the K beam-forming vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$, though the objective function $U(R_1(\{\boldsymbol{w}_i\}), \ldots, R_K(\{\boldsymbol{w}_i\}))$ is involved with all $\boldsymbol{w}_k$ coupled together. Nevertheless, this type of problems can be conveniently handled by the BSUM method [31] in a distributed and low-complexity manner, yielding an efficient algorithm for solving the outage-constrained CoBF problem (5).

B. Brief Reiview of BSUM

In this subsection, using problem (5) as an example, we briefly review the BSUM method in [31]. For ease of exposition, let us define

$$\mathcal{U}(\{\boldsymbol{w}_k\}) \triangleq U(R_1(\{\boldsymbol{w}_k\}), \ldots, R_K(\{\boldsymbol{w}_k\})).$$

The BSUM method [31] is a block-coordinate-decent (BCD) type method [34] where the block variables are updated in a

---

[1]Note that $\Phi_i(\xi_i, \{\boldsymbol{w}_k\}_{k \neq i})$ is strictly increasing w.r.t. ξ_i. Moreover, since $\Phi_i(0, \{\boldsymbol{w}_k\}_{k \neq i}) = \ln \rho_i < 0$ and $\Phi_i(\sigma_i^{-2} \ln \rho_i^{-1}, \{\boldsymbol{w}_k\}_{k \neq i}) = \sum_{k \neq i} \ln(1 + (\boldsymbol{w}_k^H \boldsymbol{Q}_{ki} \boldsymbol{w}_k)(\sigma_i^{-2} \ln \rho_i^{-1})) \geq 0$, the solution of $\Phi_i(\xi_i, \{\boldsymbol{w}_k\}_{k \neq i}) = 0$ must be positive, i.e., $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i}) > 0, \forall \{\boldsymbol{w}_k\}_{k \neq i}$, and can be efficiently obtained by bisection search.

round-robin fashion, i.e., following the Gauss-Seidel update rule. For problem (5), $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ are the K block variables. In the nth iteration, variable $\boldsymbol{w}_i$, where $i := (n - 1 \bmod K) + 1$, is updated by solving the problem

$$\boldsymbol{w}_i^{[n]} = \arg \max_{\boldsymbol{w}_i \in \mathbb{C}^{N_t}} \quad \bar{\mathcal{U}}^{(i)}\left(\boldsymbol{w}_i | \left\{\boldsymbol{w}_k^{[n-1]}\right\}\right) \tag{8a}$$

$$\text{s.t.} \quad \|\boldsymbol{w}_i\|^2 \le P_i, \tag{8b}$$

and $\boldsymbol{w}_j^{[n]} = \boldsymbol{w}_j^{[n-1]}$ for all $j \neq i$, where $\{\boldsymbol{w}_k^{[n-1]}\}$ denote the beamforming vectors obtained in the $(n - 1)$th iteration, and $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\boldsymbol{w}_k^{[n-1]}\})$ is a surrogate function of $\mathcal{U}(\{\boldsymbol{w}_k\})$ given $\{\boldsymbol{w}_k^{[n-1]}\}$. The introduction of the surrogate function $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\boldsymbol{w}_k^{[n-1]}\})$ provides extra flexibility in the algorithm design. In particular, rather than using the original function $\mathcal{U}(\{\boldsymbol{w}_k\})$, one may choose an advisable $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\boldsymbol{w}_k^{[n-1]}\})$ that can either make problem (8) easily solvable or further lead to a closed-form solution. Hence, the BSUM method is particularly useful when the original objective function is intricate and difficult to optimize, which is the case in problem (5) since $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i})$ are implicit functions without closed-form expression. It has been shown in [31] that the BSUM method performs very well in several practical signal processing and communication applications.

Theoretically, the BSUM method has the following convergence property.

Theorem 1 *[31, Proposition 2, Theorem 2(b)]*: *The iterates* $(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]})$ *converge to the set of stationary points of problem (5) as long as*

$$\mathcal{U}(\{\boldsymbol{w}_k\}) \text{ is differentiable in } \{\boldsymbol{w}_k\}; \tag{9a}$$

$$\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \le \mathcal{U}(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i}); \tag{9b}$$

$$\bar{\mathcal{U}}^{(i)}(\bar{\boldsymbol{w}}_i | \{\bar{\boldsymbol{w}}_k\}) = \mathcal{U}(\{\bar{\boldsymbol{w}}_k\}); \tag{9c}$$

$$\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \text{ is continuous in } (\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}); \tag{9d}$$

$$\text{problem (8) has a unique solution}, \tag{9e}$$

for all $\|\boldsymbol{w}_i\|^2 \le P_i$, $\|\bar{\boldsymbol{w}}_k\|^2 \le P_k$, $i, k = 1, \ldots, K$.

Condition (9a) requires that the system utility function $U(R_1, \ldots, R_K)$ is differentiable, e.g., the weighted sum-rate utility, the proportional fairness utility and the harmonic mean utility.[2] Conditions (9b) and (9c) imply that $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ is a universal lower bound of $\mathcal{U}(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ and it is tight locally when $\boldsymbol{w}_i = \bar{\boldsymbol{w}}_i$.

If all the K beamforming vectors are treated as one block variable $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)$, then the BSUM method reduces to the successive upper bound minimization (SUM) method [31]. In Section IV, we will use this SUM method to devise another algorithm for problem (5) with the weighted sum rate utility.

C. DBSUM for Problem (5)

As seen, to apply the BSUM method to our problem (5), one of the key steps is to construct appropriate surrogate functions $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$, $i = 1, \ldots, K$, that satisfy conditions in (9b)–(9e). It turns out that this is not a trivial task since there is no explicit expression for $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i})$. To overcome this, we notice that, in (6), $R_i(\{\boldsymbol{w}_k\})$ has some nice monotonicity and concavity (respectively convexity) w.r.t. $\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i$ (respectively $\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k$), as stated in the following lemma.

Lemma 1: *For each* $i \in \{1, \ldots, K\}$, *the function* $R_i(\{\boldsymbol{w}_k\})$ *in (6) is strictly increasing and strictly concave w.r.t.* $\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i$, *while it is nonincreasing and convex w.r.t. each* $\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k$ *where* $k \in \{1, \ldots, K\} \setminus \{i\}$.

The proof is given in Appendix A. Based on Lemma 1, we propose the following surrogate function for updating $\boldsymbol{w}_i$.

$$\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \triangleq$$
$$U\left(\bar{R}_1^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}), \ldots, \bar{R}_K^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})\right) - \frac{c}{2}\|\boldsymbol{w}_i - \bar{\boldsymbol{w}}_i\|^2, \tag{10}$$

where $c > 0$ is a penalty parameter and

$$\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \triangleq$$
$$\begin{cases} \log_2\left(1 + \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})(2\Re\{\bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i\} - \bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ii} \bar{\boldsymbol{w}}_i)\right), j = i, \\ R_j(\{\bar{\boldsymbol{w}}_k\}) + \frac{\partial R_j(\{\bar{\boldsymbol{w}}_k\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}\left(\boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i - \bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ij} \bar{\boldsymbol{w}}_i\right), j \neq i, \end{cases} \tag{11}$$

where $\Re\{x\}$ denotes the real part of $x \in \mathbb{C}$. A closed-form expression of the partial derivative $\frac{\partial R_j(\{\bar{\boldsymbol{w}}_k\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}$ in (11) is given by (12), shown at bottom of the page, where $\bar{I}_{kj} \triangleq \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{kj} \bar{\boldsymbol{w}}_k$ for all j, k, and the second equality therein is obtained by the implicit function theorem [35] (for computing $\frac{\partial \xi_j(\{\boldsymbol{w}_k\}_{k \neq j})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}$).

[2]We should mention that the BSUM method [31] can also handle non-differentiable problems, but it requires additional regularity assumption on the objective function. The non-differentiable MMF rate utility $U(R_1, \ldots, R_K) = \min_{i \in \{1, \ldots, K\}} R_i / \alpha_i$ unfortunately does not satisfy the regularity assumption. Alternative approach to handling the MMF rate utility problem will be discussed in Section III-D.

$$\frac{\partial R_j(\{\bar{\boldsymbol{w}}_k\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i} = \frac{\partial \log_2(1 + \xi_j \cdot \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{jj} \bar{\boldsymbol{w}}_j)}{\partial \xi_j}\bigg|_{\xi_j = \xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})} \times \frac{\partial \xi_j(\{\boldsymbol{w}_k\}_{k \neq j})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}\bigg|_{\boldsymbol{w}_k = \bar{\boldsymbol{w}}_k, \forall k \neq i}$$

$$= \frac{\bar{I}_{jj}}{\ln 2 \cdot (1 + \xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})\bar{I}_{jj})} \times \frac{-\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})}{1 + \bar{I}_{ij}\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})} \frac{1}{\sigma_j^2 + \sum_{\ell \neq j} \bar{I}_{\ell j}(1 + \bar{I}_{\ell j}\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j}))^{-1}}$$

$$= \frac{-\bar{I}_{jj}\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})}{\ln 2 \cdot (1 + \xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})\bar{I}_{jj})}\left[(1 + \bar{I}_{ij}\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})) \cdot \left(\sigma_j^2 + \sum_{\ell \neq j} \frac{\bar{I}_{\ell j}}{1 + \bar{I}_{\ell j}\xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})}\right)\right]^{-1} \le 0 \tag{12}$$

Using (10) as the surrogate function, the BSUM method in (8) has the following convergence property:

Proposition 2: *Suppose that the system utility* $U(R_1, \ldots, R_K)$ *is differentiable, jointly concave, and is nondecreasing w.r.t. each* R_i. *Then, the sequence* $\{\mathcal{U}(\{\boldsymbol{w}_k^{[0]}\}), \mathcal{U}(\{\boldsymbol{w}_k^{[1]}\}), \ldots\}$ *generated by the BSUM method in (8) and (10) converges monotonically, and every limit point of the sequence* $\{(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]})\}_{n=1}^{\infty}$ *is a stationary point of problem (5).*

Proof: Let $i := (n-1 \bmod K) + 1$. Then we have

$$
\begin{aligned}
\mathcal{U}\left(\{\boldsymbol{w}_k^{[n]}\}\right) &= \mathcal{U}\left(\boldsymbol{w}_i^{[n]}, \{\boldsymbol{w}_k^{[n-1]}\}_{k \neq i}\right) \\
&\geq \bar{\mathcal{U}}^{(i)}\left(\boldsymbol{w}_i^{[n]} | \{\boldsymbol{w}_k^{[n-1]}\}\right) \\
&\geq \bar{\mathcal{U}}^{(i)}\left(\boldsymbol{w}_i^{[n-1]} | \{\boldsymbol{w}_k^{[n-1]}\}\right) \\
&= \mathcal{U}\left(\{\boldsymbol{w}_k^{[n-1]}\}\right),
\end{aligned}
\tag{13}
$$

where the first inequality comes from (9b), the second inequality comes from the optimality of $\boldsymbol{w}_i^{[n]}$ to problem (8), and the last equality results from (9c). Equation (13) implies that the system utility is nondecreasing from one iteration to another. On the other hand, due to the transmit power constraints (5b), the sequence $\{\mathcal{U}(\{\boldsymbol{w}_k^{[n]}\})\}_{n=1}^{\infty}$ is bounded. Hence, the system utility converges monotonically. Next we show that the surrogate function (10) satisfies (9b)–(9e). Therefore, by Theorem 1, every limit point of the sequence $\{(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]})\}_{n=1}^{\infty}$ is a stationary point of problem (5).

By the first-order condition of the convex function $\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i$, we have

$$
2\Re\{\bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i\} - \bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ii} \bar{\boldsymbol{w}}_i \leq \boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i, \tag{14}
$$

for all $\boldsymbol{w}_i, \bar{\boldsymbol{w}}_i$, and the equality holds when $\boldsymbol{w}_i = \bar{\boldsymbol{w}}_i$. The monotonicity of the $\log_2(\cdot)$ function together with (14) implies that $\bar{R}_i^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ in (11) satisfies

$$
\bar{R}_i^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \leq R_i(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i}), \forall \boldsymbol{w}_i, \tag{15a}
$$

$$
\bar{R}_i^{(i)}(\bar{\boldsymbol{w}}_i | \{\bar{\boldsymbol{w}}_k\}) = R_i(\{\bar{\boldsymbol{w}}_k\}). \tag{15b}
$$

Moreover, it is clear that $\bar{R}_i^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ is concave in $\boldsymbol{w}_i$.

For $j \neq i$, since $R_j(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ is convex w.r.t. $\boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i$ according to Lemma 1, the first-order approximation of $R_j(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ w.r.t. $\boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i = \bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ij} \bar{\boldsymbol{w}}_i$, i.e., $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ in (11) for $j \neq i$, satisfies

$$
\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \leq R_i(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\}_{k \neq i}), \forall \boldsymbol{w}_i, \tag{16a}
$$

$$
\bar{R}_j^{(i)}(\bar{\boldsymbol{w}}_i | \{\bar{\boldsymbol{w}}_k\}) = R_i(\{\bar{\boldsymbol{w}}_k\}). \tag{16b}
$$

Since $\frac{\partial R_j(\{\bar{\boldsymbol{w}}_k\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}$ is non-positive (see (12)), $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ in (11) for $j \neq i$ is concave in $\boldsymbol{w}_i$. Besides, by the fact that $\xi_i(\cdot)$ is a continuously differentiable function (see Proposition 1), $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ is continuous in $(\boldsymbol{w}_i, \{\bar{\boldsymbol{w}}_k\})$, for all $j = 1, \ldots, K$.

The surrogate function $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ in (10) thereby has the following properties. First, from (15), (16), continuity of $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \forall j$, and the monotonicity of $U(R_1, \ldots, R_K)$,

we conclude that $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ in (10) satisfies the conditions (9b)–9(d). Second, the composition of $U(R_1, \ldots, R_K)$ and $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$, $j = 1, \ldots, K$, i.e., $U(\bar{R}_1^{(i)}(\cdot | \cdot), \ldots, \bar{R}_K^{(i)}(\cdot | \cdot))$, is concave in $\boldsymbol{w}_i$ due to that $\bar{R}_j^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\})$ is concave for $j = 1, \ldots, K$ and that $U(R_1, \ldots, R_K)$ is concave and nondecreasing [36, Section 3.2.4]. Moreover, the quadratic penalty $-\frac{c}{2}\|\boldsymbol{w}_i - \bar{\boldsymbol{w}}_i\|^2$ is strongly concave. Hence,

$$
\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i | \{\bar{\boldsymbol{w}}_k\}) \text{ in (10) is strongly concave,} \tag{17}
$$

implying that (9e) holds true. Therefore, we obtain from Theorem 1 that every limit point of the sequence $\{(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]})\}_{n=1}^{\infty}$ is a stationary point of problem (5). ∎

As indicated in (17), the surrogate function given in (10) is (strongly) concave. Hence, problem (8) is a convex problem which is efficiently solvable. In particular, there is only a simple 2-norm constraint in problem (8), which makes it easily implementable by using, e.g., the gradient projection method [34, Section 2.3.1]. More importantly, the BSUM method can be implemented in a distributed manner, as only one user is involved at each iteration. Information required for solving (8) can be obtained through message exchange between users. This leads to the proposed DBSUM algorithm as detailed in Algorithm 1.

Algorithm 1: DBSUM algorithm for handling problem (5)

1: **Given** a set of beamformers $\{\boldsymbol{w}_i^{[0]}\}$ satisfying (5b), and set $n := 0$; Transmitter i sends the quantity $(\boldsymbol{w}_i^{[0]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[0]}$ to transmitter j, $\forall j \neq i$, $i = 1, \ldots, K$.
2: **repeat**
3:     $n := n + 1$;
4:     $i := (n-1 \bmod K) + 1$;
5:     For all $j \neq i$, transmitter j computes $R_j(\{\boldsymbol{w}_k^{[n-1]}\})$ and $\frac{\partial R_j(\{\boldsymbol{w}_k^{[n-1]}\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}$ by (6) and (12), respectively, and sends them to transmitter i;
6:     Transmitter i solves (8) using (10) and (11) to obtain $\boldsymbol{w}_i^{[n]}$, and sends the quantity $(\boldsymbol{w}_i^{[n]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[n]}$ to transmitter j, $\forall j \neq i$;
7:     $\boldsymbol{w}_k^{[n]} := \boldsymbol{w}_k^{[n-1]}$, $\forall k \neq i$;
8: **until** the predefined stopping criterion is met.
9: **Output** $\{\boldsymbol{w}_i^{[n]}\}$ as an approximate solution of problem (5).

D. MMF Rate Utility Maximization

Unfortunately, the MMF rate utility $U(R_1, \ldots, R_K) = \min_{i \in \{1, \ldots, K\}} R_i / \alpha_i$ is not differentiable, and thus the DBSUM algorithm (Algorithm 1) cannot directly be applied. To resolve this issue, we consider the *log-sum-exp* approximation of the min function [36]; specifically, it is known that, for arbitrary $a_1, a_2, \ldots, a_N \in \mathbb{R}$,

$$
\min_{n \in \{1, \ldots, N\}} a_n \geq -\frac{1}{\gamma} \log_2 \left(\sum_{n=1}^{N} 2^{-\gamma a_n} \right) \geq \min_{n \in \{1, \ldots, N\}} a_n - \frac{\log_2 N}{\gamma} \tag{18}
$$

where γ can be any positive real value. The inequalities in (18) show that $-\frac{1}{\gamma} \log_2 (\sum_{n=1}^{N} 2^{-\gamma a_n})$ can be used as an approximation of $\min_{n \in \{1, \ldots, N\}} a_n$, and the approximation error is no

larger than $\frac{\log_2 N}{\gamma}$. By (18), we approximate the MMF rate utility as

$$\min_{i \in \{1,\dots,K\}} \frac{R_i}{\alpha_i} \approx -\frac{1}{\gamma} \log_2 \left(\sum_{i=1}^{K} 2^{-\gamma R_i / \alpha_i} \right) \triangleq \bar{U}(R_1, \dots, R_K).$$

It is readily to see that $\bar{U}(R_1, \dots, R_K)$ is differentiable, jointly concave in $(R_1, \dots, R_K)$, and is strictly increasing w.r.t. each R_i, $i = 1, \dots, K$. Therefore, the DBSUM algorithm can be applied.

IV. DISTRIBUTED WMMSE ALGORITHM FOR WEIGHTED SUM RATE MAXIMIZATION

In the previous section, the DBSUM algorithm for problem (5) updates the beamforming vectors in the Gauss-Seidel manner, though it can handle a general utility function. In this section, we focus on the weighted sum rate (WSR) utility $U(R_1, \dots, R_K) = \sum_{i=1}^{K} \alpha_i R_i$ and further propose a Jacobi-type distributed algorithm where the beamforming vectors are updated in parallel at each iteration. The idea behind is a judicious combination of the SUM method (i.e., the BSUM method with only one block) [31] and the WMMSE reformulation [13]. To proceed, let us rewrite (5) with the WSR utility here

$$\max_{\boldsymbol{w}_i \in \mathbb{C}^{N_t}, i=1,\dots,K} \quad \mathcal{U}_{wsr}(\{\boldsymbol{w}_k\}) \tag{19a}$$

$$\text{s.t.} \quad \|\boldsymbol{w}_i\|^2 \leq P_i, i = 1, \dots, K, \tag{19b}$$

where

$$\mathcal{U}_{wsr}(\{\boldsymbol{w}_k\}) \triangleq \sum_{i=1}^{K} \alpha_i \log_2(1 + \xi_i(\{\boldsymbol{w}_k\}_{k \neq i}) \boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i).$$

Note that the WMMSE method in [13] is not directly applicable to problem (19) as the objective function of (19) does not have an explicit expression. Therefore, we will first design a surrogate objective function for (19), followed by applying the WMMSE method to further obtain another surrogate function that is amendable to parallel implementation.

Let us recall the function $\Phi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i})$ in (7). Given any feasible point $\{\bar{\boldsymbol{w}}_k\}$ satisfying (19b), $\Phi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i})$ has an upper bound as follows

$$\Phi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i}) = \ln \rho_i + \sigma_i^2 \zeta_i + \sum_{k \neq i} \ln(1 + \boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k \zeta_i)$$

$$\leq \ln \rho_i + \sigma_i^2 \zeta_i + \sum_{k \neq i} \ln(1 + \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i)$$

$$+ \sum_{k \neq i} \frac{\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k \zeta_i - \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i}{1 + \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i}$$

$$= \ln \rho_i + \sum_{k \neq i} \ln(1 + \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i) - \sum_{k \neq i} \frac{\bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i}{1 + \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i}$$

$$+ \left(\sigma_i^2 + \sum_{k \neq i} \frac{\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k}{1 + \bar{\boldsymbol{w}}_k^H \mathbf{Q}_{ki} \bar{\boldsymbol{w}}_k \bar{\zeta}_i} \right) \cdot \zeta_i$$

$$\triangleq \Psi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \tag{20}$$

where $\bar{\zeta}_i = \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})$, and the inequality is due to the first-order approximation of the concave logarithm function, i.e., $\ln(y) \leq \ln(x) + \frac{y-x}{x}$ $\forall x, y \geq 0$. Note that

$\Psi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ is a locally tight upper bound of $\Phi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i})$; moreover, similar to $\Phi_i(\xi_i, \{\boldsymbol{w}_k\}_{k \neq i})$, $\Psi_i(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ is continuously differentiable w.r.t. $(\zeta_i, \{\boldsymbol{w}_k\}_{k \neq i})$, and is strictly increasing w.r.t. ζ_i. As a result, there exists a unique continuously differentiable function, denoted by $\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$, such that

$$\Psi_i\left(\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}), \{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}\right) = 0,$$

for all $\{\boldsymbol{w}_k\}_{k \neq i}$. In particular, it follows from (20) that $\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ has a closed-form expression as

$$\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) =$$

$$\gamma_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \left(\sigma_i^2 + \sum_{j \neq i} \frac{\boldsymbol{w}_j^H \mathbf{Q}_{ji} \boldsymbol{w}_j}{1 + \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})} \right)^{-1} \tag{21}$$

where

$$\gamma_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \triangleq \sum_{j \neq i} \frac{\bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \cdot \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})}{1 + \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \cdot \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})} - \ln \rho_i$$

$$- \sum_{j \neq i} \ln(1 + \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \cdot \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}))$$

$$= \sum_{j \neq i} \frac{\bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \cdot \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})}{1 + \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \cdot \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})}$$

$$+ \sigma_i^2 \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \qquad \text{(by (7))}$$

$$> 0, \forall \{\bar{\boldsymbol{w}}_k\}_{k \neq i}. \tag{22}$$

By (21) and (22), one can see that $\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) > 0$ for all feasible $\{\boldsymbol{w}_k\}_{k \neq i}$ and $\{\bar{\boldsymbol{w}}_k\}_{k \neq i}$. Moreover, from (20) and (21), $\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i})$ is a locally tight lower bound of $\xi_i(\{\boldsymbol{w}_k\}_{k \neq i})$, i.e.,

$$\zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \leq \xi_i(\{\boldsymbol{w}_k\}_{k \neq i}), \tag{23a}$$

$$\zeta_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) = \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}), \tag{23b}$$

for all $\|\boldsymbol{w}_k\|^2 \leq P_k$, $\|\bar{\boldsymbol{w}}_k\|^2 \leq P_k$, $k \neq i$. Therefore, the following function

$$\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\} | \{\bar{\boldsymbol{w}}_k\}) \triangleq$$

$$\sum_{i=1}^{K} \alpha_i \log_2(1 + \zeta_i(\{\boldsymbol{w}_k\}_{k \neq i} | \{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \cdot \boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i). \tag{24}$$

serves as a locally tight lower bound of the WSR utility $\mathcal{U}_{wsr}(\{\boldsymbol{w}_k\})$ in (19a). By defining

$$\bar{\mathbf{Q}}_{ii}(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \triangleq \gamma_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \cdot \mathbf{Q}_{ii}, \tag{25a}$$

$$\bar{\mathbf{Q}}_{ji}(\{\bar{\boldsymbol{w}}_k\}_{k \neq i}) \triangleq \left(1 + \bar{\boldsymbol{w}}_j^H \mathbf{Q}_{ji} \bar{\boldsymbol{w}}_j \xi_i(\{\bar{\boldsymbol{w}}_k\}_{k \neq i})\right)^{-1} \mathbf{Q}_{ji}, \tag{25b}$$

for $i, j = 1, \dots, K$, $j \neq i$, and by (21), one can further express $\tilde{\mathcal{U}}_{wsr}(\cdot | \cdot)$ as

$$\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\} | \{\bar{\boldsymbol{w}}_k\}) =$$

$$\sum_{i=1}^{K} \alpha_i \log_2 \left(1 + \frac{\boldsymbol{w}_i^H \bar{\mathbf{Q}}_{ii} \boldsymbol{w}_i}{\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j} \right), \tag{26}$$

where $\bar{\mathbf{Q}}_{ji}(\{\bar{\boldsymbol{w}}_k\}_{k\neq i})$ are denoted by $\bar{\mathbf{Q}}_{ji}$ for all $i, j = 1, \ldots, K$, for notational simplicity. It is interesting to note that $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ in (26) virtually corresponds to the WSR of an IFC, and the ratio in $\log_2(\cdot)$ is the associated *signal-to-interference-plus-noise-ratio* (SINR). We can hence adopt an idea similar to the iterative WMMSE method in [13], which exploits the relation between the SINR and the minimum mean-square error (MMSE) of the estimated symbols for developing efficient precoder optimization algorithm, to further obtain a lower bound of $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ that is separable over $\{\boldsymbol{w}_k\}$.

Different from [13] wherein the objective function is exactly the WSR of a MIMO IFC, the function $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ actually corresponds to the WSR of K independent MIMO channels with channel matrices $\bar{\mathbf{Q}}_{ii}^{1/2} \in \mathbb{C}^{N_t \times N_t}$, where $(\bar{\mathbf{Q}}_{ii}^{1/2})^H \bar{\mathbf{Q}}_{ii}^{1/2} = \bar{\mathbf{Q}}_{ii}$ for $i = 1, \ldots, K$. The interference terms $\{\boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j\}_{j \neq i}$ are incorporated into the AWGN of the ith MIMO channel, i.e., the additive noise of the ith MIMO channel is given by $\boldsymbol{n}_i \sim \mathcal{CN}(\mathbf{0}, (\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j) \cdot \mathbf{I}_{N_t})$ where $\mathbf{I}_{N_t}$ is the $N_t \times N_t$ identity matrix.

Suppose that the ith transmitter sends the information signal s_i to the ith receiver via transmit beamforming $\boldsymbol{w}_i$, and the receiver estimates s_i by linear decoder $\boldsymbol{y}_i$. Then, the MMSE of the estimation is given by

$$\min_{\boldsymbol{y}_i \in \mathbb{C}^{N_t}} \left|1 - \boldsymbol{y}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \boldsymbol{w}_i\right|^2 + \left(\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j\right) \boldsymbol{y}_i^H \boldsymbol{y}_i$$

$$= 1 - \frac{\boldsymbol{w}_i^H \bar{\mathbf{Q}}_{ii} \boldsymbol{w}_i}{\sigma_i^2 + \sum_{j=1}^{K} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j}$$

$$= \left(1 + \frac{\boldsymbol{w}_i^H \bar{\mathbf{Q}}_{ii} \boldsymbol{w}_i}{\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j}\right)^{-1}. \qquad (27)$$

---

[3]For the SUM method, convergence is guaranteed without the need of unique solution to problem (32); see [31, Theorem 1].

The optimal $\boldsymbol{y}_i$ to (27) can be shown to be

$$\boldsymbol{y}_i = \bar{\mathbf{Q}}_{ii}^{1/2} \boldsymbol{w}_i \sigma_i^2 + \sum_{j=1}^{K} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j. \qquad (28)$$

Then, we can further obtain a lower bound of $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ as presented in (29), shown at the bottom of the page, where

$$\bar{\boldsymbol{y}}_i \triangleq \frac{\bar{\mathbf{Q}}_{ii}^{1/2} \bar{\boldsymbol{w}}_i}{\sigma_i^2 + \sum_{j=1}^{K} \bar{\boldsymbol{w}}_j^H \bar{\mathbf{Q}}_{ji} \bar{\boldsymbol{w}}_j}, i = 1, \ldots, K, \qquad (30)$$

and the second inequality in (29) is obtained by the fact that $-\ln(y) \geq -\ln(x) - \frac{y-x}{x} \; \forall x, y > 0$. Moreover, by (27), (28) and (30), one can show that the lower bound $\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ is actually locally tight to $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ when $\boldsymbol{w}_k = \bar{\boldsymbol{w}}_k$ $\forall k = 1, \ldots, K$, i.e., $\bar{\mathcal{U}}_{wsr}(\{\bar{\boldsymbol{w}}_k\}|\{\bar{\boldsymbol{w}}_k\}) = \tilde{\mathcal{U}}_{wsr}(\{\bar{\boldsymbol{w}}_k\}|\{\bar{\boldsymbol{w}}_k\})$. Combining this result with the fact that $\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ is a locally tight lower bound of the original WSR $\mathcal{U}_{wsr}(\{\boldsymbol{w}_k\})$ (cf. (23)), we conclude that $\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ is also a locally tight lower bound of $\mathcal{U}_{wsr}(\{\boldsymbol{w}_k\})$, satisfying

$$\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\}) \leq \mathcal{U}_{wsr}(\{\boldsymbol{w}_k\}), \qquad (31a)$$

$$\bar{\mathcal{U}}_{wsr}(\{\bar{\boldsymbol{w}}_k\}|\{\bar{\boldsymbol{w}}_k\}) = \mathcal{U}_{wsr}(\{\bar{\boldsymbol{w}}_k\}), \qquad (31b)$$

$$\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\}) \text{ is continuous in } (\{\boldsymbol{w}_k\}, \{\bar{\boldsymbol{w}}_k\}), \qquad (31c)$$

for all $\|\boldsymbol{w}_k\|^2 \leq P_k, \|\bar{\boldsymbol{w}}_k\|^2 \leq P_k, k = 1, \ldots, K$.

Therefore, we can apply the SUM method [31] (i.e., BSUM with one block variable $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)$) to problem (19), by using $\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ in (29) as the surrogate function. Specifically, according to SUM, the beamforming vectors are iteratively updated as

$$(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]}) = \arg \max_{\substack{\|\boldsymbol{w}_i\|^2 \leq P_i, \\ i=1,\ldots,K}} \bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\boldsymbol{w}_k^{[n-1]}\}) \qquad (32)$$

By (31) and by [31, Theorem 1], the sequence generated by (32) is guaranteed to converge:[3]

$$\tilde{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$$

$$= \sum_{i=1}^{K} -\alpha_i \log_2\left(\left(1 + \frac{\boldsymbol{w}_i^H \bar{\mathbf{Q}}_{ii} \boldsymbol{w}_i}{\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j}\right)^{-1}\right) \qquad \text{(by (26))}$$

$$\geq \sum_{i=1}^{K} -\alpha_i \log_2\left(\left|1 - \bar{\boldsymbol{y}}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \boldsymbol{w}_i\right|^2 + \left(\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j\right) \bar{\boldsymbol{y}}_i^H \bar{\boldsymbol{y}}_i\right) \qquad \text{(by (27))}$$

$$\geq \sum_{i=1}^{K} -\alpha_i \log_2\left(\left|1 - \bar{\boldsymbol{y}}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \bar{\boldsymbol{w}}_i\right|^2 + \left(\sigma_i^2 + \sum_{j \neq i} \bar{\boldsymbol{w}}_j^H \bar{\mathbf{Q}}_{ji} \bar{\boldsymbol{w}}_j\right) \bar{\boldsymbol{y}}_i^H \bar{\boldsymbol{y}}_i\right)$$

$$+ \frac{\alpha_i}{\ln 2}\left(1 - \frac{|1 - \bar{\boldsymbol{y}}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \boldsymbol{w}_i|^2 + \left(\sigma_i^2 + \sum_{j \neq i} \boldsymbol{w}_j^H \bar{\mathbf{Q}}_{ji} \boldsymbol{w}_j\right) \bar{\boldsymbol{y}}_i^H \bar{\boldsymbol{y}}_i}{|1 - \bar{\boldsymbol{y}}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \bar{\boldsymbol{w}}_i|^2 + \left(\sigma_i^2 + \sum_{j \neq i} \bar{\boldsymbol{w}}_j^H \bar{\mathbf{Q}}_{ji} \bar{\boldsymbol{w}}_j\right) \bar{\boldsymbol{y}}_i^H \bar{\boldsymbol{y}}_i}\right)$$

$$\triangleq \bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\}), \qquad (29)$$

Proposition 3: *Every limit point of $\{(\boldsymbol{w}_1^{[n]}, \ldots, \boldsymbol{w}_K^{[n]})\}_{n=1}^{\infty}$ generated by (32) is a stationary point of problem (19).*

Unlike the DBSUM algorithm (Algorithm 1), implementation of (32) can be completely parallel with only a small amount of messages exchanged among the transmitters. Specifically, because both the surrogate function $\bar{\mathcal{U}}_{wsr}(\{\boldsymbol{w}_k\}|\{\bar{\boldsymbol{w}}_k\})$ and the constraint set are separable over the beamforming vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$, problem (32) can be decomposed into K parallel subproblems as (see (29))

$$\boldsymbol{w}_i^{[n]} = \arg \min_{\|\boldsymbol{w}_i\|^2 \leq P_i} \left\{ \eta_i |1 - \bar{\boldsymbol{y}}_i^H \bar{\mathbf{Q}}_{ii}^{1/2} \boldsymbol{w}_i|^2 \right.$$
$$\left. + \sum_{j \neq i} \eta_j \left(\boldsymbol{w}_i^H \bar{\mathbf{Q}}_{ij} \boldsymbol{w}_i \right) \bar{\boldsymbol{y}}_j^H \bar{\boldsymbol{y}}_j \right\} \quad (33)$$

for $i = 1, \ldots, K$, where

$$\eta_j \triangleq \alpha_j \ln 2 \left[|1 - \bar{\boldsymbol{y}}_j^H \bar{\mathbf{Q}}_{jj}^{1/2} \bar{\boldsymbol{w}}_j|^2 + \left(\sigma_j^2 + \sum_{k \neq j} \bar{\boldsymbol{w}}_k^H \bar{\mathbf{Q}}_{kj} \bar{\boldsymbol{w}}_k \right) \bar{\boldsymbol{y}}_j^H \bar{\boldsymbol{y}}_j \right]^{-1},$$

for $j = 1, \ldots, K$. In addition, problem (33) can be solved very efficiently, e.g., using the gradient projection method [34, Section 2.3.1] or the Lagrange dual method [36]. Finally, we summarize the proposed DWMMSE algorithm for problem (19) in Algorithm 2.

Algorithm 2: DWMMSE algorithm for problem (19)

1: **Input** a set of beamformers $\{\boldsymbol{w}_i^{[0]}\}$ satisfying (19b);

2: Set $n := 0$;

3: **repeat**

4:   $n := n + 1$;

5:   Each transmitter i obtains $\boldsymbol{w}_i^{[n]}$ by solving (33), and sends $(\boldsymbol{w}_i^{[n]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[n]}$ to transmitter j for all $j \neq i$;

6:   After receiving the quantities $(\boldsymbol{w}_i^{[n]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[n]} \, \forall i \neq j$, each transmitter j sends

$$\theta_{ij} = \frac{\eta_j \bar{\boldsymbol{y}}_j^H \bar{\boldsymbol{y}}_j}{1 + \bar{\boldsymbol{w}}_i^H \mathbf{Q}_{ij} \bar{\boldsymbol{w}}_i \cdot \xi_j(\{\bar{\boldsymbol{w}}_k\}_{k \neq j})}$$

   to transmitter i for all $i \neq j$, where $\bar{\boldsymbol{w}}_k = \boldsymbol{w}_k^{[n]}, \forall k$, and $\theta_{ij} \mathbf{Q}_{ij} = \eta_j \bar{\mathbf{Q}}_{ij} \bar{\boldsymbol{y}}_j^H \bar{\boldsymbol{y}}_j$ (cf. (25b)), $\forall i, j, i \neq j$;

7: **until** the predefined stopping criterion is met.

8: **Output** $\{\boldsymbol{w}_i^{[n]}\}$ as an approximate solution of (19).

Remark 1: It is worth mentioning that the proposed DBSUM algorithm and DWMMSE algorithm are computationally much more efficient than the DSCA algorithm proposed in [29]. Specifically, we solve the subproblem (8) in DBSUM algorithm and the subproblem (33) in DWMMSE algorithm by the gradient projection method and the Lagrange dual method, respectively, while the subproblems in DSCA needs to be implemented by the interior-point method due to the involved

structure. It is known that the gradient projection method converges linearly if the objective function is Lipschitz continuous and strongly convex (concave) [37]. We have shown that the surrogate function $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i|\{\bar{\boldsymbol{w}}_k\})$ is strongly concave (cf. (17)). Hence, if $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i|\{\bar{\boldsymbol{w}}_k\})$ is also Lipschitz continuous, e.g., when $U(R_1, \ldots, R_K)$ is the weighted sum rate utility, the computational complexity for solving problem (8) by gradient projection method is $\mathcal{O}(KN_t^2 \log \epsilon_1^{-1})$, where $\mathcal{O}(KN_t^2)$ is the complexity order of computing the gradient of $\bar{\mathcal{U}}^{(i)}(\boldsymbol{w}_i|\{\bar{\boldsymbol{w}}_k\})$ and $\epsilon_1 > 0$ is the solution accuracy of the gradient projection method. Therefore, the overall complexity of the DBSUM algorithm is

$$\kappa_1 K \cdot \mathcal{O} \left(KN_t^2 \log \epsilon_1^{-1} \right), \quad (34)$$

where κ_1 is the number of round-robin iterations run by the DBSUM algorithm. On the other hand, the computational complexity for solving problem (33) by Lagrange dual method is $\mathcal{O}(N_t^3 + KN_t^2 + N_t^2 \log \epsilon_2^{-1})$, where $\mathcal{O}(N_t^3)$ is due to the computation of matrix inversion, $\mathcal{O}(KN_t^2)$ and $\mathcal{O}(N_t^2)$ are due to matrix-vector multiplication, and $\epsilon_2 > 0$ is the solution accuracy of the Lagrange dual method. Therefore, the overall complexity of the DWMMSE algorithm is

$$\kappa_2 K \cdot \mathcal{O} \left(N_t^3 + KN_t^2 + N_t^2 \log \epsilon_2^{-1} \right), \quad (35)$$

where κ_2 is the number of iterations run by the DWMMSE algorithm. The overall complexity of the DSCA algorithm is given by [1]

$$\kappa_3 K \cdot \mathcal{O} \left((N_t^{6.5} + K^{3.5}) \log \epsilon_3^{-1} \right), \quad (36)$$

where κ_3 is the number of iterations run by the DSCA algorithm, and $\epsilon_3 > 0$ is the solution accuracy of the interior-point method for solving the subproblems therein.

To compare (34), (35) and (36), as can be seen from Section VI and [29], the iteration numbers κ_1, κ_2 and κ_3 of the three algorithms are similar in practice and of the same order. So their practical computational complexity highly depends on the per (round-robin) iteration complexity. It is seen from (34)–(36) that the per iteration complexity of the DBSUM algorithm and the DWMMSE algorithm is much lower than that of the DSCA algorithm. In particular, the per iteration complexity of the DBSUM algorithm and the DWMMSE algorithm only increases linearly with the number of users, K, while the per iteration complexity of the DSCA algorithm increases in the rate of $K^{3.5}$, implying that the DBSUM algorithm and the DWMMSE algorithm have much better scalability w.r.t. the cardinality of coordinated cluster. We will further justify the above complexity analysis by simulations in Section VI.

Remark 2: The communication overhead induced by information exchange between the transmitters is an important issue in the implementation of distributed algorithms. To address this, we compare the communication overhead of the two proposed algorithms with the scheme where full CDI is exchanged between the transmitters so that the CoBF problem (3) can be handled independently by each transmitter. Assuming that the transmitters are connected by backhaul links, e.g., optical fibers, and

that the transmitters exchange information in a point-to-point fashion, the communication overhead induced by exchanging full CDI between the transmitters is $K^2(K-1)N_t^2$ real values [29]. For the DBSUM algorithm, each transmitter i needs to acquire two real scalars, i.e., $R_j(\{\boldsymbol{w}_k^{[n-1]}\})$ and $\frac{\partial R_j(\{\boldsymbol{w}_k^{[n-1]}\})}{\partial \boldsymbol{w}_i^H \mathbf{Q}_{ij} \boldsymbol{w}_i}$, from transmitter j for all $j \neq i$ in order to update its beamformer $\boldsymbol{w}_i$. After updating $\boldsymbol{w}_i$ by solving problem (8), transmitter i needs to inform transmitter j of the scalar $(\boldsymbol{w}_i^{[n]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[n]}$ for all $j \neq i$. Therefore, the communication overhead associated with updating $\boldsymbol{w}_i$ once is $3(K-1)$ real values, and hence the total communication overhead of the DBSUM algorithm is $\kappa_1 \times K \times 3(K-1) = 3\kappa_1 K(K-1)$ real values. In one iteration (one parallel update of all beamformers $\{\boldsymbol{w}_k\}$ of the DWMMSE algorithm, each transmitter i needs to inform transmitter j of and acquire from transmitter j the two real scalars $(\boldsymbol{w}_i^{[n]})^H \mathbf{Q}_{ij} \boldsymbol{w}_i^{[n]}$ and θ_{ij} (cf. step 6 of Algorithm 2), respectively. Hence, the total communication overhead of the DWMMSE algorithm is $\kappa_2 \times K \times 2(K-1) = 2\kappa_2 K(K-1)$ real values. One can see that, if $\kappa_1 < KN_t^2/3$, then the DBSUM algorithm has a smaller amount of communication overhead than exchanging full CDI between transmitters. Analogously, the communication overhead of DWMMSE is smaller than exchanging full CDI if $\kappa_2 < KN_t^2/2$. As we show in Section VI, the DBSUM algorithm and the DWMMSE algorithm can converge within $KN_t^2/3$ and $KN_t^2/2$ iterations, respectively.

V. Outer Approximation by Polyblock Optimization

The DBSUM algorithm and the DWMMSE algorithm (that are based on BSUM and SUM methods [31], respectively) presented in the previous two sections are so called "inner" approximation methods [38] since, at each iteration, the approximate beamforming solutions are restrictively feasible and provide lower bounds to problem (3). In this section, we consider an "outer" approximation method that instead solves a constraint-relaxed version of problem (3), thus providing upper bounds to the optimal value of problem (3). The motive is that the proposed DBSUM and WMMSE algorithms can be benchmarked against such a method, as the approximation errors of the proposed algorithms are no larger than the gap between the outer and inner approximation methods. Compared with the exhaustive search method which is not feasible when the number of users is large, the outer approximation method is computationally more efficient.

Our approach is based on the polyblock outer approximation (POA) algorithm, which is used for solving the *monotonic optimization problems* [32]. To be self-contained, a review of the POA algorithm is given in Appendix B. Roughly speaking, the POA algorithm systematically constructs a sequence of optimization problems which has a structured feasible set (called polyblock; see Definition 1 in Appendix B) that contains the feasible set of the original problem. The structured feasible set shrinks at every iteration and converges to the true feasible set of the original problem. Thereby, the objective values of the constructed problems converge to the true optimal value from above asymptotically.

The POA algorithm has been used as an outer approximation method for the utility maximization problem in IFC under different assumptions about the CSI at the transmitters [20]–[22],

[39]. While the POA algorithm can also be applied to our problem (3), it will involve solving a sequence of NP-hard subproblems and thus is not computationally feasible. To see this, let us define the achievable rate region of (3) as (cf. (4), (7))

$$\mathcal{R}(\{\boldsymbol{w}_k\}) \triangleq$$
$$\left\{ [R_1, \ldots, R_K]^T \succeq \mathbf{0} \,\middle|\, \Phi_i\left(\frac{2^{R_i}-1}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}, \{\boldsymbol{w}_k\}_{k\neq i}\right) \leq 0, \forall i \right\}, \quad (37)$$

and write problem (3) compactly as

$$\max_{\substack{R_i \geq 0, \\ i=1,\ldots,K}} \quad U(R_1, \ldots, R_K) \tag{38a}$$

$$\text{s.t.} \quad [R_1, \ldots, R_K]^T \in \mathcal{R} \triangleq \bigcup_{\substack{\|\boldsymbol{w}_i\|^2 \leq P_i, \\ i=1,\ldots,K}} \mathcal{R}(\{\boldsymbol{w}_k\}). \tag{38b}$$

By the fact that $\Phi_i\left(\frac{2^{R_i}-1}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}, \{\boldsymbol{w}_k\}_{k\neq i}\right)$ is increasing w.r.t. R_i, one can easily verify that $\mathcal{R}(\{\boldsymbol{w}_k\})$ is a normal set; thus $\mathcal{R} \subseteq \mathbb{R}_+^K$, which is the union of normal sets, is also a normal set [32, Proposition 3]. As a result, problem (38) is a monotonic optimization problem. However, different from [20]–[22], [39], directly applying the POA algorithm (Algorithm 4 in Appendix B) to problem (38) results in prohibitively high computational complexity. In particular, both step 3 and step 7 of Algorithm 4 for problem (38) corresponds to solving a problem in the form

$$
\begin{aligned}
&\max_{\beta \geq 0} \quad U(\beta v_1^\star, \ldots, \beta v_K^\star) \\
&\quad \text{s.t.} \quad [\beta v_1^\star, \ldots, \beta v_K^\star]^T \in \mathcal{R} \\
&= \max_{\substack{\beta \geq 0, \boldsymbol{w}_i \in \mathbb{C}^{N_t}, \\ i=1,\ldots,K}} \quad \beta \\
&\quad \text{s.t.} \quad \Phi_i\left(\frac{2^{\beta v_i^\star}-1}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}, \{\boldsymbol{w}_k\}_{k\neq i}\right) \leq 0, \\
&\quad\quad\quad \|\boldsymbol{w}_i\|^2 \leq P_i, i=1,\ldots,K,
\end{aligned} \tag{39}
$$

where $\boldsymbol{b}^\star = [v_1^\star, \ldots, v_K^\star]^T \succeq \mathbf{0}$ is a given point, and the equality is due to the fact that the utility $U(\cdot)$ is nondecreasing. As seen, problem (39) is equivalent to problem (3) with the MMF rate utility, which, however, is NP-hard in general (when $N_t \geq 2$) as proved in [30, Theorem 3]. Hence, it is inefficient to use the POA algorithm to solve problem (38).

To overcome this issue, we instead consider a relaxed convex approximation problem. Let us consider a lower bound of $\Phi_i\left(\frac{2^{R_i}-1}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}, \{\boldsymbol{w}_k\}_{k\neq i}\right)$ (cf. (7)) as follows

$$
\begin{aligned}
&\Phi_i\left(\frac{2^{R_i}-1}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}, \{\boldsymbol{w}_k\}_{k\neq i}\right) \\
&\geq \ln \rho_i + \ln\left(1 + \frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}\right) \\
&\quad + \sum_{k\neq i} \ln\left(1 + \frac{(2^{R_i}-1)\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}\right) \\
&= \ln\left[\rho_i \times \left(1 + \frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}\right)\right. \\
&\quad \left. \times \prod_{k\neq i}\left(1 + \frac{(2^{R_i}-1)\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}\right)\right], \quad (40)
\end{aligned}
$$

for $i = 1, \ldots, K$, where the inequality is owing to $x \geq \ln(1+x) \ \forall x \geq 0$. Moreover, since the terms $\frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i}$ and $\frac{(2^{R_i}-1)\boldsymbol{w}_k^H \mathbf{Q}_{ki} \boldsymbol{w}_k}{\boldsymbol{w}_i^H \mathbf{Q}_{ii} \boldsymbol{w}_i} \ \forall k \neq i$ are non-negative, we can further obtain

$$
\begin{aligned}
&\ln\left[\rho_i \cdot \left(1 + \frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \mathbf{Q}_{ii}\boldsymbol{w}_i}\right) \prod_{k \neq i}\left(1 + \frac{(2^{R_i}-1)\boldsymbol{w}_k^H \mathbf{Q}_{ki}\boldsymbol{w}_k}{\boldsymbol{w}_i^H \mathbf{Q}_{ii}\boldsymbol{w}_i}\right)\right] \\
&\geq \ln\left[\rho_i \cdot \left(1 + \frac{(2^{R_i}-1)\sigma_i^2}{\boldsymbol{w}_i^H \mathbf{Q}_{ii}\boldsymbol{w}_i} + \sum_{k \neq i}\frac{(2^{R_i}-1)\boldsymbol{w}_k^H \mathbf{Q}_{ki}\boldsymbol{w}_k}{\boldsymbol{w}_i^H \mathbf{Q}_{ii}\boldsymbol{w}_i}\right)\right] \\
&= \ln\left[\rho_i \cdot \left(1 + \frac{\sigma_i^2 + \sum_{k \neq i}\mathrm{tr}(\boldsymbol{w}_k^H \mathbf{Q}_{ki}\boldsymbol{w}_k)}{(2^{R_i}-1)^{-1}\mathrm{tr}(\boldsymbol{w}_i^H \mathbf{Q}_{ii}\boldsymbol{w}_i)}\right)\right]
\end{aligned}
\tag{41}
$$

for $i = 1, \ldots, K$, where $\mathrm{tr}(\cdot)$ represents the trace of a matrix. By using the lower bound in (41), we obtain the following problem which has a relaxed constraint set comparing to problem (3)

$$
\max_{\substack{\boldsymbol{w}_i \in \mathbb{C}^{N_t}, R_i \geq 0, \\ i=1,\ldots,K}} U(R_1, \ldots, R_K) \tag{42a}
$$

$$
\text{s.t.} \quad \frac{\sigma_i^2 + \sum_{k \neq i}\mathrm{tr}(\boldsymbol{w}_k \boldsymbol{w}_k^H \mathbf{Q}_{ki})}{(2^{R_i}-1)^{-1}\mathrm{tr}(\boldsymbol{w}_i \boldsymbol{w}_i^H \mathbf{Q}_{ii})} \leq \frac{1-\rho_i}{\rho_i}, \tag{42b}
$$

$$
\mathrm{tr}(\boldsymbol{w}_i \boldsymbol{w}_i^H) \leq P_i, i = 1, \ldots, K. \tag{42c}
$$

Furthermore, we consider the *semidefinite relaxation* (SDR) technique [40], by which we relax the rank-one $\boldsymbol{w}_i \boldsymbol{w}_i^H$ to a PSD matrix $\mathbf{W}_i \succeq \mathbf{0}$, for all $i = 1, \ldots, K$. The resultant problem can be expressed as

$$
\max_{\substack{R_i \geq 0, \\ i=1,\ldots,K}} U(R_1, \ldots, R_K) \tag{43a}
$$

$$
\text{s.t.} \quad [R_1, \ldots, R_K]^T \in \tilde{\mathcal{R}} \triangleq \bigcup_{\substack{\mathrm{tr}(\mathbf{W}_i) \leq P_i, \\ \mathbf{W}_i \succeq \mathbf{0}, \forall i}} \tilde{\mathcal{R}}(\{\mathbf{W}_k\}), \tag{43b}
$$

where

$$
\tilde{\mathcal{R}}(\{\mathbf{W}_k\}) \triangleq
$$
$$
\left\{ [R_1, \ldots, R_K]^T \succeq \mathbf{0} \ \Bigg| \ \frac{\sigma_i^2 + \sum_{k \neq i}\mathrm{tr}(\mathbf{W}_k \mathbf{Q}_{ki})}{(2^{R_i}-1)^{-1}\mathrm{tr}(\mathbf{W}_i \mathbf{Q}_{ii})} \leq \frac{1-\rho_i}{\rho_i}, \forall i \right\}.
\tag{44}
$$

Note that $\mathcal{R} \subseteq \tilde{\mathcal{R}} \subseteq \mathbb{R}_+^K$, i.e., problem (43) is a relaxed problem of problem (38). Problem (43) is a monotonic optimization problem as $\tilde{\mathcal{R}}$ can be verified to be normal. Moreover, compared to (38), problem (43) can be handled by the POA algorithm in a more efficient manner. Specifically, step 3 and step 7 of Algorithm 4, which is in Appendix B, for problem (43) now correspond to solving

$$
\max_{\substack{\beta \geq 0, \mathbf{W}_i \in \mathbb{C}^{N_t \times N_t}, \\ i=1,\ldots,K}} \beta
$$

$$
\text{s.t.} \quad \frac{\sigma_i^2 + \sum_{k \neq i}\mathrm{tr}(\mathbf{W}_k \mathbf{Q}_{ki})}{(2^{\beta v_i^\star}-1)^{-1}\mathrm{tr}(\mathbf{W}_i \mathbf{Q}_{ii})} \leq \frac{1-\rho_i}{\rho_i},
$$
$$
\mathrm{tr}(\mathbf{W}_i) \leq P_i, \mathbf{W}_i \succeq \mathbf{0}, i = 1, \ldots, K, \tag{45}
$$

which can be shown efficiently solvable by a bisection method [36]. In Algorithm 4, we summarize the POA algorithm for

solving problem (43) to obtain an upper bound of the optimal utility value of problem (3).

VI. SIMULATION RESULTS

In this section, we evaluate the performance of Algorithm 1 and Algorithm 2 by simulations. The noise powers at all receivers are assumed to be the same, i.e., $\sigma_1^2 = \cdots = \sigma_K^2 \triangleq \sigma^2$, and all the power constraints are set to one, i.e., $P_1 = \cdots = P_K = 1$. The channel covariance matrices $\{\mathbf{Q}_{ik}\}$ are randomly generated with full column rank (i.e., $\mathrm{rank}(\mathbf{Q}_{ki}) = N_t$), and with the maximal eigenvalues of $\{\mathbf{Q}_{ik}\}$ normalized to $\lambda_{\max}(\mathbf{Q}_{ii}) = 1$, $\lambda_{\max}(\mathbf{Q}_{ik}) = \eta$ for all $k \neq i$, $i = 1, \ldots, K$. The parameter $\eta \in (0, 1]$, thereby, represents the relative cross-link interference level. The tolerable outage probabilities are set to 10% for all receivers, i.e., $\epsilon_1 = \cdots = \epsilon_K = 0.1$. The stopping conditions of Algorithm 1 and Algorithm 2 are

$$
\left| \mathcal{U}(\{\boldsymbol{w}_k^{[n]}\}) - \mathcal{U}(\{\boldsymbol{w}_k^{[n-K]}\}) \right| < 10^{-3} \left| \mathcal{U}(\{\boldsymbol{w}_k^{[n-K]}\}) \right|; \tag{46a}
$$

$$
\left| \mathcal{U}(\{\boldsymbol{w}_k^{[n]}\}) - \mathcal{U}(\{\boldsymbol{w}_k^{[n-1]}\}) \right| < 10^{-3} \left| \mathcal{U}(\{\boldsymbol{w}_k^{[n-1]}\}) \right|, \tag{46b}
$$

respectively. Note that the DSCA and SCA algorithms in [29] are also subject to the same stopping conditions as in (46), respectively. The four algorithms (DBSUM, DWMMSE, DSCA and SCA) are all initialized by randomly generated unit-norm complex vectors, i.e., $\|\boldsymbol{w}_i^{[0]}\| = 1$, for all $i = 1, \ldots, K$. Besides, we also run the POA algorithm (Algorithm 3) as it can yield an upper bound to problem (3). The subproblem involved in step 3 and the one in step 7 are handled by the convex solver CVX [41], and Algorithm 3 is stopped if it either has spent 200 iterations or has reached the solution accuracy of $\delta = 10^{-3}$. All simulation results are averaged over 500 realizations of CDI $\{\mathbf{Q}_{ik}\}$.

Algorithm 3: POA algorithm for solving problem (43)

1: **Initialization**: Set the solution accuracy as $\delta \geq 0$, and set $n := 0$.
2: Set $\mathcal{V}[0] := \{\boldsymbol{b}^\star[0] \triangleq [v_1^\star[0], \ldots, v_K^\star[0]]^T\}$, where $v_i^\star[0] = \log_2(1 + \ln(1/\rho_i)P_i\lambda_{\max}(\mathbf{Q}_{ii})/\sigma_i^2)$, is the maximal achievable rate of user i, for $i = 1, \ldots, K$;
3: Solve problem (45) with $\boldsymbol{b}^\star = \boldsymbol{b}^\star[0]$ by bisection to obtain $\beta^\star[0]$, and set $\tilde{\boldsymbol{b}}[0] = \beta^\star[0]\boldsymbol{b}^\star[0]$;
4: **while** $U(v_1^\star[n], \ldots, v_K^\star[n]) - U(\tilde{v}_1[n], \ldots, \tilde{v}_K[n]) > \delta$
5: $\quad n := n + 1$;
6: $\quad$ Set $\mathcal{V}[n] = (\mathcal{V}[n-1]\backslash\{\boldsymbol{b}^\star[n-1]\}) \bigcup \{\boldsymbol{b}^\star[n-1] - (v_i^\star[n-1] - \tilde{v}_i[n-1])\boldsymbol{e}_i\}_{i=1}^K$, where $\boldsymbol{e}_i$ is the ith column of the $K \times K$ identity matrix;
7: $\quad$ Find $\boldsymbol{b}^\star[n] = \arg\max_{\boldsymbol{b} \in \mathcal{V}[n]} U(v_1, \ldots, v_K)$ followed by solving problem (45) with $\boldsymbol{b}^\star = \boldsymbol{b}^\star[n]$ by bisection to obtain $\tilde{\boldsymbol{b}}[n] = \beta^\star[n]\boldsymbol{b}^\star[n]$;
8: **end while**
9: **Output** $U(v_1^\star[n], \ldots, v_K^\star[n])$ as the approximation of the optimal value of (43).

Example 1: We demonstrate the efficacy of Algorithm 1, i.e., the DBSUM algorithm, by comparing it with the DSCA algorithm in [29] and the benchmark POA algorithm. We first consider the cases of $K = 2$ and $K = 3$, and the number of
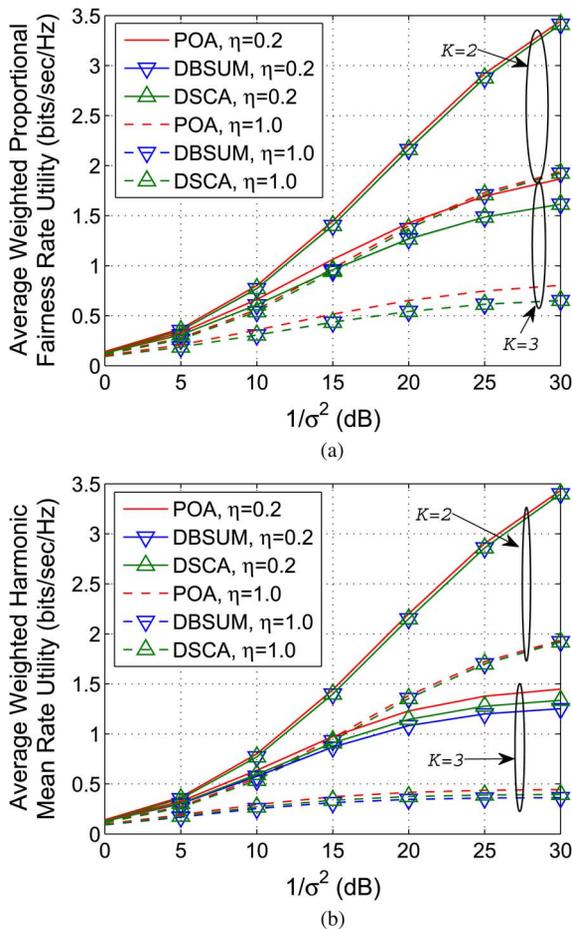
Fig. 1. Performance comparison for the proposed DBSUM algorithm (Algorithm 1) and the DSCA algorithm, for $K = 2$, $(\alpha_1, \alpha_2) = \left(\frac{1}{2}, \frac{1}{2}\right)$, $N_t = 4$, and for $K = 3$, $(\alpha_1, \alpha_2, \alpha_3) = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right)$, $N_t = 4$; (a) average proportional fairness utility and (b) average harmonic mean utility versus $1/\sigma^2$.

transmit antennas is set to $N_t = 4$. The priority weights are set as $(\alpha_1, \alpha_2) = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $(\alpha_1, \alpha_2, \alpha_3) = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right)$ for the $K = 2$ and $K = 3$ cases, respectively. Fig. 1(a) shows some simulation results for the weighted proportional fairness rate utility. One can observe from Fig. 1(a) that the DBSUM algorithm and the DSCA algorithm almost yield the same proportional fairness rate for both $K = 2$ and $K = 3$, and for both $\eta = 0.2$ (the weak interference scenario) and $\eta = 1.0$ (the strong interference scenario). It can also be observed that, for the case of $K = 2$, the DBSUM algorithm and the DSCA algorithm almost achieve the performance upper bound obtained by the POA algorithm, implying that both of them can achieve near optimal performance. For the case of $K = 3$, a non-negligible performance gap between the POA upper bound and the DBSUM and DSCA algorithms can be observed.[4] Nevertheless, both the DBSUM and the DSCA algorithms can achieve at least 80% of the upper bound, indicating that the performance loss must be within 20% compared with the global optimum to problem (3).

Fig. 1(b) displays some simulation results for the weighted harmonic mean rate utility. One can observe that, for the case

---

[4]We found in simulations that under this setting the POA algorithm in general cannot reach the preset solution accuracy within 200 iterations. So the performance gap might be reduced if one allows more iterations for the POA algorithm.

---

of $K = 2$, the DBSUM and the DSCA almost achieve the optimal performance; while for the case of $K = 3$, the DSCA algorithm performs slightly better than the DBSUM algorithm, though both algorithms achieve at least 85% of the optimal harmonic mean rate.

Example 2: In Fig. 2, we demonstrate the efficacy of the DBSUM algorithm for handling the MMF rate utility. Since the log-sum-exp approximation is used, we denote it by DBSUM-LSE in Fig. 2. We consider a 4-user MISO IFC under a medium interference level $\eta = 0.5$ [Fig. 2(a)] and a strong interference level $\eta = 1.0$ [Fig. 2(b)], respectively. The user priority weights are set to be $\alpha_1 = \cdots = \alpha_4 = \frac{1}{4}$, and $\gamma = 5$ is used in the log-sum-exp approximation (see (18)). Note that the DSCA algorithm is not able to handle the MMF rate function, so we instead compare DBSUM-LSE with the centralized SCA algorithm [29]. It is also worthwhile to note that, for the MMF formulation, the POA algorithm reduces to solving problem (45) only once, with $[v_1^\star, \ldots, v_K^\star]^T = [\alpha_1, \ldots, \alpha_K]^T$. From both Fig. 2(a) and Fig. 2(b), one can see that the SCA algorithm performs slightly better than the DBSUM-LSE algorithm at low SNR, whereas the two algorithms perform comparably at high SNR. By comparing with the POA algorithm, both DBSUM-LSE and SCA algorithms achieve at least 80% of the optimal MMF rate. It can also be observed that the achievable MMF rate saturates at high SNR due to the strict user fairness requirement; however, it can be improved as the number of transmit antennas increases.

Example 3: In this example, we consider the sum rate utility, and compare the performance and complexity of the DBSUM algorithm, the DWMMSE algorithm (Algorithm 2) and the DSCA algorithm. To demonstrate the scalability of the DBSUM algorithm and the DWMMSE algorithm, we consider scenarios for multiple users ($K = 2, 3, \ldots, 6$) and multiple transmit antennas ($N_t = 2, 4, 8$). The SNR and relative cross-link interference level are respectively fixed to $1/\sigma^2 = 10$ dB and $\eta = 0.5$.

In Fig. 3, it can be observed that the DBSUM algorithm and the DWMMSE algorithm yield nearly the same system throughput, which increases with the number of users and the number of transmit antennas. However, the performance of the DSCA algorithm drastically degrades when $K \geq 4$. The reason for this might be that the DSCA algorithm is relatively easier to get trapped in some local maximum when $K \geq 4$. In Fig. 3, the curve denoted by TDMA represents the achieved system throughput by time-division multiple access. One can see from this figure that allowing all the users to access the spectrum simultaneously attains higher spectral efficiency than TDMA even when only CDI is available at the transmitters. As also observed, the performance gain of the spectrum sharing policy over the TDMA policy increases with the number of users.

We illustrate in Fig. 4(a) and Fig. 4(b) the evolution of the normalized sum rate (the sum rate achieved after certain number of iterations normalized by the convergent sum rate) for the DBSUM algorithm and the DWMMSE algorithm, respectively. It is observed that the DBSUM algorithm and the DWMMSE algorithm converge in less than $KN_t^2/3$ and $KN_t^2/2$ iterations, respectively, implying that the distributed implementations indeed save the backhaul communication resources compared with exchanging full CDI between the transmitters as discussed
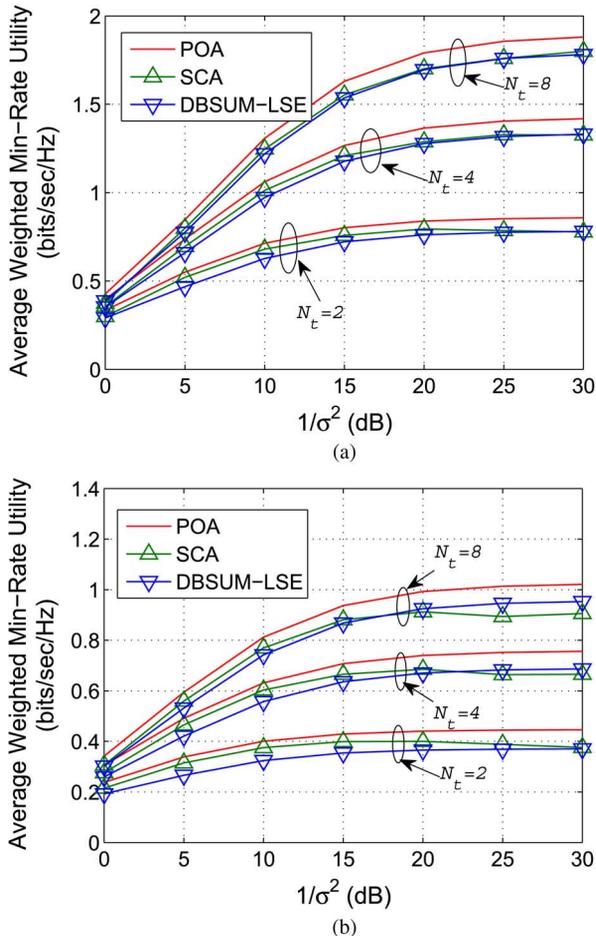
Fig. 2. Simulation results of average achievable weighted min-rate utility versus $1/\sigma^2$, for (a) $\eta = 0.5$, (b) $\eta = 1.0$, where $K = 4$, $\alpha_1 = \cdots = \alpha_4 = 1/4$, $N_t = 2, 4, 8$.
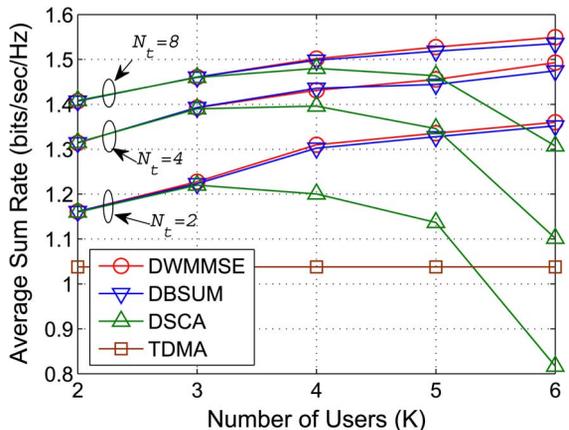

Fig. 3. Performance comparison for DWMMSE, DBSUM, and DSCA algorithms, where $1/\sigma^2 = 10$ dB, $\eta = 0.5$, $\alpha_1 = \cdots = \alpha_K = 1$, $\text{rank}(\mathbf{Q}_{ki}) = N_t$ for all k, i.

Fig. 4. Convergence curves for (a) DBSUM algorithm and (b) DWMMSE algorithm, where $1/\sigma^2 = 10$ dB, $\eta = 0.5$, $\alpha_1 = \cdots = \alpha_K = 1$, $\text{rank}(\mathbf{Q}_{ki}) = N_t$ for all k, i.

Fig. 5. Complexity comparison for DWMMSE, DBSUM, and DSCA algorithms, where $1/\sigma^2 = 10$ dB, $\eta = 0.5$, $\alpha_1 = \cdots = \alpha_K = 1$, $\text{rank}(\mathbf{Q}_{ki}) = N_t$ for all k, i.

in Remark 2. Furthermore, the DBSUM algorithm and the DWMMSE algorithm achieve more than 95% of the convergent sum rate within much less than $KN_t^2/3$ iterations and $KN_t^2/2$ iterations, respectively, which suggests that these two distributed algorithms can achieve a good tradeoff between the performance and the induced backhaul burden.

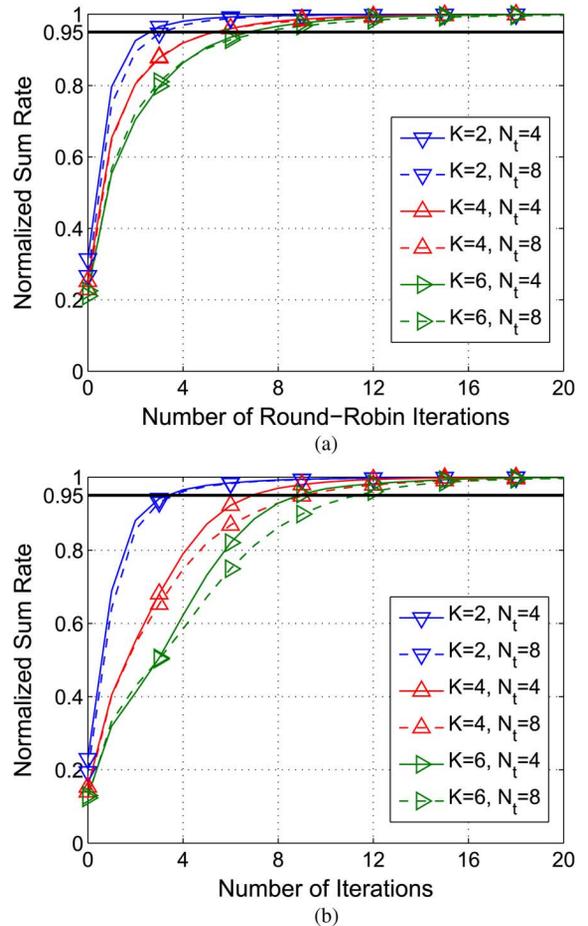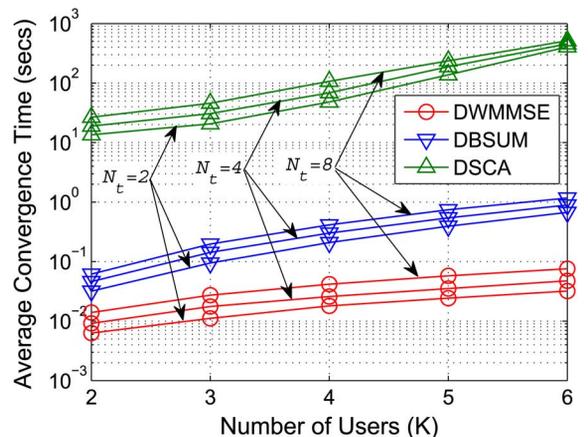In Fig. 5, we compare the computation load of the DBSUM algorithm, the DWMMSE algorithm and the DSCA algorithm

in terms of the average computation time per realization (in seconds). In our simulations, the convex subproblems involved in the DSCA algorithm (i.e., [29, Eqn. (36)]) and the DBSUM algorithm (i.e., (8)) are handled by CVX and the gradient projection method, respectively; while the subproblem (33) in the DWMMSE algorithm is solved by the Lagrange dual method [36] (see [13, Problem (14)] for the details). It can be observed that the average computation time of the DBSUM algorithm and

the DWMMSE algorithm increases at a slower rate than that of the DSCA algorithm w.r.t. the number of users, demonstrating that the DBSUM and DWMMSE algorithms have better scalability. Apart from that, we see from Fig. 5 that the DBSUM algorithm is $10^2 \sim 10^3$ faster than the DSCA algorithm, and the DWMMSE algorithm is about ten times faster than the DBSUM algorithm.[5] These results are consistent with the analysis about the complexity orders of the three algorithms given in Remark 1.

VII. CONCLUSIONS

We have presented two efficient distributed algorithms for handling the NP-hard rate outage constrained CoBF design problem in (3), namely, the DBSUM algorithm (Algorithm 1) and the DWMMSE algorithm (Algorithm 2). The former is a Gauss-Seidel type algorithm, which can handle problem (3) with general utility functions, while the latter is a Jocobi-type algorithm specifically designed for the weighted sum rate maximization. For the performance evaluation of the proposed two algorithms, we have also presented a POA algorithm (Algorithm 3) to obtain an upper bound to the optimal utility value of problem (3). The presented simulation results have shown that the proposed DBSUM and DWMMSE algorithms significantly outperform the existing DSCA algorithm in both efficacy and computational efficiency, and yield promising approximation performance as the performance gap below the benchmark POA algorithm is small (less than 20%).

APPENDIX A
PROOF OF LEMMA 1

For ease of exposition, let us define $I_{ki} \triangleq \boldsymbol{w}_i^H \mathbf{Q}_{ki} \boldsymbol{w}_i$ for $i, k = 1, \ldots, K$, and set

$$\tilde{\Phi}_i(\tilde{\xi}_i, \{I_{ki}\}_{k \neq i}) = \Phi(\tilde{\xi}_i, \{\boldsymbol{w}_k\}_{k \neq i}), \quad \text{(see (7))}$$
$$\tilde{\xi}_i(\{I_{ki}\}_{k \neq i}) = \xi_i(\{\boldsymbol{w}_k\}_{k \neq i}),$$
$$\tilde{R}_i(\{I_{ki}\}_k) = R_i(\{\boldsymbol{w}_k\}),$$

for all $i, k = 1, \ldots, K$. Hence, our goal is to show that $\tilde{R}_i(\{I_{kI}\}_k)$ is strictly increasing and strictly concave w.r.t. I_{ii} while is nonincreasing and convex w.r.t. $I_{ki}, k \neq i$, for each $i = 1, \ldots, K$.

Since $\tilde{\xi}_i(\{I_{ki}\}_{k \neq i}) = \xi_i(\{\boldsymbol{w}_k\}_{k \neq i}) > 0$ for any $\{\boldsymbol{w}_k\}_{k \neq i}$, it can be directly inferred from the strict monotoninicity and strict concavity of $\log_2(\cdot)$ that $\tilde{R}_i(\{I_{ki}\}_k)$ is strictly increasing and strictly concave w.r.t. I_{ii}. To prove the monotonicity and convexity of $\tilde{R}_i(\{I_{ki}\}_{k \neq i})$ w.r.t. $I_{ki}, k \neq i$, we need the following lemma:

Lemma 2: *For all $I_{\ell i} \geq 0$, $\ell \neq i$, $\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})$ is strictly decreasing while $I_{\ell i} \cdot \tilde{\xi}_i(\{I_{ki}\}_{k \neq i})$ is strictly increasing w.r.t. $I_{\ell i}$, for $i = 1, \ldots, K$.*

Proof: By definition, we know that

$$\tilde{\Phi}_i(\tilde{\xi}_i(\{I_{ki}\}_{k \neq i}), \{I_{ki}\}_{k \neq i}) =$$
$$\ln \rho_i + \sigma_i^2 \tilde{\xi}_i(\{I_{ki}\}_{k \neq i}) + \sum_{k \neq i} \ln(1 + I_{ki}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})) = 0,$$

for all $I_{ki} \geq 0$, $k \neq i$. Suppose that $I_{\ell i} < I'_{\ell i}$. Then,

$$0 = \ln \rho_i + \tilde{\xi}'_i \sigma_i^2 + \ln(1 + \tilde{\xi}'_i I'_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}'_i I_{ki})$$
$$= \ln \rho_i + \tilde{\xi}_i \sigma_i^2 + \ln(1 + \tilde{\xi}_i I_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}_i I_{ki})$$
$$< \ln \rho_i + \tilde{\xi}_i \sigma_i^2 + \ln(1 + \tilde{\xi}_i I'_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}_i I_{ki}),$$

where we denote $\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})$ and $\xi_i(\{I_{ki}\}_{k \neq i, k \neq \ell}, I'_{\ell i})$ by $\tilde{\xi}_i$ and $\tilde{\xi}'_i$ for notational simplicity. Since $\tilde{\Phi}_i(\tilde{\xi}_i, \{I_{ki}\}_{k \neq i})$ is a strictly increasing function of $\tilde{\xi}_i$, we have $\tilde{\xi}_i > \tilde{\xi}'_i$ by the first line and third line of the above inequality. Hence, $\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})$ is strictly decreasing w.r.t. $I_{\ell i}$ for all $\ell \neq i$. Furthermore, by the fact that $\tilde{\xi}_i > \tilde{\xi}'_i$, we can obtain

$$0 = \ln \rho_i + \tilde{\xi}_i \sigma_i^2 + \ln(1 + \tilde{\xi}_i I_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}_i I_{ki})$$
$$= \ln \rho_i + \tilde{\xi}'_i \sigma_i^2 + \ln(1 + \tilde{\xi}'_i I'_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}'_i I_{ki})$$
$$< \ln \rho_i + \tilde{\xi}'_i \sigma_i^2 + \ln(1 + \tilde{\xi}'_i I'_{\ell i}) + \sum_{k \neq i, k \neq \ell} \ln(1 + \tilde{\xi}_i I_{ki}),$$

which implies $I'_{\ell i}\tilde{\xi}'_i > I_{\ell i}\tilde{\xi}_i$ and completes the proof. ∎

By Lemma 2 and the monotonicity of the logarithmic function, it can be seen that $\tilde{R}_i(\{I_{ki}\}_{k \neq i})$ is nonincreasing w.r.t. I_{ki} for all $k \neq i$. We prove the convexity of $\tilde{R}_i(\{I_{ki}\}_k)$ w.r.t. $I_{ki}, k \neq i$ by showing that $\partial \tilde{R}_i(\{I_{ki}\}_k)/\partial I_{ki}$ is nondecreasing w.r.t. I_{ki} for all $k \neq i$, i.e., $\partial^2 \tilde{R}_i(\{I_{ki}\}_k)/\partial I_{ki}^2 \geq 0$ for all $k \neq i$. Let $\ell \in \{1, \ldots, i-1, i+1, \ldots, K\}$. By (12), we can explicitly express $\partial \tilde{R}_i(\{I_{ki}\}_k)/\partial I_{ki}$ as (A.1), shown at the bottom of the next page. By Lemma 2, we can see that $I_{ii}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})/(1 + I_{ii}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})) \geq 0$ is nonincreasing w.r.t. $I_{\ell i}$ while $(1 + I_{ji}\xi_i(\{I_{ki}\}_{k \neq i}))^{-1}$ is nondecreasing and $(1 + I_{\ell i}\xi_i(\{I_{ki}\}_{k \neq i}))$ is strictly increasing w.r.t. $I_{\ell i}$. Therefore, $\partial \tilde{R}_i(\{I_{ki}\}_{k \neq i})/\partial I_{\ell i}$ is nondecreasing w.r.t. $I_{\ell i}$, and hence $\tilde{R}_i(\{I_{ki}\}_{k \neq i})$ is convex w.r.t. $I_{ki}, \forall k \neq i$. ∎

APPENDIX B
MONOTONIC OPTIMIZATION BY POLYBLOCK OUTER APPROXIMATION ALGORITHM

Monotonic optimization refers to maximizing a nondecreasing function over an intersection of so called *normal sets* [32]. By definition, a nonnegative set $\mathcal{D} \subseteq \mathbb{R}_+$ is called normal if for any two points $\mathbf{d}_1 \succeq \mathbf{d}_2 \succeq \mathbf{0}$, $\mathbf{d}_1 \in \mathcal{D}$ implies $\mathbf{d}_2 \in \mathcal{D}$. Let $f : \mathbb{R}^N \to \mathbb{R}$ be a $\mathbb{R}_+^N$-nondecreasing function and $\mathcal{D} \subseteq \mathbb{R}_+^N$ be a compact normal set. Then, the monotonic optimization problem can be formulated as

$$\max_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x}). \tag{A.2}$$

According to [32], this class of problems can be optimally solved by a POA algorithm which is briefly reviewed in this section. Before presenting the POA algorithm, some essential definitions are given as follows.

Definition 1: *A set is called a polyblock if it is the union of a finite number of boxes, where a box associated with a vertex*

$\boldsymbol{b} \in \mathbb{R}_+^N$ is referred to the hyperrectangle $\mathcal{B}(\boldsymbol{b}) = \{\boldsymbol{x} \in \mathbb{R}_+^N | \boldsymbol{0} \preceq \boldsymbol{x} \preceq \boldsymbol{b}\}$.

Definition 2: *A vertex $\boldsymbol{b} \in \mathcal{P}$ is called a proper vertex of the polyblock $\mathcal{P}$ if there is no vertex $\boldsymbol{b}' \in \mathcal{P}$ such that $\boldsymbol{b}' \succeq \boldsymbol{b}$ and $\boldsymbol{b}' \neq \boldsymbol{b}$.*

The main effort of the POA algorithm lies in constructing a sequence of polyblocks $\{\mathcal{P}[0], \mathcal{P}[1], \ldots\}$ such that

$$\mathcal{P}[0] \supseteq \mathcal{P}[1] \supseteq \cdots \supseteq \mathcal{P}[n] \supseteq \cdots \supseteq \mathcal{D}, \qquad \text{(A.3a)}$$

$$\lim_{n \to \infty} \max_{\boldsymbol{x} \in \mathcal{P}[n]} f(\boldsymbol{x}) = \max_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x}). \qquad \text{(A.3b)}$$

In general, the initial polyblock can simply be a single box associated with a vertex $\boldsymbol{b}^\star[0]$, i.e., $\mathcal{P}[0] = \mathcal{B}(\boldsymbol{b}^\star[0])$, such that $\mathcal{D} \subseteq \mathcal{B}(\boldsymbol{b}^\star[0])$. Given the polyblock $\mathcal{P}[n-1]$ at the $(n-1)$th iteration, the polyblock $\mathcal{P}[n]$ for iteration n can be constructed as follows. Let $\mathcal{V}[n-1]$ denote the set of proper vertices of $\mathcal{P}[n-1]$. Firstly, we find a point $\boldsymbol{b}^\star[n-1] \in \mathcal{V}[n-1]$ that maximizes $f(\boldsymbol{x})$ over $\mathcal{V}[n-1]$, and hence maximizes $f(\boldsymbol{x})$ over $\mathcal{P}[n-1]$ according to the monotonicity of $f(\boldsymbol{x})$. Specifically, we find $\boldsymbol{b}^\star[n-1]$ such that

$$\boldsymbol{b}^\star[n-1] \in \arg\max_{\boldsymbol{b} \in \mathcal{V}[n-1]} f(\boldsymbol{b}) \subseteq \arg\max_{\boldsymbol{x} \in \mathcal{P}[n-1]} f(\boldsymbol{x}). \quad \text{(A.4)}$$

Problem (A.4) can be solved by enumerating all the vertexes in $\mathcal{V}[n-1]$. Secondly, we search for the intersection of the right-upper boundary of $\mathcal{D}$ and the ray from the origin to $\boldsymbol{b}^\star[n-1]$, i.e.,

$$\tilde{\boldsymbol{b}}[n-1] = \beta^\star[n-1]\boldsymbol{b}^\star[n-1], \beta^\star[n-1] = \arg\max_{\beta\boldsymbol{b}^\star[n-1] \in \mathcal{D}} \beta. \tag{A.5}$$

Problem (A.5) can be solved by bisecting over β, which entails checking the feasibility of $\beta\boldsymbol{b}^\star[n-1] \in \mathcal{D}$ iteratively. Thirdly, using $\boldsymbol{b}^\star[n-1]$ and $\tilde{\boldsymbol{b}}[n-1]$, we generate N new vertices by

$$\boldsymbol{b}[n,i] = \boldsymbol{b}^\star[n-1] - (v_i^\star[n-1] - \tilde{v}_i[n-1])\boldsymbol{b}_i, i = 1, \ldots, N, \tag{A.6}$$

where $v_i^\star[n-1]$ and $\tilde{v}_i[n-1]$ are the ith elements of $\boldsymbol{b}^\star[n-1]$ and $\tilde{\boldsymbol{b}}[n-1]$, respectively, and $\boldsymbol{e}_i$ is the ith column of the $N \times N$ identity matrix. Then, a new vertex set $\mathcal{V}[n]$ is obtained as

$$\mathcal{V}[n] = \mathcal{V}[n-1] \bigcup \{\boldsymbol{b}[n,1], \ldots, \boldsymbol{b}[n,N]\} \backslash \{\boldsymbol{b}^\star[n-1]\}, \quad \text{(A.7)}$$

Algorithm 4: POA algorithm for solving problem (A.2)

1: **Initialization**: Set the solution accuracy as $\delta \geq 0$, and set $n := 0$.
2: Set $\mathcal{V}[0] := \{\boldsymbol{b}^\star[0]\}$, where $\boldsymbol{b}^\star[0]$ can be any vector such that $\mathcal{D} \subseteq \mathcal{B}(\boldsymbol{b}^\star[0])$;
3: Compute $\tilde{\boldsymbol{b}}[0]$ by (A.5), and set $\bar{\boldsymbol{b}}[0] := \tilde{\boldsymbol{b}}[0]$;
4: **while** $f(\boldsymbol{b}^\star[n]) - f(\bar{\boldsymbol{b}}[n]) > \delta$
5: $n := n + 1$;
6: Set $\mathcal{V}[n] = \mathcal{V}[n-1] \bigcup \{\boldsymbol{b}[n,1], \ldots, \boldsymbol{b}[n,N]\} \backslash \{\boldsymbol{b}^\star[n-1]\}$, where $\boldsymbol{b}[n,i]$, $i = 1, \ldots, N$, are given by (A.6);
7: Compute $\boldsymbol{b}^\star[n]$ and $\tilde{\boldsymbol{b}}[n]$ by (A.4) and (A.5), respectively;
8: Set $\bar{\boldsymbol{b}}[n] := \arg\max_{\boldsymbol{b} \in \{\bar{\boldsymbol{b}}[n-1], \tilde{\boldsymbol{b}}[n]\}} f(\boldsymbol{b})$;
9: **end while**
10: **Output** $f(\boldsymbol{b}^\star[n])$ as the approximation of the optimal value of (A.2).

---

which leads to a new polyblock for the nth iteration

$$\mathcal{P}[n] = \bigcup_{\boldsymbol{b} \in \mathcal{V}[n]} \mathcal{B}(\boldsymbol{b}). \tag{A.8}$$

Notice that $\mathcal{P}[n] \subseteq \mathcal{P}[n-1]$ since $\boldsymbol{b}^\star[n-1] \succeq \boldsymbol{b}[n,i]$ for all $i = 1, \ldots, N$. Besides, by (A.5) and by the fact that $\mathcal{D}$ is normal, one can infer that the intersection of $\mathcal{D}$ and $\mathcal{P}[n-1] \backslash \mathcal{P}[n]$ must be empty,[6] implying that $\mathcal{D} \in \mathcal{P}[n]$. As a result, the polyblocks $\{\mathcal{P}[0], \mathcal{P}[1], \ldots, \mathcal{P}[n], \ldots\}$ generated in this manner indeed satisfy (A.3a). In addition, it has been shown in [32, Theorem 1] that (A.3b) also holds true. Thus, by (A.4), the sequence $\{f(\boldsymbol{b}^\star[n])\}_{n=0}^\infty$ monotonically converges to the optimal value of problem (A.2) from above. On the other hand, let $\bar{\boldsymbol{b}}[n] = \arg\max_{\boldsymbol{b} \in \{\bar{\boldsymbol{b}}[n-1], \tilde{\boldsymbol{b}}[n]\}} f(\boldsymbol{b})$, where $\tilde{\boldsymbol{b}}[n] \in \mathcal{D}$ for all $n \geq 0$. Then the sequence $\{f(\bar{\boldsymbol{b}}[n])\}_{n=0}^\infty$ will also monotonically converge to the optimal value of problem (A.2) from below [32, Theorem 1]. Therefore, the gap between $f(\boldsymbol{b}^\star[n])$ and $f(\bar{\boldsymbol{b}}[n])$ can be used as an estimate of the difference between $f(\bar{\boldsymbol{b}}[n])$ and the optimal

---

[6]A brief proof is as follows. From (A.7), we see that $\mathcal{P}[n-1] \backslash \mathcal{P}[n] = \{\boldsymbol{x} | \tilde{\boldsymbol{b}}[n-1] \prec \boldsymbol{x} \preceq \boldsymbol{b}^\star[n-1]\}$. If the intersection of $\mathcal{D}$ and $\mathcal{P}[n-1] \backslash \mathcal{P}[n]$ is not empty, then there must exist a point $\bar{\boldsymbol{x}}$ such that $\bar{\boldsymbol{x}} \in \mathcal{D}$ and $\tilde{\boldsymbol{b}}[n-1] \prec \bar{\boldsymbol{x}} \preceq \boldsymbol{b}^\star[n-1]$. This implies that there exists $\beta \in (\beta^\star[n-1], 1]$ such that $\tilde{\boldsymbol{x}} = \beta\boldsymbol{b}^\star[n-1] \preceq \bar{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}} \in \mathcal{D}$ (since $\mathcal{D}$ is normal), which however contradicts with the optimality of $\beta^\star[n-1]$ to problem (A.5).

$$\frac{\partial \tilde{R}_i(\{I_{ki}\}_k)}{\partial I_{\ell i}}$$

$$= \frac{-I_{ii}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})}{\ln 2 \cdot (1 + \tilde{\xi}_i(\{I_{ki}\}_{k \neq i})I_{ii})} \left[(1 + I_{\ell i}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})) \cdot \left(\sigma_i^2 + \sum_{j \neq i} \frac{I_{ji}}{1 + I_{ji}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})} \right) \right]^{-1}$$

$$= \frac{-I_{ii}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})}{\ln 2 \cdot (1 + I_{ii}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i}))} \left[(1 + I_{\ell i}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})) \cdot \left(\sigma_i^2 + \sum_{j \neq i, j \neq \ell} \frac{I_{ji}}{1 + I_{ji}\tilde{\xi}_i(\{I_{ki}\}_{k \neq i})} \right) + I_{\ell i} \right]^{-1}. \tag{A.1}$$

value of (A.2), serving as a stopping criterion for the POA algorithm. Finally, the POA algorithm for problem (A.2) is summarized in Algorithm 4.

ACKNOWLEDGMENT

REFERENCES

[1] W.-C. Li, T.-H. Chang, C. Lin, and C.-Y. Chi, "Outage constrained weighted sum rate maximization for MISO interference channel by pricing-based optimization," in *Proc. IEEE ICASSP*, Vancouver, BC, May 26–31, 2013, pp. 4799–4803.

[2] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 44–50, Nov. 2012.

[3] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

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

[5] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of MIMO interference channels in the low interference regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2565–2581, May 2011.

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

[7] D. Cai, T. Quek, C. W. Tan, and S. Low, "Max-min SINR coordinated multipoint downlink transmission-duality and algorithms," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5384–5395, Oct. 2012.

[8] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 728–738, 2011.

[9] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming design approaching Pareto boundary with max-min fairness," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2921–2933, 2012.

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

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

[12] S.-J. Kim and G. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[13] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[14] D. H. N. Nguyen and T. Le-Ngoc, "Multiuser downlink beamforming in multicell wireless systems: A game theoretical approach," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3326–3338, Jul. 2011.

[15] P. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Multicell MISO downlink weighted sum-rate maximization: A distributed approach," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 556–570, Feb. 2013.

[16] M.-Y. Hong and Z.-Q. Luo, "Signal Processing and Optimal Resource Allocation for the Interference Channel," Academic Press Library in Signal Process., 2013 [Online]. Available: http://arxiv.org/abs/1206.5144

[17] E. G. Larsson and E. A. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[18] E. G. Larsson, E. A. Jorswieck, J. Lindblom, and R. Mochaourab, "Game theory and the flat-fading Gaussian interference channel," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 18–27, Sep. 2009.

[19] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Distributed resource allocation schemes: Pricing algorithms for power control and beamformer design in interference networks," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 53–63, Sep. 2009.

[20] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2159–2168, Jul. 2010.

[21] W. Utschick and J. Brehmer, "Monotonic optimization framework for coordinated beamforming in multicell networks," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1899–1909, Apr. 2012.

[22] L. Liu, R. Zhang, and K.-C. Chua, "Achieving global optimality for weighted sum-rate maximization in the K-user Gaussian interference channel with multiple antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1933–1945, May 2012.

[23] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[24] C. W. Tan, "Optimal power control in Rayleigh-fading heterogeneous networks," in *Proc. IEEE INFOCOM*, Shanghai, Apr. 10–15, 2011, pp. 2552–2560.

[25] Y. Huang, C. W. Tan, and B. Rao, "Outage balancing in multiuser MISO networks: Network duality and algorithms," in *Proc. IEEE GLOBECOM*, Anaheim, CA, USA, Dec. 3–7, 2012, pp. 3918–3923.

[26] S. Ghosh, B. D. Rao, and J. R. Zeidler, "Outage-efficient strategies for multiuser MIMO networks with channel distribution information," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6312–6324, Dec. 2010.

[27] J. Lindblom, E. Karipidis, and E. G. Larsson, "Achievable outage rate regions for the MISO interference channel," *IEEE Wireless Commun. Lett.*, vol. 2, no. 4, pp. 439–442, Aug. 2013.

[28] J. Park, Y. Sung, D. Kim, and H. V. Poor, "Outage probability and outage-based robust beamforming for MIMO interference channels with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3561–3573, Oct. 2012.

[29] W.-C. Li, T.-H. Chang, C. Lin, and C.-Y. Chi, "Coordinated beamforming for multiuser MISO interference channel under rate outage constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1087–1103, Mar. 2013.

[30] W.-C. Li, T.-H. Chang, and C.-Y. Chi, "Multicell coordinated beamforming with rate outage constraint—Part I: Complexity analysis," *IEEE Trans. Signal Process.* vol. 63, no. 11, pp. 2749–2762, 2015.

[31] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimiz.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[32] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM J. Optimiz.*, vol. 11, no. 2, pp. 464–494, 2000.

[33] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, pp. 556–567, Oct. 2000.

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

[35] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem: History, Theory, Applications*. Boston, MA: Birkhäuser, 2002.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

[37] J. C. Dunn, "Global and asymptotic convergence rate estimates for a class of projected gradient processes," *SIAM J. Contr. Optimiz.*, vol. 19, no. 3, pp. 368–400, 1981.

[38] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operat. Res.*, vol. 26, pp. 681–683, 1978.

[39] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, pp. 2508–2523, May 2012.

[40] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20–34, May 2010.

[41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," [Online]. Available: http://cvxr.com/cvx Apr. 2011

Wei-Chiang Li received the B.S. degree in electrical engineering from the National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2009.

Currently, he is pursuing the Ph.D. degree in communications engineering at the National Tsing Hua University, Hsinchu. His research interests are in wireless communications and signal processing.
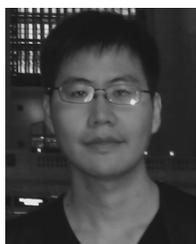
Tsung-Hui Chang (S'07–M'08) received the B.S. degree in electrical engineering and the Ph.D. degree in communications engineering from the National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2003 and 2008, respectively.

Since September 2012, he has been with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan, as an Assistant Professor. Before joining NTUST, he held research positions with NTHU (2008–2011), and the University of California at Davis, USA (2011–2012). He was also a Visiting Scholar with the University of Minnesota, Twin Cities, USA, the Chinese University of Hong Kong and Xidian University, Xian, China. His research interests are widely in signal processing and optimization problems in data communications, smart grid, and machine learning.

Dr. Chang received the 2014 Young Researcher Award of NTUST. He currently serves as an Associate Editor of the IEEE Transactions on Signal Processing and IEEE Transactions on Signal and Information Processing Over Networks.

Chong-Yung Chi (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, USA, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA. He has been a Professor with the Department of Electrical Engineering since 1989 and the Institute of Communications Engineering (ICE) since 1999, and National Tsing Hua University, Hsinchu, Taiwan. He has published more than 200 technical papers, including more than 75 journal papers, four book chapters, and more than 130 peer-reviewed conference papers, as well as a graduate-level textbook, *Blind Equalization and System Identification* (New York: Springer-Verlag, 2006). His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

Dr. Chi is been a Technical Program Committee member for many IEEE sponsored and cosponsored workshops, symposiums, and conferences on signal processing and wireless communications, including Co-Organizer and General Co-Chairman of the 2001 IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), and Co-Chair of Signal Processing for Communications (SPC) Symposium, ChinaCOM 2008, and Lead Co-Chair of SPC Symposium (ChinaCOM 2009). He was also the Chairman of ICE during 2002–2005. He was an Associate Editor (AE) of the IEEE Transactions on Signal Processing (May 2001–April 2006), IEEE Transactions on Circuits and Systems II (January 2006–December 2007), IEEE Transactions on Circuits and Systems I (January 2008–December 2009), AE of the IEEE Signal Processing Letters (June 2006–May 2010), and a member of the Editorial Board of *Elsevier Signal Processing* (June 2005–May 2008), and an editor (July 2003–December 2005) as well as a Guest Editor (2006) of EURASIP *Journal on Applied Signal Processing*. He was a member of the Signal Processing Theory and Methods Technical Committee (SPTM-TC) (2005–2010), IEEE Signal Processing Society. Currently, he is a member of the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC) and a member of the Sensor Array and Multichannel Technical Committee (SAM-TC), IEEE Signal Processing Society.