Coordinated Beamforming for MISO Interference Channel: Complexity Analysis and Efficient Algorithms*

Ya-Feng Liu†, Yu-Hong Dai†, and Zhi-Quan Luo‡

Abstract

In a cellular wireless system, users located at cell edges often suffer significant out-of-cell interference. Assuming each base station is equipped with multiple antennas, we can model this scenario as a multiple-input single-output (MISO) interference channel. In this paper we consider a coordinated beamforming approach whereby multiple base stations jointly optimize their downlink beamforming vectors in order to simultaneously improve the data rates of a given group of cell edge users. Assuming perfect channel knowledge, we formulate this problem as the maximization of a system utility (which balances user fairness and average user rates), subject to individual power constraints at each base station. We show that, for the single carrier case and when the number of antennas at each base station is at least two, the optimal coordinated beamforming problem is NP-hard for both the harmonic mean utility and the proportional fairness utility. For general utilities, we propose a cyclic coordinate descent algorithm, which enables each transmitter to update its beamformer locally with limited information exchange, and establish its global convergence to a stationary point. We illustrate its effectiveness in computer simulations by using the space matched beamformer as the benchmark.

Index Terms

* Ya-Feng Liu and Yu-Hong Dai are supported by the National Natural Science Foundation of China, Grant number 10831006, and the CAS Grant kjcx-yw-s7-03. Zhi-Quan Luo is supported by the Army Research Office, Grant number W911NF-09-1-0279, and by the National Science Foundation, Grant number CMMI-0726336.

† State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing, 100190, China. (e-mail: {yafliu,dyh}@lsec.cc.ac.cn)

‡ Department of Electrical Computer Science Engineering, University of Minnesota, Minneapolis, 55455, USA. (e-mail: luozq@umn.edu)

MISO interference channel, coordinated beamforming, complexity, cyclic coordinate descent algorithm, global convergence.

I. INTRODUCTION

In a conventional wireless cellular system, base stations from different cells communicate with their respective remote terminals independently. Signal processing is performed on an intra-cell basis, while the out-of-cell interference is treated as background noise. This architecture often causes undesirable service outages to users situated near cell edges where the out-of-cell interference can be severe. Since the conventional intra-cell signal processing can not effectively mitigate the impact of inter-cell interference, we are led to consider coordinated base station beamforming across multiple cells in order to improve the services to edge users. In this paper, we focus on the downlink scenario where the base stations are equipped with multiple antennas and model it as a MISO interference channel. We consider joint optimal beamforming across multiple base stations to simultaneously improve the data rates of a given group of cell edge users. Assuming that the channel state information (CSI) is known, we formulate this problem as the maximization of a system utility (which balances user fairness and average user rates), subject to individual power constraints at each base station. We show that, for the single carrier case and when the number of antennas at each base station is at least two, the optimal coordinated beamforming problem is NP-hard for both the harmonic mean utility and the proportional fairness utility. This NP-hardness result is in contrast to the single antenna case for which the same optimization problem is convex for both the harmonic mean and proportional fairness utility functions [1]. For the min-rate utility, this problem is known to also be solvable in polynomial time [1], [2].

In addition to the complexity analysis, we propose a practical iterative cyclic coordinate descent algorithm for the multi-cell coordinated beamforming problem by exploiting the separability of power constraints. We prove the global convergence of this cyclic coordinate descent algorithm (to a stationary point). Numerical experiments are also presented to illustrate the effectiveness of the proposed algorithm.

A. Related Work

Downlink beamforming has been studied extensively in the single cell setup [3], [4]. For the multi-cell interference channel, the reference [5] considered coordinated beamforming for the minimization of total weighted transmit power across the base stations subject to individual signal-to-interference-plus-noise-ratio (SINR) constraints at the remote users. It turns out this problem can be transformed into a convex second order conic programming (SOCP) and efficiently solved. However, the maximization of weighted sum rates for a multi-cell interference channel under individual power constraints is NP-hard even for the single antenna and the single carrier case [1]. In fact, more is known about the single antenna interference channel case. For instance, if the system utility is changed into either the geometric mean rate (i.e., proportional fairness), the harmonic mean rate, or the min-rate, the corresponding utility maximization problem (for the single tone case) can be converted to a convex optimization problem and solved efficiently with global optimality [1], [6]. However, when the number of tones is more than two, all of the aforementioned power control problems are NP-hard. The focus of this paper is to study the multi-antenna case (MISO interference channel), analyze the complexity of the corresponding utility maximization problems, and propose a practical algorithm to solve them.

In addition to the aforementioned utility based formulations, various base station cooperation techniques have been proposed to mitigate inter-cell interferences, including multi-point coordinated transmission, or network multi-input multi-output (MIMO) transmission [7]–[16]. For example, distributed or decentralized approaches are proposed for coordinated transmitter beamforming design in MISO interference channel in [7], [10], [13], [14], [16], [17], some of which are based on dual uplink channels.

The references [11], [12] show that coordination enables the cellular network to enjoy a greater spectral efficiency. Most of these cooperative techniques require each base station to have not only full/partial CSI but also the knowledge of actual independent data streams to all remote terminals. With the complete sharing of data streams and CSI, the multi-cell scenario is effectively reduced to a single cell interference management problem with either total [18] or per-group-of-antenna power constraints [19], [20]. Among the major drawbacks of these techniques (in comparison to the utility based approaches) are their stringent requirement on base station coordination, the large demand on the communication bandwidth of backhaul links, as well as the heavy computational load associated with the increasing number of cells [21], [22]. The references [8], [15], [23], [24] provided characterizations of the achievable rate region and proved the existence of a unique Nash equilibrium which is inefficient in the sense that the achievable rates are bounded by a constant, regardless of the available transmit power. See [23], [25] for more recent results of the MISO channel.

Notation: We adopt the following notations in this paper. We use lower case boldface to denote column vectors. For any vector $\mathbf{h}$, we denote its transpose and Hermitian transpose by $\mathbf{h}^T$ and $\mathbf{h}^\dagger$, respectively. Also, we use (x, y) to represent a two-dimensional row vector and $\|\cdot\|$ to represent the Euclidean norm. Finally, for a multi-variable function $f(\mathbf{x})$, we let $\nabla f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x})$ denote its gradient and Hessian.

II. Problem Formulation

Consider a cellular system in which there are K base stations each equipped with L transmit antennas. The K base stations wish to transmit respectively to K mobile receivers each having only a single antenna. Each base station can direct a beam to its intended receiver in such a way that the resulting interference to the other mobile units is small. Consider the single carrier case, and let $\mathbf{h}_{jk} \in \mathbb{C}^L$ denote L-dimensional complex channel vector between base station j and receiver k. Let $\mathbf{v}_k \in \mathbb{C}^L$ denote the beamforming vector used by base station k, while s_k is a complex scalar denoting the information signal for user k with $\mathbb{E}|s_k|^2 = 1$. The transmit vector of the j-th base station is $\mathbf{v}_j s_j$. Then the signal received by user k can be described as

$$y_k = \sum_{j=1}^{K} \mathbf{h}_{jk}^{\dagger} \mathbf{v}_j s_j + z_k,\ 1 \le k \le K, \tag{1}$$

where z_k is the additive white Gaussian noise (AWGN) with variance $\sigma_k^2/2$ per real dimension. Treating interference as noise, we can write the SINR of each user as

$$\mathrm{SINR}_k = \frac{|\mathbf{h}_{kk}^{\dagger} \mathbf{v}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_{jk}^{\dagger} \mathbf{v}_j|^2},\ 1 \le k \le K. \tag{2}$$

Adopting a utility, we can formulate the optimal coordinated downlink beamforming problem as

$$\begin{aligned} \max \quad & H(r_1, r_2, ..., r_K) \\ \text{s.t.} \quad & r_k = \log\left(1 + \frac{|\mathbf{h}_{kk}^{\dagger} \mathbf{v}_k|^2}{\sigma_k^2 + \sum_{j \neq k} |\mathbf{h}_{jk}^{\dagger} \mathbf{v}_j|^2}\right), \\ & \|\mathbf{v}_k\|^2 \le P_k,\ 1 \le k \le K, \end{aligned} \tag{3}$$

where P_k denotes the power budget of base station k and $H(\cdot)$ denotes the system utility which may be any of the following

- Weighted sum-rate utility: $H_1 = \frac{1}{K}\sum_{k=1}^{K} w_k r_k$, with weight $w_k \ge 0$.
- Proportional fairness utility: $H_2 = \left(\prod_{k=1}^{K} r_k\right)^{1/K} \Leftrightarrow \frac{1}{K}\sum_{k=1}^{K} \log r_k$.
- Harmonic mean utility: $H_3 = K/\left(\sum_{k=1}^{K} r_k^{-1}\right)$.
- Min-rate utility: $H_4 = \min_{1 \le k \le K} r_k$.

According to [23], [24], problem (3) can be written in a more general form as

$$\begin{aligned} \max \quad & H(r_1, r_2, ..., r_K) \\ \text{s.t.} \quad & r_k = \log\left(1 + \frac{\mathbf{h}_{kk}^\dagger \mathbf{V}_k \mathbf{h}_{kk}}{\sigma_k^2 + \sum_{j\neq k} \mathbf{h}_{jk}^\dagger \mathbf{V}_j \mathbf{h}_{jk}}\right), \\ & \text{Trace}(\mathbf{V}_k) \leq P_k,\ \mathbf{V}_k \succcurlyeq 0,\ 1 \leq k \leq K, \end{aligned} \tag{4}$$

where $\mathbf{V}_k$ is the transmit covariance matrix at transmitter k. The results in [23], [24] state that problem (4) has a rank-one optimal solution for each $\mathbf{V}_k$. This implies that problem (3) and problem (4) are equivalent. We focus on formulation (3) in this paper.

The above beamforming problem (3) can be nonconvex in general due to the nonlinear equality constraints. The tuple of optimal transmit rates $(r_1, r_2, ..., r_K)$ corresponding to problem (3) (with any choice of the four mentioned system utilities) lie on the boundary of the achievable rate region. See [8], [9], [15], [23], [24] for various efforts to characterize the achievable rate region of the interference channel.

In practice, the choice of utilities depends on a suitable compromise between system performance (total rates achievable) and user fairness. The sum-rate utility H_1 focuses entirely on system performance, while the min-rate utility H_4 places the highest emphasis on user fairness. The other two choices H_2 and H_3 represent an appropriate tradeoff between the two extremes.

III. Complexity Analysis

In this section, we investigate the complexity status of the optimal coordinated downlink beamforming problem (3) for various choices of system utilities. We provide a complete analysis on when the problem is NP-hard and also identify subclasses of the (general NP-hard) problem that are solvable in polynomial time.

Generally speaking, convex optimization problems are relatively easy to solve, provided that there is a fast way to evaluate the objective function and its subgradient and to determine the feasibility of a candidate solution. More precisely, for any convex optimization problem and any $\epsilon > 0$, the so-called ellipsoid algorithm [26] can be used to find an ϵ-optimal solution (i.e., a feasible solution whose objective value is within ϵ from being globally optimal) with a complexity that is polynomial in the problem dimension and $\log(1/\epsilon)$. In contrast, nonconvex optimization problems are generally difficult to solve as they require exponential effort. However, not all nonconvex problems are hard since the lack of convexity may be due to inappropriate formulation. In fact, there are many nonconvex optimization problems which admit a convex reformulation. Thus, convexity is a useful but unreliable test of the computational tractability of an optimization problem. A more robust tool is the computational complexity theory which we briefly describe below.

A. Computational Complexity Theory: A Brief Background

Generically, an optimization problem can be described by the minimization of an objective function over a feasible region. A decision version of the minimization problem is to decide if the feasible region contains a vector at which the objective function value is below a given threshold. The answer to the decision problem is binary, true or false, and there is no need to identify what the solution is. The decision version is typically easier to solve than the original optimization problem which requires the determination of an (globally) optimal solution. The size of an optimization problem instance is defined as the length of a binary string required to describe the objective function and the feasible region [26]. We say an algorithm solves the decision version of an optimization problem if for each instance of the problem, the algorithm correctly gives "true" or "false" answer. We can define the running time of an algorithm as the maximum number of basic computational steps (e.g., number of elementary bit operations) required to solve the decision version of an optimization problem of a given size. Typically, the algorithm's running time is a function of the problem size.

In the computational complexity theory [27], [28], there are two important classes of optimization problems, P and NP. The class P contains optimization problems which are solvable (or decidable) by an algorithm whose running time grows at most as a polynomial function of the input size. The class NP, which stands for *Nondeterministic Polynomial* time, consists of decision version of optimization problems whose "true" instances can be verified in polynomial time, assuming the availability of a feasible solution that meets the threshold requirement. More formally, we say a nondeterministic algorithm solves a decision version of an optimization problem if we can verify each "true" instance of the problem using a sequence of nondeterministic steps (i.e., involving random guesses). If the number of nondeterministic steps is polynomial, then the algorithm is said to have a nondeterministic polynomial running time. For example, for any $n \times n$ symmetric matrix $\mathbf{Q}$ with integer entries and an integer threshold value L, consider the problem of deciding if there exists a binary vector $\mathbf{x} \in \{-1, 1\}^n$ such that $\mathbf{x}^T\mathbf{Q}\mathbf{x} \leq L$. A nondeterministic algorithm to solve this problem is to guess a binary vector $\mathbf{x}$ and then check if $\mathbf{x}^T\mathbf{Q}\mathbf{x} \leq L$ indeed holds. Such a binary vector $\mathbf{x}$ exists for all "true" instances of the problem, and the verification process requires polynomially many steps although some steps may involve a random guess (of a component of $\mathbf{x}$). In this case, the problem is solvable in nondeterministic polynomial time.

The class NP contains precisely those decision version of optimization problems that are solvable in nondeterministic polynomial time. Clearly, P is contained in NP. It is widely conjectured that P $\neq$ NP, or equivalently, there are problems in NP which are not solvable in (deterministic) polynomial time.

NP-complete problems are the most difficult problems in NP, in the sense that if any one of them is in P, so is every other problem in NP. There are many well known NP-complete problems such as the traveling salesman problem and the 3-colorability problem. The latter problem is to decide if the nodes of a given graph can be colored in three colors so that each adjacent pair of nodes are colored differently. The 3-colorability problem is clearly in NP since we can easily check if a guessed coloring scheme meets the requirement. There is no known polynomial time algorithm to solve the 3-colorability problem. In fact, if this problem is solvable in polynomial time (i.e., in P), then every problem in NP is solvable in polynomial time, or equivalently P = NP. A problem $\mathcal{P}$ is said to be NP-hard if it is at least as hard as those NP-complete problems, which means that the polynomial time solvability of $\mathcal{P}$ would imply every NP-complete problem is in P. A NP-hard problem may not be in NP. For example, the binary quadratic minimization problem $\min_{\mathbf{x}\in\{-1,1\}^n} \mathbf{x}^T\mathbf{Q}\mathbf{x}$ is NP-hard, since it is not known to be in NP, and is at least as hard as the NP-complete problem of deciding if there exists a binary vector $\mathbf{x}$ such that $\mathbf{x}^T\mathbf{Q}\mathbf{x} \leq L$, where the threshold value L is given.

To prove a problem $\mathcal{P}$ is NP-complete, we need to show two things. First, we verify the problem is in NP. This step is usually easy. Second, we need to show $\mathcal{P}$ is at least as difficult as a known NP-complete problem. This can be accomplished by a standard technique called polynomial time transformation. In a polynomial time transformation, we pick a known NP-complete problem and show that it is equivalent to a *special* case of $\mathcal{P}$. More precisely, we take an arbitrary instance of a known NP-complete problem, construct a special instance (with polynomial size) of $\mathcal{P}$, and then establish the equivalence of the two instances. To show a problem is NP-hard, we simply ignore the first step, as there is no need to show $\mathcal{P}$ is in NP. If a known NP-complete problem is polynomially transformed to a special case of problem $\mathcal{P}$, then a polynomial time algorithm for $\mathcal{P}$ would also solve the NP-complete problem, which is not possible unless P = NP.

B. Maximization of the Weighted Sum-Rate Utility

Consider the system utility $H_1 = \frac{1}{K}\sum_{k=1}^{K} w_k r_k$. In the single antenna case ($L = 1$), the original system optimization problem (3) becomes

$$\begin{aligned} \max \quad & \frac{1}{K}\sum_{k=1}^{K} w_k r_k \\ \text{s.t.} \quad & r_k = \log\left(1 + \frac{x_k}{\gamma_k + \sum_{j\neq k}\alpha_{jk}x_j}\right), \\ & 0 \leq x_k \leq P_k,\ 1 \leq k \leq K, \end{aligned} \tag{5}$$

where $x_k = \|\mathbf{v}_k\|^2,\ \alpha_{jk} = \|\mathbf{h}_{jk}\|^2/\|\mathbf{h}_{kk}\|^2$ and $\gamma_k = \sigma_k^2/\|\mathbf{h}_{kk}\|^2$. Problem (5) is known to be NP-hard [1] even when the weights w_k are all equal, and the proof is based on a polynomial time transformation from the maximum independent set problem [27] (which is known to be NP-complete). Thus, the general case of $L \geq 1$ is also NP-hard. However, some subclasses of problem (5) can still be polynomial time solvable. For instances, distributed polynomial time algorithms have been proposed in [25], [29]–[32] to solve the sum-rate maximization problem for various *special* channels.

C. Maximization of the Harmonic Mean Utility

We now study the complexity status of problem (3) defined by the harmonic mean utility.

Theorem 3.1 (Harmonic Mean Utility): For the harmonic mean utility $H_3 = K/\left(\sum_{k=1}^{K} r_k^{-1}\right)$, the optimal coordinated downlink beamforming problem can be transformed into a convex optimization problem when $L = 1$, but is NP-hard when $L \geq 2$.

When there is only a single transmit antenna ($L = 1$), reference [1] shows that the harmonic mean utility maximization can be transformed into an equivalent convex problem. We thus focus on the case $L \geq 2$. Notice that the harmonic mean utility maximization problem is a continuous optimization problem. To show its NP-hardness, we need to transform a known NP-complete discrete problem to it. To facilitate this transformation, it is necessary to induce certain discrete structure to its solutions. This is accomplished by using the concavity of the harmonic mean utility with respect to each beamforming vector $\mathbf{v}_k$. In particular, Lemma 3.2 shows that we can constrain the optimal beamforming vectors to be taken from two orthogonal vectors $\mathbf{h}_a$ or $\mathbf{h}_b$.

The NP-hardness proof of Theorem 3.1 is based on a transformation from a variant of the 3-SAT [28] problem. To describe this variant, we need to define the UNANIMITY property and the NAE (stands for "not-all-equal") property of a disjunctive clause[1].

Definition 3.1: For a given truth assignment to a set of Boolean variables, a disjunctive clause is said to be *UNANIMOUS* if all literals in the clause have the same value (whether it is the *True* or the *False* value). Otherwise it is said to be satisfied in the *NAE* (Not-All-Equal) sense.

Definition 3.2: The *MAX-UNANIMITY* problem is described as follows: given a positive integer M and m disjunctive clauses defined over n Boolean variables, we ask whether there exists a truth assignment such that the number of *unanimous* disjunctive clauses is at least M. When the number of literals in each clause is two, we denote the corresponding problem by *MAX-2UNANIMITY.* When each clause contains three literals, the problem of determining whether there exists a truth assignment under which all clauses are satisfied in the NAE sense is called *NAE-SAT*.

The NAE-SAT problem is known to be NP-complete [28]. The next lemma says that the MAX-2UNANIMITY problem is also NP-complete (the proof is provided in Appendix A).

Lemma 3.1: MAX-2UNANIMITY is NP-complete.

We are now ready to prove Theorem 3.1.

Proof: Let the utility in problem (3) be H_3. Consider an instance of MAX-2UNANIMITY with clauses $c_1, c_2, ..., c_m$ defined over Boolean variables $x_1, x_2, ..., x_n$ and an integer M. Let $\mathbf{h}_a = (1,0)^T$, $\mathbf{h}_b = (0,1)^T$ and $\mathbf{h} = (\sqrt{N}, 0)^T$, where N is a large positive number (to be specified later). We write each clause $c_j = \alpha_j \vee \beta_j$, with $\alpha_j,\ \beta_j$ taken from $\{x_1, x_2, ..., x_n, \bar{x}_1, \bar{x}_2, ..., \bar{x}_n\}$. Let us define two mappings

$$\pi,\ \tau : \{1, 2, ..., m\} \mapsto \{\pm 1, \pm 2, ..., \pm n\}$$

such that

$$\pi(j) = \begin{cases} i, & \text{if } \alpha_j = x_i, \\ -i, & \text{if } \alpha_j = \bar{x}_i, \end{cases} \quad \text{and } \tau(j) = \begin{cases} i, & \text{if } \beta_j = x_i, \\ -i, & \text{if } \beta_j = \bar{x}_i. \end{cases}$$

For instance, if $c_4 = x_3 \vee \bar{x}_5$, then we have $\alpha_4 = x_3,\ \beta_4 = \bar{x}_5$, with $\pi(4) = 3$ and $\tau(4) = -5$. For $i = \pm 1, \pm 2, ..., \pm n$, we define

$$\mathbf{h}_i = \begin{cases} \mathbf{h}_a, & \text{if } i > 0, \\ \mathbf{h}_b, & \text{if } i < 0. \end{cases}$$

Given any instance of MAX-2UNANIMITY, we construct the following (6) as an instance of (3) (the inverse of harmonic mean utility minimization is equivalent to harmonic mean utility maximization) with a total of $K = 4n + 2m$ users. Herein, each Boolean variable x_i corresponds to four users, including user

[1] Recall that for a given set of Boolean variables, a literal is defined as either a Boolean variable or its negation, while a disjunctive clause refers to a logical expression consisting of the logical "OR" of literals.

$4i$ (called "variable user") and user $4i-1, 4i-2$ and $4i-3$ (called "auxiliary variable users"); while each clause c_j corresponds to a pair of users, i.e., user $4n+2j$ and $4n+2j-1$ (called "clause users"). In (6), each (variable, auxiliary variable or clause) user k is associated with a transmitter beamforming vector $\mathbf{v}_k, k = 1, 2, ..., 4n+2m$.

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\left(\frac{1}{r_{4i}}+\frac{1}{r_{4i-1}}+\frac{1}{r_{4i-2}}+\frac{1}{r_{4i-3}}\right)+\sum_{j=1}^{m}\left(\frac{1}{r_{4n+2j}}+\frac{1}{r_{4n+2j-1}}\right)\\
\text{s.t.} \quad & r_{4i}=\log\left(1+|(\sqrt{0.9},\sqrt{0.9})\mathbf{v}_{4i}|^2\right),\ r_{4i-1}=\log\left(1+\frac{|(1,0)\mathbf{v}_{4i-1}|^2}{|(1,1)\mathbf{v}_{4i}|^2}\right),\ 1\le i\le n,\\
& r_{4i-2}=\log\left(1+\frac{|(10,0)\mathbf{v}_{4i-2}|^2}{|(1,0)\mathbf{v}_{4i}|^2}\right),\ r_{4i-3}=\log\left(1+\frac{|(10,0)\mathbf{v}_{4i-3}|^2}{|(0,1)\mathbf{v}_{4i}|^2}\right),\ 1\le i\le n,\\
& r_{4n+2j}=\log\left(1+\frac{|\mathbf{h}^\dagger\mathbf{v}_{4n+2j}|^2}{1+|\mathbf{h}^\dagger_{\pi(j)}\mathbf{v}_{4|\pi(j)|}|^2+|\mathbf{h}^\dagger_{\tau(j)}\mathbf{v}_{4|\tau(j)|}|^2}\right),\ 1\le j\le m,\\
& r_{4n+2j-1}=\log\left(1+\frac{|\mathbf{h}^\dagger\mathbf{v}_{4n+2j-1}|^2}{1+|\mathbf{h}^\dagger_{-\pi(j)}\mathbf{v}_{4|\pi(j)|}|^2+|\mathbf{h}^\dagger_{-\tau(j)}\mathbf{v}_{4|\tau(j)|}|^2}\right),\ 1\le j\le m,\\
& \|\mathbf{v}_k\|^2\le 1,\ 1\le k\le 4n+2m.
\end{aligned}
\tag{6}
$$

The n "variable users" $4i, i = 1, 2, ..., n$, in (6) communicate interference free. Their channel vectors are $(\sqrt{0.9}, \sqrt{0.9})^T$ and their noise power are 1. The $3n$ "auxiliary variable users" $4i-1, 4i-2, 4i-3, i = 1, 2, ..., n$, do suffer from crosstalk interference from "variable user" $4i$. That is, the interference channel vectors from "variable user" $4i$ are $(1, 1)^T, (1, 0)^T$ and $(0, 1)^T$; the direct channel vectors are $(1, 0)^T, (10, 0)^T$ and $(10, 0)^T$; the self noise power are zero. For the "clause users" $4n+2j$ and $4n+2j-1,\ j = 1, 2, ..., m$: their direct channel vectors are $\mathbf{h}$; their noise powers are 1; their interference channel vectors are $\mathbf{h}_{\pi(j)}, \mathbf{h}_{\tau(j)}$ and $\mathbf{h}_{-\pi(j)}, \mathbf{h}_{-\tau(j)}$, respectively. Take the clause $c_1 = x_2 \vee \bar{x}_3$ as an illustrative example and the corresponding "clause users" are $4n+2$ and $4n+1$. Since $\pi(1) = 2$ and $\tau(1) = -3$, the two "clause users" experience interferences from "variable users" $4|\pi(1)| = 8$ and $4|\tau(1)| = 12$. The interferences are $|\mathbf{h}_a^\dagger \mathbf{v}_8|^2 + |\mathbf{h}_b^\dagger \mathbf{v}_{12}|^2$ and $|\mathbf{h}_b^\dagger \mathbf{v}_8|^2 + |\mathbf{h}_a^\dagger \mathbf{v}_{12}|^2$, respectively.

The correspondence between MAX-2UNANIMITY problem and problem (6) is listed as Table I. Notice that r_{4n+2j} can be obtained from clause c_j according to Table I and $r_{4n+2j-1}$ can be obtained from r_{4n+2j} by swapping $\mathbf{h}_a$ with $\mathbf{h}_b$.

We first fix some easy variables of (6) to simplify the problem. Since each beamforming vector $\mathbf{v}_{4n+2j}$, $\mathbf{v}_{4n+2j-1}$, $j = 1, 2, ..., m$, and $\mathbf{v}_{4i-l}$, $i = 1, 2, ..., n$, $l = 1, 2, 3$, appears exactly once, it follows by optimality that they must match the corresponding channel vectors. That is, $\mathbf{v}^*_{4n+2j} = \mathbf{v}^*_{4n+2j-1} = \mathbf{v}^*_{4i-1} = \mathbf{v}^*_{4i-2} = \mathbf{v}^*_{4i-3} = \mathbf{h}_a$. It remains to determine the optimal beamforming vectors $\mathbf{v}^*_{4i}$, $i = 1, 2, ..., n$. For this purpose, we need the following key lemma whose proof is relegated to Appendix B.

TABLE I

VARIABLE CORRESPONDENCE

MAX-2UNANIMITY	Problem (6)
variable x_i	beamforming vector $\mathbf{v}_{4i}$
clause c_j	rates r_{4n+2j} and $r_{4n+2j-1}$
literal x_i	interference $\|\mathbf{h}_a^\dagger \mathbf{v}_{4i}\|^2$
literal $\bar{x}_i$	interference $\|\mathbf{h}_b^\dagger \mathbf{v}_{4i}\|^2$

Lemma 3.2: When $N \geq 2(\mathrm{e}^{200m} - 1)$, the optimal beamforming vectors $\{\mathbf{v}_{4i}^*\}$ of (6) must be either $\mathbf{h}_a$ or $\mathbf{h}_b$.

From Lemma 3.2, the first sum in the objective function of (6) equals nC regardless of $\mathbf{v}_{4i}^* = \mathbf{h}_a$ or $\mathbf{h}_b$, where

$$C \triangleq \frac{1}{\log 1.9} + \frac{1}{\log 101} + \frac{1}{\log 2} = \frac{1}{r_{4i}} + \frac{1}{r_{4i-1}} + \frac{1}{r_{4i-2}} + \frac{1}{r_{4i-3}}$$

is a constant. Thus, we only need to consider the second sum of (6). Notice that the value of each term in the second sum of (6) only depends on whether clause c_j is unanimous, i.e.,

$$\frac{1}{r_{4n+2j}} + \frac{1}{r_{4n+2j-1}} = \begin{cases} \dfrac{1}{\log(1+N/3)} + \dfrac{1}{\log(1+N/1)}, & \text{if } c_j \text{ is unanimous,} \\ \dfrac{1}{\log(1+N/2)} + \dfrac{1}{\log(1+N/2)}, & \text{if } c_j \text{ is not.} \end{cases}$$

Since

$$\frac{1}{\log(1+N/3)} + \frac{1}{\log(1+N)} < \frac{2}{\log(1+N/2)}$$

from Claim 1 in Appendix B, it follows that the second sum of (6) will be smaller if more clauses are satisfied unanimously. Therefore, the minimum of (6) is only related to the maximum number of unanimous clauses of the given MAX-2UNANIMITY instance. Specifically, the minimum of (6) is no more than

$$nC + \frac{M}{\log(1+N/3)} + \frac{M}{\log(1+N)} + \frac{2(m-M)}{\log(1+N/2)} \tag{7}$$

if and only if there exists a truth assignment such that at least M clauses are made unanimous for the given MAX-2UNANIMITY instance. Thus, we have transformed MAX-2UNANIMITY problem to the problem of checking if problem (6) will have an optimal value below the above threshold (7).

Finally, given any instance of MAX-2UNANIMITY, we can construct problem (6) in polynomial time. Since MAX-2UNANIMITY is NP-complete (Lemma 3.1), it follows that problem (3) with harmonic mean utility is NP-hard. ■

A few remarks are in order. First, it follows from the proof of Theorem 3.1 that even if the optimal transmit power levels are known (i.e., $\|\mathbf{v}_k\|^2 \leq P_k$ is replaced with $\|\mathbf{v}_k\|^2 = P_k$), the problem of finding the optimal beamforming directions of harmonic mean rate maximization problem is still NP-hard. Second, we have set the noise powers of users $4i-1, 4i-2, 4i-3, i = 1, 2, ..., n,$ to zero in (6). These settings simplify the proof and do not reduce any generality. We could have used small noise power values in the proof (even though some extra argument is needed), since there is a positive gap between the global optimal value and the local optimal values of (6). Finally, our proof actually implies that there is a *positive* probability (measure) that a randomly generated MISO coordinated beamforming problem under the harmonic mean utility is NP-hard. In particular, by continuity, all slightly perturbed versions of the constructed instance (6) (i.e., channel vectors, noise/transmit powers are slightly changed) will be equivalent to MAX-2UNANIMITY problem. This is because there is a positive (and constant) jump in the global optimal value of the constructed example when the optimal value of the corresponding MAX-2UNANIMITY instance increases by one. When channel conditions change slightly, this one-to-one correspondence between the optimal values and the property of the discrete jump in the optimal value of the constructed MISO problem remains valid.

D. Maximization of the Proportional Fairness Utility

Like the harmonic mean utility, we have the following hardness result.

Theorem 3.2 (Proportional Fairness Utility): For the proportional fairness utility $H_2 = \left(\prod_{k=1}^{K} r_k\right)^{1/K}$, the optimal coordinated downlink beamforming problem can be transformed into a convex optimization problem when $L = 1$, but is NP-hard when $L \geq 2$.

Proof: The first part of Theorem 3.2 is proved in [1]. For the second part, the argument is similar to that of Theorem 3.1. We only give a proof outline here.

First, we have the following lemma whose proof is provided in Appendix C.

Lemma 3.3: The function $f(x) = \log\log\left(1 + \frac{1}{\sigma^2 + x}\right)$ is strictly convex in $x \geq 0$ for any σ.

Second, given any MAX-2UNANIMITY instance, an instance (8) of problem (3) with utility $\log H_2$ (equivalent to proportional fairness utility maximization) and $3n+2m$ users is constructed as follows:

$$\begin{aligned}
\max \quad & \sum_{i=1}^{n} (\log r_{3i} + \log r_{3i-1} + \log r_{3i-2}) + \sum_{j=1}^{m} (\log r_{3n+2j} + \log r_{3n+2j-1}) \\
\text{s.t.} \quad & r_{3i} = \log\left(1 + |(\sqrt{0.1}, \sqrt{0.1})\mathbf{v}_{3i}|^2\right), \; r_{3i-1} = \log\left(1 + \frac{|(1,0)\mathbf{v}_{3i-1}|^2}{|(1,0)\mathbf{v}_{3i}|^2}\right), \; 1 \le i \le n, \\
& r_{3i-2} = \log\left(1 + \frac{|(1,0)\mathbf{v}_{3i-2}|^2}{|(0,1)\mathbf{v}_{3i}|^2}\right), \; 1 \le i \le n, \\
& r_{3n+2j} = \log\left(1 + \frac{|\mathbf{h}^\dagger \mathbf{v}_{3n+2j}|^2}{1 + |\mathbf{h}^\dagger_{\pi(j)} \mathbf{v}_{3|\pi(j)|}|^2 + |\mathbf{h}^\dagger_{\tau(j)} \mathbf{v}_{3|\tau(j)|}|^2}\right), \; 1 \le j \le m, \\
& r_{3n+2j-1} = \log\left(1 + \frac{|\mathbf{h}^\dagger \mathbf{v}_{3n+2j-1}|^2}{1 + |\mathbf{h}^\dagger_{-\pi(j)} \mathbf{v}_{3|\pi(j)|}|^2 + |\mathbf{h}^\dagger_{-\tau(j)} \mathbf{v}_{3|\tau(j)|}|^2}\right), \; 1 \le j \le m, \\
& \|\mathbf{v}_k\|^2 \le 1, \; 1 \le k \le 3n+2m.
\end{aligned} \tag{8}$$

Notice that each solution to (8) must have $\mathbf{v}^*_{3i-1} = \mathbf{v}^*_{3i-2} = \mathbf{h}_a, \; i = 1, 2, ..., n$, and $\mathbf{v}^*_{3n+2j} = \mathbf{v}^*_{3n+2j-1} = \mathbf{h}_a, \; j = 1, 2, ..., m$. Moreover, we consider the following parametric optimization problem (similar to (18) in the harmonic mean case):

$$\begin{aligned}
\max \quad & \log r_3 + \log r_2 + \log r_1 \\
\text{s.t.} \quad & r_3 = \log\left(1 + |(\sqrt{0.1}, \sqrt{0.1})\mathbf{v}_3|^2\right), \\
& r_2 = \log\left(1 + 1/(\sigma^2 + |(1,0)\mathbf{v}_3|^2)\right), \\
& r_1 = \log\left(1 + 1/(\sigma^2 + |(0,1)\mathbf{v}_3|^2)\right), \\
& \|\mathbf{v}_3\| = t,
\end{aligned} \tag{9}$$

where $\sigma > 0$ is a constant and t is the parameter. The global maxima of (9) should be $(t, 0)^T$ or $(0, t)^T$ when σ is small. Furthermore, the optimum value of (9) is an increasing function with respect to $t \in [0, 1]$. Using an argument similar to that of the harmonic mean case for (6), each globally optimal beamforming solution $\mathbf{v}^*_{3i}$ of (8) should be either $\mathbf{h}_a$ or $\mathbf{h}_b$ when $N \ge 3(\mathrm{e}^{6m} - 1)$. When restricted to solutions of the form $\mathbf{v}^*_{3i} = \mathbf{h}_a$ or $\mathbf{h}_b$, the maximum of (8) is only linearly related to the maximum number of unanimous clauses of the given MAX-2UNANIMITY instance. Thus, maximizing the number of unanimous clauses is the same as solving (8). According to Lemma 3.1, it follows that the optimal coordinated downlink beamforming problem with utility H_2 is also NP-hard. ■

E. Maximization of the Min-Rate Utility

Let the system utility function in (3) be given by $H = H_4$. In this case, the problem can be solved in polynomial time for arbitrary L and K [1], [2]. Specifically, letting

$$r = \min_{1 \le k \le K} \{r_k\},$$

the min-rate utility maximization problem becomes

$$\begin{aligned} \max \quad & r \\ \text{s.t.} \quad & r \le \log\left(1 + \frac{|\mathbf{h}_{kk}^{\dagger}\mathbf{v}_k|^2}{\sigma_k^2 + \sum_{j \ne k} |\mathbf{h}_{jk}^{\dagger}\mathbf{v}_j|^2}\right), \\ & \|\mathbf{v}_k\|^2 \le P_k,\ 1 \le k \le K. \end{aligned} \tag{10}$$

Given a $r > 0$, we can efficiently check if there exists $\mathbf{v}_k, k = 1, 2, ..., K$, such that the constraints in (10) are satisfied. This feasibility problem is a second order cone programming, which can be solved efficiently using interior-point methods. The following theorem is a generalization of the result of [2], which deals with the single-cell case.

Theorem 3.3 (Min-Rate Utility): For the min-rate utility, the optimal coordinated downlink beamforming problem can be solved in polynomial time with arbitrary K and L under the (very wild) assumption that $\min_k \left\{P_k\|\mathbf{h}_{kk}\|^2/\sigma_k^2\right\} \le R$.

Proof: We propose the following polynomial time algorithm for (10) based on the bisection technique.

A Polynomial Time Algorithm for Min-Rate Utility Maximization	
Step 1.	Initialization: Choose r_ℓ and r_u such that the optimal r_{opt} lies in $[r_\ell, r_u]$ and a tolerance ϵ.
Step 2.	If $r_u - r_\ell \le \epsilon$, stop, else go to **Step 3**.
Step 3.	Let $r_{\text{mid}} = (r_\ell + r_u)/2$ and solve an SOCP problem to check the feasibility problem of (10) with $r = r_{\text{mid}}$. If feasible, set $r_\ell = r_{\text{mid}}$, else set $r_u = r_{\text{mid}}$ and go to **Step 2**.

According to standard analysis of path-following interior-point methods, **Step 3** can be finished in $O(K^{3.5}L^{3.5})$ time. As for the initial choices of r_ℓ and r_u, we can let $\bar{\mathbf{v}}_k$ be the space matched beamformer, i.e., $\bar{\mathbf{v}}_k = \mathbf{h}_{kk}\sqrt{P_k}/\|\mathbf{h}_{kk}\|$ and

$$r_\ell = \min_k \log\left(1 + \frac{|\mathbf{h}_{kk}^{\dagger}\bar{\mathbf{v}}_k|^2}{\sigma_k^2 + \sum_{j \ne k} |\mathbf{h}_{jk}^{\dagger}\bar{\mathbf{v}}_j|^2}\right),$$

$$r_u = \min_k \log\left(1 + \frac{|\mathbf{h}_{kk}^\dagger \bar{\mathbf{v}}_k|^2}{\sigma_k^2}\right).$$

It takes $\log_2\left((r_u - r_\ell)/\epsilon\right)$ iterations to reach tolerance ϵ. Thus, a total of $O(K^{3.5}L^{3.5}\log_2\left((r_u - r_\ell)/\epsilon\right))$ arithmetic operations are needed in the worst case. Since

$$r_u - r_\ell \le \min_k \log(1 + P_k\|\mathbf{h}_{kk}\|^2/\sigma_k^2) \le R,$$

we have

$$K^{3.5}L^{3.5}\log_2\left((r_u - r_\ell)/\epsilon\right) \le K^{3.5}L^{3.5}N(R,\epsilon),$$

where $N(R,\epsilon)$ is the smallest integer which is greater than $\log_2(R/\epsilon)$. Therefore, the above algorithm has a polynomial time worst case complexity. ■

The algorithm described above can easily be extended to the weighted min-rate maximization problems. In [30], [31], the weighted min-rate maximization is related to the weighted sum MSE minimization and the weighted sum-rate maximization via the Friedland-Karlin spectral radius minimax theorem in nonnegative matrix theory. A brief sketch of the Friedland-Karlin inequalities can be found in [33].

Table II summarizes the complexity status of the optimal coordinated downlink beamforming problem (3) for different choices of utilities.

TABLE II

COMPLEXITY STATUS OF THE OPTIMAL COORDINATED DOWNLINK BEAMFORMING PROBLEM IN THE MISO INTERFERENCE CHANNEL

Class \ Utility	Weighted Sum-Rate	Proportional Fairness	Harmonic Mean	Min-Rate
$L=1$, any K	NP-hard [1]	Convex [1]	Convex [1]	Poly. Time Solvable [1], [2]
$L\ge 2$, any K	NP-hard	NP-hard	NP-hard	Poly. Time Solvable [1], [2]

IV. A CYCLIC COORDINATE DESCENT ALGORITHM

In this section, we consider how to solve the coordinated beamforming problem (3) with a general utility $\rho(\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_K)$, i.e., it can be the inverse of the harmonic mean utility function $\sum_{k=1}^K 1/r_k$. In particular, problem (3) is changed into

$$\begin{aligned} \min \quad & \rho(\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_K) \\ \text{s.t.} \quad & \|\mathbf{v}_k\|^2 \le P_k,\ 1\le k\le K. \end{aligned} \tag{11}$$

Since this problem is NP-hard in general (proved in Section III), we are led to develop efficient algorithms to find a high quality approximate solution or a stationary point for it. Due to variable *separability* in the constraints of (11) and our desire for distributed implementations, we propose to solve (11) by *cyclicly* adjusting the beamforming vector $\mathbf{v}_k$ while assuming the beamforming vectors $\{\mathbf{v}_j : j \neq k\}$ are fixed. In other words, we solve a sequence of per-base station problems

$$\begin{aligned} \min_{\mathbf{v}_k} \quad & \rho(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K) \\ \text{s.t.} \quad & \|\mathbf{v}_k\|^2 \leq P_k. \end{aligned} \tag{12}$$

A. An Inexact Cyclic Coordinate Descent Algorithm

The cyclic coordinate descent algorithm is also known as the nonlinear Gauss-Seidel iteration [34]. There are several studies of this type of algorithms [34]–[38] and its applications in engineering [39]. However, most of these studies require either the convexity of objective function or exact solution of subproblem (12), which not only is costly but also may result in algorithm divergence [35]. Below we consider a general differentiable optimization problem with separable constraints

$$\begin{aligned} \min \quad & f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K) \\ \text{s.t.} \quad & \mathbf{x}_k \in X_k, \ 1 \leq k \leq K, \end{aligned} \tag{13}$$

where the feasible set $X := \prod_{k=1}^{K} X_k$ is separable, bounded and closed. We propose an easily implementable cyclic coordinate descent algorithm which simply requires a sufficient decrease in the objective of (12) at each iteration. The algorithm can be applied to solve the utility maximization problem (3) with $H = H_1$, H_2 and H_3 and have the same convergence properties because they have smooth objective functions and a separable feasible region (the feasible region is in the Cartesian product form). But the same can not be said about H_4 since it is non-differentiable.

Interestingly, the pricing algorithm introduced in [17] can be viewed as a partially linearized version of our cyclic coordinate descent algorithm. Specifically, under the power constraint ($\|\mathbf{v}_k\|^2 \leq P_k$): the proposed algorithm tries to allocate resources of the k-th transmitter by maximizing the summation of all users' utility functions; while the pricing algorithm lets transmitter k maximize its own utility function plus the summation of the first order approximation of all other users' utility functions at the current point $I_{kj} = |\mathbf{h}_{kj}^{\dagger}\mathbf{v}_k|^2$, where I_{kj} denotes the received interference at the j-th receiver from the k-th transmitter.

The next result shows that the above inexact cyclic coordinate descent algorithm converges to a KKT point of (13). The proof of this result is relegated to Appendix D.

Theorem 4.1: Suppose $f(\mathbf{x})$ is twice continuously differentiable and bounded below, and the feasible set X is convex, separable and compact. Then every accumulation point of the sequence $\{\mathbf{x}^i\}$ generated by the inexact cyclic coordinate descent algorithm is a stationary point of (13).

An Inexact Cyclic Coordinate Descent Algorithm

Step 1. Initialization: choose $\mathbf{x}^1 = [\mathbf{x}_1^1, \mathbf{x}_2^1, ..., \mathbf{x}_K^1]$ and a tolerance $\epsilon > 0$.

Step 2. Iteration $i \geq 1$: Let $\mathbf{z}_0^{i+1} = \mathbf{x}^i$.

For $k = 1, 2, ..., K$,

- Compute the gradient projection direction for the component $\mathbf{x}_k$ according to

$$\mathbf{d}_k^{i+1} = P_{X_k}[\mathbf{x}_k^i - \nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})] - \mathbf{x}_k^i, \tag{14}$$

where $P_{X_k}[\cdot]$ denotes the orthogonal projection to X_k and $\mathbf{z}_k^{i+1}$ is defined as

$$\mathbf{z}_k^{i+1} = \left(\mathbf{x}_1^{i+1}, ..., \mathbf{x}_k^{i+1}, \mathbf{x}_{k+1}^i, ..., \mathbf{x}_K^i\right). \tag{15}$$

- Determine a stepsize α_k^{i+1} using the backtracking line search [40].
- Update $\mathbf{x}_k^{i+1} = \mathbf{x}_k^i + \alpha_k^{i+1}\mathbf{d}_k^{i+1}$.

End (For)

Step 3. Termination: If $\|\mathbf{x}^{i+1} - \mathbf{x}^i\| \leq \epsilon$, then stop; else set $i = i + 1$ and go to **Step 2**.

The separability of the constraints is necessary for the algorithm's convergence. The following example (taken from [37]) shows that, without the separability, the algorithm can get stuck at an uninteresting point:

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 + x_2 \geq 2. \end{aligned}$$

This strongly convex function has a unique global solution at $x_1^* = x_2^* = 1$. However, if the initial point is $(1.5, 0.5)$, the cyclic coordinate descent algorithm will be stuck.

Specializing the inexact cyclic coordinate descent algorithm to the coordinated beamforming problem (11), we need to perform a projected gradient descent iteration for the subproblem (12). In this case, we have a ball constraint $V_k = \{\mathbf{v} \mid \|\mathbf{v}\|^2 \leq P_k\}$ and the corresponding projection is straightforward

$$P_{V_k}(\mathbf{v}) = \begin{cases} \mathbf{v}, & \text{if } \|\mathbf{v}\|^2 \leq P_k, \\ \dfrac{\sqrt{P_k}\mathbf{v}}{\|\mathbf{v}\|}, & \text{if } \|\mathbf{v}\|^2 > P_k. \end{cases} \tag{16}$$

The proposed optimization procedure can be implemented in a distributed fashion to solve the MISO downlink beamforming problem. At the initial step, each base station needs to know the system utility function and the local CSI for all channels originating from that transmitter (either through feedback or reverse-link estimation [41]). The only information to be exchanged are the numerator and denominator of SINR terms at K receivers. In subsequent iterations, a base station updates its beamforming vector by solving (12) inexactly using a gradient projection algorithm. Take minus sum-rate utility minimization as an example. Denote the received desired signal power and received interference-plus-noise power at receiver k by $\hat{P}_k$ and $\hat{I}_k$. Then we have $\text{SINR}_k = \hat{P}_k/\hat{I}_k$ and the utility function in (12) becomes

$$\rho(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K) = -\sum_{k=1}^{K} \log(1 + \text{SINR}_k).$$

Since transmitter k knows $\hat{P}_j, \hat{I}_j\,(j = 1, 2, ..., K)$, its local CSI $\mathbf{h}_{kj}\,(j = 1, 2, ..., K)$ and the utility function ρ, we have $\nabla_{\mathbf{v}_k}\rho(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K) = -\sum_{j=1}^{K} \mathbf{g}_k(j)$, where

$$\mathbf{g}_k(j) = \begin{cases} \dfrac{2\mathbf{h}_{kk}^{\dagger}\mathbf{v}_k\mathbf{h}_{kk}}{(1 + \text{SINR}_k)\,\hat{I}_k}, & \text{if } j = k, \\[2ex] \dfrac{-2\hat{P}_j\mathbf{h}_{kj}^{\dagger}\mathbf{v}_k\mathbf{h}_{kj}}{(1 + \text{SINR}_j)\,\hat{I}_j^2}, & \text{if } j \neq k. \end{cases}$$

Combining (14) and (16), the search direction $\mathbf{d}_k$ is obtained (as transmitter k knows its own power budget P_k). Furthermore, the step α_k can be determined using the backtracking search technique because we can compute the utility function and its gradient at point $\mathbf{v}_k + \alpha\mathbf{d}_k$ for any trial step α. After that, all receivers measure individual SINR terms and send the SINR information (both the numerator and denominator of SINR) to the next base station. The inexact cyclic coordinate descent algorithm enables each transmitter to update its beamformer with only limited information exchange.

As a variant, we can also use the so-called Barzilai-Borwein (BB) [42] projection step for the subproblem (12) to replace the standard gradient projection step. In particular, at i-th iteration, the BB gradient projection direction $\mathbf{d}_{BB}^i$ is given by

$$\begin{cases} \mathbf{d}_{BB}^i = P_{V_k}\left(\mathbf{v}_k^i - \alpha_{BB}^i\nabla_{\mathbf{v}_k}\rho(\mathbf{v}^i)\right) - \mathbf{v}_k^i, \\[1ex] \alpha_{BB}^i = \dfrac{\|\mathbf{s}^{i-1}\|^2}{(\mathbf{s}^{i-1})^T\mathbf{y}^{i-1}}, \end{cases}$$

where $\mathbf{s}^{i-1} = \mathbf{v}_k^i - \mathbf{v}_k^{i-1}$, $\mathbf{y}^{i-1} = \nabla_{\mathbf{v}_k}\rho(\mathbf{v}^i) - \nabla_{\mathbf{v}_k}\rho(\mathbf{v}^{i-1})$. The BB (projection) method is known to have better numerical performance than the standard gradient (projection) method. The R-linear convergence of BB method has been established in [43] for strongly convex quadratic functions.

V. Numerical Simulations

To evaluate the effectiveness of the proposed inexact cyclic coordinate descent algorithm, we have conducted numerical simulations for a 7-cell network with one user per cell as shown in Fig. 1. Each

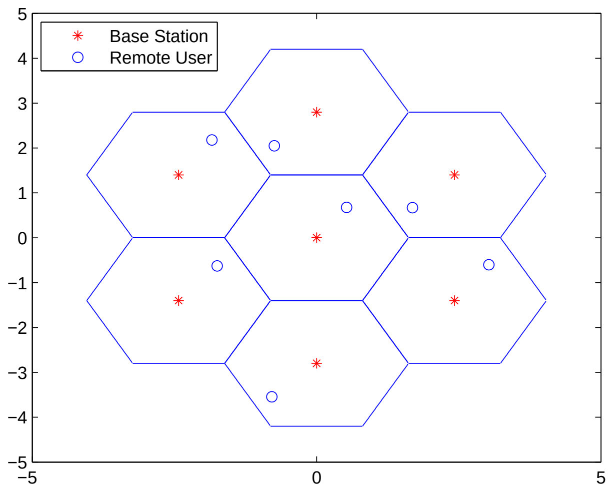

Fig. 1. A wireless network with seven base stations and a single user per cell.

base station is equipped with L antennas. Similar to [5], standard WiMax parameters are used in all the simulations; see Table III, where d is the distance in kilometers. The location of each remote user is chosen randomly within its cell but at least 0.5km away from the corresponding base station. The proposed algorithm is initialized to the space matched beamformer in all simulations.

TABLE III

STANDARD WIMAX PARAMETERS

Model or Parameters	Values
noise power spectral density	-162 dBm/Hz
path loss model	$128.1 + 37.6\log_{10}(d)$
log-normal shadowing	8 dB
distance between neighboring base stations	2.8km
antenna gain	15 dBi

Figure 2 plots the iteration process of BB projection method for problem (3) with harmonic mean utility. It can be seen that most of improvement is achieved in the first 1-2 iterations, making the method attractive for practical implementations.

For a two-user MISO channel, a parametrization of the achievable rate region boundary was given in [8], [24]. We can use this parametrization to compute the global optimal beamformer of the coordinated MISO

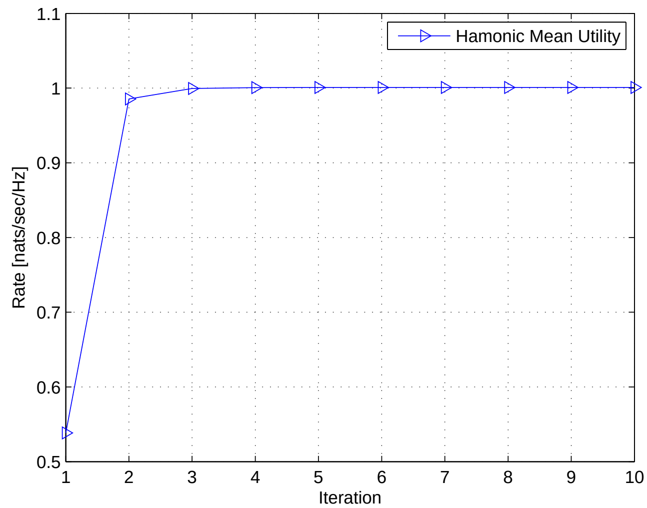

Fig. 2. A typical iteration process of the inexact cyclic coordinate descent algorithm with $K = 7, L = 4$ and $P = 30$ dBm.

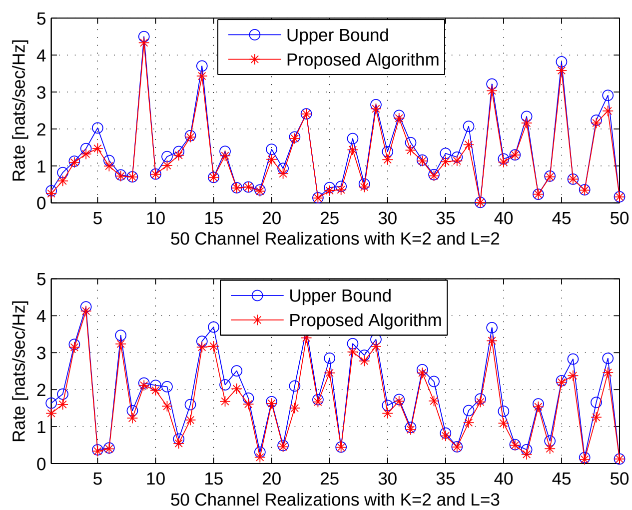

Fig. 3. Performance comparison of the proposed algorithm and the parametrization method with $K = 2, P = 30$ dBm.

downlink beamforming problem by searching along the rate region boundary. In Fig. 3, the performance of the proposed algorithm is compared against the global optimum for 50 randomly generated two-user MISO channels. It can be seen that the proposed algorithm either achieves, or nearly achieves, the global optimality.

The individual rate distribution is shown in Fig. 4 over 500 channel realizations, where $K = 7, L = 4$

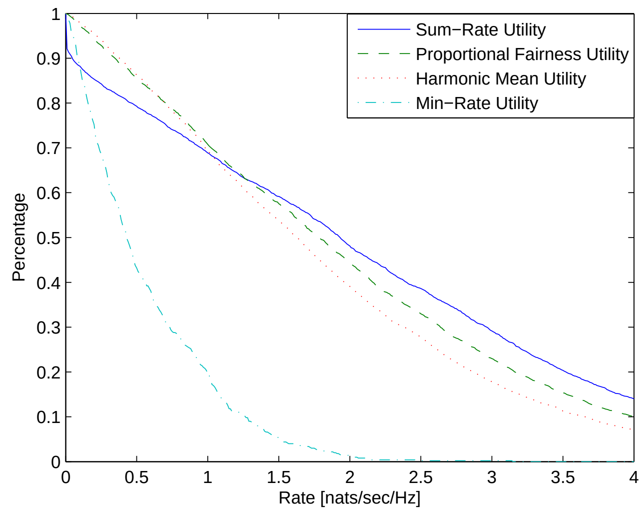

Fig. 4. Individual rate distribution with $K = 7, L = 4$ and $P = 30$ dBm.

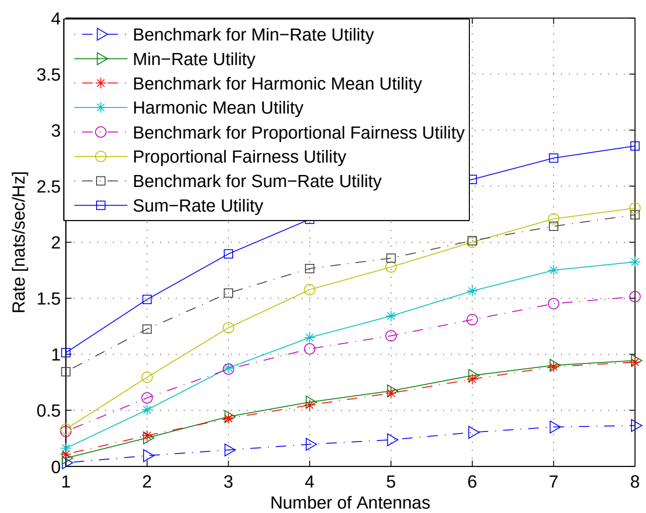

Fig. 5. Performance comparison of coordinated beamformers and space matched beamformers with $K = 7, P = 30$ dBm.

and $P = 30$dBm. It plots the percentage of users at or above the given rate level with the four utilities. We can observe from Fig. 4 that sum-rate maximization achieves the over efficiency of the network. Its basic idea is serving only users with good channel states, so about 10% of users are switched off. Min-rate maximization balances transmit rates of all users at the same level and achieves max-min fairness among the users. But the overall system performance is severely degraded when there are users with bad channel attenuations. Proportional fairness and harmonic mean utilities indeed tradeoff user fairness and system performance, i.e., they improve system performance remarkably (compared to min-rate utility) while keep serving all users in the network (compared to sum-rate utility).

Figure 5 shows the performance of four different utility functions with coordinated beamformers achieved by the proposed algorithm and a fixed transmit power $P = 30$ dBm versus different number of transmit antennas. Each point in Fig. 5 is obtained by averaging over 500 independent channel realizations. Space matched beamformers are used as the benchmark. It can be seen that the transmit rates improve significantly over the benchmark solution. When the number of users ($K = 7$) and the transmit power ($P = 30$ dBm) are fixed, an increase in the system utility is observed with the number of transmit antennas.

VI. Conclusion

Coordinated transmit beamforming is a promising approach for interference mitigation in the MISO interference channel. A major design challenge is to find, for a given channel state, a globally optimal beamforming strategy under an appropriate utility criterion. In the single carrier case with a single antenna per transmitter, maximizing the (weighted) sum-rate is known to be NP-hard. However, the same problem is polynomial time solvable when the proportional fairness, harmonic mean or min-rate utility is used. It turns out the situation with multiple transmit antennas is rather different. For instance, when each transmitter is equipped with two antennas, the corresponding joint beamforming design problem becomes NP-hard under either the proportional fairness or the harmonic mean criterion. These complexity results suggest that we should abandon effort to find globally optimal beamformers for a general MISO interference channel unless the min-rate utility is used. In the latter case the problem remains polynomial time solvable.

Motivated by these complexity results, we propose a simple distributed inexact cyclic coordinate descent algorithm to find a (locally optimal) beamforming strategy. Our algorithm exploits the separable structure of the power constraints, and is provably globally convergent to a KKT solution. The algorithm requires only local CSI and an exchange of SINR information at each iteration. Numerical experiments with WiMax system parameters show that the proposed algorithm is both effective and efficient, providing significant rate gain over the space matched beamforming strategy.

VII. Acknowledgment

The authors wish to thank Professor Wei Yu of University of Toronto for his help in numerical simulations and the anonymous reviewers for their useful comments.

APPENDIX A
PROOF OF LEMMA 3.1

We construct a polynomial time transformation from NAE-SAT problem. It can be checked that MAX-2UNANIMITY problem is in NP. Given a disjunctive clause with three literals, $c = x \vee y \vee z$, let us construct the following six clauses, each involving only two literals:

$$R(c): \; x \vee \bar{y}, \; x \vee \bar{z}, \; y \vee \bar{x}, \; y \vee \bar{z}, \; z \vee \bar{x}, \; z \vee \bar{y}. \tag{17}$$

It can be checked that $R(c)$ has the following properties:

1) The number of unanimous clauses (i.e., all literals having the same value) in $R(c)$ is at most four.
2) The clause c is satisfied in the NAE sense if and only if the number of *unanimous* clauses in $R(c)$ is four.

Now given any instance ϕ of NAE-SAT, we construct a corresponding instance $R(\phi)$ of MAX-2UNANIMITY as follows: for each clause $c = \alpha \vee \beta \vee \gamma$ of ϕ, we add to $R(\phi)$ the six clauses in (17), with $x, \; y, \; z$ replaced with the literals $\alpha, \; \beta, \; \gamma$, respectively. In this way, if ϕ has m clauses, then $R(\phi)$ will have $6m$ clauses. Let $M = 4m$. Then properties 1 and 2 imply that all clauses in ϕ are simultaneously satisfied in the NAE sense if and only if at least $M = 4m$ clauses in $R(\phi)$ can be made *unanimous*. In particular, suppose that $4m$ clauses in $R(\phi)$ are unanimous under a given truth assignment. Since, by property 1, each group $R(c)$ of six clauses can have at most four unanimous clauses, it follows that exactly four clauses must be unanimous in each group. By property 2, this further implies that each clause in ϕ is satisfied in the NAE sense. Conversely, any truth assignment that satisfies a clause c in the NAE sense will give rise to four unanimous clauses. Thus, if all m clauses in ϕ are satisfied in the NAE sense, then there will be $4m$ unanimous clauses in $R(\phi)$. Finally, this transformation is in polynomial time.

APPENDIX B
PROOF OF LEMMA 3.2

The proof consists of establishing three claims.

Claim 1: The function $\log^{-1}\left(1 + \left(\sigma^2 + x\right)^{-1}\right)$ is strictly concave in $x \geq 0$ for any $\sigma \neq 0$. Furthermore, $\mathbf{h}_a$ and $\mathbf{h}_b$ are the only global minima for the optimization problem

$$\begin{aligned} \min \quad & \log^{-1}\left(1 + \frac{N}{\sigma^2 + x}\right) + \log^{-1}\left(1 + \frac{N}{\sigma^2 + y}\right) \\ \text{s.t.} \quad & x + y = 1, \; x \geq 0, \; y \geq 0, \end{aligned}$$

where $N > 0$.

To establish Claim 1, we first show strict concavity of $r(x) = \log^{-1}\left(1 + \left(\sigma^2 + x\right)^{-1}\right)$. Since

$$r'(x) = \frac{r^2(x)}{(1+\sigma^2+x)\,(\sigma^2+x)},$$
$$r''(x) = \frac{2r^2(x)\left(r(x) - \left(1/2+\sigma^2+x\right)\right)}{\left(1+\sigma^2+x\right)^2\left(\sigma^2+x\right)^2},$$

it follows that $r''(x) < 0$ is equivalent to

$$g(x) = \log\left(1 + \left(\sigma^2+x\right)^{-1}\right) - \left(1/2+\sigma^2+x\right)^{-1} > 0.$$

Let $z = 1/(\sigma^2 + x)$ and consider $h(z) = \log(1+z) - 2z/(2+z)$. Due to

$$h(0) = 0 \quad \text{and} \quad h'(z) = \frac{z^2}{(1+z)(z+2)^2} > 0,\ \forall\ z > 0,$$

it follows that $g(x) = h(z) > 0$ for all $x \geq 0$, implying strict concavity of $r(x)$ over the interval $(0, \infty)$. Since affine transformation does not change strict concavity of a function, this implies that $\log^{-1}\left(1 + N/\left(\sigma^2+x\right)\right)$ is also strictly concave for $x \geq 0$. Finally, since the minimum of a strictly concave function over a polytope is always attained at a vertex [44], we have established the claim.

To establish Lemma 3.2, let us consider the following parametric optimization problem in $\mathbb{R}^2$

$$\begin{aligned}
\min \quad & \frac{1}{r_4} + \frac{1}{r_3} + \frac{1}{r_2} + \frac{1}{r_1} \\
\text{s.t.} \quad & r_4 = \log\left(1 + 0.9|(1,1)\mathbf{v}|^2\right), \\
& r_3 = \log\left(1 + 1/|(1,1)\mathbf{v}|^2\right), \\
& r_2 = \log\left(1 + 100/|(1,0)\mathbf{v}|^2\right), \\
& r_1 = \log\left(1 + 100/|(0,1)\mathbf{v}|^2\right), \\
& \|\mathbf{v}\| = t,
\end{aligned} \tag{18}$$

where $\mathbf{v} = (v_1, v_2)^T$ and $t \geq 0$ is the parameter.

Claim 2: Let $f(t)$ denote the minimum value of (18). The following properties hold:

1) $f(t)$ is a strictly decreasing function in $[0, 1]$,
2) When $t \in (0.95, 1]$, the global minima of (18) are $(t, 0)^T$ and $(0, t)^T$,
3) $f'(t)$ is an increasing function in $[0.95, 1]$.

Let us argue that Claim 2 is true. First, if part 2) is true, then we can check part 3) directly by first computing $f(t)$ as the objective value of (18) at the solutions $(t, 0)^T$ or $(0, t)^T$, and then verifying that $f''(t) > 0$ for $t \in [0.95, 1]$. We omit the details of computations for space reason. So we only need to argue parts 1) and 2).

When t is fixed, the KKT condition for problem (18) can be written as

$$\left\{\begin{array}{rcl}\phi(v_1,v_2)+\psi(v_1) &=& \lambda v_1,\\ \phi(v_1,v_2)+\psi(v_2) &=& \lambda v_2,\\ v_1^2+v_2^2 &=& t^2,\end{array}\right. \tag{19}$$

where $\phi(x,y)$ and $\psi(x)$ are given as

$$\left\{\begin{array}{rcl}\phi(x,y) &=& \dfrac{-1.8\,(x+y)}{\log^2\left(1+0.9\,(x+y)^2\right)\left(1+0.9\,(x+y)^2\right)}+\dfrac{2}{\log^2\left(1+1/\,(x+y)^2\right)\left((x+y)^2+1\right)(x+y)},\\ \psi(x) &=& \dfrac{200}{\log^2\left(1+100/x^2\right)\left(100+x^2\right)x},\end{array}\right. \tag{20}$$

and λ is the Lagrangian multiplier associated with the constraint $\|\mathbf{v}\|=t$ in (18). It can be seen that $\sqrt{2}/2(t,t)^T$ is always a solution to (19). Moreover, $(t,0)^T$ and $(0,t)^T$ are the (only) non-differentiable points of (18). Suppose $(\bar{v}_1(t),\bar{v}_2(t))^T$, where $\bar{v}_1(t)\neq\bar{v}_2(t)$, are other KKT solutions to (19) (if there are any). Notice that the global minimum $f(t)$ is always attained at a KKT point or a non-differentiable point of (18). Hence,

$$f(t)=\min_{\{A,B,C\}}\{f_A(t),f_B(t),f_C(t)\},$$

where $f_A(t)$, $f_B(t)$ and $f_C(t)$ are the objective value of (18) at $\sqrt{2}/2(t,t)^T$, $(t,0)^T$ and $(\bar{v}_1(t),\bar{v}_2(t))^T$, respectively.

Next, we claim that $f(t)$ is a decreasing function in $t\in[0,1]$. We prove this by examining the monotonicity of the component functions $f_A(t)$, $f_B(t)$ and $f_C(t)$. By (19), we have

$$\lambda=\frac{\psi(\bar{v}_1(t))-\psi(\bar{v}_2(t))}{\bar{v}_1(t)-\bar{v}_2(t)}<0,$$

where the last step follows from the fact that $\psi(x)$ is a decreasing function for $x\in[0,1]$. Therefore, we conclude from the standard sensitivity analysis [40] of duality multipliers that $f_C(t)$ decreases as t increases in $[0,1]$. For $t\in[0,0.75]$, it can be checked that all of $f_i(t), i=A,B,C$, are decreasing functions. Thus, $f(t)$ decreases monotonically in $[0,0.75]$. For $t\in(0.75,0.85)$, the global minimizer of (18) is neither $\sqrt{2}/2(t,t)^T$ nor $(t,0)^T$ (or $(0,t)^T$) as we always can find a point at which the objective function is smaller than $f_A(t)$ and $f_B(t)$. As a result, $f(t)=f_C(t)$ decreases in $t\in(0.75,0.85)$. For $t\in[0.85,1]$, we have known $f_B(t)\leq f_A(t)$, so $f(t)=\min\{f_B(t),f_C(t)\}$. As both of $f_B(t)$ and $f_C(t)$ are decreasing functions in $t\in[0.85,1]$, we know $f(t)$ decreases in $[0.85,1]$. This completes the proof that $f(t)$ is monotonically decreasing in $[0,1]$.

We next prove part 2) of Claim 2: namely for $t\in[0.95,1]$, the global minima of (18) are $(t,0)^T$ and $(0,t)^T$. We can use $v_2=\sqrt{t^2-v_1^2}$ to transform (18) into an unconstrained univariate optimization problem, where $-t \leq v_1 \leq t$. The global minimizer $(v_1, v_2)^T$ of (18) should satisfy that $v_1 v_2 \geq 0$, else we can find $(\bar{v}_1, \bar{v}_2)^T$ such that $\bar{v}_1 + \bar{v}_2 = v_1 + v_2$, $|\bar{v}_1| < |v_1|$ and $|\bar{v}_2| < |v_2|$, at which the objective would be lower. Thus, we only need to consider the case $v_1 v_2 \geq 0$. Due to symmetry of (18), we only consider the case $0 \leq v_1 \leq \sqrt{2}t/2$. It can be checked that for any $t \in [0.95, 1]$, the objective function of (18) increases as v_1 increases in $[0, \sqrt{2}t/2]$. Hence, $(0, t)^T$ and $(t, 0)^T$ are the global minima for (18) with $t \in [0.95, 1]$.

Claim 3: When $N \geq 2(\mathrm{e}^{200m} - 1)$, the global minima $\mathbf{v}_{4i}^*$ of (6) must have unit-norm $\|\mathbf{v}_{4i}^*\| = 1, i = 1, 2, ..., n$.

By symmetry, we only need to prove $\|\mathbf{v}_4^*\| = 1$. Let us consider parametric optimization problem (21) in $\mathbf{v}_4$ with the other variables $\mathbf{v}_{4i}$, $i = 2, 3, ..., n$, fixed, and t as the parameter.

$$\begin{aligned}
\min_{\mathbf{v}_4} \quad & \left(\frac{1}{r_4} + \frac{1}{r_3} + \frac{1}{r_2} + \frac{1}{r_1}\right) + \sum_{j=1}^{m} \left(\frac{1}{r_{4n+2j}} + \frac{1}{r_{4n+2j-1}}\right) \\
\text{s.t.} \quad & r_4 = \log\left(1 + 0.9|(1,1)\mathbf{v}_4|^2\right), \ r_3 = \log\left(1 + 1/|(1,1)\mathbf{v}_4|^2\right), \\
& r_2 = \log\left(1 + 100/|(1,0)\mathbf{v}_4|^2\right), \ r_1 = \log\left(1 + 100/|(0,1)\mathbf{v}_4|^2\right), \ \|\mathbf{v}_4\| = t, \\
& r_{4n+2j} = \log\left(1 + N/\left(1 + |\mathbf{h}_{\pi(j)}^\dagger \mathbf{v}_{4|\pi(j)|}|^2 + |\mathbf{h}_{\tau(j)}^\dagger \mathbf{v}_{4|\tau(j)|}|^2\right)\right), \ 1 \leq j \leq m, \\
& r_{4n+2j-1} = \log\left(1 + N/\left(1 + |\mathbf{h}_{-\pi(j)}^\dagger \mathbf{v}_{4|\pi(j)|}|^2 + |\mathbf{h}_{-\tau(j)}^\dagger \mathbf{v}_{4|\tau(j)|}|^2\right)\right), \ 1 \leq j \leq m.
\end{aligned} \tag{21}$$

Let $g(t)$ be the optimum value of (21) and $g_1(t)$, $g_2(t)$ denote respectively the values of

$$\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \frac{1}{r_4} \quad \text{and} \quad \sum_{j=1}^{m} \left(\frac{1}{r_{4n+2j}} + \frac{1}{r_{4n+2j-1}}\right)$$

when evaluated at the global minima of (21). By definition, $f(t) \leq g_1(t)$ for all $t \in [0, 1]$. We claim that $f(t) = g_1(t)$ for $t \in [0.95, 1]$. To verify this, we consider optimization problem

$$\begin{aligned}
\min_{\mathbf{v}_4} \quad & \sum_{j=1}^{m} \left(\frac{1}{r_{4n+2j}} + \frac{1}{r_{4n+2j-1}}\right) \\
\text{s.t.} \quad & \|\mathbf{v}_4\| = t, \\
& r_{4n+2j} = \log\left(1 + N/\left(1 + |\mathbf{h}_{\pi(j)}^\dagger \mathbf{v}_{|4\pi(j)|}|^2 + |\mathbf{h}_{\tau(j)}^\dagger \mathbf{v}_{|4\tau(j)|}|^2\right)\right), \ 1 \leq j \leq m, \\
& r_{4n+2j-1} = \log\left(1 + N/\left(1 + |\mathbf{h}_{-\pi(j)}^\dagger \mathbf{v}_{4|\pi(j)|}|^2 + |\mathbf{h}_{-\tau(j)}^\dagger \mathbf{v}_{4|\tau(j)|}|^2\right)\right), \ 1 \leq j \leq m.
\end{aligned} \tag{22}$$

Define $|\mathbf{h}_a^\dagger \mathbf{v}_4|^2 = x$, then $|\mathbf{h}_b^\dagger \mathbf{v}_4|^2 = t^2 - x$, and the objective function of (22) is only dependent on x, while its constraint $\|\mathbf{v}_4\| = t$ can be transformed to $0 \leq x \leq t^2$. Using Claim 1, we see that the objective function of (22) is strictly concave in x. As a result, the globally optimal x can only be attained at the end points of the constraint interval $[0, t^2]$. In other words, the global minimizer of (22) must be either

$(t, 0)^T$ or $(0, t)^T$. Since the global minima of (18) are also $(t, 0)^T$ and $(0, t)^T$ when $t \in [0.95, 1]$, it follows that the global minimizer of (21) is either $(t, 0)^T$ or $(0, t)^T$, if $t \in [0.95, 1]$. This implies that $f(t) = g_1(t)$ in $[0.95, 1]$.

Next, we prove that if $N \geq 2(\mathrm{e}^{200m} - 1)$, then $g(1) < g(t)$ for $t \in [0, 1)$. Due to the choice of N, it can be checked that

$$f'(1) + \frac{2m}{\log^2(1 + N/3)} \leq 0, \quad f(0.95) \geq f(1) + \frac{2m}{\log(1 + N/2)}.$$

Then, we have

1) For any $t \in [0.95, 1]$,

$$\begin{aligned} g'(t) &= g_1'(t) + g_2'(t) = f'(t) + g_2'(t) \\ &\leq f'(1) + g_2'(t) \\ &< f'(1) + \frac{2m}{\log^2(1 + N/3)} \leq 0. \end{aligned}$$

The second equality is due to the fact $f(t) = g_1(t)$ for $t \in [0.95, 1]$. The first inequality holds since $f'(t)$ is an increasing function in $[0.95, 1]$ (cf. part 3) of Claim 2), while the second inequality follows from

$$g_2'(t) < \frac{2m}{\log^2(1 + N/3)}, \text{ for all } t \in [0.95, 1].$$

The nonpositiveness of $g'(t)$ over $t \in [0.95, 1]$ implies $g(t) > g(1)$, for $t \in [0.95, 1)$.

2) For any $t \in [0, 0.95]$, we have

$$\begin{aligned} g(t) &> g_1(t) \geq f(t) \\ &\geq f(0.95) \geq f(1) + \frac{2m}{\log(1 + N/2)} \\ &\geq g_1(1) + g_2(1) = g(1), \end{aligned}$$

where the third inequality holds because $f(t)$ is a strictly decreasing function in $[0, 1]$ (cf. part 1) of Claim 2) and the last inequality is due to $2m/\log(1 + N/2) \geq g_2(1)$ and $f(t) = g_1(t)$ for $t \in [0.95, 1]$.

Combining the above two steps shows $g(t) > g(1)$ for all $t \in [0, 1)$. It follows that the minimum of $g(t)$ over $[0, 1]$ is attained at $t = 1$. This establishes Claim 3.

Finally, notice that Claim 3 implies that (6) is equivalent to (21) with $t = 1$, when $N \geq 2(\mathrm{e}^{200m} - 1)$. Since the global minimizer of (21) must be either $(0, t)^T$ or $(t, 0)^T$ for $t \in [0.95, 1]$, it follows that the solutions $\mathbf{v}_{4i}^*, i = 1, 2, ..., n$, to (6) should be either $\mathbf{h}_a$ or $\mathbf{h}_b$ when $N \geq 2(\mathrm{e}^{200m} - 1)$.

APPENDIX C
PROOF OF LEMMA 3.3

Let $r(x) = \log(1 + 1/(b + x))$, then we have $f(x) = \log r(x)$, and

$$r'(x) = -1/((b+1+x)(b+x)),\ \ f'(x) = r'(x)/r(x),$$

$$f''(x) = (g(x) - 1)/\big((b+1+x)^2(b+x)^2 r^2(x)\big),$$

where $g(x) = (2b + 2x + 1)r(x)$ and $b = \sigma^2 \geq 0$. Let $y = x + b$, then $g(x)$ becomes $h(y) = (2y + 1)\log(1 + 1/y)$. It suffices to prove that $h(y) \geq 1$ for all $y \geq 0$. Since

$$h'(y) = 2\log(1 + 1/y) - (2y+1)/(y(y+1)),$$

$$h''(y) = 1/\big(y^2(y+1)^2\big) > 0,$$

we know $h'(y)$ is increasing. Since $\lim_{y\to+\infty} h'(y) = 0$, it follows that $h'(y) \leq 0$ and $h(y)$ is a decreasing function. Notice that $\lim_{y\to+\infty} h(y) = 2$. Thus, we have $h(y) \geq 2 > 1$ for all $y \geq 0$. This further implies that $\log\log(1 + 1/(b + x))$ is strictly convex for $x \geq 0$.

APPENDIX D
CONVERGENCE OF THE INEXACT CYCLIC COORDINATE DESCENT ALGORITHM

We first need to estimate the step length.

Claim 1: Suppose $c_1 \in (0, 1)$, $\nabla h(\mathbf{y}^i)^T\mathbf{d}^i < 0$ and $\max\limits_{\mathbf{y}\in Y} \|\nabla^2 h(\mathbf{y})\| \leq B$, where $h(\mathbf{y})$ is a multi-variable function associated with $\mathbf{y}$ and Y denotes the feasible region. Consider the backtracking line search [40] whereby $\alpha_i := \gamma^\ell$, with $\gamma \in (0, 1)$ and $\ell \geq 0$ being the smallest integer satisfying

$$h(\mathbf{y}^i + \gamma^\ell\mathbf{d}^i) \leq h(\mathbf{y}^i) + c_1\gamma^\ell\nabla h(\mathbf{y}^i)^T\mathbf{d}^i. \tag{23}$$

Then we have

$$1 \geq \alpha_i \geq \min\left\{1, \frac{2\gamma(c_1-1)\nabla h(\mathbf{y}^i)^T\mathbf{d}^i}{B\|\mathbf{d}^i\|^2}\right\}, \tag{24}$$

and

$$\begin{aligned} & h(\mathbf{y}^i) - h(\mathbf{y}^i + \alpha_i\mathbf{d}^i) \\ \geq & \min\left\{-c_1\nabla h(\mathbf{y}^i)^T\mathbf{d}^i,\ \frac{2\gamma c_1(1-c_1)\left(\nabla h(\mathbf{y}^i)^T\mathbf{d}^i\right)^2}{B\|\mathbf{d}^i\|^2}\right\}. \end{aligned} \tag{25}$$

Let us argue Claim 1 holds . Suppose that the step $\alpha_i = 1$ is not accepted. In this case, α_i will be the largest step satisfying the sufficient decrease condition (23), implying

$$h(\mathbf{y}^i + \alpha_i\gamma^{-1}\mathbf{d}^i) > h(\mathbf{y}^i) + c_1\alpha_i\gamma^{-1}\nabla h(\mathbf{y}^i)^T\mathbf{d}^i. \tag{26}$$

By Taylor expansion, we have

$$\begin{aligned} h(\mathbf{y}^i + \alpha_i\gamma^{-1}\mathbf{d}^i) = & \; h(\mathbf{y}^i) + \alpha_i\gamma^{-1}\nabla h(\mathbf{y}^i)^T\mathbf{d}^i + \frac{\alpha_i^2\gamma^{-2}}{2}\mathbf{d}^{iT}\nabla^2 h(\xi)\mathbf{d}^i \\ \leq & \; h(\mathbf{y}^i) + \alpha_i\gamma^{-1}\nabla h(\mathbf{y}^i)^T\mathbf{d}^i + \frac{\alpha_i^2\gamma^{-2}}{2}B\|\mathbf{d}^i\|^2, \end{aligned} \tag{27}$$

where $\xi = \mathbf{y}^i + s\alpha_i\gamma^{-1}\mathbf{d}^i$ with certain $s \in (0,1)$. Combining (26) and (27) yields (24). Substituting (24) into (23), we immediately obtain (25), which establishes Claim 1.

We now proceed with the proof of Theorem 4.1. The basic idea is based on the contradiction principle. More exactly, if there is no convergence, we can find one descent direction which can provide a sufficient decrease in the objective function and then obtain a contradiction.

At first, since the iterates $\{\mathbf{x}^i\}$ lie in a compact set X, there must exist an accumulation point for $\{\mathbf{x}^i\}$. Let $\bar{\mathbf{x}}$ denote an accumulation point such that

$$\bar{\mathbf{x}} = \lim_{i\in\mathcal{I}_0, i\to\infty} \mathbf{x}^i$$

for some subsequence indexed by $\mathcal{I}_0$. Since the feasible set X is closed, $\bar{\mathbf{x}}$ must also be feasible. Furthermore, since the projection mapping and the function f are both continuous, it follows that

$$\begin{aligned} \lim_{i\in\mathcal{I}_0, i\to\infty} \mathbf{d}_1^{i+1} &= \lim_{i\in\mathcal{I}_0, i\to\infty} P_{X_1}(\mathbf{x}_1^i - \nabla_{\mathbf{x}_1} f(\mathbf{z}_0^{i+1})) - \mathbf{x}_1^i \\ &= P_{X_1}[\bar{\mathbf{x}}_1 - \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}})] - \bar{\mathbf{x}}_1 \triangleq \bar{\mathbf{d}}_1 \end{aligned}$$

and

$$\lim_{i\in\mathcal{I}_0, i\to\infty} f(\mathbf{x}^i) = f(\bar{\mathbf{x}}).$$

Notice that the function values $\{f(\mathbf{x}^i)\}$ are decreasing and bounded below, then $f(\mathbf{x}^i) \to f(\bar{\mathbf{x}})$. Since $f(\mathbf{x})$ is twice continuously differentiable and the feasible set X is bounded, it follows that

$$B = \max_{k=1,2,\ldots,K} \max_{\mathbf{x}_k \in X_k} \|\nabla^2_{\mathbf{x}_k} f(\mathbf{x})\| < \infty.$$

Because the projection operator is non-expansive and $\nabla f(\mathbf{x})$ is continuous in a bounded region, we obtain from (14) that

$$\begin{aligned} \|\mathbf{d}_k^{i+1}\| &= \|P_{X_k}[\mathbf{x}_k^i - \nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})] - P_{X_k}[\mathbf{x}_k^i]\| \\ &\leq \|[\mathbf{x}_k^i - \nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})] - \mathbf{x}_k^i\| = \|\nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})\|. \end{aligned}$$

Denoting $\max_{\mathbf{x}\in X} \|\nabla f(\mathbf{x})\| \triangleq M$, hence we have

$$\|\mathbf{d}_k^{i+1}\| \leq \|\nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})\| \leq M < +\infty. \tag{28}$$

The lower and upper bounds on $\|\nabla_{\mathbf{x}_k} f(\mathbf{z}_{k-1}^{i+1})\|$ will be useful in estimating the decrease in the objective function.

We proceed by contradiction and suppose $\bar{\mathbf{x}}$ is not a KKT point. Then $\bar{\mathbf{d}} = P_X[\bar{\mathbf{x}} - \nabla f(\bar{\mathbf{x}})] - \bar{\mathbf{x}} \neq 0$ so that $\delta = \|\bar{\mathbf{d}}\| > 0$. Let

$$k^* = \min\{k \mid \delta_k = \|\bar{\mathbf{d}}_k\| > 0\},$$

and suppose $k^* > 1$ without loss of generality. By definition, we have

$$\lim_{i\in\mathcal{I}_0, i\to\infty} \|\mathbf{d}_k^{i+1}\| = \delta_k = 0, \text{ for } k < k^*.$$

Recall the definition (15) of $\mathbf{z}_k^{i+1}$. Since

$$\begin{aligned} \|\mathbf{z}_1^{i+1} - \bar{\mathbf{x}}\| &= \|\mathbf{x}_1^i + \alpha_1^{i+1}\mathbf{d}_1^{i+1} - \bar{\mathbf{x}}_1\| \\ &\leq \|\mathbf{x}_1^i - \bar{\mathbf{x}}_1\| + \alpha_1^{i+1}\|\mathbf{d}_1^{i+1}\| \\ &\leq \|\mathbf{x}_1^i - \bar{\mathbf{x}}_1\| + \|\mathbf{d}_1^{i+1}\| \underset{i\in\mathcal{I}_0, i\to\infty}{\longrightarrow} 0, \end{aligned}$$

we have $\lim_{i\in\mathcal{I}_0, i\to\infty} \mathbf{z}_1^{i+1} = \bar{\mathbf{x}}$. In general, the same argument shows that

$$\lim_{i\in\mathcal{I}_0, i\to\infty} \mathbf{z}_k^{i+1} = \bar{\mathbf{x}}, \ \forall\ k < k^*.$$

Consequently, there holds

$$\begin{aligned} &\lim_{i\in\mathcal{I}_0, i\to\infty} \nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1}) = \nabla_{\mathbf{x}_{k^*}} f(\bar{\mathbf{x}}), \\ &\lim_{i\in\mathcal{I}_0, i\to\infty} \mathbf{d}_{k^*}^{i+1} = \bar{\mathbf{d}}_{k^*}. \end{aligned} \tag{29}$$

Let us use $\bar{\theta}_{k^*}$ to denote the angle between $\bar{\mathbf{d}}_{k^*}$ and $\nabla_{\mathbf{x}_{k^*}} f(\bar{\mathbf{x}})$. Since $\bar{\mathbf{y}}_{k^*} = \bar{\mathbf{x}}_{k^*} - \nabla_{\mathbf{x}_{k^*}} f(\bar{\mathbf{x}}) \notin X_{k^*}$, it follows from the property of projection that

$$(\bar{\mathbf{x}}_{k^*} - P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}])^T (\bar{\mathbf{y}}_{k^*} - P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}]) \leq 0,$$

which further implies

$$\begin{aligned} & \|P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}] - \bar{\mathbf{x}}_{k^*}\|^2 \\ \leq\ & (\bar{\mathbf{x}}_{k^*} - P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}])^T(\bar{\mathbf{x}}_{k^*} - \bar{\mathbf{y}}_{k^*}) \\ =\ & \|P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}] - \bar{\mathbf{x}}_{k^*}\|\|\nabla_{\mathbf{x}_{k^*}} f(\bar{\mathbf{x}})\| \cos\bar{\theta}_{k^*}. \end{aligned}$$

Canceling the factor $\|P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}] - \bar{\mathbf{x}}_{k^*}\| \neq 0$ from both sides yields

$$\begin{aligned} \delta_{k^*} = \|\bar{\mathbf{d}}_{k^*}\| &= \|P_{X_{k^*}}[\bar{\mathbf{y}}_{k^*}] - \bar{\mathbf{x}}_{k^*}\| \\ &\leq \|\nabla_{\mathbf{x}_{k^*}} f(\bar{\mathbf{x}})\| \cos\bar{\theta}_{k^*} \\ &\leq M \cos\bar{\theta}_{k^*}, \end{aligned}$$

where the last inequality follows from (28). Thus, we have

$$\cos\bar{\theta}_{k^*} \geq \frac{\delta_{k^*}}{M}. \tag{30}$$

Now we can use (28), (29) and (30) to conclude

$$\|\nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1})\| \geq \|\mathbf{d}_{k^*}^{i+1}\| \geq \frac{\delta_{k^*}}{2} \quad \text{and} \quad \cos\theta_{k^*}^{i+1} \geq \frac{\delta_{k^*}}{2M},$$

for all $i \in \mathcal{I}_0$ and $i \geq i_0$, where i_0 is a sufficiently large integer and $\theta_{k^*}^{i+1}$ denotes the angle between $\nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1})$ and $\mathbf{d}_{k^*}^{i+1}$.

Now we can use (23)-(25) of Claim 1 to obtain a contradiction. In particular, we consider

$$\begin{aligned}
+\infty \;&>\; \sum_{i}\left(f(\mathbf{x}^i)-f(\mathbf{x}^{i+1})\right) \geq \sum_{i\in\mathcal{I}_0}\left(f(\mathbf{x}^i)-f(\mathbf{x}^{i+1})\right) \geq \sum_{i\in\mathcal{I}_0}\left(f(\mathbf{z}_{k^*-1}^{i+1})-f(\mathbf{z}_{k^*}^{i+1})\right)\\
&\geq \sum_{i\in\mathcal{I}_0}\left(\frac{2\gamma c_1(1-c_1)\left(\nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1})^T\mathbf{d}_{k^*}^{i+1}\right)^2}{B\|\mathbf{d}_{k^*}^{i+1}\|^2}\right)\\
&= \sum_{i\in\mathcal{I}_0}\left(\frac{2\gamma c_1(1-c_1)\left(\|\nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1})\|\|\mathbf{d}_{k^*}^{i+1}\|\cos\theta_{k^*}^{i+1}\right)^2}{B\|\mathbf{d}_{k^*}^{i+1}\|^2}\right)\\
&= \sum_{i\in\mathcal{I}_0}\left(\frac{2\gamma c_1(1-c_1)\left(\|\nabla_{\mathbf{x}_{k^*}} f(\mathbf{z}_{k^*-1}^{i+1})\|\cos\theta_{k^*}^{i+1}\right)^2}{B}\right)\\
&\geq \sum_{i\geq i_0,\; i\in\mathcal{I}_0}\left(\frac{\gamma c_1(1-c_1)\delta_{k^*}^4}{8BM^2}\right) = +\infty,
\end{aligned} \tag{31}$$

which is a contradiction. The fourth inequality is due to Claim 1. Therefore, $\bar{\mathbf{x}}$ is a stationary point.

References

[1] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57-73, Feb. 2008.

[2] A. Wiesel, Y. C. Eldar, and S. S. Shitz, "Linear precoding via conic programming for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161-176, Jan. 2006.

[3] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmitter beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437-1450, Otc. 1998.

[4] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE Veh. Tech. Conf.*, Sept. 2002, vol. 1, pp. 87-91.

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

[6] Z.-Q. Luo and S. Zhang, "Duality gap estimation and polynomial time approximation for optimal spectrum management," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2675-2689, Jul. 2009.

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

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

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

[10] O. Somekh, O. Simeone, Y. Bar-Ness, and A. Haimovich, "Distributed multi-cell zero-forcing beamforming in cellular downlink channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2006, pp. 1-6.

[11] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 56-61, Aug. 2006.

[12] G. Foschini, M. Karakayali, and R. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," in *IEE Proc. Commun.*, Aug. 2006, vol. 153, no. 4, pp. 548-555.

[13] B. O. Lee, H. W. Je, I. Sohn, O.-S. Shin, and K. B. Lee, "Interference-aware decentralized precoding for multicell MIMO TDD systems," in *Proc. IEEE GLOBECOM*, Nov. 2008, pp. 1-5.

[14] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Distributed multicell and multiantenna precoding: Characterization and performance evaluation," in *Proc. IEEE GLOBECOM*, Nov. 2009, pp. 1-6.

[15] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059-1069, Sept. 2008.

[16] B. L. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491-5499, Dec. 2008.

[17] D. Schmidt, C. Shi, R. Berry, M. 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, Sept. 2009.

[18] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912-1921, Aug. 2003.

[19] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361-374, Jan. 2006.

[20] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646-2660, Jun. 2007.

[21] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Ann. Telecommun.*, vol. 63, pp. 253-269, May 2008.

[22] S. Jing, D. Tse, J. Soriaga, J. Hou, J. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, no. 5, Jan. 2008, article ID 586878.

[23] X. Shang, B. Chen, and H. V. Poor, "Multi-user MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region," submitted to *IEEE Trans. Inf. Theory*, *http://arxiv.org/abs/0907.0505,* Apr. 2009.

[24] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," submitted to *IEEE Trans. Signal Process.*, *http://arxiv.org/abs/1004.4492,* Apr. 2010.

[25] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of MIMO interference channels in the low interference regime," submitted to *IEEE Trans. Inf. Theory*, *http://arxiv.org/abs/0909.2074,* Sept. 2009.

[26] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization.* Philadelphia, U.S.A.: SIAM-MPS Series on Optimization, SIAM Publications, 2001.

[27] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* San Francisco, U.S.A.: W. H. Freeman and Company, 1979.

[28] C. H. Papadimitriou, *Computational Complexity.* Mass., U.S.A.: Addison-Wesley, 1994.

[29] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Noisy-interference sum-rate capacity of parallel Gaussian interference channel," submitted to *IEEE Trans. Inf. Theory*, *http://arxiv.org/abs/0903.0595,* Mar. 2009.

[30] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing MSE on multiuser downlink: Optimality, fast algorithms and equivalence via max-min SIR," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 2669-2673.

[31] C. W. Tan, M. Chiang, and R. Srikant, "Fast algorithms and performance bounds for sum rate maximization in wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 1350-1358.

[32] S. Friedland and C. W. Tan, "Maximizing sum rates in Gaussian interference-limited channels," submitted to *Numer. Anal.*, *http://arxiv.org/abs/0806.2860v2,* Jul. 2008.

[33] C. W. Tan, S. Friedland, and S. H. Low, "Spectrum management in multiuser cognitive wireless networks: Optimality and algorithms," to appear in *IEEE J. Sel. Areas Commun.*, *http://www.math.uic.edu/ friedlan/CR_TanFriedlandLowJSAC.pdf.*

[34] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Mass. U.S.A.: Athena Scientific, 1997.

[35] M. J. D. Powell, "On search direction for minimization algorithms," *Math. Program.*, vol. 4, no. 1, pp. 193-201, Jan. 1973.

[36] J. C. Bezdek, R. J. Hathaway, R. E. Howard, C. A. Wilson, and M. P. Windham, "Local convergence analysis of a grouped variable version of coordinate descent," *J. Optim. Theory and Appl.*, vol. 54, no. 3, pp. 471-477, Sept. 1987.

[37] M. V. Solodov, "On the convergence of constrained parallel variable distribution algorithms," *SIAM J. Optim.*, vol. 8, no. 1, pp. 187-196, Feb. 1998.

[38] Z.-Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Oper. Theory and Appl.*, vol. 72, no. 1, pp. 7-35, Jan. 1992.

[39] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460-473, Jul. 1972.

[40] J. Nocedal and S. J. Wright, *Numerical Optimization*. Germany: Springer, 1999.

[41] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbirarily correlated Rician MIMO channels with Rician ditribubance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807-1820, Mar. 2010.

[42] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141-148, 1988.

[43] Y.-H. Dai and L.-Z. Liao, "R-linear convergence of the Barzilai and Borwein gradient method," *IMA J. Numer. Anal.*, vol. 22, no. 1, pp. 1-10, Jan. 2002.

[44] H. L. Hoffman, "A method for globally minimizing concave functions over convex sets," *Math. Program.*, vol. 20, no. 1, pp. 22-32, Dec. 1981.