2013

Linear transceiver design for a MIMO interfering broadcast channel achieving max–min fairness

Meisam Razaviyayn
Mingyi Hong
Zhi-Quan Luo

Linear Transceiver Design for a MIMO Interfering Broadcast Channel Achieving Max-Min Fairness

Meisam Razaviyayn, Mingyi Hong, and Zhi-Quan Luo

Abstract

We consider the problem of linear transceiver design to achieve max-min fairness in a downlink MIMO multicell network. This problem can be formulated as maximizing the minimum rate among all the users in an interfering broadcast channel (IBC). In this paper we show that when the number of antennas is at least two at each of the transmitters and the receivers, the min rate maximization problem is NP-hard in the number of users. Moreover, we develop a low-complexity algorithm for this problem by iteratively solving a sequence of convex subproblems, and establish its global convergence to a stationary point of the original minimum rate maximization problem. Numerical simulations show that this algorithm is efficient in achieving fairness among all the users.

I. Introduction

We consider the linear transceiver design problem in a MIMO-IBC, in which a set of Base Stations (BSs) send data to their intended users. Both the BSs and the users are equipped with multiple antennas, and they share the same time/frequency resource for transmission. The objective is to maximize the minimum rate among all the users in the network, in order to achieve network-wide fairness.

Providing max-min fairness has long been considered as an important design criterion for wireless networks. Hence various algorithms that optimize the min-rate utility in different network settings have been proposed in the literature. References [20], [21] are early works that studied the max-min signal to interference plus noise ratio (SINR) power control problem and a related SINR feasibility problem in a scalar interference channel (IC). It was shown in [20], [21] that for randomly generated scalar ICs, with probability one there exists a unique optimal solution to the max-min problem. The proposed algorithm with an additional binary search can be used to solve the max-min fairness problem efficiently. Recently reference [17] derived a set of algorithms based on nonlinear Perron-Frobenius theory for the same network setting. Differently from [20], [21], the proposed algorithms can also deal with individual users' power constraints.

Apart from the scalar IC case, there have been many published results [1], [3], [4], [8], [14], [18], [19] on the min rate maximization problem in a multiple input single output (MISO) network, in which the BSs are equipped with multiple antennas and the users are only equipped with a single antenna. Reference [19] utilized the nonnegative matrix theory to study the related power control problem when the beamformers are known and fixed. When optimizing the transmit power and the beamformers jointly,

M. Razaviyayn, M. Hong, and Z.-Q. Luo are with the Department of Electrical and Computer Engineering University of Minnesota, Minneapolis, MN 55455, USA

This research is supported in part by a research gift from Huawei Technologies Inc, and by the Army Research Office, grant number W911NF-09-1-0279.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

the corresponding min-rate utility maximization problem is non-convex. Despite the lack of convexity, the authors of [1] showed that a semidefinite relaxation is tight for this problem, and the optimal solution can be constructed from the solution to a reformulated semidefinite program. Furthermore, the authors of [18] showed that this max-min problem can be solved by a sequence of second order cone programs (SOCP). Reference [14] identified an interesting uplink downlink duality property, in which the downlink min-rate maximization problem can be solved by alternating between a downlink power update and a uplink receiver update. In a related work [3], the authors made an interesting observation that in a single cell MISO network, the global optimum of this problem can be obtained by solving a (simpler) weighted sum inverse SINR problem with a set of appropriately chosen weights. However, this observation is only true when the receiver noise is negligible. The authors of [4] extended their early results [17] to the MISO setting with a single BS and multiple users. A fixed-point algorithm that alternates between power update and beamformer updates was proposed, and the nonlinear Perron-Frobenius theory was applied to prove the convergence of the algorithm.

Unlike the MISO case, the existing work on the max-min problem for MIMO networks is rather limited; see [4] and [9]. Both of these studies consider a MIMO network in which a *single stream* is transmitted for each user. In particular, the author of [9] showed that finding the global optimal solution for this problem is intractable (NP-hard) when the number of antennas at each transmitter/receiver is at least *three*. They then proposed an efficient algorithm that alternates between updating the transmit and the receive beamformers to find a local optimal solution. The key observation is that when the users' receive beamformers are fixed, finding the set of optimal transmit beamformers can be again reduced to a sequence of SOCP and solved efficiently. For more discussion of the max-min and its related resource allocation problems in interfering wireless networks, we refer the readers to a recent survey [7].

In this paper, we consider a MIMO interfering broadcast network whereby there are *multiple* users associated with each BS, and all the users and the BSs are equipped with multiple antennas. Such a setting is more general than those studied in all the works cited above. Moreover, we do not restrict the number of transmitted data streams for each user. Recent works that deal with linear transceiver design in this type of network include [5], [16]. However, these works aim at optimizing differentiable system utilities such as the weighted sum rate (WSR) utility that excludes the min-rate utility considered in this work. To the best of our knowledge, there is no known algorithm that can effectively compute a high quality solution for the max-min problem in the general context considered in this work.

The main contributions of this paper are summarized as follows. First, we show that in the considered general setting, when there are at least *two* antennas at each transmitters and the receivers, the min-rate maximization problem is NP-hard in the number of users. This result is a generalization of that presented in [9], in which the NP-hardness results require more than *three* antennas at the users and BSs. We further provide a reformulation of the original max-min problem by generalizing the framework developed in [16], and design an algorithm that computes an approximate solution to the max-min problem. The proposed algorithm has the following desirable features: *i)* it is computationally efficient, as in each step

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

a convex optimization problem whose solution can be obtained easily in a closed form; *ii)* it is guaranteed to converge to a stationary solution of the original problem.

The rest of the paper is organized as follows. In Section II, we provide the system model. In Section III, we give detailed analysis of the complexity of the considered problem. In Section IV–V, we reformulate the problem into an equivalent form, and propose an algorithm that iteratively optimizes the transformed problem. In Section VI, numerical experiments are provided.

Notations: For a symmetric matrix $\mathbf{X}$, $\mathbf{X} \succeq 0$ signifies that $\mathbf{X}$ is positive semi-definite. We use $\mathrm{Tr}(\mathbf{X})$, $|\mathbf{X}|$ and $\mathbf{X}^H$ to denote the trace, determinant and hermitian of a matrix, respectively. We use $\mathbb{R}^{N\times M}$ and $\mathbb{C}^{N\times M}$ to denote the set of real and complex $N \times M$ matrices. We use the notation j to denote the imaginary unit, with $j^2 = -1$. We use the expression: $0 \leq a \perp b \geq 0$ to indicate $a \geq 0, b \geq 0, a \times b = 0$.

II. System Model and Problem Formulation

Consider a cellular network with K cells. Let us assume that the cell assignment of the users has been done and each BS is interested in serving the users in its own cell. More precisely, we assume each BS k, $k = 1, 2, \ldots, K$, is equipped with M_k transmit antennas and serves I_k number of users in cell k. Let us use the notation i_k to denote the i-th user in cell k and N_{i_k} to denote the number of receive antennas of user i_k. Also define $\mathcal{I}$ and $\mathcal{I}_k$ to be the set of all users and the set of users in cell k, respectively:

$$\mathcal{I} = \{i_k \mid 1 \leq k \leq K, 1 \leq i \leq I_k\}, \quad \mathcal{I}_k = \{i_k \mid 1 \leq i \leq I_k\}.$$

Let $\mathcal{K}$ to be the set of all BSs $\mathcal{K} = \{1, 2, \ldots, K\}$. Throughout, we use i, m to denote the index for the users, and use k, ℓ to denote the index for the BSs.

For the standard linear channel model, the received signal of user i_k can be written as

$$\mathbf{y}_{i_k} = \underbrace{\mathbf{H}_{i_k k}\mathbf{x}_{i_k}}_{\text{desired signal}} + \underbrace{\sum_{m\neq i, m=1}^{I_k} \mathbf{H}_{i_k k}\mathbf{x}_{m_k}}_{\text{intracell interference}} + \underbrace{\sum_{\ell\neq k, \ell=1}^{K} \sum_{m=1}^{I_j} \mathbf{H}_{i_k \ell}\mathbf{x}_{m_\ell} + \mathbf{n}_{i_k}}_{\text{intercell interference plus noise}},$$

where $\mathbf{x}_{i_k} \in \mathbb{C}^{M_k\times 1}$ and $\mathbf{y}_{i_k} \in \mathbb{C}^{N_{i_k}\times 1}$ are respectively the transmitted and received signal of user i_k. The matrix $\mathbf{H}_{i_k j} \in \mathbb{C}^{N_{i_k}\times M_j}$ represents the channel response from transmitter j to receiver i_k, while $\mathbf{n}_{i_k} \in \mathbb{C}^{N_{i_k}\times 1}$ denotes the complex additive white Gaussian noise with distribution $\mathcal{CN}(0, \sigma_{i_k}^2\mathbf{I})$ at receiver i_k.

For practical considerations, we focus on optimal *linear* transmit and receive strategies that can maximize a system utility. Specifically, let BS k use a beamforming matrix $\mathbf{V}_{i_k}$ to send the signal vector $\mathbf{s}_{i_k}$ to receiver i_k, and suppose receiver i_k estimates the transmitted data vector $\mathbf{s}_{i_k}$ by using a linear beamforming matrix $\mathbf{U}_{i_k}$, i.e.,

$$\mathbf{x}_{i_k} = \mathbf{V}_{i_k}\,\mathbf{s}_{i_k}, \quad \hat{\mathbf{s}}_{i_k} = \mathbf{U}_{i_k}^H\mathbf{y}_{i_k}, \quad \forall\, i_k \in \mathcal{I},$$

where the data vector $\mathbf{s}_{i_k} \in \mathbb{C}^{d_{i_k}\times 1}$ is normalized so that $\mathbb{E}[\mathbf{s}_{i_k}\mathbf{s}_{i_k}^H] = \mathbf{I}$, and $\hat{\mathbf{s}}_{i_k}$ is the estimate of $\mathbf{s}_{i_k}$ at i-th receiver in cell k. $\mathbf{V}_{i_k} \in \mathbb{C}^{M_k\times d_{i_k}}$ and $\mathbf{U}_{i_k} \in \mathbb{C}^{N_{i_k}\times d_{i_k}}$ are respectively the transmit and receive

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

beamforming matrices used for serving the i-th user in cell k. Let $\mathbf{Q}_{i_k} \in \mathbb{R}^{M_k \times M_k}$, $\mathbf{Q}_{i_k} \triangleq \mathbf{V}_{i_k}\mathbf{V}_{i_k}^H$, denote the transmit covariance matrix for user i_k. Let $\mathbf{V} \triangleq \{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}$ and $\mathbf{Q} \triangleq \{\mathbf{Q}_{i_k}\}_{i_k \in \mathcal{I}}$.

The mean squared error (MSE) matrix for user i_k can be written as

$$\begin{aligned}\mathbf{E}_{i_k} &\triangleq \mathbb{E}_{\mathbf{s},\mathbf{n}}\left[(\hat{\mathbf{s}}_{i_k} - \mathbf{s}_{i_k})(\hat{\mathbf{s}}_{i_k} - \mathbf{s}_{i_k})^H\right] \\ &= (\mathbf{I} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k k}\mathbf{V}_{i_k})(\mathbf{I} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k k}\mathbf{V}_{i_k})^H \\ &+ \sum_{m_\ell \neq i_k} \mathbf{U}_{i_k}^H \mathbf{H}_{i_k \ell}\mathbf{V}_{m_\ell}\mathbf{V}_{m_\ell}^H \mathbf{H}_{i_k \ell}^H \mathbf{U}_{i_k} + \sigma_{i_k}^2 \mathbf{U}_{i_k}^H \mathbf{U}_{i_k}.\end{aligned} \tag{1}$$

Treating interference as noise, the rate of the i-th user in cell k is given by

$$R_{i_k} = \log\det\left(\mathbf{I} + \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H \mathbf{H}_{i_k k}^H \left(\sigma_{i_k}^2 \mathbf{I} + \sum_{m_\ell \neq i_k} \mathbf{H}_{i_k \ell}\mathbf{V}_{m_\ell}\mathbf{V}_{m_\ell}^H \mathbf{H}_{i_k \ell}^H\right)^{-1}\right) \tag{2}$$

$$= \log\det\left(\mathbf{I} + \mathbf{H}_{i_k k}\mathbf{Q}_{i_k}\mathbf{H}_{i_k k}^H \left(\sigma_{i_k}^2 \mathbf{I} + \sum_{m_\ell \neq i_k} \mathbf{H}_{i_k \ell}\mathbf{Q}_{m_\ell}\mathbf{H}_{i_k \ell}^H\right)^{-1}\right). \tag{3}$$

We will occasionally use the notations $R_{i_k}(\mathbf{V})$ (resp. $R_{i_k}(\mathbf{Q})$) to make their dependencies on $\mathbf{V}$ (resp. $\mathbf{Q}$) explicit.

The problem of interest is to find the transmit beamformers $\mathbf{V} = \{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}$ such that a utility of the system is maximized, while each BS k's power budget of the form $\sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k$ is satisfied. Note that P_k denotes the power budget of transmitter k. In this work, our focus is on the max-min utility function, i.e., we are interested in solving the following problem

$$\begin{aligned}\max_{\{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}} \quad & \min_{i_k \in \mathcal{I}} \quad R_{i_k}(\mathbf{V}) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k, \quad \forall\, k \in \mathcal{K}.\end{aligned} \tag{P}$$

Similar to [18], one can solve (P) by solving a series of problems of the following type for different values of γ:

$$\begin{aligned}\min_{\{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \\ \text{s.t.} \quad & R_{i_k}(\mathbf{V}) \geq \gamma, \quad \forall\, i_k \in \mathcal{I} \\ & \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k, \quad \forall\, k \in \mathcal{K}.\end{aligned} \tag{4}$$

The above problem is to minimize the total power consumption in the network subject to quality of service (QoS) constraints. In what follows, we first study the complexity status of problem (P) and (4). Then, we propose an efficient algorithm for designing the beamformers based on the maximization of the worst user performance in the system.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

III. NP-Hardness of Optimal Beamformer Design

In this section, we analyze the complexity status of problem (P) and (4). In the single input single output (SISO) case where $M_k = N_{i_k} = 1, \forall\ k \in \mathcal{K}, \forall\ i_k \in \mathcal{I}$, it has been shown that problem (P) and problem (4) can be solved in polynomial time, see [10] and the references therein. Furthermore, it is shown that in the multiple input single output (MISO) case where $M_k > N_{i_k} = 1, \forall\ k \in \mathcal{K}, \forall\ i_k \in \mathcal{I}$, both problems are still polynomial time solvable [1], [12]. In this section, we consider the MIMO case where $M_k \geq 2$, and $N_{i_k} \geq 2$. We show that unlike the above mentioned special cases, both problems (P) and (4) are NP-hard.

In fact, it is sufficient to show that for a *simpler* MIMO IC network with K transceiver pairs and with each node equipped with at least two antennas, solving the max-min problem (P) and the min-power problem (4) are both NP-hard. For convenience, we rewrite the max-min beamformer design problem in this K user MIMO IC as an equivalent[1] covariance maximization form

$$\begin{array}{ll} \max\limits_{(\lambda,\mathbf{Q})} & \lambda \\ \text{s.t.} & \lambda \leq R_k(\mathbf{Q}),\ \mathrm{Tr}(\mathbf{Q}_k) \leq 1, \mathbf{Q}_k \succeq 0, \quad \forall\ k = 1, \cdots, K. \end{array} \tag{5}$$

where $R_k(\mathbf{Q}) = \log\det\left(\mathbf{I} + \mathbf{H}_{kk}\mathbf{Q}_k\mathbf{H}_{kk}^H(\sigma_k^2\mathbf{I} + \sum_{j\neq k}\mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H)^{-1}\right)$. Note that λ is the slack variable that is introduced to represent the objective value of the problem. The first step towards proving the desired complexity result is to recognize certain special structures in the optimal solutions of the problem (5).

Let us consider a 3-user MIMO IC with two antennas at each node. Suppose $\sigma_k^2 = 1$ for all k and the channels are given as

$$\mathbf{H}_{ii} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \forall\ i = 1,2,3 \quad \text{and} \quad \mathbf{H}_{im} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}, \quad \forall\ i \neq m, \quad i,m = 1,2,3. \tag{6}$$

Our first result characterizes the global optimal solutions for problem (5) in this special network.

Lemma 1 *Suppose $K = 3$ and the channels are given as* (6). *Let $\mathcal{S} = \{(\lambda^*, \mathbf{Q}_1^*, \mathbf{Q}_2^*, \mathbf{Q}_3^*)\}$ denote the set of optimal solutions of the problem* (5). *Then $\mathcal{S}$ can be expressed as*

$$\mathcal{S} = \{(1, \mathbf{Q}_a^*, \mathbf{Q}_a^*, \mathbf{Q}_a^*), (1, \mathbf{Q}_b^*, \mathbf{Q}_b^*, \mathbf{Q}_b^*), (1, \mathbf{Q}_c^*, \mathbf{Q}_c^*, \mathbf{Q}_c^*), (1, \mathbf{Q}_d^*, \mathbf{Q}_d^*, \mathbf{Q}_d^*)\}, \tag{7}$$

where $\mathbf{Q}_a^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{Q}_b^* = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{Q}_c^* = \begin{bmatrix} 0.5 & 0.5j \\ -0.5j & 0.5 \end{bmatrix}$, *and* $\mathbf{Q}_d^* = \begin{bmatrix} 0.5 & -0.5j \\ 0.5j & 0.5 \end{bmatrix}$.

The proof of this lemma can be found in the Appendix A. Next we proceed to consider a 5-user interference channel with two antennas at each node. Again suppose $\sigma^2 = 1$ and the channels are given

[1]The equivalence is in the sense that for every optimal solution $\{\mathbf{V}^*\}$ of (P) with $M_k = d_k$, there exists $\lambda^* \geq 0$ so that by defining $\mathbf{Q}_k^* = \mathbf{V}_k^*\mathbf{V}_k^{*H}, \forall\ k$, the point $\{\lambda^*, \mathbf{Q}^*\}$ is an optimal solution of (5). Conversely, if $\{\lambda^*, \mathbf{Q}^*\}$ is an optimal solution of (5) and $\mathbf{Q}_k^* = \mathbf{V}_k^*\mathbf{V}_k^{*H}, \forall\ k$, then $\mathbf{V}^*$ is an optimal solution of (P).

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

as

$$\mathbf{H}_{ii} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \forall\, i = 1,2,3 \quad \text{and} \quad \mathbf{H}_{im} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}, \quad \forall\, i \neq m, \quad i,m = 1,2,3; \tag{8}$$

$$\mathbf{H}_{ii} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \forall\, i = 4,5, \quad \mathbf{H}_{4m} = \begin{bmatrix} 1 & j \\ 0 & 0 \end{bmatrix}, \forall\, m = 1,2,3, \quad \mathbf{H}_{5m} = \begin{bmatrix} j & 1 \\ 0 & 0 \end{bmatrix}, \forall\, m = 1,2,3; \tag{9}$$

$$\mathbf{H}_{im} = 0,\ \forall\, i = 1,2,3,\ \forall\, m = 4,5; \quad \mathbf{H}_{im} = 0,\ \forall\, i \neq m,\ i,m = 4,5. \tag{10}$$

Our next result characterizes the global optimal solutions for the problem (5) for this special case.

Lemma 2 *Suppose $K = 5$ and the channels are given as* (8)–(10). *Let $\mathbf{Q}_a^*,\ \mathbf{Q}_b^*$ be defined in Lemma 1. Denote the set of optimal solutions of the problem* (5) *as $\mathcal{T}$. Then $\mathcal{T}$ can be expressed as*

$$\mathcal{T} = \left\{(1, \mathbf{Q}_a^*, \mathbf{Q}_a^*, \mathbf{Q}_a^*, \mathbf{Q}_a^*, \mathbf{Q}_a^*), (1, \mathbf{Q}_b^*, \mathbf{Q}_b^*, \mathbf{Q}_b^*, \mathbf{Q}_b^*, \mathbf{Q}_b^*)\right\}. \tag{11}$$

Proof: First of all, it is not hard to see that by selecting each of the values in the optimal set $\mathcal{T}$, we get the objective value of $\lambda^* = 1$. Therefore, it suffices to show that for any other feasible point, we get lower objective value. To show this, we first notice that the first three users form an interference channel which is exactly the same as the one in Lemma 1. Therefore, in order to get the minimum rate of one, we need to use one of the optimal solutions in $\mathcal{S}$ in Lemma 1 for $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$. Furthermore, it is not hard to see that using either $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3) = (\mathbf{Q}_c^*, \mathbf{Q}_c^*, \mathbf{Q}_c^*)$ or $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3) = (\mathbf{Q}_d^*, \mathbf{Q}_d^*, \mathbf{Q}_d^*)$ would cause high interference to either user 4 or user 5 and prevent them from achieving the communication rate of one. Therefore, the only optimal solutions are the ones in the set $\mathcal{T}$. ■

Using Lemma 2, we can discretize the variables in the max-min problem and use it to prove the NP-hardness of the problem. In fact, for any 5 users similar to the ones in Lemma 2, there are only two possible strategies that can maximize the minimum rate of communication: either we should transmit on the first antenna or transmit on the second antenna. This observation will be crucial in establishing our NP-hardness result.

Theorem 1 *For a K-cell MIMO interference channel where each transmit/receive node is equipped with at least two antennas, the problem of designing covariance matrices to achieve max-min fairness is NP-hard in K. More specifically, solving the following problem is NP-hard*

$$\begin{aligned} \max_{\{\mathbf{Q}_i\}_{i=1}^K} \quad & \min_k \quad \log\det\left(\mathbf{I} + \mathbf{H}_{kk}\mathbf{Q}_k\mathbf{H}_{kk}^H\big(\sigma_k^2\mathbf{I} + \sum_{j\neq k}\mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H\big)^{-1}\right) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{Q}_k) \leq P_k, \mathbf{Q}_k \succeq 0,\ k = 1, \cdots, K. \end{aligned} \tag{12}$$

This theorem is proved based on a polynomial time reduction from the 3-satisfiability (3-SAT) problem which is known to be NP-complete [6]. The 3-SAT problem is described as follows. Given M disjunctive clauses $c_1, \cdots, c_M$ defined on N Boolean variables $x_1 \cdots, x_N$, i.e., $c_m = y_{m1} \vee y_{m2} \vee y_{m3}$ with $y_{mi} \in \{x_1, \cdots, x_N, \bar{x}_1, \cdots, \bar{x}_N\}$, the problem is to check whether there exists a truth assignment for the Boolean

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

variables such that all the clauses are satisfied simultaneously. The details of the proof of the theorem can be found in Appendix B.

Corollary 1 *Under the same set up as in Theorem 1, problem* (4) *is NP-hard.*

To see why the above corollary holds, we assume the contrary. Then a binary search procedure for λ would imply a polynomial time algorithm for (P), which would contradict the NP-hardness result of Theorem 1.

IV. The Max-Min Problem and Its Equivalent Reformulation

The complexity results established in the previous section suggests that it is generally not possible to solve the max-min problem (P) to its global optimality in a time that grows polynomially in K. Guided by this insight, we reset our goal to that of designing computationally efficient algorithms that can compute a high quality solution for (P). To this end, we first provide an equivalent reformulation of problem (P), which will be used later for our algorithm design.

Introducing a slack variable λ, the problem (P) can be equivalently written as

$$\begin{aligned} \max_{\{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}},\lambda} \quad & \lambda \\ \text{s.t.} \quad & R_{i_k}(\mathbf{V}) \geq \lambda, \forall\ i_k \in \mathcal{I} \\ & \sum_{i\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H] \leq P_k,\ \forall\ k \in \mathcal{K}. \end{aligned} \tag{P1}$$

In order to further simplify the problem, we need to introduce the following lemma.

Lemma 3 *The rate of user i_k in* (2) *can also be represented as*

$$R_{i_k} = \max_{\mathbf{U}_{i_k},\mathbf{W}_{i_k}} \log\det\left(\mathbf{W}_{i_k}\right) - \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) + d_{i_k}, \tag{13}$$

where $\mathbf{E}_{i_k}$ is the MSE value of user i_k given by (1)*.*

Proof: First, by checking the first order optimality condition of (13) with respect to $\mathbf{U}_{i_k}$, we get

$$\mathbf{W}_{i_k}^H\left(\mathbf{J}_{i_k}\mathbf{U}_{i_k}^* - \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\right) = 0 \implies \mathbf{U}_{i_k}^* = \mathbf{J}_{i_k}^{-1}\mathbf{H}_{i_k k}\mathbf{V}_{i_k},$$

where $\mathbf{J}_{i_k} = \sigma_{i_k}^2\mathbf{I} + \sum_{\ell_j\in\mathcal{I}}\mathbf{H}_{i_k j}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_k j}^H$ and $\mathbf{U}_{i_k}^*$ is the optimal solution of (13). By plugging in the optimal value $\mathbf{U}_{i_k}^*$ in (1), we obtain $\mathbf{E}_{i_k}^{opt} = \mathbf{I} - \mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{J}_{i_k}^{-1}\mathbf{H}_{i_k k}\mathbf{V}_{i_k}$. Hence plugging $\mathbf{E}_{i_k}^{opt}$ in (13) yields

$$\begin{aligned} & \max_{\mathbf{U}_{i_k},\mathbf{W}_{i_k}} \quad \log\det\left(\mathbf{W}_{i_k}\right) - \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) + d_{i_k} \\ & = \max_{\mathbf{W}_{i_k}} \quad \log\det\left(\mathbf{W}_{i_k}\right) - \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}^{opt}\right) + d_{i_k}. \end{aligned} \tag{14}$$

The first order optimality condition of (14) with respect to $\mathbf{W}_{i_k}$ implies $\mathbf{W}_{i_k}^* = \left(\mathbf{E}_{i_k}^{opt}\right)^{-1}$.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

By plugging in the optimal $\mathbf{W}_{i_k}^*$ in (14), we can write

$$\begin{aligned}
&\max_{\mathbf{U}_{i_k},\mathbf{W}_{i_k}} \log\det\left(\mathbf{W}_{i_k}\right) - \mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\right) + d_{i_k} \\
&= -\log\det(\mathbf{E}_{i_k}^{opt}) \\
&= -\log\det\left(\mathbf{I} - \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{J}_{i_k}^{-1}\right) \\
&= \log\det\left(\mathbf{J}_{i_k}\left(\mathbf{J}_{i_k} - \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\right)^{-1}\right),
\end{aligned}$$

which is the rate of user i_k in (2). ■

Using the observation in Lemma 3, we consider the following reformulated problem of (P)

$$\begin{aligned}
\min_{\mathbf{U},\mathbf{V},\mathbf{W}} \quad & \max_{i_k\in\mathcal{I}} \mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} \\
\text{s.t.} \quad & \sum_{i\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H] \le P_k,\ \forall\ k\in\mathcal{K}.
\end{aligned} \tag{Q}$$

Again introducing the slack variable λ, the above problem is equivalent to

$$\begin{aligned}
\max_{\mathbf{V},\mathbf{U},\mathbf{W},\lambda} \quad & \lambda \\
\text{s.t.} \quad & \mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} \le -\lambda,\ \forall\ i_k\in\mathcal{I} \\
& \sum_{i_k\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H] \le P_k,\ \forall\ k\in\mathcal{K}.
\end{aligned} \tag{Q1}$$

In the above formulation, we have introduced the weight matrix variables $\mathbf{W}_{i_k}\in\mathbb{C}^{d_{i_k}\times d_{i_k}}$ and the receive beamformer variables $\mathbf{U}_{i_k}\in\mathbb{C}^{d_{i_k}\times N_{i_k}},\ i_k\in\mathcal{I}$. Note that for fixed receive and transmit beamformers, the MSE matrix for user i_k is a function of $\mathbf{V}\triangleq\{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}}$ and $\mathbf{U}_{i_k}$, and is defined in (1). In the following analysis, $\mathbf{E}_{i_k}(\mathbf{U}_{i_k},\mathbf{V})$ will be occasionally used to make the dependency of the MSE matrix on the transmit/recieve beamformers explicit. For notational simplicity, we further define $\mathbf{W}\triangleq\{\mathbf{W}_{i_k}\}_{i_k\in\mathcal{I}}$ and $\mathbf{U}\triangleq\{\mathbf{U}_{i_k}\}_{i_k\in\mathcal{I}}$. The feasible set of beamformers of BS k is $\mathcal{V}_k\triangleq\{\{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}_k}:\sum_{i\in\mathcal{I}_k}\mathrm{Tr}[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H]\le P_k\}$; Define the feasible set for $\mathbf{V}$ as $\mathcal{V}\triangleq\prod_{k\in\mathcal{K}}\mathcal{V}_k$.

At this point, the precise relationship of the problems (P1) and (Q1) (or their equivalent problems (P) and (Q)) is still not clear. In the following, we provide a series of results that reveal an intrinsic equivalence relationship of these two problems.

To proceed, the following definitions are needed. The minimum mean squared error (MMSE) receiver for user i_k is defined as

$$\mathbf{U}_{i_k}^{\mathrm{mmse}} = \left(\sum_{\ell=1}^{K}\sum_{m\in\mathcal{I}_\ell}\mathbf{H}_{i_k,\ell}\mathbf{V}_{m_\ell}\mathbf{V}_{m\ell}^H\mathbf{H}_{i_k,\ell}^H + \sigma_{i_k}^2\mathbf{I}\right)^{-1}\mathbf{H}_{i_k,k}\mathbf{V}_{i_k} \triangleq \Psi_{i_k}(\mathbf{V}). \tag{15}$$

When the MMSE receiver is used, the corresponding MSE matrix $\mathbf{E}_{i_k}^{\mathrm{mmse}}$ (which is a function of $\mathbf{V}$) is given by

$$\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}) = \mathbf{I} - \mathbf{V}_{i_k}^H\mathbf{H}_{i_k,k}^H\mathbf{U}_{i_k}^{\mathrm{mmse}} \succ 0. \tag{16}$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Define the inverse of the MSE matrix as

$$\Upsilon_{i_k}(\mathbf{V}) \triangleq \left(\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V})\right)^{-1} \succ 0. \tag{17}$$

Let $\Psi(\mathbf{V}) \triangleq \{\Psi_{i_k}(\mathbf{V})\}_{i_k \in \mathcal{I}}$ and $\Upsilon(\mathbf{V}) \triangleq \{\Upsilon_{i_k}(\mathbf{V})\}_{i_k \in \mathcal{I}}$.

Our next result shows that there is a connection between the stationary solutions (or every KKT point) of problem (P1) and the stationary solutions of problem (Q1). Moreover, the same connection holds for the global optimal solutions of the two problems.

Proposition 1 *Let $(\lambda^*, \mathbf{V}^*)$ be an arbitrary KKT point of problem (P1). Then $(\lambda^*, \mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\lambda^*, \mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$ is a KKT point of (Q1). Conversely, suppose $(\lambda^*, \mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$ is a KKT point of problem (Q1), then $(\lambda^*, \mathbf{V}^*)$ must be a KKT point of problem (P1). Moreover, the triple $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$ is a global optimal solution of problem (Q1) if and only if $\mathbf{V}^*$ is a global optimal solution of problem (P1).*

The proof of Proposition 1 is rather involved and thus relegated to the Appendix C.

V. The Proposed Algorithm

In this section, we utilize the above equivalence relationship to design a simple algorithm for problem (P1). Our strategy is to compute a stationary solution of problem (Q1) instead.

We propose to update the variables $\mathbf{U}, \mathbf{V}$, and $\mathbf{W}$ alternately. More specifically, let us use n as the iteration index. Then the proposed algorithm alternates among the following three steps

$$\mathbf{V}^{n+1} \in \Phi(\mathbf{U}^n, \mathbf{W}^n), \quad \mathbf{U}^{n+1} = \Psi(\mathbf{V}^{n+1}), \quad \mathbf{W}^{n+1} = \Upsilon(\mathbf{V}^{n+1})$$

where $\Psi(\cdot)$ and $\Upsilon(\cdot)$ are respectively given in (15) and (17). The mapping $\Phi(\cdot)$ is defined as

$$\mathbf{V} \in \Phi(\mathbf{U}, \mathbf{W}) \quad \Longleftrightarrow \quad \mathbf{V} \in \arg\min_{\mathbf{V} \in \mathcal{V}} \max_{i_k \in \mathcal{I}} \mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}] - \log\det(\mathbf{W}_{i_k}) - d_{i_k}.$$

In words, every element in the range of the map $\Phi(\mathbf{W}, \mathbf{U})$ is an optimal solution to the problem

$$\begin{aligned} &\min_{\mathbf{V}} \max_{i_k \in \mathcal{I}} \mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} \\ &\text{s.t.} \quad \mathbf{V} \in \mathcal{V}. \end{aligned} \tag{Q-V}$$

In the following, we will proceed to obtain the solution to the problem (Q-V). Introducing a slack variable γ, the problem (Q-V) can be equivalently written as

$$\begin{aligned} \min_{\mathbf{V}, \gamma} \quad & \gamma \\ \text{s.t.} \quad & \mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} \leq \gamma, \ \forall\ i_k \in \mathcal{I} \\ & \mathbf{V} \in \mathcal{V}. \end{aligned}$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Utilizing the definition of the MSE matrix in (1), we can see that this problem is a convex problem, as the objective is linear, and all the constraints are convex (in fact, quadratic). Thus, this problem can be solved in a centralized way using conventional optimization package. The overall algorithm is summarized in Table I.

Theorem 2 *The iterates generated by the alternating algorithm given in Table I converge to the set of KKT solutions of the problem* (P1)*. In other words,*

$$\lim_{n\to\infty} \quad d(\mathbf{V}^n, \mathcal{S}) = 0$$

where $\mathcal{S}$ is the set of KKT points of (P1) *and* $d(\mathbf{V}, \mathcal{S}) \triangleq \inf_{\mathbf{U}\in\mathcal{S}} \|\mathbf{V} - \mathbf{U}\|$.

Proof: Let us define the value of the objective function of problem (Q) as $G(\mathbf{V}, \mathbf{U}, \mathbf{W})$. Due to equivalence, $G(\mathbf{V}, \mathbf{U}, \mathbf{W})$ can also represent the value of the objective function of problem (Q1). First, we observe that the sequence $\{G(\mathbf{V}^n, \mathbf{U}^n, \mathbf{W}^n)\}_{n=1}^{\infty}$ monotonically decreases and converges. Denote its limit as $\bar{G}$. Due to the compactness of the set $\mathcal{V}$, the iterates $\{\mathbf{V}^n\}_{n=1}^{\infty}$ must have a cluster point $\bar{\mathbf{V}}$. Let $\{\mathbf{V}^{n_t}\}_{t=1}^{\infty}$ be the subsequence converging to $\bar{\mathbf{V}}$. Since the maps $\Upsilon(\cdot)$ and $\Psi(\cdot)$ are continuous, we must have

$$\lim_{t\to\infty} \quad (\mathbf{V}^{n_t}, \mathbf{U}^{n_t}, \mathbf{W}^{n_t}) = (\bar{\mathbf{V}}, \bar{\mathbf{U}}, \bar{\mathbf{W}}) \triangleq (\bar{\mathbf{V}}, \Psi(\bar{\mathbf{V}}), \Upsilon(\bar{\mathbf{V}}))$$

First we will show that in the limit we have: $\bar{\mathbf{V}} \in \Phi(\bar{\mathbf{W}}, \bar{\mathbf{U}})$. Due to the optimality of $\mathbf{V}^{n_t+1}$ and monotonic decrease of the objective function, we have that

$$G(\mathbf{V}^{n_{t+1}}, \mathbf{U}^{n_{t+1}}, \mathbf{W}^{n_{t+1}}) \leq G(\mathbf{V}^{n_t+1}, \mathbf{U}^{n_t}, \mathbf{W}^{n_t}) \leq G(\mathbf{V}, \mathbf{U}^{n_t}, \mathbf{W}^{n_t}), \quad \forall\ \mathbf{V} \in \mathcal{V},\ \forall\ n_t.$$

Taking the limit of both sides, we have that[2]

$$\bar{G} = G(\bar{\mathbf{V}}, \bar{\mathbf{U}}, \bar{\mathbf{W}}) \leq G(\mathbf{V}, \bar{\mathbf{U}}, \bar{\mathbf{W}}), \quad \forall\ \mathbf{V} \in \mathcal{V}.$$

Consequently, we must have $\bar{\mathbf{V}} \in \Phi(\bar{\mathbf{W}}, \bar{\mathbf{U}})$.

The next step is to establish that $(\bar{\mathbf{V}}, \bar{\mathbf{U}}, \bar{\mathbf{W}}) = (\bar{\mathbf{V}}, \Psi(\bar{\mathbf{V}}), \Upsilon(\bar{\mathbf{V}}))$ is a KKT solution of (Q1). Firstly, the fact that $\bar{\mathbf{V}} \in \Phi(\bar{\mathbf{W}}, \bar{\mathbf{U}})$ implies that $\bar{\mathbf{V}}$ is a global optimal solution of the following convex problem

$$\begin{aligned} &\min_{\mathbf{V},\lambda} \lambda \\ \text{s.t.}\ \ &\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k}, \mathbf{V})] - \log\det(\bar{\mathbf{W}}_{i_k}) - d_{i_k} \leq \lambda,\ \forall\ i_k \in \mathcal{I} \\ &\mathbf{V} \in \mathcal{V}. \end{aligned}$$

Consequently $(\bar{\mathbf{V}}, \bar{\lambda})$ must satisfy the following optimality conditions (where $\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\epsilon}}$ are the associated

[2]Note that taking the limit inside the objective value $G(\cdot)$ is legitimate, as the objective function of problem (Q) is continuous (albeit nonsmooth).

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Lagrangian multipliers)

$$-\sum_{i_k\in\mathcal{I}}\bar{\mu}_{i_k}\nabla_{\mathbf{V}_{m_\ell}}\left(\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})]\right)-2\bar{\epsilon}_j\bar{\mathbf{V}}_{\ell_j}=0,\forall\ m_\ell\in\mathcal{I}$$

$$\sum_{i_k\in\mathcal{I}}\bar{\mu}_{i_k}=1$$

$$0\leq\bar{\mu}_{i_k}\perp-\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})_{i_k}]+\log\det(\bar{\mathbf{W}}_{i_k})+d_{i_k}-\bar{\lambda}\geq 0,\forall\ i_k\in\mathcal{I}$$

$$0\leq\bar{\epsilon}_k\perp P_k-\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}[\bar{\mathbf{V}}_{i_k}\bar{\mathbf{V}}_{i_k}^H]\geq 0,\forall\ k\in\mathcal{K}.$$

Similarly, using the fact that $\bar{\mathbf{U}}=\Psi(\bar{\mathbf{V}})$ and $\bar{\mathbf{W}}=\Upsilon(\bar{\mathbf{V}})$, we have that $\bar{\mathbf{U}}$ and $\bar{\mathbf{W}}$ must satisfy

$$\nabla_{\mathbf{U}_{i_k}}\left(\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})]\right)=0,\forall\ i_k\in\mathcal{I}$$

$$\nabla_{\mathbf{W}_{i_k}}\left(\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})]-\log\det(\bar{\mathbf{W}}_{i_k})\right)=0,\ \forall\ i_k\in\mathcal{I}$$

which in turn implies that the following conditions are true

$$-\bar{\mu}_{i_k}\nabla_{\mathbf{U}_{i_k}}\left(\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})]\right)=0,\forall\ i_k\in\mathcal{I}$$

$$-\bar{\mu}_{i_k}\nabla_{\mathbf{W}_{i_k}}\left(\mathrm{Tr}[\bar{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\bar{\mathbf{U}}_{i_k},\bar{\mathbf{V}})]-\log\det(\bar{\mathbf{W}}_{i_k})\right)=0,\ \forall\ i_k\in\mathcal{I}.$$

In conclusion, we have that $(\bar{\mathbf{V}},\bar{\mathbf{U}},\bar{\mathbf{W}})=(\bar{\mathbf{V}},\Psi(\bar{\mathbf{V}}),\Upsilon(\bar{\mathbf{V}}))$ along with the slack variable $\bar{\lambda}$ and the multipliers $(\bar{\boldsymbol{\mu}},\bar{\boldsymbol{\epsilon}})$ satisfy the KKT condition for problem (Q1) (as expressed in (30)–(35) in the Appendix C). This result implies that $(\bar{\mathbf{V}},\bar{\mathbf{U}},\bar{\mathbf{W}})=(\bar{\mathbf{V}},\Psi(\bar{\mathbf{V}}),\Upsilon(\bar{\mathbf{V}}))$ is a KKT solution to problem (Q1). Applying the result in Proposition 1, we conclude that $\bar{\mathbf{V}}$ must be a KKT point of the original problem (P1).
So far we have proved that any cluster point of the iterates is a KKT point of (P1). Since the feasible set $\mathcal{V}$ is compact, we have $\lim_{n\to\infty}\quad d(\mathbf{V}^n,\mathcal{S})=0$, and this completes the proof. ■

Several remarks regarding the above results and the existing results in [16] are in order.

Remark 1: The original max-min problem (P) has *nonsmooth* objective functions. Consequently the proof of the equivalence relationship of problem (P) and (Q) (i.e., Proposition 1) is very different from the cases presented in [16]. In particular, in the reformulated problem (Q1), fixing the variable $\mathbf{V}$ and solving for variables $\mathbf{U}$ and $\mathbf{W}$ generally admits *multiple solutions*. This is because at optimality, it is possible that not all the constraints on $\mathbf{U}$ and $\mathbf{W}$ variables in (Q1) are *active*. For those constraints that are inactive, their corresponding $\mathbf{U}_{i_k}$ and $\mathbf{W}_{i_k}$ can take multiple values.

The possibility of the existence of multiple solutions for problem (Q1) when fixing $\mathbf{V}$ has the following consequence: *a)* The proof of the equivalence relationship of the stationary solutions becomes much involved (see the proof of Proposition 1); *b)* There is no longer an one to one relationship between the stationary solutions of problem (P1) and (Q1). Instead, one stationary solution of problem (P1) may correspond to a set of stationary solutions of problem (Q1).

Remark 2: The proof of convergence of the alternating directions algorithm becomes more involved. Different from that of [16], the conventional convergence analysis for the block coordinate descent (BCD) algorithm no longer applies in this context. This is because the proof for the conventional BCD algorithm

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

requires that at least 2 of the 3 subproblems involving block variables must have *unique* solutions (see, e.g., [2]), which is clearly not the case here. Furthermore, the BCD algorithm requires that the objective function is continuously differentiable and the constraints are *separable* among the block variables. However, in our case the constraints of problem (Q1) are *coupled* among different block variables .

VI. Simulation Results

In this section, we present some numerical experiments comparing four different approaches for the beamformer design in the interfering broadcast channel. The first approach for designing the beamformers is the simple "*WMMSE*" algorithm proposed in [16] for maximizing the weighted sum rate of the system. Since the sum rate utility function is not a fair utility function among the users, we also consider the proportional fairness (geometric mean) utility function of the users. We use the framework in [16], [13] for maximizing the geometric mean utility function of the system and the resulting plots are denoted by the label "*GWMMSE*".

Another way of designing the beamformers for maximizing the performance of the worst user in the system is to approximate the max-min utility function. One proposed approximation for the max-min utility function could be (see [11]): $\min_{i_k} R_{i_k} \approx \log\left(\sum_{i_k \in \mathcal{I}} \exp(-R_{i_k})\right)$. Therefore instead of solving problem (P), we may maximize the above approximation of the objective by solving the following optimization problem

$$\begin{aligned} \max_{\mathbf{V}} \quad & \sum_{i_k \in \mathcal{I}} \exp(-R_{i_k}) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k, \quad \forall\, k \in \mathcal{K}. \end{aligned} \tag{18}$$

If we restrict ourselves to the case of $d_{i_k} = 1, \forall\, i_k \in \mathcal{I}$, then $\mathbf{E}_{i_k}$ in (1) becomes a scalar and thus we can denote it by e_{i_k}. Using the relation (13) and plugging in the optimal value for the matrix $\mathbf{W}_{i_k}$ yields $R_{i_k} = \log(e_{i_k}^{-1})$. Plugging in this relation in (18), we obtain the equivalent optimization form of (18):

$$\begin{aligned} \min_{\mathbf{V}} \quad & \sum_{i_k \in \mathcal{I}} e_{i_k} \\ \text{s.t.} \quad & \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k, \quad \forall\, k, \end{aligned} \tag{19}$$

which is the well-known sum MSE minimization problem and we use the algorithm in [15] to solve (19). The corresponding plots of this method are labeled by "*MMSE*" in our figures.

In our simulations, the first four plots are averaged over 50 channel realizations. In each channel realization, the channel coefficients are drawn from the zero mean unit variance i.i.d. Gaussian distribution.

In the first numerical experiment, we consider $K = 4$ BSs, each equipped with $M = 6$ antennas. There are $I = 3$ users in each cell where each of them is equipped with $N = 2$ antennas. Figure 1 and

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Figure 2 respectively represent the rate cumulative rate distribution function and the minimum rate in the system. The SNR level is set to 20dB in Figure 1. As these figures show, our proposed method yields substantially more fair rate allocation in the system.

In our second set of numerical experiments in Figure 3 and Figure 4, we explore the system with $K = 5$ cells where each BS serves $I = 3$ users. The number of transmit and receive antennas are respectively $M = 3$ and $N = 2$.

Figure 5 and Figure 6 show the convergence rate of the algorithm while a user is joining the system. In these plots, there are 5 cells and 2 users in each cell initially and at iteration 4, another user is added to one of the cells. When the extra user is added to the system, the power for the users in the same cell is reduced by a factor of $\frac{2}{3}$ and the rest of the power is used to serve the joined user initially. The precoder of the joined user is initialized randomly. Figure 5 shows the objective function of (Q) during the iterations while Figure 6 demonstrates the minimum rate of the users in the system versus the iteration number.

Figure 7 and Figure 8 represent the performance and the convergence rate of the algorithm when the channel is changing during the iterations. At iteration 7, the channel is changed by a Rayleigh fade with power 0.1. As it can be seen from the plots, the algorithm converges fast and it adapts to the new channel after a few iterations.

VII. Appendix

A. Proof of Lemma 1

First of all, it can be observed that choosing $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3 = \mathbf{Q}_a^*$ yields an objective value of $\lambda^* = 1$; the same result holds for the case of $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3 = \mathbf{Q}_b^*$, $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3 = \mathbf{Q}_c^*$, and $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3 = \mathbf{Q}_d^*$.

Let $(\lambda, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3) \in \mathcal{S}$ be an optimal solution. Clearly, at least one of the users must transmit with full power, for otherwise we could simultaneously scale $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ to get a better objective function. Without loss of generality, let us assume that user 1 is transmitting with full power, i.e., $\mathrm{Tr}(\mathbf{Q}_1) = 1$. Using eigenvalue decomposition of $\mathbf{Q}_1$, we can write $\mathbf{Q}_1 = \alpha\mathbf{a}\mathbf{a}^H + \beta\mathbf{b}\mathbf{b}^H$, where $\mathbf{a}$ and $\mathbf{b}$ are the orthonormal eigenvectors of $\mathbf{Q}_1$ and the scalars $\alpha, \beta \geq 0$ are the eigenvalues of $\mathbf{Q}_1$ with $\alpha + \beta = 1$.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Since canceling the interference results in higher rate of communication, we have

$$\begin{aligned}
R_2 &= \log\det\left(\mathbf{I}+\mathbf{Q}_2\left(\mathbf{I}+\sum_{m\neq 2}\mathbf{H}_{2m}\mathbf{Q}_m\mathbf{H}_{2m}^H\right)^{-1}\right)\\
&\leq \log\det\left(\mathbf{I}+\mathbf{Q}_2\left(\mathbf{I}+\mathbf{H}_{21}(\alpha\mathbf{a}\mathbf{a}^H+\beta\mathbf{b}\mathbf{b}^H)\mathbf{H}_{21}^H\right)^{-1}\right)\\
&= \log\det\left(\mathbf{I}+\mathbf{Q}_2\left(\mathbf{I}+4\alpha\ \underline{\mathbf{a}}\ \underline{\mathbf{a}}^H+4\beta\ \underline{\mathbf{b}}\ \underline{\mathbf{b}}^H\right)^{-1}\right)\\
&= \log\det\left(\mathbf{I}+\mathbf{Q}_2\left(\frac{1}{1+4\alpha}\ \underline{\mathbf{a}}\ \underline{\mathbf{a}}^H+\frac{1}{1+4\beta}\ \underline{\mathbf{b}}\ \underline{\mathbf{b}}^H\right)\right)\\
&\leq \log\det\left(\mathbf{I}+\frac{1}{\mathrm{Tr}(\mathbf{Q}_2)}\mathbf{Q}_2\left(\frac{1}{1+4\alpha}\ \underline{\mathbf{a}}\ \underline{\mathbf{a}}^H+\frac{1}{1+4\beta}\ \underline{\mathbf{b}}\ \underline{\mathbf{b}}^H\right)\right),
\end{aligned} \tag{20}$$

where $\underline{\mathbf{a}}=\frac{1}{2}\mathbf{H}_{21}\mathbf{a}$ and $\underline{\mathbf{b}}=\frac{1}{2}\mathbf{H}_{21}\mathbf{b}$. The last inequality is due to the fact that $\mathrm{Tr}(\mathbf{Q}_2)\leq 1$. Clearly, $\underline{\mathbf{a}}^H\underline{\mathbf{b}}=0$ and $\|\underline{\mathbf{a}}\|=\|\underline{\mathbf{b}}\|=1$.

Let us use the eigenvalue decomposition $\frac{\mathbf{Q}_2}{\mathrm{Tr}(\mathbf{Q}_2)}=\theta\mathbf{c}\mathbf{c}^H+(1-\theta)\mathbf{d}\mathbf{d}^H$, for some $\theta\in[0,1]$ and some orthonormal vectors $\mathbf{c}$ and $\mathbf{d}$. Utilizing the fact that determinant is the product of the eigenvalues and trace is the sum of the eigenvalues, we can further simplify the inequality in (20) as

$$\begin{aligned}
R_2 &\leq \log\Bigg\{1+\mathrm{Tr}\left[\left(\theta\mathbf{c}\mathbf{c}^H+(1-\theta)\mathbf{d}\mathbf{d}^H\right)\left(\frac{1}{1+4\alpha}\underline{\mathbf{a}}\ \underline{\mathbf{a}}^H+\frac{1}{1+4\beta}\underline{\mathbf{b}}\ \underline{\mathbf{b}}^H\right)\right]\\
&\qquad+\det\left[\left(\theta\mathbf{c}\mathbf{c}^H+(1-\theta)\mathbf{d}\mathbf{d}^H\right)\left(\frac{1}{1+4\alpha}\underline{\mathbf{a}}\ \underline{\mathbf{a}}^H+\frac{1}{1+4\beta}\underline{\mathbf{b}}\ \underline{\mathbf{b}}^H\right)\right]\Bigg\}\\
&= \log\left[1+\frac{\theta x}{1+4\alpha}+\frac{\theta(1-x)}{1+4\beta}+\frac{(1-\theta)(1-x)}{1+4\alpha}+\frac{(1-\theta)x}{1+4\beta}+\frac{\theta(1-\theta)}{(1+4\alpha)(1+4\beta)}\right]\\
&\leq \max_{(x,\theta,\alpha,\beta)\in\mathcal{Y}}\ \log\left[1+\frac{\theta x}{1+4\alpha}\frac{\theta(1-x)}{1+4\beta}+\frac{(1-\theta)(1-x)}{1+4\alpha}+\frac{(1-\theta)x}{1+4\beta}+\frac{\theta(1-\theta)}{(1+4\alpha)(1+4\beta)}\right],
\end{aligned} \tag{21}$$

where $x\triangleq|\mathbf{c}^H\mathbf{a}|^2$, $\mathcal{Y}\triangleq\{(x,\theta,\alpha,\beta)\mid\alpha+\beta=1,\ 0\leq\alpha,\beta,x\leq 1\}$. Since the function in (21) is linear in x, it suffices to only check the boundary points $x=0$ and $x=1$ in order to find the maximum. The claim is that the maximum in (21) takes the value of 1, and it is achieved at both boundary points.

First consider the boundary point $x=1$. We have

$$R_2\leq\max_{(\theta,\alpha,\beta)\in\mathcal{X}}\quad f(\theta,\alpha,\beta), \tag{22}$$

where $\mathcal{X}\triangleq\{(\theta,\alpha,\beta)\mid\alpha+\beta=1,\ 0\leq\alpha,\beta\}$ and

$$f(\theta,\alpha,\beta)\triangleq\log\left(1+\frac{\theta}{1+4\alpha}+\frac{1-\theta}{1+4\beta}+\frac{\theta(1-\theta)}{(1+4\alpha)(1+4\beta)}\right) \tag{23}$$

We are interested in finding the set of optimal solutions of (23). In particular, we want to characterize $\mathcal{S}_1=\{(\theta^*,\alpha^*,\beta^*)\}$ defined by

$$\mathcal{S}_1\triangleq\arg\max_{(\theta,\alpha,\beta)\in\mathcal{X}}\quad f(\theta,\alpha,\beta).$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

In what follows, we will prove that $\mathcal{S}_1 = \{(0,1,0),(1,0,1)\}$.

First we observe that $f(0,1,0) = f(1,0,1) = 1$. Now, we show that $f(\theta,\alpha,\beta) < 1$, for all $(\theta,\alpha,\beta) \in \mathcal{X}$ such that $0 < \theta < 1$. Assume the contrary that there exists an optimal point $(\theta^*,\alpha^*,\beta^*)$ such that $0 < \theta^* < 1$. Using the first order optimality condition $\frac{\partial}{\partial\theta} f(\theta^*,\alpha^*,\beta^*) = 0$, we obtain

$$\theta^* = \frac{4\beta^* - 4\alpha^* + 1}{2}.$$

Combining with $0 < \theta^* < 1$ yields

$$-\frac{1}{4} < \beta^* - \alpha^* < \frac{1}{4}. \tag{24}$$

Plugging in the value of optimal $\theta^* = \frac{4\beta^*-4\alpha^*+1}{2}$ in $f(\cdot)$ and simplifying the equations, we obtain

$$f(\theta^*,\alpha^*,\beta^*) = \log\left(1 + \frac{13 + 16(\beta^*-\alpha^*)^2}{4(1+4\alpha^*)(1+4\beta^*)}\right).$$

Combining with (24) yields

$$\begin{aligned} f(\theta^*,\alpha^*,\beta^*) &\le \log\left(1 + \frac{14}{4(1+4\alpha^*)(1+4\beta^*)}\right) \\ &\le \log\left(1 + \frac{14}{4(1+4\alpha^*+4\beta^*)}\right) \\ &= \log\left(1 + \frac{14}{20}\right) < 1, \end{aligned}$$

which contradicts the fact that $\max_{(\theta,\alpha,\beta)\in\mathcal{X}} f(\theta,\alpha,\beta) = 1$. Therefore, the optimal θ only happens at the boundary and we have

$$\{(0,1,0),(1,0,1)\} = \arg\max_{(\theta,\alpha,\beta)\in\mathcal{X}} f(\theta,\alpha,\beta).$$

Similarly, for the case when $x = 0$, we can see that the optimal solution set is $\{(0,0,1),(1,1,0)\}$.

Using these optimal values yields $R_2 \le 1$. Note that in order to have equality $R_2 = 1$, we must have $\mathrm{Tr}(\mathbf{Q}_2) = 1$ and

$$(x,\theta,\alpha,\beta) \in \{(1,0,1,0),(1,1,0,1),(0,0,0,1),(0,1,1,0)\}.$$

Let us choose the optimal solution $(x,\theta,\alpha,\beta) = (1,0,1,0)$. Therefore,

$$\mathbf{Q}_1 = \mathbf{a}\mathbf{a}^H, \quad \mathbf{Q}_2 = \mathbf{d}\mathbf{d}^H, \quad x = |\mathbf{c}^H\mathbf{a}|^2 = 1,$$

which yields $\underline{\mathbf{a}}^H\mathbf{d} = 0$. Repeating the above argument for user 2 and user 3, we get $\mathbf{Q}_3 = \mathbf{g}\mathbf{g}^H$ with $\underline{\mathbf{a}}^H\mathbf{g} = 0$. Since $\mathbf{d}$ and $\mathbf{g}$ are both orthogonal to $\underline{\mathbf{a}}$, we obtain $\mathbf{d} = \exp^{j\phi_d}\mathbf{g}$. Repeating the above argument for the other pair of users yields

$$\mathbf{a} = \exp^{j\phi_a}\mathbf{g} \quad \text{and} \quad \underline{\mathbf{a}}^H\mathbf{a} = 0,$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

where the last relations imply that $\mathbf{a}, \mathbf{d}$, and $\mathbf{g}$ are the same up to the phase rotation and they belong to the following set (after the proper phase rotation)

$$\mathbf{a} \in \left\{ [1 \quad 0]^H, [0 \quad 1]^H, \frac{1}{\sqrt{2}}[j \quad 1]^H, \frac{1}{\sqrt{2}}[1 \quad j]^H \right\}.$$

Each of these points gives us one of the optimal covariance matrices in (7).

B. Proof of Theorem 1

Proof: The proof is based on a polynomial time reduction from the 3-satisfiability (3-SAT) problem which is known to be NP-complete. We first consider an instance of the 3-SAT problem with n variables $x_1, x_2, \ldots, x_n$ and m clauses $c_1, c_2, \ldots, c_m$. For each variable x_i, we consider 5 users $\mathcal{X}_{1i}, \mathcal{X}_{2i}, \ldots, \mathcal{X}_{5i}$ in our interference channel. Each user is equipped with two antennas, and the channels between the users are specified as in (8)–(10). For each clause c_j, $j = 1, 2, \ldots, m$, we consider one user $\mathcal{C}_j$ in the system with two antennas. In summary, we totally have $5n + m$ users in the system. Set the noise power $\sigma^2 = 1$ and the power budget $P_k = 1$ for all users. We define the channel between the users $\mathcal{C}_i$ and $\mathcal{C}_j$ to be zero for all $j \neq i$. Furthermore, we assume that the channel between the transmitter and receiver of user $\mathcal{C}_i$ is given by

$$\mathbf{H}_{\mathcal{C}_i\mathcal{C}_i} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Let us also assume that *i)* there is no interference among the blocks of users that correspond to different variables and *ii)* there is no interference from the transmitter of user $\mathcal{C}_j$ to the receivers of users $\mathcal{X}_{1i}, \ldots, \mathcal{X}_{5i}$ for all $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$. Consider a clause $c_j : y_{j1} + y_{j2} + y_{j3}$, where $y_{j1}, y_{j2}, y_{j3} \in \{x_1, x_2, \ldots, x_n, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n\}$ with $\overline{x}_i$ denoting the negation of x_i. We use the following rules to define the channels from the transmitter of user $\mathcal{X}_{ki}$ to the receiver of user $\mathcal{C}_j$:

- If the variable x_i appears in c_j, we define the channel from the transmitter of $\mathcal{X}_{1i}$ to the receiver of $\mathcal{C}_j$ to be $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.
- If the variable $\bar{x}_i$ appears in c_j, we define the channel from the transmitter of $\mathcal{X}_{1i}$ to the receiver of $\mathcal{C}_j$ to be $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.
- If x_i does not appear in c_j, we define the channel from the transmitter of $\mathcal{X}_{1i}$ to the receiver of $\mathcal{C}_j$ to be zero.
- The channel from transmitters of users $\mathcal{X}_{2i}, \mathcal{X}_{3i}, \mathcal{X}_{4i}, \mathcal{X}_{5i}$ to the receiver of user $\mathcal{C}_j$ is zero for all $i = 1, \ldots, n$ and $j = 1, 2, \ldots, m$.

[[[As an example, draw a figure for a simple clause.]]] Now we claim that the 3-SAT problem is satisfiable if and only if solving the problem (12) for the corresponding interference channel leads to the optimum value of one. To prove this fact, let us assume that the optimum value of (12) is one. According to the Lemma 2, the only way to get the rate of one for users $\mathcal{X}_{kj}$, $k = 1, \ldots, 5$, $j = 1, \ldots, n$,

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

is to transmit with full power either on the first antenna or on the second antenna. Now, based on the optimal solution of (12), we can determine the solution of the 3-SAT problem. In particular, if user $\mathcal{X}_{1i}$ is transmitting on the first antenna, we set $x_i = 0$. Otherwise, if it transmits on the second antenna, we set $x_i = 1$. By assigning values to all the variables in this way, we claim that all clauses are satisfied. We prove by contradiction. Assume the contrary that there exists a clause c_j that is not satisfied, i.e., all the corresponding variables are zero. Therefore, user $\mathcal{C}_j$ gets interference on the first receive antenna from all three users corresponding to the variables appearing in $\mathcal{C}_j$. As the result, the interference power is 3. Since the noise power is one and the received signal power is 3, the SINR level for user $\mathcal{C}_j$ is $\frac{3}{1+3}$ which contradicts the fact that the minimum rate in the system is one.

Now we prove the other direction. Let us assume that the 3-SAT problem is satisfiable. We claim that the optimal value of (12) is one. Since in each block of 5 users the optimum value is one, it suffices to show that the objective value of one is achievable. Now, we design the covariance matrices based on the solution of the 3-SAT problem. If $x_i = 0$, we transmit with full power on the first antenna of users $\mathcal{X}_{1i}, \mathcal{X}_{2i}, \ldots, \mathcal{X}_{5i}$. If $x_i = 1$, we allocate full power for transmission on the second antenna of users $\mathcal{X}_{1i}, \mathcal{X}_{2i}, \ldots, \mathcal{X}_{5i}$. With this allocation, each user $\mathcal{X}_{ki}$, $k = 1, \ldots, 5$, $i = 1, \ldots, n$, gets the rate of one. For all users $\mathcal{C}_j$, $j = 1, 2, \ldots, m$, we transmit with full power on the first antenna. Since 3-SAT problem is satisfiable with the given boolean allocation of the variables, for each clause $\mathcal{C}_j$ at one of the corresponding variables are one. Therefore, the interference level at the receiver of user $\mathcal{C}_j$ is at most 2. Since the received signal power at the receiver of user $\mathcal{C}_j$ is one, the SINR level is at least $\frac{3}{1+2} = 1$ which yields the rate of communication $R_{\mathcal{C}_j} \geq 1$. Thus, all users $\mathcal{C}_j, j = 1, \ldots, m$, have rate at least one; which completes the proof of our claim. As the result, checking whether the objective value of one is achievable for (12) is equivalent to solving the instance of 3-SAT problem. Thus, problem (12) is NP-hard. ■

C. Proof of Proposition 1

Proof: The Lagrangian of problem (P1) can be expressed as

$$L(\mathbf{V}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \lambda + \sum_{i_k \in \mathcal{I}} \mu_{i_k} \left(R_{i_k}(\mathbf{V}) - \lambda\right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H]) \tag{25}$$

where $\boldsymbol{\mu} \triangleq \{\mu_{i_k}\}_{i_k \in \mathcal{I}}$ and $\boldsymbol{\epsilon} \triangleq \{\epsilon_k\}_{k \in \mathcal{K}}$ are the set of associated optimal Lagrangian multipliers. Suppose $(\lambda^*, \mathbf{V}^*)$ is a KKT point of (P1), and $\{\mu_{i_k}^*\}_{i_k \in \mathcal{I}}$ and $\{\epsilon_k^*\}_{k \in \mathcal{K}}$ are the set of associated optimal Lagrangian

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

multipliers. The KKT optimality condition for problem (P1) can be written as

$$\nabla_{\mathbf{V}_{\ell_j}} L(\mathbf{V}^*, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = \sum_{i_k \in \mathcal{I}} \mu^*_{i_k} \nabla_{\mathbf{V}_{\ell_j}} R_{i_k}(\mathbf{V}^*) - 2\epsilon^*_j \mathbf{V}^*_{\ell_j} = \mathbf{0}, \forall\ \ell_j \in \mathcal{I}_k \tag{26}$$

$$\sum_{i_k \in \mathcal{I}} \mu^*_{i_k} = 1 \tag{27}$$

$$0 \le \mu^*_{i_k} \perp R_{i_k}(\mathbf{V}^*) - \lambda^* \ge 0, \forall\ i_k \in \mathcal{I} \tag{28}$$

$$0 \le \epsilon^*_k \perp P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}^*_{i_k} (\mathbf{V}^*_{i_k})^H] \ge 0, \forall\ k \in \mathcal{K}. \tag{29}$$

Similarly, the Lagrangian of problem (Q1) can be expressed as

$$\widehat{L}(\mathbf{V}, \mathbf{U}, \mathbf{W}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \lambda - \sum_{i_k \in \mathcal{I}} \mu_{i_k} \left(\mathrm{Tr}[\mathbf{W}_{i_k} \mathbf{E}_{i_k}(\mathbf{U}_{i_k}, \mathbf{V})] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} + \lambda\right)$$
$$+ \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}^H_{i_k}]).$$

Let $(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda})$ be a KKT solution of problem (Q1), and let $\{\widehat{\mu}_{i_k}\}_{i_k \in \mathcal{I}_k}$ and $\{\widehat{\epsilon}_k\}_{k \in \mathcal{K}}$ be the set of associated optimal Lagrangian multipliers. The KKT optimality condition for problem (Q1) is as follows.

$$\nabla_{\mathbf{V}_{m_\ell}} \widehat{L}(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\epsilon}}) = -\sum_{i_k \in \mathcal{I}} \widehat{\mu}_{i_k} \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})]\right) - 2\widehat{\epsilon}_\ell \widehat{\mathbf{V}}_{m_\ell} = 0, \forall\ m_\ell \in \mathcal{I} \tag{30}$$

$$\nabla_{\mathbf{U}_{i_k}} \widehat{L}(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\epsilon}}) = -\widehat{\mu}_{i_k} \nabla_{\mathbf{U}_{i_k}} \left(\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})]\right) = 0, \forall\ i_k \in \mathcal{I} \tag{31}$$

$$\nabla_{\mathbf{W}_{i_k}} \widehat{L}(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\epsilon}}) = -\widehat{\mu}_{i_k} \nabla_{\mathbf{W}_{i_k}} \left(\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})] - \log\det(\widehat{\mathbf{W}}_{i_k})\right) = 0,\ \forall\ i_k \in \mathcal{I} \tag{32}$$

$$\sum_{i_k \in \mathcal{I}} \widehat{\mu}_{i_k} = 1 \tag{33}$$

$$0 \le \widehat{\mu}_{i_k} \perp -\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})_{i_k}] + \log\det(\widehat{\mathbf{W}}_{i_k}) + d_{i_k} - \widehat{\lambda} \ge 0, \forall\ i_k \in \mathcal{I} \tag{34}$$

$$0 \le \widehat{\epsilon}_k \perp P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\widehat{\mathbf{V}}_{i_k} \widehat{\mathbf{V}}^H_{i_k}] \ge 0, \forall\ k \in \mathcal{K} \tag{35}$$

The claim is that if $\mathbf{V}^*$ satisfies the KKT system (26)–(29), then the set of solutions $(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*), \lambda^*)$, $(\widehat{\mu}_{i_k}, \widehat{\epsilon}_k) = (\mu^*_{i_k}, \epsilon^*_k)$ must satisfy the KKT system (30)–(35). The proof for this claim consists of three steps.

Step 1: It is easy to observe that condition (33) and (35) are satisfied. We can then verify that by letting $\widehat{\mathbf{U}} = \Psi(\widehat{\mathbf{V}})$ and $\widehat{\mathbf{W}} = \Upsilon(\widehat{\mathbf{V}})$, we have that

$$\nabla_{\mathbf{U}_{i_k}} \left(\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})]\right) = 0, \quad \nabla_{\mathbf{W}_{i_k}} \left(\mathrm{Tr}[\widehat{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widehat{\mathbf{U}}_{i_k}, \widehat{\mathbf{V}})] - \log\det(\widehat{\mathbf{W}}_{i_k})\right) = 0.$$

Consequently, conditions (31)–(32) are satisfied.

Step 2: We then show that condition (30) is satisfied.

For a set of given multipliers $\{\mu^*_{i_k}\}_{i_k \in \mathcal{I}}$, define the following two index sets

$$\bar{\mathcal{A}} \triangleq \{i_k | \mu^*_{i_k} = 0\}; \quad \mathcal{A} \triangleq \{i_k | \mu^*_{i_k} > 0\}.$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

In words, the set $\mathcal{A}$ includes the users for which the rate constraints in (P1) are active. Notice that due to the constraint (27), set set $\mathcal{A}$ must be nonempty, i.e., $|\mathcal{A}| > 0$.

According to the above defined index sets, we partition all the users' receive beamformers into two parts $\mathbf{U}_{\mathcal{A}} \triangleq \{\mathbf{U}_{i_k}\}_{i_k \in \mathcal{A}}; \quad \mathbf{U}_{\bar{\mathcal{A}}} \triangleq \{\mathbf{U}_{i_k}\}_{i_k \in \bar{\mathcal{A}}}$. Define the sets $\mathbf{W}_{\mathcal{A}}$, $\mathbf{W}_{\bar{\mathcal{A}}}$, $\Psi_{\mathcal{A}}(\cdot)$, $\Psi_{\bar{\mathcal{A}}}(\cdot)$, $\Upsilon_{\mathcal{A}}(\cdot)$, $\Upsilon_{\bar{\mathcal{A}}}(\cdot)$ and $\boldsymbol{\mu}_{\mathcal{A}}$ and $\boldsymbol{\mu}_{\bar{\mathcal{A}}}$ similarly. Define the reduced Lagrangian function as

$$\begin{aligned} L_{\mathcal{A}}(\mathbf{V}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon}) &= \lambda + \sum_{i_k \in \mathcal{A}} \mu_{i_k} \left(R_{i_k}(\mathbf{V}) - \lambda\right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H]) \\ \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}, \mathbf{W}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon}) &= \lambda - \sum_{i_k \in \mathcal{A}} \mu_{i_k} \left(\mathrm{Tr}[\mathbf{W}_{i_k} \mathbf{E}_{i_k}(\mathbf{U}_{i_k}, \mathbf{V})] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} + \lambda\right) \\ &\quad - \sum_{k \in \mathcal{K}} \epsilon_k (\sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H] - P_k) \end{aligned}$$

A key observation is that $\widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}, \mathbf{W}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon})$ is only a function of $\mathbf{U}_{\mathcal{A}}$ and $\mathbf{W}_{\mathcal{A}}$, but not of $\mathbf{U}_{\bar{\mathcal{A}}}$ and $\mathbf{W}_{\bar{\mathcal{A}}}$. Consequently, we can express it as $\widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda; \boldsymbol{\mu}, \boldsymbol{\epsilon})$

Step 2.1: We show the following key identity. If $\mathbf{U} = \Psi(\mathbf{V})$ and $\mathbf{W} = \Upsilon(\mathbf{V})$, then we have

$$\nabla_{\mathbf{V}} L_{\mathcal{A}}(\mathbf{V}, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = \nabla_{\mathbf{V}} \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*).$$

Notice the fact that $\mathbf{U}_{i_k} = \Psi_{i_k}(\mathbf{V}), i_k \in \mathcal{A}$ and $\mathbf{W}_{i_k} = \Upsilon(\mathbf{V}), i_k \in \mathcal{A}$ are the *unique* solutions to the following two problems, respectively

$$\begin{aligned} &\max_{\mathbf{U}_{\mathcal{A}}} \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) \\ &\max_{\mathbf{W}_{\mathcal{A}}} \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*). \end{aligned}$$

This claim can be easily checked using the first order optimality conditions of the respective problems. We note here that the uniqueness of the solutions comes from the fact that for all $i_k \in \mathcal{A}$, $\mu_{i_k}^* > 0$, and the fact that $\mathbf{W}_{i_k}$ and $\mathbf{U}_{i_k}$ are the unique solutions to the following problems, respectively.

$$\begin{aligned} &\arg\max_{\mathbf{U}_{i_k}} - \left(\mathrm{Tr}[\mathbf{W}_{i_k} \mathbf{E}_{i_k}(\mathbf{U}_{i_k}, \mathbf{V})]\right) \\ &\arg\max_{\mathbf{W}_{i_k}} - \left(\mathrm{Tr}[\mathbf{W}_{i_k} \mathbf{E}_{i_k}(\mathbf{U}_{i_k}, \mathbf{V})] - \log\det(\mathbf{W}_{i_k})\right). \end{aligned}$$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Moreover, plugging $\mathbf{U}_{\mathcal{A}} = \Psi_{\mathcal{A}}(\mathbf{V})$ and $\mathbf{W}_{\mathcal{A}} = \Upsilon_{\mathcal{A}}(\mathbf{V})$ into $\widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$, we obtain

$$\begin{aligned}
&\widehat{L}_{\mathcal{A}}(\mathbf{V}, \Psi_{\mathcal{A}}(\mathbf{V}), \Upsilon_{\mathcal{A}}(\mathbf{V}), \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) \\
&= \lambda - \sum_{i_k \in \mathcal{A}} \mu_{i_k}^* \left(\mathrm{Tr}[\mathbf{W}_{i_k} \mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V})] - \log\det(\mathbf{W}_{i_k}) - d_{i_k} + \lambda \right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H]) \\
&= \lambda - \sum_{i_k \in \mathcal{A}} \mu_{i_k}^* \left(d_{i_k} - \log\det((\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}))^{-1}) - d_{i_k} + \lambda \right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H]) \\
&= \lambda + \sum_{i_k \in \mathcal{A}} \mu_{i_k}^* \left(\log\det((\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}))^{-1}) - \lambda \right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H]) \\
&= \lambda + \sum_{i_k \in \mathcal{A}} \mu_{i_k}^* \left(R_{i_k}(\mathbf{V}) - \lambda \right) + \sum_{k \in \mathcal{K}} \epsilon_k (P_k - \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H])
\end{aligned}$$

where in the last equality we have used a well known relationship between the MSE matrix and the achievable rate: $R_{i_k}(\mathbf{V}) = -\log\det(\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}))$. Consequently, we obtain

$$\mathbf{U}_{\mathcal{A}} = \Psi_{\mathcal{A}}(\mathbf{V}),\ \mathbf{W}_{\mathcal{A}} = \Upsilon_{\mathcal{A}}(\mathbf{V}) \Longrightarrow \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = L_{\mathcal{A}}(\mathbf{V}, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*).$$

Combine the above result and the uniqueness of $\mathbf{U}_{i_k} = \Psi_{i_k}(\mathbf{V}), i_k \in \mathcal{A}$ and $\mathbf{W}_{i_k} = \Upsilon(\mathbf{V}), i_k \in \mathcal{A}$, we can apply Dankin's Min-Max Theorem (See [2, Proposition B 2.5]) to obtain the desired equality

$$\nabla_{\mathbf{V}} L_{\mathcal{A}}(\mathbf{V}, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = \nabla_{\mathbf{V}} \widehat{L}_{\mathcal{A}}(\mathbf{V}, \mathbf{U}_{\mathcal{A}}, \mathbf{W}_{\mathcal{A}}, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*). \tag{36}$$

This concludes our proof of Step 2.1.

Step 2.2: We show that condition (26) implies condition (30).

Notice the fact that for all $i_k \in \bar{\mathcal{A}}$, $\mu_{i_k}^* = 0$, and the fact that $\nabla_{\mathbf{V}_{i_k}} R_{i_k}(\mathbf{V}^*)$ takes finite value for all $\mathbf{V}^* \in \mathcal{V}$. Then condition (26) is equivalent to the following condition

$$\nabla_{\mathbf{V}_{\ell_j}} L_{\mathcal{A}}(\mathbf{V}^*, \lambda; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = 0, \forall\ \ell_j \in \mathcal{I}. \tag{37}$$

We have the following series of equalities

$$\begin{aligned}
&\nabla_{\mathbf{V}_{m_\ell}} \widehat{L}(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) \\
&= -\sum_{i_k \in \mathcal{I}} \mu_{i_k}^* \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}_{i_k^*} \mathbf{E}_{i_k}(\mathbf{U}_{i_k}^*, \mathbf{V}^*)] \right) - 2\epsilon_\ell^* \mathbf{V}_{m_\ell}^* \\
&\overset{(a)}{=} -\sum_{i_k \in \mathcal{A}} \mu_{i_k}^* \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}_{i_k}^* \mathbf{E}_{i_k}(\mathbf{U}_{i_k}^*, \mathbf{V}^*)] \right) - 2\epsilon_\ell^* \mathbf{V}_{m_\ell}^* \\
&= \nabla_{\mathbf{V}_{m_\ell}} \widehat{L}_{\mathcal{A}}(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) \\
&= 0
\end{aligned}$$

where in (a) we have again used the fact that $\mu_{i_k}^* = 0$ for all $i_k \in \bar{\mathcal{A}}$, $\mathcal{A} \bigcup \bar{\mathcal{A}} = \mathcal{I}$, and the fact that $\nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}_{i_k}^* \mathbf{E}_{i_k}(\mathbf{U}_{i_k}^*, \mathbf{V}^*)] \right)$ takes finite value for all $\mathbf{U}^* = \Upsilon_{i_k}(\mathbf{V}^*)$, $\mathbf{W}^* = \Psi_{i_k}(\mathbf{V}^*)$, and all $\mathbf{V}^* \in \mathcal{V}$; the last equality is due to (36) and (37). This shows that (30) is true.

Step 3: In this step, we show that condition (28) implies (34).

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

Let $\mathbf{U}_{i_k} = \Psi_{i_k}(\mathbf{V})$ and $\mathbf{W}_{i_k} = \Upsilon_{i_k}(\mathbf{V})$, we have that

$$\begin{aligned}
&-\mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}(\mathbf{U}_{i_k},\mathbf{V})_{i_k}] + \log\det(\mathbf{W}_{i_k}) + d_{i_k} - \lambda \\
&= -\mathrm{Tr}[\mathbf{W}_{i_k}\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V})] + \log\det(\mathbf{W}_{i_k}) + d_{i_k} - \lambda \\
&= -d_{i_k} + \log\det\left((\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}))^{-1}\right) + d_{i_k} - \lambda \\
&= R_{i_k}(\mathbf{V}) - \lambda
\end{aligned} \tag{38}$$

In (28) we have that $0 \le \mu_{i_k}^* \perp R_{i_k}(\mathbf{V}^*) - \lambda^* \ge 0$. This condition combined with (38) ensures (34) is true for $(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*), \lambda^*)$.

In conclusion, we have shown that if $(\mathbf{V}^*, \lambda^*)$ and $(\boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$ satisfy the KKT system (26)–(29), then $(\widehat{\mathbf{V}}, \widehat{\mathbf{U}}, \widehat{\mathbf{W}}, \widehat{\lambda}) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*), \lambda^*)$, $(\widehat{\mu}_{i_k}, \widehat{\epsilon}_k) = (\mu_{i_k}^*, \epsilon_k^*)$ satisfy the KKT system (30)–(35).

We now establish the correspondence between the global optimal solutions of the two problems. The proof has two main steps.

Step 1: We first argue that for every KKT solution $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}})$ of problem (Q1), there is a corresponding solution $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$ that is also a KKT solution. Furthermore, it achieves the same objective value as $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}})$.

Again consider the equivalent reformulation (Q1). Let $(\boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$ denote a set of optimal multiplier corresponds to solution $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}, \lambda^*)$, that is, together they satisfy the KKT system (30)–(35). Define the index sets $\mathcal{A}$ and $\bar{\mathcal{A}}$ as $\bar{\mathcal{A}} \triangleq \{i_k | \mu_{i_k}^* = 0\}$; $\mathcal{A} \triangleq \{i_k | \mu_{i_k}^* > 0\}$. We will show that the solution $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$, along with the optimal slack variable λ^* and the multipliers $(\boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$ must also satisfy the KKT system (30)–(35).

Firstly it is easy to see that the conditions (35) and (33) are satisfied.

We then show that the conditions (31)–(32) are satisfied. Observe that for $i_k \in \mathcal{A}$, conditions (31)–(32) imply that

$$\begin{aligned}
\nabla_{\mathbf{U}_{i_k}}\left(\mathrm{Tr}[\widetilde{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\widetilde{\mathbf{U}}_{i_k},\mathbf{V}^*)]\right) &= 0 \\
\nabla_{\mathbf{W}_{i_k}}\left(\mathrm{Tr}[\widetilde{\mathbf{W}}_{i_k}\mathbf{E}_{i_k}(\widetilde{\mathbf{U}}_{i_k},\mathbf{V}^*)] - \log\det(\widetilde{\mathbf{W}}_{i_k})\right) &= 0
\end{aligned}$$

which in turn imply that the solutions for the above problems are uniquely given as

$$\widetilde{\mathbf{U}}_{i_k} = \Psi_{i_k}(\mathbf{V}^*), \forall\ i_k \in \mathcal{A}, \quad \widetilde{\mathbf{W}}_{i_k} = \Upsilon_{i_k}(\mathbf{V}^*), \forall\ i_k \in \mathcal{A}. \tag{39}$$

Thus, setting $\widetilde{\mathbf{U}}_{\mathcal{A}} = \Psi_{\mathcal{A}}(\mathbf{V}^*)$ and $\widetilde{\mathbf{W}}_{\mathcal{A}} = \Upsilon_{\mathcal{A}}(\mathbf{V}^*)$ ensures condition (31)–(32) for all $i_k \in \mathcal{A}$. Alternatively, for $i_k \in \bar{\mathcal{A}}$, due to the fact that $\mu_{i_k}^* = 0$, the conditions (31)–(32) is also satisfied.

We then show that condition (30) is satisfied for solution $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*, \lambda^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*), \lambda^*)$.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

This is shown by utilizing the following series of equalities similarly as in the proof of Proposition 1:

$$\begin{aligned}
&\nabla_{\mathbf{V}_{m_\ell}} \widehat{L}(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) \\
&= -\sum_{i_k \in \mathcal{I}} \mu^*_{i_k} \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}^*_{i_k} \mathbf{E}_{i_k}(\mathbf{U}^*_{i_k}, \mathbf{V}^*)]\right) - 2\epsilon^*_\ell \mathbf{V}^*_{m_\ell} \\
&\overset{(a)}{=} -\sum_{i_k \in \mathcal{A}} \mu^*_{i_k} \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}^*_{i_k} \mathbf{E}_{i_k}(\mathbf{U}^*_{i_k}, \mathbf{V}^*)]\right) - 2\epsilon^*_\ell \mathbf{V}^*_{m_\ell} \\
&\overset{(b)}{=} -\sum_{i_k \in \mathcal{A}} \mu^*_{i_k} \nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\widetilde{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widetilde{\mathbf{U}}_{i_k}, \mathbf{V}^*)]\right) - 2\epsilon^*_\ell \mathbf{V}^*_{m_\ell} \\
&\overset{(c)}{=} \nabla_{\mathbf{V}_{m_\ell}} \widehat{L}_{\mathcal{A}}(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}, \lambda^*; \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*) = 0
\end{aligned}$$

where in (a) (resp. in (c)) we have again used the fact that $\mu^*_{i_k} = 0$ for all $i_k \in \bar{\mathcal{A}}$, $\mathcal{A} \bigcup \bar{\mathcal{A}} = \mathcal{I}$, and the fact that $\nabla_{\mathbf{V}_{m_\ell}} \left(\mathrm{Tr}[\mathbf{W}^*_{i_k} \mathbf{E}_{i_k}(\mathbf{U}^*_{i_k}, \mathbf{V}^*)]\right)$ takes finite value for all feasible $\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*$ that satisfies the KKT system (30)–(35); (b) is due to (39); the last equality is due to the assumption that $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}}, \lambda^*)$ together with $(\boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$ satisfy (30).

Next, it is straightforward to see that when $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$, then for all $i_k \in \mathcal{I}$,

$$\begin{aligned}
&-\mathrm{Tr}[\mathbf{W}^*_{i_k} \mathbf{E}_{i_k}(\mathbf{U}^*_{i_k}, \mathbf{V}^*)] + \log\det(\mathbf{W}^*_{i_k}) + d_{i_k} - \lambda^* \\
&\geq -\mathrm{Tr}[\widetilde{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widetilde{\mathbf{U}}_{i_k}, \mathbf{V}^*)] + \log\det(\widetilde{\mathbf{W}}_{i_k}) + d_{i_k} - \lambda^*.
\end{aligned} \tag{40}$$

This result implies that the feasibility part of (34) is satisfied. In order to show that the complementarity part of (34) is also satisfied, it is sufficient to show that for all $i_k \in \mathcal{A}$ (i.e., for all i_k such that $\mu^*_{i_k} > 0$), (40) achieves strict equality. This is guaranteed by (39).

So far we have shown that $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*, \lambda^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*), \lambda^*)$ along with $(\boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$ satisfy the KKT system (30)–(35). The last step we need to show is this solution achieves the same objective value as $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}})$.

For this purpose, observe that due to the fact that $\sum_{i_k \in \mathcal{I}_k} \mu^*_{i_k} = 1$, we must have $|\mathcal{A}| > 0$. Due to complementarity condition (29), at least one of the constraints

$$-\mathrm{Tr}[\Upsilon_{i_k}(\mathbf{V}^*) \mathbf{E}_{i_k}(\Psi_{i_k}(\mathbf{V}^*), \mathbf{V}^*)] + \log\det(\Upsilon_{i_k}(\mathbf{V}^*)) + d_{i_k} - \lambda^* \geq 0$$

is active. This implies that $\min_{i_k \in \mathcal{I}} - \left(\mathrm{Tr}[\Upsilon_{i_k}(\mathbf{V}^*) \mathbf{E}_{i_k}(\Psi_{i_k}(\mathbf{V}^*), \mathbf{V}^*)] - \log\det(\mathbf{W}^*_{i_k}) - d_{i_k}\right) = \lambda^*$. Similarly, we must also have $\min_{i_k \in \mathcal{I}} - \left(\mathrm{Tr}[\widetilde{\mathbf{W}}_{i_k} \mathbf{E}_{i_k}(\widetilde{\mathbf{U}}_{i_k}, \mathbf{V}^*)] - \log\det(\widetilde{\mathbf{W}}_{i_k}) - d_{i_k}\right) = \lambda^*$.

We conclude that for every KKT solution $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}})$ of problem (Q1), $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*) = (\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*))$ is also a KKT solution, and it achieves the same objective value as $(\mathbf{V}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{W}})$.

Step 2: Now we are ready to argue the equivalence of problem (P1) and (Q1). Let us use $f(\mathbf{V})$ and $\bar{f}(\mathbf{V}, \mathbf{U}, \mathbf{W})$ to denote the objective value of problem (P1) and (Q1), respectively.

Firstly, we can check by simple substitution that for any given $\mathbf{V} \in \mathcal{V}$, $f(\mathbf{V}) = \bar{f}(\mathbf{V}, \Psi(\mathbf{V}), \Upsilon(\mathbf{V}))$.

Suppose $\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*$ is a global optimal solution of problem (Q1), but $\mathbf{V}^*$ is not a global optimal solution of (P1). Then there must exist a solution $\widetilde{\mathbf{V}}$ such that $f(\widetilde{\mathbf{V}}) > f(\mathbf{V}^*)$. From the first part of the

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

proof we have that $\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*)$ is also a KKT solution and it achieves the same objective value as $\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*$. Consequently $\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*)$ is also a global optimal solution for (Q1). Using the fact that $f(\mathbf{V}) = \bar{f}(\mathbf{V}, \Psi(\mathbf{V}), \Upsilon(\mathbf{V}))$, we conclude that

$$\bar{f}(\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*)) = f(\mathbf{V}^*) < f(\widetilde{\mathbf{V}}) = \bar{f}(\widetilde{\mathbf{V}}, \Psi(\widetilde{\mathbf{V}}), \Upsilon(\widetilde{\mathbf{V}})) \tag{41}$$

However this contradicts the global optimality of the solution $\mathbf{V}^*, \Psi(\mathbf{V}^*), \Upsilon(\mathbf{V}^*)$ for problem (Q1). The reverse direction can be argued similarly. ■

REFERENCES

[1] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook of Antennas in Wireless Communications*, 2001, cRC Press.

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

[3] H. Boche and M. Schubert, "Resource allocation in multiantenna systems-achieving max-min fairness by optimizing a sum of inverse SIR," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1990 – 1997, june 2006.

[4] D. Cai, T. Quek, and C. W. Tan, "A unified analysis of max-min weighted sinr for mimo downlink system," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3850 –3862, aug. 2011.

[5] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[6] 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.

[7] M. Hong and Z.-Q. Luo, "Signal processing and optimal resource allocation for the interference channel," *EURASIP E-Reference Signal Processing*, 2012, accepted, available at http://arxiv.org.

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

[9] ——, "Max-min fairness linear transceiver design for a multi-user MIMO interference channel," in *the Proceedings of the international conference on Communicaitons 2011*, 2011.

[10] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, 2008.

[11] E. Polak, J. O. Royset, and R. S. Womersley, "Algorithms with adaptive smoothing for finite minimax problems," *Journal of Optimization Theory and Applications*, vol. 119, pp. 459–484, 2003.

[12] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437 –1450, oct 1998.

[13] M. Razaviyayn, H. Baligh, A. Callard, and Z.-Q. Luo, "Joint transceiver design and user grouping in a MIMO interfering broadcast channel," *45th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2011.

[14] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18 – 28, jan. 2004.

[15] S. Serbetli and A. Yener, "Transceiver optimization for multiuser mimo systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 214 – 226, jan. 2004.

[16] 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 Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[17] C. W. Tan, M. Chiang, and R. Srikant, "Fast algorithms and performance bounds for sum rate maximization in wireless networks," in *The Proceedings of IEEE International Conference on Computer Communication (INFOCOM)*, april 2009, pp. 1350 –1358.

[18] A. Wiesel, Y. C. Eldar, and S. Shamai(Shitz), "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, 2006.

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

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

[20] J. Zander, “Distributed cochannel interference control in cellular radio systems,” *IEEE Transactions on Vehicular Technology*, vol. 41, no. 3, pp. 305 –311, aug 1992.

[21] ——, “Performance of optimal transmitter power control in cellular system,” *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, pp. 57 –63, feb 1992.

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

TABLE I
PSEUDO CODE OF THE PROPOSED ALGORITHM

Set $n = 0$. Initialize $\mathbf{V}^0, \mathbf{U}^0$, and $\mathbf{W}^0$ randomly such that the power budget constraints are satisfied.

repeat

  $\mathbf{V}^{n+1} \in \Phi(\mathbf{U}^n, \mathbf{W}^n)$

  $\mathbf{U}^{n+1} = \Psi(\mathbf{V}^{n+1})$

  $\mathbf{W}^{n+1} = \Upsilon(\mathbf{V}^{n+1})$

  $n \leftarrow n + 1$

until some convergence criterion is met

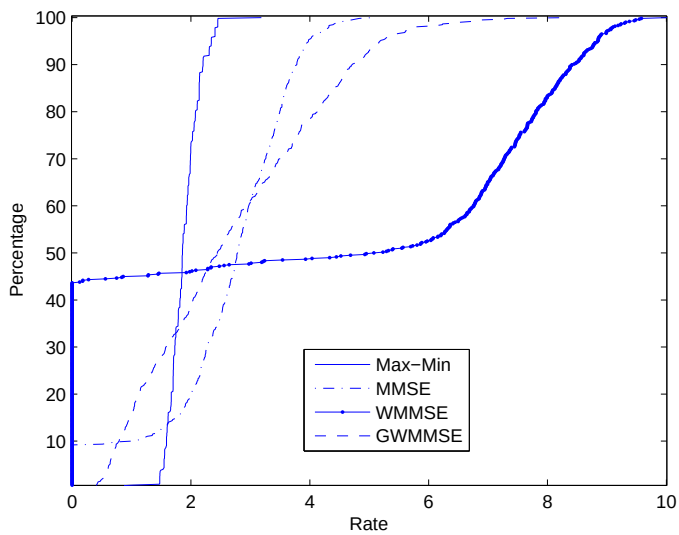

Fig. 1. Rate CDF: $K = 4, I = 3, M = 6, N = 2, d = 1$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

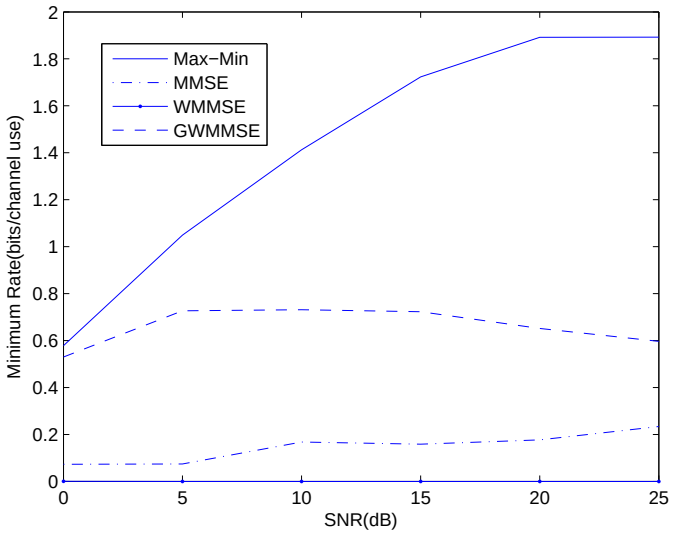

Fig. 2. Minimum rate in the system versus SNR: $K = 4, I = 3, M = 6, N = 2, d = 1$

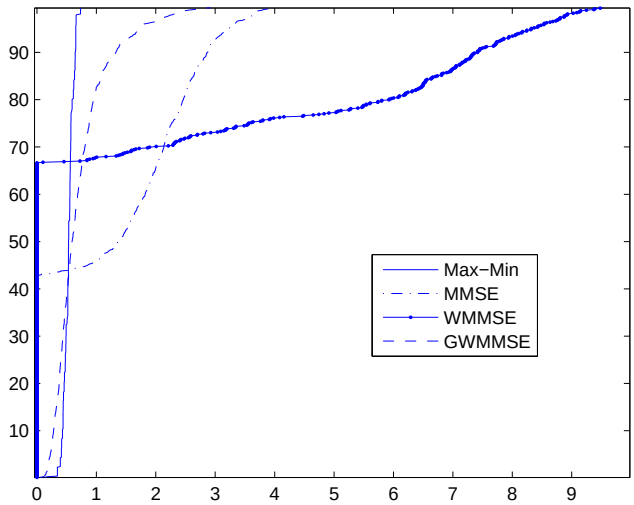

Fig. 3. Rate CDF: $K = 5, I = 3, M = 3, N = 2, d = 1$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

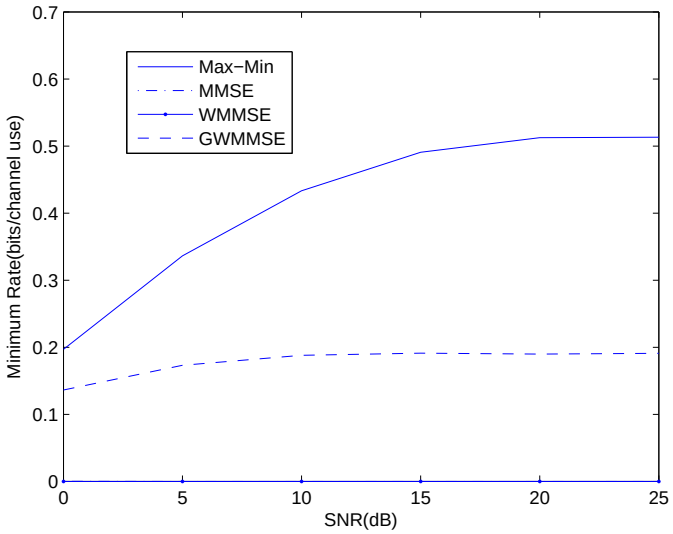

Fig. 4. Minimum rate in the system versus SNR: $K = 5, I = 3, M = 3, N = 2, d = 1$

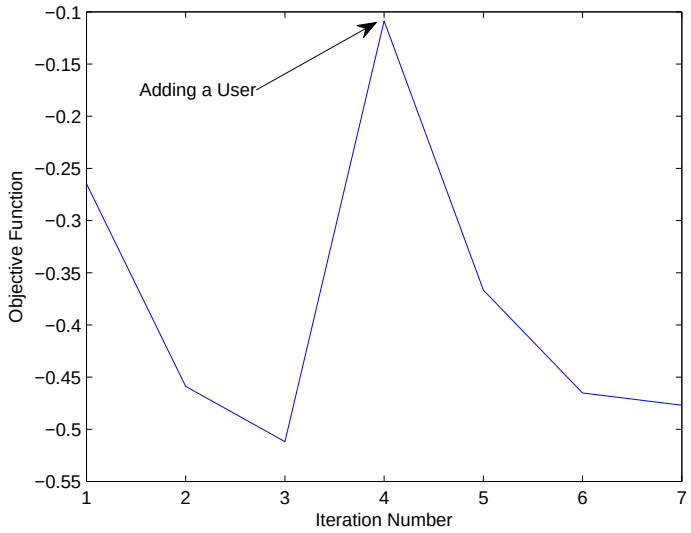

Fig. 5. WMMSE objective function while adding a User: $K = 5, M = 3, N = 2, d = 1$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

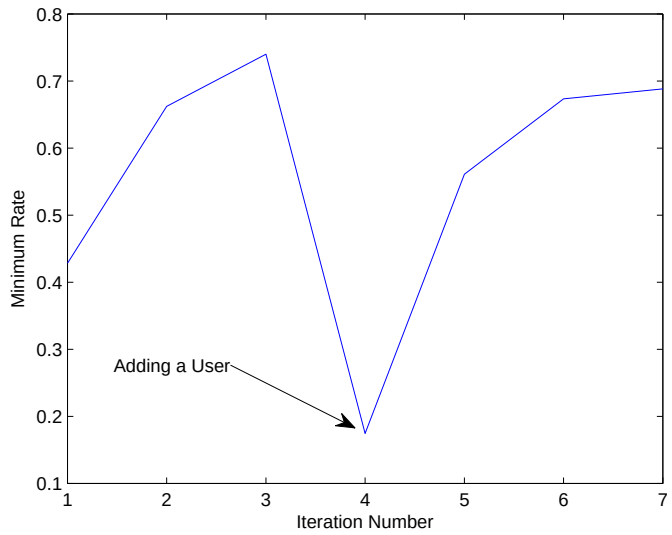

Fig. 6. Minimum rate while adding a User $K = 5, I = 3, M = 3, N = 2, d = 1$

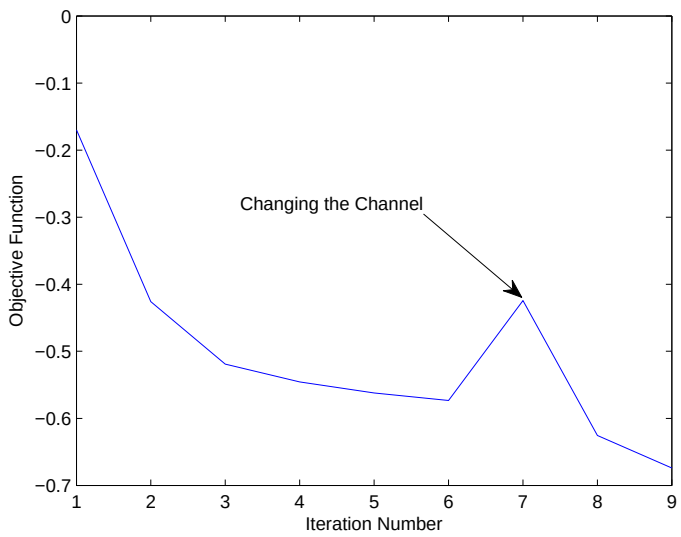

Fig. 7. WMMSE objective function while changing the channel: $K = 5, I = 2, M = 3, N = 2, d = 1$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.

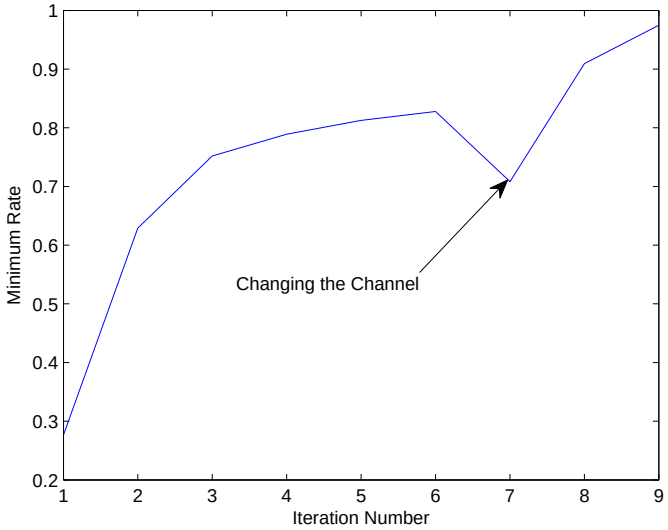

Fig. 8. Minimum rate while changing the channel: $K = 5, I = 2, M = 3, N = 2, d = 1$

This is a manuscript of an article from Signal Processing 93 (2013): 3327, doi: 10.1016/j.sigpro.2013.02.017. Posted with permission.