Optimization of a Class of Non-Convex Objectives on the Gaussian MIMO Multiple Access Channel: Algorithm Development and Convergence Analysis

Daniel Calabuig
Institute of Telecommunications and Multimedia Applications
Universidad Politécnica de Valencia
Valencia, Spain
Email: dacaso@iteam.upv.es

Ramy H. Gohary and Halim Yanikomeroglu
Department of Systems and Computer Engineering
Carleton University
Ottawa, Canada
Email: {gohary, halim}@sce.carleton.ca

***Abstract*—In this paper we develop an algorithm for computing the optimal transmission parameters, which include the transmission covariance, the time-shares and the user-orderings that minimize a particular class of objectives defined over the capacity region of Gaussian multiple antenna multiple access channels. This class includes objectives that are twice-differentiable, non-increasing and convex in the users' rates, but not necessarily convex in the aforementioned transmission parameters. As such, this class includes design objectives that are non-convex and that, without the proposed algorithm, are difficult to solve in general. The proposed algorithm is iterative with polynomial complexity per iteration and with convergence to the global optimal guaranteed. The utility of this algorithm is illustrated via a numerical example for maximizing proportional fairness.**

***Index Terms*—Multiple access channels, time-sharing, optimization, convergence analysis, proportional fairness**

I. Introduction

The Gaussian multiple access channel (GMAC) model arises in various uplink communication scenarios including cellular systems when multiple users communicate with a base station, and satellite systems when multiple ground stations communicate with a satellite [1]. The capacity region of general multiple access channels was obtained in [2] and [3]. Particularizing these results to Gaussian channels, it was shown that corner points on the boundary of the capacity region are achieved when the signal of each user is Gaussian distributed with an appropriate covariance and the receiver uses successive interference cancellation (SIC) to decode the users' signals sequentially [4]. Other points on the boundary of the capacity region can be obtained by time-sharing, whereby each decoding order and collection of users' covariance matrices are used during a fraction of the signalling duration.

Achieving particular points within the capacity region of the GMAC was considered in [5], [6] and [7], when perfect channel state information (CSI) is available at the receiver and the transmitters. In particular, in [5], an iterative water-filling algorithm is considered, which, with a sum power constraint per user, yields a rate vector on the sum-capacity facet and the input covariance matrices that achieve it. In [6], the input covariance matrices of all users are optimized to maximize the sum-capacity, and those matrices are subsequently used with time-sharing to achieve a fairness criterion. In [7], two cases are considered: the case of small number of users, which gives rise to a scenario in which time-sharing is feasible, and the case of large number of users, which gives rise to a scenario in which time-sharing is infeasible. In [7], fairness is not directly addressed. However, the points to be achieved are those at which the weighted sum of the rates of a given subset of users is maximized while the rates of the remaining users are restricted to prescribed values.

In this paper, we focus on the GMAC scenario with perfect CSI at the receiver and transmitters. In contrast with previous works, we consider the joint optimization of the transmission parameters, which include the transmission covariance, the time-shares and the user-orderings that minimize a particular class of objectives defined over the capacity region of multiple antenna GMACs. This class includes objectives that are twice-differentiable, nonincreasing and convex in the users' rates, but not necessarily convex in the aforementioned transmission parameters. As such, this class includes design objectives that are non-convex and that, without the proposed algorithm, are difficult to solve in general. The proposed algorithm is iterative with polynomial complexity per iteration and with convergence to the global optimal guaranteed. The utility of this algorithm is illustrated via a numerical example for maximizing proportional fairness among users.

II. System model and optimization

The GMAC is composed of K users transmitting to one base station. The number of transmit antennas of the k-th user is N_k, $k = 1, \ldots, K$, and the number of receive antennas at the base station is N_{R}. The received signal is given by

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{z},$$

where $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{R}} \times N_k}$ is the channel matrix of the k-th user and $\boldsymbol{x}_k \in \mathbb{C}^{N_k}$ is its transmitted signal. The Gaussian noise at the base station is denoted by $\boldsymbol{z} \in \mathbb{C}^{N_{\mathrm{R}}}$, which, without loss of generality, is assumed to satisfy $\mathrm{E}[\boldsymbol{z}\boldsymbol{z}^{\dagger}] = \boldsymbol{I}$.

978-1-4799-4903-8/14/$31.00 ©2014 IEEE

Let $\boldsymbol{Q}_k = \mathrm{E}[\boldsymbol{x}_k \boldsymbol{x}_k^\dagger]$ be the covariance matrix of the signal of user k, and let $\bar{\boldsymbol{Q}} = \boldsymbol{Q}_1 \oplus \cdots \oplus \boldsymbol{Q}_K$ be the composite covariance matrix, that is, $\bar{\boldsymbol{Q}}$ is block diagonal with the matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K$ along the diagonal. We consider systems with L power constraints which can be expressed as $g_\ell(\bar{\boldsymbol{Q}}) \leq 0,\ \ell = 1, \ldots, L$. Let

$$\mathcal{P} = \{\bar{\boldsymbol{Q}} | \bar{\boldsymbol{Q}} \succeq 0, g_\ell(\bar{\boldsymbol{Q}}) \leq 0,\ \ell = 1, \ldots, L\} \tag{1}$$

be the set of all feasible $\bar{\boldsymbol{Q}}$. This set can be assumed to be bounded, which implies that the transmit powers are finite.

Corner points of the GMAC capacity region can be achieved by using an SIC receiver. To achieve a particular corner, the receiver orders the users and decodes their signals sequentially. To decode the signal of a particular user, the receiver treats the signals of the users interfering with it as additive noise. After decoding, the signal of that user is stripped off from the signals of the remaining users. Since the interference observed in decoding the signal of a particular user depends on the ordering, maximizing a given objective requires the receiver to determine the optimal user ordering. Non-corner points of the GMAC capacity region can be achieved by time-sharing.

Let $\pi_1, \cdots, \pi_{K!}$ be the set of all $K!$ permutations, where $\pi_i(j)$ refers to the user in the j-th position of the i-th ordering. When the receiver uses π_i for decoding the users' signals, each user $k \in \{1, \ldots, K\}$ is able to achieve the following rate:

$$r_{ki}\left(\bar{\boldsymbol{Q}}\right) = \log \frac{\left|\boldsymbol{I} + \sum_{j \geq \pi_i^{-1}(k)} \boldsymbol{H}_{\pi_i(j)} \boldsymbol{Q}_{\pi_i(j)} \boldsymbol{H}_{\pi_i(j)}^\dagger\right|}{\left|\boldsymbol{I} + \sum_{j > \pi_i^{-1}(k)} \boldsymbol{H}_{\pi_i(j)} \boldsymbol{Q}_{\pi_i(j)} \boldsymbol{H}_{\pi_i(j)}^\dagger\right|}. \tag{2}$$

Our focus herein is to minimize a nonincreasing function $f : \mathbb{R}^K \to \mathbb{R}$, which is convex in the GMAC users' rates, but not necessarily convex in the transmission paramters. We consider the cases in which $\mathcal{P}$ is convex in the input covariance martrices. For those cases, it was shown in [8] that, to find the optimal transmission parameters, it suffices to find a vector $\boldsymbol{\beta}^* \in \mathbb{R}^{K!}$ and a composite covariance matrix $\bar{\boldsymbol{Q}}^*$ that solve

$$\min_{\boldsymbol{\beta} \in \mathcal{S}, \bar{\boldsymbol{Q}} \in \mathcal{P}} \quad f\left(\boldsymbol{\rho}\left(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}\right)\right), \tag{3}$$

where $\mathcal{S}$ is the $K!$ dimensional unit simplex, i.e.,

$$\mathcal{S} \triangleq \left\{\boldsymbol{\beta} \in \mathbb{R}^{K!} \middle| \sum_{i=1}^{K!} \beta_i = 1,\ \beta_i \geq 0,\ \forall i\right\} \tag{4}$$

and the k-th entry of $\boldsymbol{\rho}\left(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}\right)$ is given by

$$\rho_k\left(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}\right) = \sum_{i=1}^{K!} \beta_i r_{ki}\left(\bar{\boldsymbol{Q}}\right). \tag{5}$$

III. Algorithm and Convergence Analysis

In this section we develop and analyze an algorithm that solves (3). This algorithm relies on the following result [8].

Lemma 1. *For the problem in* (3)*, let* $w_0 = 0$ *and let*

1) *the objective* f *be continuously differentiable and convex in the users' rates (not necessarily convex in* $\boldsymbol{\beta}$ *and* $\bar{\boldsymbol{Q}}$*);*
2) *the weights* $\{w_k\}_{k=1}^K$ *be given by*

$$w_k = -\left.\frac{\partial f(\boldsymbol{x})}{\partial x_k}\right|_{\boldsymbol{x} = \boldsymbol{\rho}(\boldsymbol{\beta}^*, \bar{\boldsymbol{Q}}^*)} \qquad k = 1, \ldots, K; \quad \textit{and} \tag{6}$$

3) *the users be labelled so that* $w_1 \leq \cdots \leq w_K$.

Then, the time-sharing vector $\boldsymbol{\beta}^*$ *and the composite covariance matrix* $\bar{\boldsymbol{Q}}^*$ *are optimum if and only if for each strictly positive element of* $\boldsymbol{\beta}^*$*, say* β_i^**,*

1) *the permutation* $\pi_i(\cdot)$ *follows the increasing order of the weights* $\{w_k\}_{k=1}^K$*, and*
2) *the composite covariance matrix* $\bar{\boldsymbol{Q}}^*$ *solves*

$$\max_{\bar{\boldsymbol{Q}} \in \mathcal{P}} \sum_{k=1}^{K} (w_k - w_{k-1}) \log\left|\boldsymbol{I} + \sum_{j \geq k} \boldsymbol{H}_j \boldsymbol{Q}_j \boldsymbol{H}_j^\dagger\right|. \tag{7}$$

First, we note that this result does not impose any constraints on $\mathcal{P}$. For instance, $\mathcal{P}$ can be nonconvex. Second, we note that this result cannot be readily used to obtain the optimal transmission parameters, $\boldsymbol{\beta}^*$ and $\bar{\boldsymbol{Q}}^*$. This is because the gradient of the objective at the optimum rate vector, and hence the optimal weights in (6), are not known *a priori*.

A. Proposed algorithm

The proposed algorithm solves (3) iteratively. At each iteration t, the algorithm uses $\bar{\boldsymbol{Q}}^*(t-1)$ and $\boldsymbol{\beta}^*(t-1)$, obtained at the previous iteration, to obtain $\bar{\boldsymbol{Q}}^*(t)$ and $\boldsymbol{\beta}^*(t)$. In particular, the algorithm uses (6) to compute the weights $\{w_k(t)\}_{k=1}^K$ at the rate vector $\boldsymbol{x} = \boldsymbol{\rho}(\boldsymbol{\beta}^*(t-1), \bar{\boldsymbol{Q}}^*(t-1))$. Subsequently, Lemma 1 is used to find the vector $\boldsymbol{\beta}^*(t)$ and the composite covariance matrix $\bar{\boldsymbol{Q}}^*(t)$. We will show, in Section III-B, that $f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big)$ converges to the optimum of (3).

It remains to show how $\boldsymbol{\beta}^*(t)$ and $\bar{\boldsymbol{Q}}^*(t)$ are obtained. We begin by obtaining $\bar{\boldsymbol{Q}}^*(t)$. To do that, let $\bar{\boldsymbol{Q}}^o(t)$ be the solution of (7) for the weights $\{w_k(t)\}_{k=1}^K$. We will choose $\bar{\boldsymbol{Q}}^*(t)$ to be a convex combination of $\bar{\boldsymbol{Q}}^*(t-1)$ and $\bar{\boldsymbol{Q}}^o(t)$. In particular,

$$\bar{\boldsymbol{Q}}^*(t) = \varepsilon(t)\bar{\boldsymbol{Q}}^o(t) + (1 - \varepsilon(t))\bar{\boldsymbol{Q}}^*(t-1), \tag{8}$$

where $\varepsilon(t)$ solves

$$\min_{\delta \in [0,1]} f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t-1), \delta\bar{\boldsymbol{Q}}^o(t) + (1-\delta)\bar{\boldsymbol{Q}}^*(t-1)\big)\big). \tag{9}$$

Despite being generally non-convex, this optimization problem can be easily solved by exhaustive search over $\delta \in [0, 1]$.

From the definition of $\bar{\boldsymbol{Q}}^*(t)$ and $\varepsilon(t)$, we have

$$f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t-1), \bar{\boldsymbol{Q}}^*(t)\big)\big) \leq f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t-1), \bar{\boldsymbol{Q}}^*(t-1)\big)\big),\ \forall t. \tag{10}$$

After computing $\bar{\boldsymbol{Q}}^*(t)$, the vector $\boldsymbol{\beta}^*(t)$ is chosen to solve the following convex optimization problem,

$$\min_{\boldsymbol{\beta} \in \mathcal{S}} \quad f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}^*(t)\big)\big). \tag{11}$$

Hence, it can be seen that

$$f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) \leq f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t-1), \bar{\boldsymbol{Q}}^*(t)\big)\big) \tag{12}$$

which, with (10), imply that the value of the objective is reduced at every iteration. Boundedness of the objective imply that the algorithm, which is summarized in Algorithm 1 below, converges. The optimality of the point to which the algorithm converges will be shown in Section III-B.

Algorithm 1 Computing optimum transmission parameters.

1: Initialize $\boldsymbol{\beta}^*(0)$ and $\bar{\boldsymbol{Q}}^*(0)$.
2: **for** $t = 1, 2, \ldots$ **do**
3: Compute a vector of weights, $\{w_k(t)\}_{k=1}^K$, as in (6) at the rate vector $\boldsymbol{x} = \boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t-1), \bar{\boldsymbol{Q}}^*(t-1)\big)$.
4: Solve the convex problem in (7) with $\{w_k(t)\}_{k=1}^K$ to obtain $\bar{\boldsymbol{Q}}^o(t)$.
5: Use (8) to update the covariance matrices at iteration t.
6: Find the optimum vector of (11), $\boldsymbol{\beta}^*(t)$.
7: **end for**

B. Convergence analysis

In this section, we will show that Algorithm 1 is guaranteed to converge to the global optimal of (3), provided that the conditions of the following theorem are satisfied.

Theorem 1. *Suppose that Algorithm 1 is used to solve the optimization problem in* (3)*, i.e.,* $\min_{\boldsymbol{\beta}\in\mathcal{S},\bar{\boldsymbol{Q}}\in\mathcal{P}} f\left(\boldsymbol{\rho}\left(\boldsymbol{\beta},\bar{\boldsymbol{Q}}\right)\right)$*, and suppose that the following conditions are satisfied:*

1) *The power constraint set $\mathcal{P}$ is convex in $\bar{\boldsymbol{Q}}$;*
2) *The objective f is second order differentiable, monotonically nonincreasing , and convex in the users' rates, $\boldsymbol{\rho}$, but not necessarily convex in $\boldsymbol{\beta}$ and $\bar{\boldsymbol{Q}}$.*

It follows that Algorithm 1 converges to the optimum $\boldsymbol{\beta}$ and $\bar{\boldsymbol{Q}}$, that is, if $\boldsymbol{x}^*$ is the optimum rate vector, then

$$\lim_{t\to\infty} f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) - f(\boldsymbol{x}^*) = 0. \tag{13}$$

Proof: The proof hinges on the following result.

Proposition 1. *Let $\boldsymbol{x}^*$ be the optimum rate vector. Then, the following inequality holds:*

$$f(\boldsymbol{x}^*) \geq f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) + \sum_{k=1}^{K} w_k(t+1)\big(\rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big) - \rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^o(t+1)\big)\big). \tag{14}$$

The proof of Proposition 1 and the rest of the proof of the theorem are provided in Appendices A and B respectively. ■

Theorem 1 identifies a class of problems for which Algorithm 1 converges to the global optimum solution. In Section IV we will use this algorithm to solve an instance of such problems. Theorem 1 tightens the results in [8], where the convergence of the algorithm is guaranteed, but not the convergence to the global optimal.

IV. Numerical example

We now provide an instance of (3) in which Algorithm 1 is used to solve a two-user GMAC optimization problem with an objective that is nonconvex in the transmission parameters. Each user has two transmit antennas and a power budget of $P = 10$ dB. The destination has two receive antennas and the channel matrices are given by

$$\boldsymbol{H}_1 = \begin{pmatrix} 0.32 & -0.06 \\ -0.72 & -0.88 \end{pmatrix} + \jmath \begin{pmatrix} -0.15 & -1.38 \\ -1.34 & -0.01 \end{pmatrix},$$

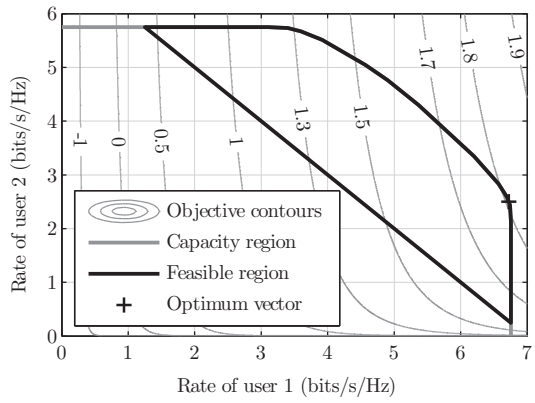

Fig. 1. Feasible rate region and optimum rate vector of the example of Section IV. The capacity region is shown in grey color.

$$\boldsymbol{H}_2 = \begin{pmatrix} -0.21 & 0.29 \\ -0.08 & 0.91 \end{pmatrix} + \jmath \begin{pmatrix} -0.65 & 0.15 \\ 0.13 & 1.39 \end{pmatrix}.$$

For these channel matrices, it can be readily verified that the GMAC capacity region is the one shown in Figure 1.

Our goal is to maximize the weighted proportional fairness [9], provided that the sum rate exceeds a given threshold R. In this case, the optimum transmission parameters solve

$$\max_{\boldsymbol{\beta},\bar{\boldsymbol{Q}}} \quad w_1 \log\left(\rho_1\left(\boldsymbol{\beta},\bar{\boldsymbol{Q}}\right)\right) + w_2 \log\left(\rho_2\left(\boldsymbol{\beta},\bar{\boldsymbol{Q}}\right)\right), \tag{15a}$$

$$\text{subject to} \quad \beta_1 + \beta_2 = 1,\ \beta_i \geq 0,\ i = 1, 2, \tag{15b}$$

$$\boldsymbol{Q}_k \succeq 0,\ \mathrm{Tr}\left(\boldsymbol{Q}_k\right) \leq P,\ k = 1, 2, \tag{15c}$$

$$\log\left|\boldsymbol{I} + \boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^\dagger + \boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^\dagger\right| \geq R. \tag{15d}$$

We note that this problem is highly non-convex in $\boldsymbol{\beta}$ and $\bar{\boldsymbol{Q}}$. However, since the objective is to maximize a concave function in the users' rates, Lemma 1 can be used to characterize the optimum transmission parameters. Unfortunately, although the objective is nonincreasing, its second order derivative is not bounded when the rates approach zero. To circumvent this difficulty, we add the constraints $\mathrm{Tr}\left(\boldsymbol{Q}_k\right) \geq \gamma$, $k = 1, 2$, for a small $\gamma > 0$. These constraints ensure that the conditions of Theorem 1 are satisfied and hence, Algorithm 1 is guaranteed to converge to the optimum solution of (15).

Figure 1 shows the feasible rate region when $R = 7$ bits/s/Hz and the objective contours for $w_1 = 0.9$ and $w_2 = 0.1$. The rate vector that maximizes the weighted proportional fairness is marked by '+'.

Figure 2 illustrates the convergence behaviour of the proposed algorithm. In particular, this figure shows the upper bound (18) on the error at each iteration. As shown in this figure, five iterations suffice for the error to be less than 10^{-8}.

Another instance of (3), which goal is to minimize the total completion time, and that can be solved with Algorithm 1, is provided in [8].

V. Conclusion

In this paper we considered the optimization of a class of non-linear objective functions defined over the multiple

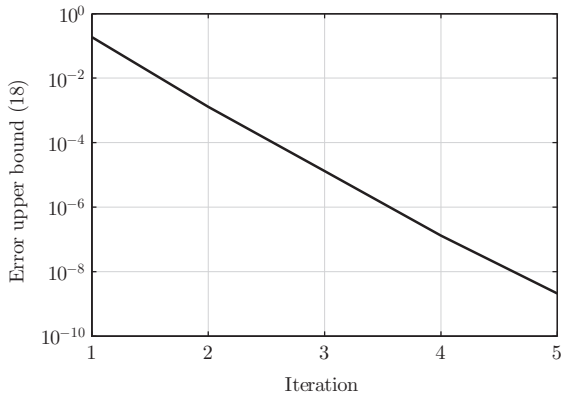

Fig. 2. Error upper bound (cf. (18)) after each iteration.

antenna GMAC capacity region. We developed an efficient algorithm for computing the optimal transmission parameters, which include the transmission covariance, the time-shares and the user-orderings. This class includes objectives that are twice-differentiable, nonincreasing and convex in the users' rates, but not necessarily convex in the transmission parameters. As such, this class includes design objectives that are non-convex and that, without the proposed algorithm, are difficult to solve in general. The proposed algorithm is iterative with polynomial complexity per iteration and with convergence to the global optimal guaranteed. The utility of this algorithm is illustrated via a numerical example for maximizing proportional fairness among users.

Appendix A
Proof of Proposition 1

Since f is convex, we have that [10, Proposition B.3]

$$f(\boldsymbol{x}^*) - f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) \geq \sum_{k=1}^{K} \big(x_k^* - \rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) \frac{\partial f(\boldsymbol{x})}{\partial x_k}\bigg|_{\boldsymbol{x}=\boldsymbol{\rho}\left(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\right)},$$

and, recalling that the weights $\{w_k(t+1)\}_{k=1}^K$ are computed as in (6) at $\boldsymbol{x} = \boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)$, we conclude that

$$f(\boldsymbol{x}^*) - f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) \geq \sum_{k=1}^{K} \big(\rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big) - x_k^*\big) w_k(t+1). \quad (16)$$

Since $\bar{\boldsymbol{Q}}^o(t+1)$ is the composite covariance matrix that solves (7) for the weights $\{w_k(t+1)\}_{k=1}^K$ (cf. step 2 of the algorithm in Section III-A), and the orderings used in $\boldsymbol{\beta}^*(t)$ follow the increasing order of these weights, it follows that

$$\sum_{k=1}^{K} w_k(t+1) x_k^* \leq \sum_{k=1}^{K} w_k(t+1) \rho_k\left(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^o(t+1)\right). \quad (17)$$

Note that the right hand side of (17) is the maximum of the objective in (7) at iteration $t+1$. Substituting (17) into the right hand side of (16), yields (14).

Appendix B
Proof of Theorem 1

To prove this theorem we will use Proposition 1 to upper bound the difference of the objective in each iteration and at the optimum. In particular, from (14) we have that

$$\sum_{k=1}^{K} w_k(t+1)\big(\rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^o(t+1)\big) - \rho_k\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) \geq f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) - f(\boldsymbol{x}^*) \geq 0, \quad (18)$$

where the second inequality follows from the optimality of $\boldsymbol{x}^*$. We will proof the theorem by showing that the left hand side of (18) goes to zero as t goes to ∞. To do so, we define

$$h(\delta, t) = \sum_{k=1}^{K} w_k(t+1) \rho_k\big(\boldsymbol{\beta}^*(t), \delta \bar{\boldsymbol{Q}}^o(t+1) + (1-\delta)\bar{\boldsymbol{Q}}^*(t)\big).$$

Using this notation, the left hand side of (18) can be expressed as $h(1,t) - h(0,t)$. The proof of the theorem will be completed in two steps. First, we will show that $h(1,t) - h(0,t)$ goes to zero if $\frac{\partial h(\delta,t)}{\partial \delta}\big|_{\delta=0}$ goes to zero. Second, we will show that this derivative goes to zero as t goes to ∞.

Step 1: Recall that, in Algorithm 1, $\boldsymbol{\beta}^*(t)$ solves (11), which is a special case of the problem in (3). As such, $\boldsymbol{\beta}^*(t)$ must satisfy the first condition of Lemma 1 for the weights $\{w_k(t+1)\}_{k=1}^K$. Now, we use the following result.

Lemma 2. *Let $w_1 \leq \cdots \leq w_K$, and let $\beta_i > 0$ only if the i-th ordering follows the increasing order of these weights. Then,*

$$\sum_{k=1}^{K} w_k \rho_k(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}) = \sum_{k=1}^{K} (w_k - w_{k-1}) \log\Big|\boldsymbol{I} + \sum_{j \geq k} \boldsymbol{H}_j \boldsymbol{Q}_j \boldsymbol{H}_j^\dagger\Big|. \quad (19)$$

Proof: Let i_o be the identity ordering, that is, $\pi_{i_o}(k) = k$, $k = 1, \ldots, K$, and let $\boldsymbol{\beta}^o$ be a vector such that $\beta_{i_o}^o = 1$ and $\beta_i^o = 0$ for all $i \neq i_o$. Then, it can be shown that (19) holds for $\boldsymbol{\beta} = \boldsymbol{\beta}^o$. In the rest of cases, wherein $\beta_i > 0$ for some ordering $i \neq i_o$ that follows the increasing order of the weights, it can be shown that the weighted sum rate for the two orderings is equal, that is, $\sum_{k=1}^K w_k r_{ki}(\bar{\boldsymbol{Q}}) = \sum_{k=1}^K w_k r_{ki_o}(\bar{\boldsymbol{Q}})$. Then, using (5) and the fact that $\sum_{i=1}^{K!} \beta_i = 1$, we have that $\sum_{k=1}^K w_k \rho_k(\boldsymbol{\beta}, \bar{\boldsymbol{Q}}) = \sum_{k=1}^K w_k r_{ki_o}(\bar{\boldsymbol{Q}}) = \sum_{k=1}^K w_k \rho_k(\boldsymbol{\beta}^o, \bar{\boldsymbol{Q}})$. Hence, the equality in (19) holds. ■

Then, assuming that $w_1(t+1) \leq \cdots \leq w_K(t+1)$ and using (19), $h(\delta, t)$ can be expressed as

$$h(\delta, t) = \sum_{k=1}^{K} (w_k(t+1) - w_{k-1}(t+1)) \log\Big|\boldsymbol{I} + \sum_{j \geq k} \boldsymbol{H}_j (\delta \boldsymbol{Q}_j^o(t+1) + (1-\delta)\boldsymbol{Q}_j^*(t)) \boldsymbol{H}_j^\dagger\Big|, \quad (20)$$

where $w_0(t+1) = 0$. The function $h(\delta, t)$ can be shown to be concave in δ. Since $\bar{\boldsymbol{Q}}^o(t+1)$ solves (7) for the weights $\{w_k(t+1)\}_{k=1}^K$, then $\delta = 1$ solves $\max_{\delta \in [0,1]} h(\delta, t)$.

Then, from the concavity of h [10, Proposition B.3] and the optimality of $\delta = 1$, we have that

$$\frac{\partial h(\delta,t)}{\partial \delta}\Big|_{\delta=0} \geq h(1,t) - h(0,t) \geq 0, \tag{21}$$

which implies that $\lim_{t\to\infty} h(1,t) - h(0,t) = 0$ if $\lim_{t\to\infty} \frac{\partial h(\delta,t)}{\partial \delta}\big|_{\delta=0} = 0$.

Step 2: To show that $\lim_{t\to\infty} \frac{\partial h(\delta,t)}{\partial \delta}\big|_{\delta=0} = 0$, we begin by noting that the convexity of f implies that $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) + (\boldsymbol{x}^* - \boldsymbol{x})^\intercal \nabla f(\boldsymbol{x})$ for all $\boldsymbol{x}$. Since f is continuously differentiable and the capacity region is bounded, then the norms of $\boldsymbol{x}^*$, $\boldsymbol{x}$, and $\nabla f(\boldsymbol{x})$ are finite. This implies that $f(\boldsymbol{x}^*)$ is bounded below if $f(\boldsymbol{x}) > -\infty$ for at least one $\boldsymbol{x}$. From (10) and (12), and the boundedness of f, Algorithm 1 must converge to a point. Our goal now is to show that this point is optimal. To do so, we note that

$$\lim_{t\to\infty} f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) - f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t+1), \bar{\boldsymbol{Q}}^*(t+1)\big)\big) = 0. \tag{22}$$

This, along with (10) and (12), implies that

$$\lim_{t\to\infty} f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\big)\big) - f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t+1)\big)\big) = 0. \tag{23}$$

To proceed with the analysis, we define the function

$$\hat{h}(\delta,t) = f\big(\boldsymbol{\rho}\big(\boldsymbol{\beta}^*(t), \delta\bar{\boldsymbol{Q}}^o(t+1) + (1-\delta)\bar{\boldsymbol{Q}}^*(t)\big)\big). \tag{24}$$

Since $\bar{\boldsymbol{Q}}^*(t+1) = \varepsilon(t+1)\bar{\boldsymbol{Q}}^o(t+1) + (1-\varepsilon(t+1))\bar{\boldsymbol{Q}}^*(t)$, cf. (8), then, (23) can be expressed as

$$\lim_{t\to\infty} \hat{h}(0,t) - \hat{h}(\varepsilon(t+1),t) = 0. \tag{25}$$

Using this observation in computing $\frac{\partial \hat{h}(\delta,t)}{\partial \delta}\Big|_{\delta=0}$ yields

$$\frac{\partial \hat{h}(\delta,t)}{\partial \delta}\Big|_{\delta=0} = \sum_{k=1}^{K} \frac{\partial f(\boldsymbol{x})}{\partial x_k}\Big|_{\boldsymbol{x}=\boldsymbol{\rho}\left(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\right)} \times \frac{\partial \rho_k\big(\boldsymbol{\beta}^*(t), \delta\bar{\boldsymbol{Q}}^o(t+1) + (1-\delta)\bar{\boldsymbol{Q}}^*(t)\big)}{\partial \delta}\Big|_{\delta=0}. \tag{26}$$

Recall that (cf. Algorithm 1) for $k = 1, \ldots, K$, $w_k(t+1) = -\frac{\partial f(\boldsymbol{x})}{\partial x_k}\Big|_{\boldsymbol{x}=\boldsymbol{\rho}\left(\boldsymbol{\beta}^*(t), \bar{\boldsymbol{Q}}^*(t)\right)}$. Hence,

$$0 \leq \frac{\partial h(\delta,t)}{\partial \delta}\Big|_{\delta=0} = -\frac{\partial \hat{h}(\delta,t)}{\partial \delta}\Big|_{\delta=0}, \tag{27}$$

where the first inequality follows from (21). Using (27), to show that $\lim_{t\to\infty} \frac{\partial h(\delta,t)}{\partial \delta}\big|_{\delta=0} = 0$ it suffices to show that $\lim_{t\to\infty} \frac{\partial \hat{h}(\delta,t)}{\partial \delta}\big|_{\delta=0} = 0$. We will proceed by contradiction. In particular, suppose that $\lim_{t\to\infty} \frac{\partial \hat{h}(\delta,t)}{\partial \delta}\big|_{\delta=0} \neq 0$, then there exists a $\gamma > 0$ such that for all t_0 we can find some $t \geq t_0$ with $\frac{\partial \hat{h}(\delta,t)}{\partial \delta}\big|_{\delta=0} \leq -\gamma$. Since the second order derivative of f is bounded, we can find $D > 0$ such that, for all δ and t, $\frac{\partial^2 \hat{h}(\delta,t)}{\partial \delta^2} \leq D$. Integrating both sides of this inequality twice and using the fact that $\frac{\partial \hat{h}(\delta,t)}{\partial \delta}\Big|_{\delta=0} \leq -\gamma$ yields

$$\hat{h}(\delta,t) \leq \frac{D}{2}\delta^2 - \gamma\delta + \hat{h}(0,t), \tag{28}$$

Recalling that $\varepsilon(t)$ solves (9), we have $\hat{h}(\varepsilon(t+1),t) \leq \hat{h}(\delta,t)$ for all $\delta \in [0,1]$. Then, from (28),

$$\hat{h}(\varepsilon(t+1),t) - \hat{h}(0,t) \leq \min_{\delta\in[0,1]} \frac{D}{2}\delta^2 - \gamma\delta. \tag{29}$$

Note that $\delta = 0$ cannot minimize the right hand side of (29). Since the derivative at $\delta = 0$ equals $-\gamma$ which is strictly negative, we conclude that

$$\min_{\delta\in[0,1]} \frac{D}{2}\delta^2 - \gamma\delta < 0. \tag{30}$$

Using (25) and recalling that $\hat{h}(\varepsilon(t+1),t) \leq \hat{h}(0,t)$, we have that, for all $\lambda > 0$, there exists a t_0 such that for all $t \geq t_0$

$$\hat{h}(0,t) - \hat{h}(\varepsilon(t+1),t) < \lambda. \tag{31}$$

Using (30), we can take $\lambda = -\big(\min_{\delta\in[0,1]} \frac{D}{2}\delta^2 - \gamma\delta\big)$, which implies that the inequality in (31) contradicts the one in (29), which shows that indeed $\lim_{t\to\infty} \frac{\partial \hat{h}(\delta,t)}{\partial \delta}\big|_{\delta=0} = 0$.

Acknowledgments

The work of the first author was supported by a Marie Curie International Outgoing Fellowship (IOF) of the European Commission under the project COMIC (253990). The work of the second and third authors is supported in part by Huawei Canada Co., Ltd., and in part by the Ontario Ministry of Economic Development and Innovations ORF-RE (Ontario Research Fund-Research Excellence) program.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*, (Tsahkadsor, Armenia, USSR), pp. 103–105, Sept. 1971.

[3] H. Liao, *Multiple Access Channels*. PhD thesis, Univ. Hawaii, 1972.

[4] A. D. Wyner, "Recent results in the shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, pp. 2–10, Jan. 1974.

[5] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 145–152, Jan. 2004.

[6] M. A. Maddah-Ali, A. Mobasher, and A. Khandani, "Fairness in multiuser systems with polymatroid capacity region," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2128–2138, May 2009.

[7] W. Mesbah and H. Alnuweiri, "Joint rate, power, and decoding order optimization of multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 1681–1686, June 2011.

[8] D. Calabuig, R. Gohary, and H. Yanikomeroglu, "Optimum transmission through the Gaussian multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory*, (Istanbul, Turkey), pp. 201–205, July 2013.

[9] F. P. Kelly, A. K. Malullo, and D. K. H. Tan, "Rate control in communications networks: Shadow prices, proporotional fairness and stablity," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, Mar. 1998.

[10] D. P. Bertsekas, *Nonlinear Programming*. Nashua, NH: Athena Scientific, 2nd ed., 1999.