Energy Efficiency in MIMO Interference Channels: Social Optimality and Max-Min Fairness

Yang Yang[1] and Marius Pesavento[2]
1. Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, L-1855 Luxembourg.
2. Communication Systems Group, Technische Universität Darmstadt, Darmstadt 64283, Germany.
Email: yang.yang@uni.lu, pesavento@nt.tu-darmstadt.de

***Abstract*—In this paper, we consider the energy efficiency optimization problem in MIMO multi-cell systems, where all users suffer from inter-cell interference. To solve this multi-objective optimization problem, we consider both socially optimal solutions and max-min fairness solutions. We propose two novel iterative algorithms that converge to socially optimal solutions and max-min fairness solutions. The proposed algorithms have the following advantages: 1) fast convergence as the structure of the original optimization problem is preserved as much as possible in the approximate problem solved in each iteration, and 2) efficient implementation as each approximate problem is natural for parallel computation and/or its solution has a closed-form expression. The advantages of the proposed algorithms are also illustrated numerically.**

***Index Terms*—Energy Efficiency, Interference Channel, MIMO, Successive Pseudoconvex optimization**

I. Introduction

With the advent of 5G by 2020, the number of connected devices is predicted to reach 50 billions with a targeted 10-fold increase of the data rates. The increase in the data rate is expected to be achieved at the same or even a lower level of energy consumption. Therefore the so-called energy efficiency (EE) of the network is a key performance indicator that attracts extensive attention and imposes stringent requirements on transmission schemes enhancing the EE.

In this paper, we adopt the definition of EE as the ratio between achievable transmission rate and the consumed power, and study the EE optimization problem in a downlink MIMO multi-cell system, where the users served by different base station (BS) are sharing the same frequency resources and suffer from inter-cell interference. Since each BS has its own energy efficiency, the joint optimization of all BSs' energy efficiency is essentially a multi-objective optimization problem, which is usually addressed from three aspects: i) global energy efficiency (GEE), which is the ratio between the sum transmission rate and the total consumed power; ii) sum energy efficiency (SEE), also known as the socially optimal solution, which is the sum of all BSs' individual energy efficiency; and iii) max-min fairness, where the minimum energy efficiency among all BSs is maximized.

Since a special instance of the GEE and the SEE, namely, sum rate maximization in such an interference-limited system, is a nonconvex problem and NP-hard [1], most studies (with [2] as an exception) focus on efficient iterative algorithms that can find a stationary point. For the GEE maximization problem, many algorithms are developed, see [2–8] and the references therein. By comparison, less attention has been paid to the SEE maximization and max-min fairness problems, see [9–12] and [2, 13] and the references therein.

The SEE maximization problem in interference-limited systems is more challenging than the GEE maximization problem, because the SEE is a sum of multiple nonconvex fractional functions while the GEE is a single nonconvex fractional function, see [9, 11, 12]. The iterative algorithm proposed in [9] has two limitations. Firstly, it has a high complexity because it consists of two layers while the inner layer is an iterative block coordinate descent (BCD) type algorithm which typically converges after many iterations. Secondly, only convergence in function value is established and the convergence to a stationary point is left open. The sequential pricing algorithm proposed in [11] is a variant of the BCD algorithm. On the one hand, it is designed for MISO systems. On the other hand, the optimization problem solved in each iteration does not exhibit any convexity, making the iterative algorithm not suitable for practical implementation. In [12], both centralized and distributed algorithms are proposed, which, however, are only applicable for MISO systems.

In [2, 13], the max-min fairness solution is studied, and an iterative algorithm is proposed for the MISO and MIMO systems, respectively. The algorithms may have a high complexity, because the approximate problem solved in each iteration does not have structures that can be exploited to enable efficient parallel computation. Instead, it is solved by general purpose convex optimization solvers and the performance may not scale well with respect to the problem dimension.

To fill the above gap, we propose in this paper two novel iterative algorithms for the EE maximization problem in a MIMO multi-cell interference-limited system, one to achieve the socially optimal solution and one to achieve the max-min fairness solution. The proposed algorithms have the following notable features: 1) guaranteed convergence to a stationary point/Karush-Kuhn-Tucker (KKT) point; 2) a parallel (Jacobi) best-response type implementation with typically very fast convergence; 3) the approximate problem solved in each iteration either admits a closed-form solution or can easily be solved. In particular, the iterative algorithm achieving social optimality is based on the recently developed successive pseudoconvex approximation framework [7], where the approximate problem solved in each iteration may not even pseudoconvex. The iterative algorithm achieving max-min fairness is based on the successive upper bound minimization and the proximal point algorithm and can easily be implemented by standard dual decomposition techniques.

II. Problem Model

We consider a MIMO downlink multi-cell system where the BS in each cell is serving a single user on given frequency resource and all BSs are operating in the same frequency band. We assume the number of cells is K. We denote $\mathbf{H}_{kj}$ as the channel coefficient from the BS of cell j to user k. Assume the inter-user interference is treated as noise, the downlink transmission rate of the k-th user is:

$$r_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) = \log\det\left(\mathbf{I} + \mathbf{R}_k(\mathbf{Q}_{-k})^{-1}\mathbf{H}_{kk}\mathbf{Q}_k\mathbf{H}_{kk}^H\right), \quad (1)$$

where $\mathbf{Q}_k$ is BS k's transmit covariance matrix, $\mathbf{Q}_{-k} \triangleq (\mathbf{Q}_j)_{j\neq k}$, and $\mathbf{R}_k(\mathbf{Q}_{-k}) \triangleq \sigma_k^2\mathbf{I} + \sum_{j\neq k}\mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H$ is the covariance matrix of noise plus the inter-cell interference experienced by user k.

The work of Y. Yang is supported by the FNR projects SATSENT and CIPHY, and the work of M. Pesavento is supported by the EXPRESS Project within the DFG Priority Program CoSIP (DFG-SPP 1798).

The power consumption at BS k can be approximated by $P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k)$ [14, Section 4.3], where $P_{0,k}$ and ρ_k is the constant specifying the power consumption at the zero RF output power (i.e., $\mathbf{Q}_k = \mathbf{0}$) and the slope of the load dependent power consumption, respectively. The values of $P_{0,k}$ and ρ_k depend on the types of the cell, e.g., macro/micro cell and remote radio head [14, Table 8].

We define the EE of BS k as the ratio of the transmission rate and the consumed power:

$$f_k(\mathbf{Q}) \triangleq \frac{r_k(\mathbf{Q}_k, \mathbf{Q}_{-k})}{P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k)}, \tag{2}$$

where $\mathbf{Q} \triangleq (\mathbf{Q}_k)_{k=1}^K$. Given the coexistence of multiple users, a popular design approach for this multi-objective optimization problem aims at finding the so-called socially optimal point that maximizes the sum EE of all cells:

$$\begin{aligned} \underset{\mathbf{Q}}{\text{maximize}} \quad & f(\mathbf{Q}) \triangleq \textstyle\sum_{k=1}^K f_k(\mathbf{Q}) \\ \text{subject to} \quad & \mathbf{Q}_k \succeq \mathbf{0},\, \text{tr}(\mathbf{Q}_k) \leq P_k,\, \forall k, \end{aligned} \tag{3}$$

where P_k is the power budget for BS k. However, to achieve the socially optimal point, some BSs experiencing unfavorable channel conditions may have to excessively limit their transmission. An approach enhancing the fairness among the BSs is to maximize the minimum EE among all BSs:

$$\begin{aligned} \underset{\mathbf{Q}}{\text{maximize}} \quad & \min_{k=1,\ldots,K} f_k(\mathbf{Q}) \\ \text{subject to} \quad & \mathbf{Q}_k \succeq \mathbf{0},\, \text{tr}(\mathbf{Q}_k) \leq P_k,\, \forall k. \end{aligned} \tag{4}$$

In Section III and Section IV, we propose an iterative algorithm that can efficiently solve problem (3) and (4), respectively.

III. The Proposed Iterative Algorithm for Socially Optimal Design

In this section, we propose an iterative algorithm for problem (3) based on the successive pseudoconvex approximation framework developed in [7]. It consists of solving a sequence of successively refined approximate problems: in iteration t, the approximate problem defined around the point $\mathbf{Q}^t$ consists of maximizing an approximate function, denoted as $\tilde{f}(\mathbf{Q}; \mathbf{Q}^t)$, under the same constraints as (3).

To start with, we introduce the definition of pseudoconvex functions: a function $f(\mathbf{x})$ is said to be *pseudoconvex* if [15]

$$f(\mathbf{y}) < f(\mathbf{x}) \Longrightarrow (\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x}) < 0. \tag{5}$$

In other words, $f(\mathbf{y}) < f(\mathbf{x})$ implies $\mathbf{y} - \mathbf{x}$ is a *descent direction* of $f(\mathbf{x})$ [16]. A function $f(\mathbf{x})$ is pseudoconcave if $-f(\mathbf{x})$ is pseudoconvex. We remark that the (strong) convexity of a function implies that the function is pseudoconvex, which in turn implies that the function is quasiconvex.

The numerator functions $(r_k(\mathbf{Q}))_{k=1}^K$ in (2) are not concave in $\mathbf{Q}$, and thus the objective function $f(\mathbf{Q})$ is not even pseudoconcave. Meanwhile, the function $r_k(\mathbf{Q})$ is concave in $\mathbf{Q}_k$, and exploiting this partial concavity may notably accelerate the convergence [17]. Therefore, to design an approximate function that is easy to maximize, we approximate the nonconcave function $f(\mathbf{Q})$ with respect to $\mathbf{Q}_k$ at the point $\mathbf{Q}^t$ by the following function denoted as $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$:

$$\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t) \triangleq \frac{r_k(\mathbf{Q}_k, \mathbf{Q}_{-k}^t) + (\mathbf{Q}_k - \mathbf{Q}_k^t) \bullet \mathbf{\Pi}_k(\mathbf{Q}^t)}{P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k)}, \forall k, \tag{6}$$

where

$$\mathbf{\Pi}_k(\mathbf{Q}^t) \triangleq (P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k^t)) \sum_{j \neq k} \frac{\nabla_{\mathbf{Q}_k^*} r_j(\mathbf{Q}^t)}{P_{0,j} + \rho_j \text{tr}(\mathbf{Q}_j^t)}.$$

In (6), we set $\mathbf{Q}_{-k}$ to be $\mathbf{Q}_{-k} = \mathbf{Q}_{-k}^t$ in $r_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ and linearize $r_j(\mathbf{Q})$ with respect to $\mathbf{Q}_k$. As a result, the numerator function of $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ is concave in $\mathbf{Q}_k$, and $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$, the fractional function of a nonnegative concave function and a positive linear function, is pseudoconcave in $\mathbf{Q}_k$ [15, Chapter 9.6].

Given point $\mathbf{Q}^t$ in iteration t, we define the approximate function $\tilde{f}(\mathbf{Q}; \mathbf{Q}^t)$ to be the sum of all individual approximate functions: $\tilde{f}(\mathbf{Q}; \mathbf{Q}^t) = \sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$. Then the approximate problem is

$$\begin{aligned} \underset{\mathbf{Q}}{\text{maximize}} \quad & \textstyle\sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t) \\ \text{subject to} \quad & \mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k,\ k = 1, \ldots, K. \end{aligned} \tag{7}$$

We remark that although $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ is pseudoconcave in $\mathbf{Q}_k$, the approximate function $\sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ in (7) is not jointly pseudoconcave in $\mathbf{Q}$, because, unlike concave functions, the sum of pseudoconcave functions is not necessarily pseudoconcave [15, Chapter 9.5].

On solving the approximate problem (7). We denote as $\mathbb{B}\mathbf{Q}^t = (\mathbb{B}_k \mathbf{Q}^t)_{k=1}^K$ the solution of the approximate problem (7). Since (7) is well decoupled across different variables $(\mathbf{Q}_k)_{k=1}^K$, it can be decomposed into many subproblems, each with a much smaller dimension than (7), and they can be solved in parallel:

$$\begin{aligned} \mathbb{B}\mathbf{Q}^t &= \underset{(\mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k)_{k=1}^K}{\arg\max} \ \textstyle\sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t) \\ & \Updownarrow \\ \mathbb{B}_k \mathbf{Q}^t &= \underset{\mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k}{\arg\max} \ \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t),\ k = 1, \ldots, K, \end{aligned} \tag{8}$$

where $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ is defined in (6). Note that each subproblem in (8) is pseudoconvex, and $\mathbb{B}_k \mathbf{Q}^t$ is unique as we will explain later.

On solving the subproblem (8). Since the subproblem (8) is pseudoconvex and a fractional programming problem, we apply the Dinkelbach's algorithm to find $\mathbb{B}_k \mathbf{Q}^t$ iteratively. In particular, in iteration τ of the Dinkelbach's algorithm, the following problem is solved for a given and fixed $s_k^{t,\tau}$ ($s_k^{t,0}$ can be set to 0):

$$\begin{aligned} \underset{\mathbf{Q}_k \succeq \mathbf{0},\, \text{tr}(\mathbf{Q}_k) \leq P_k}{\text{maximize}} \quad & r_k(\mathbf{Q}_k, \mathbf{Q}_{-k}^t) + (\mathbf{Q}_k - \mathbf{Q}_k^t) \bullet \mathbf{\Pi}_k(\mathbf{Q}^t) \\ & - s_k^{t,\tau} (P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k)). \end{aligned} \tag{9}$$

The optimal point of problem (9), denoted as $\mathbf{Q}_k^\star(s_k^{t,\tau})$, has a closed form expression based on the generalized waterfilling solution [18, Lemma 2]. After $(\mathbf{Q}_k^\star(s_k^{t,\tau}))_{k=1}^K$ is obtained, $s_k^{t,\tau}$ is updated according to the following rule:

$$s_k^{t,\tau+1} = \frac{r_k(\mathbf{Q}_k^\star(s_k^{t,\tau}), \mathbf{Q}_{-k}^t) + (\mathbf{Q}_k^\star(s_k^{t,\tau}) - \mathbf{Q}_k^t) \bullet \mathbf{\Pi}_k(\mathbf{Q}^t)}{P_{0,k} + \rho_k \text{tr}(\mathbf{Q}^\star(s_k^{t,\tau}))}. \tag{10}$$

It follows from the convergence properties of the Dinkelbach's algorithm (cf. [4]) that $\lim_{\tau \to \infty} \mathbf{Q}_k^\star(s_k^{t,\tau}) = \mathbb{B}_k \mathbf{Q}^t$ for all k at a superlinear convergence rate.

Despite the lack of pseudoconcavity in the approximate function $\tilde{f}(\mathbf{Q}; \mathbf{Q}^t) = \sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ in (7), $\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t$ is still an ascent direction of the original objective function $f(\mathbf{Q})$ at $\mathbf{Q} = \mathbf{Q}^t$. To see this, we first remark that the approximate function $\tilde{f}_k(\mathbf{Q}_k; \mathbf{Q}^t)$ and the original function $f_k(\mathbf{Q})$ have identical functional value and gradient at the point $\mathbf{Q}^t$ around which $\tilde{f}_k(\mathbf{Q}; \mathbf{Q}^t)$ is defined, i.e.,

$$\begin{gathered} \textstyle\sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k^t; \mathbf{Q}^t) = \sum_{k=1}^K f_k(\mathbf{Q}^t), \\ \nabla_{\mathbf{Q}_k^*} \Big(\textstyle\sum_{j=1}^K \tilde{f}_j(\mathbf{Q}_j; \mathbf{Q}^t) \Big) \Big|_{\mathbf{Q}=\mathbf{Q}^t} = \nabla_{\mathbf{Q}_k^*} \Big(\sum_{j=1}^K f_j(\mathbf{Q}) \Big) \Big|_{\mathbf{Q}=\mathbf{Q}^t}, \forall k, \end{gathered} \tag{11}$$

while the detailed verification steps are left to the reader. Note that $\mathbb{B}_k \mathbf{Q}^t$ is unique because both $\mathbf{Q}_k^\star(s_k^{t,\tau})$ and $\lim_{\tau \to \infty} s_k^{t,\tau}$ are unique (any stationary point of a pseudoconvex optimization problem is also globally optimal). We can thus claim that either $\mathbb{B}_k\mathbf{Q}^t = \mathbf{Q}_k^t$ or

$$\tilde{f}_k(\mathbb{B}_k\mathbf{Q}^t;\mathbf{Q}^t) = \max_{\mathbf{Q}_k\succeq\mathbf{0},\text{tr}(\mathbf{Q}_k)\leq P_k} \tilde{f}_k(\mathbf{Q}_k;\mathbf{Q}^t) > \tilde{f}_k(\mathbf{Q}_k^t;\mathbf{Q}^t),$$

while the latter implies that

$$0 < (\mathbb{B}_k\mathbf{Q}^t - \mathbf{Q}_k^t)\bullet\nabla_{\mathbf{Q}_k^*}\tilde{f}_k(\mathbf{Q}_k^t;\mathbf{Q}^t) = (\mathbb{B}_k\mathbf{Q}^t - \mathbf{Q}_k^t)\bullet\nabla_{\mathbf{Q}_k^*}f(\mathbf{Q}^t), \tag{12}$$

where the inequality in (12) comes from the definition of pseudo-concave functions, cf. (5), and the equality in (12) comes from (11). Adding up (12) over all $k = 1,\ldots,K$, we obtain

$$(\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t)\bullet\nabla_{\mathbf{Q}^*}f(\mathbf{Q}^t) > 0. \tag{13}$$

Therefore $\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t$ is an ascent direction of $f(\mathbf{Q})$ at $\mathbf{Q} = \mathbf{Q}^t$.

Given the ascent direction $\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t$, we calculate the stepsize by the successive line search: given two scalars $0 < \alpha < 1$ and $0 < \beta < 1$, γ^t is set to be $\gamma^t = \beta^{m_t}$, where m_t is the smallest nonnegative integer m satisfying the following inequality:

$$f(\mathbf{Q}^t + \beta^m(\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t)) \geq f(\mathbf{Q}^t) + \alpha\beta^m\nabla f(\mathbf{Q}^t)\bullet(\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t). \tag{14}$$

Note that the successive line search is carried out over the original objective function $f(\mathbf{Q})$ defined in (3). After the stepsize γ^t is found, the variable $\mathbf{Q}^t$ is updated as

$$\mathbf{Q}^{t+1} = \mathbf{Q}^t + \gamma^t(\mathbb{B}\mathbf{Q}^t - \mathbf{Q}^t). \tag{15}$$

From (13)-(15) it can be verified that the sequence $\{f(\mathbf{Q}^t)\}_t$ is monotonically increasing. Moreover, the sequence $\{\mathbf{Q}^t\}$ has a limit point and every limit point is a stationary point of (3), whose proof follows the same line of analysis as [7, Theorem 2].

In what follows, we draw some comments on the properties of the proposed iterative algorithm.

The approximate function in (7) is constructed in the same spirit as [7, 17] by keeping as much concavity as possible, namely, $r_k(\mathbf{Q}_k,\mathbf{Q}_{-k})$ in $\mathbf{Q}_k$ and $\sum_{k=1}^K(P_{0,k}+\rho_k\text{tr}(\mathbf{Q}_k))$ in $\mathbf{Q}$, and linearizing the nonconcave functions only, namely, $\sum_{j\neq k} r_j(\mathbf{Q})$. Besides this, the fractional operator is also kept. Therefore, the proposed algorithm is of a best-response nature and expected to exhibit a fast convergence behavior, as we shall later illustrate numerically.

The proposed algorithm is the first parallel (Jacobi) best-response algorithm of its type, while the approximate function $\sum_{k=1}^K \tilde{f}_k(\mathbf{Q}_k;\mathbf{Q}^t)$ may not even be pseudoconcave. Instead, only the component functions $\tilde{f}_k(\mathbf{Q}_k;\mathbf{Q}^t)$ are pseudoconcave. This is the weakest convergence condition available in literature and the convergence of the proposed algorithm cannot be proved by existing successive convex approximate frameworks such as [17, 19] which requires the approximate function to be strictly or strongly concave.

The proposed algorithm could be implemented by a central unit which has the knowledge of the channel state information of the direct-link and cross-link channels, namely, $(\mathbf{H}_{kj})_{j,k}$. In practical networks, this central unit could be embedded in the centralized radio access network (CRAN): each BS k measures the direct-link channel $\mathbf{H}_{kk}$ and cross-link channels $(\mathbf{H}_{kj})_{j\neq k}$ and send the channel state information $(\mathbf{H}_{kj})_j$ to the central unit in the CRAN. Then the central unit implements the proposed algorithm and informs each BS k the optimal transmit covariance matrix $\mathbf{Q}_k$. The incurred latency is mainly due to the signaling exchange between the central unit and the BSs, and the execution of the proposed algorithm.

IV. The Proposed Iterative Algorithm for Max-Min Fairness Design

In this section, we propose an iterative algorithm for problem (4). We start by introducing an auxiliary variable $\mathbf{Y} = (\mathbf{Y}_k)_{k=1}^K$ and rewriting (4) into the following equivalent form:

$$\begin{aligned}\max_{\mathbf{Q}} \quad & \min_{k=1,\ldots,K} \frac{r_k^+(\mathbf{Y}_k) - r_k^-(\mathbf{Q})}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k)},\\ \text{s.t.} \quad & \mathbf{Q}_k \succeq \mathbf{0},\ \text{tr}(\mathbf{Q}_k) \leq P_k, \mathbf{Y}_k = \textstyle\sum_{j=1}^K \mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H,\ \forall k,\end{aligned} \tag{16}$$

where the numerator function of $f_k(\mathbf{Q})$, namely, $r_k(\mathbf{Q})$ defined in (1), can be written as the difference of two concave functions $r_k^+(\mathbf{Y}_k) \triangleq \log\det(\sigma_k^2\mathbf{I} + \mathbf{Y}_k)$ and $r_k^-(\mathbf{Q}) \triangleq \log\det\left(\sigma_k^2\mathbf{I} + \sum_{j\neq k}\mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H\right)$. On one hand, it is easy to see

$$r_k^+(\mathbf{Y}_k) \geq r_k^+(\mathbf{Y}_k) - c\textstyle\sum_{j=1}^K\left\|\mathbf{Y}_j - \mathbf{Y}_j^t\right\|^2 \triangleq \underline{r}_k^+(\mathbf{Y};\mathbf{Y}^t), \tag{17}$$

where c is a given positive constant and the proximal term in (17) is introduced for numerical benefit that will become clear later. On the other hand, the concave function $r_k^-(\mathbf{Q}_{-k})$ is upper bounded by its linear approximation at any feasible point, e.g., $\mathbf{Q} = \mathbf{Q}^t$:

$$\begin{aligned}r_k^-(\mathbf{Q}) \leq\, & r_k^-(\mathbf{Q}^t) + \textstyle\sum_{j\neq k}(\mathbf{Q}_j - \mathbf{Q}_j^t)\bullet\nabla_{\mathbf{Q}_j^*}r_k^-(\mathbf{Q}^t)\\ & + c\textstyle\sum_{j=1}^K\left\|\mathbf{Q}_j - \mathbf{Q}_j^t\right\|^2 \triangleq \overline{r}_k^-(\mathbf{Q};\mathbf{Q}^t),\end{aligned} \tag{18}$$

where the linear approximation in (18) is further augmented by a quadratic proximal term. Thus $\underline{r}_k^+(\mathbf{Y};\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q};\mathbf{Q}^t)$ is a global lower bound of $r_k^+(\mathbf{Y}_k) - r_k^-(\mathbf{Q})$ where equality holds at $(\mathbf{Q}^t,\mathbf{Y}^t)$:

$$r_k^+(\mathbf{Y}_k) - r_k^-(\mathbf{Q}_{-k}) \geq \underline{r}_k^+(\mathbf{Y};\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q};\mathbf{Q}^t). \tag{19}$$

In the proposed iterative algorithm, given $(\mathbf{Q}^t,\mathbf{Y}^t)$ in iteration t, we solve the following approximate problem:

$$\begin{aligned}\max_{\mathbf{Q},\mathbf{Y}} \quad & \min_k \frac{\underline{r}_k^+(\mathbf{Y};\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q};\mathbf{Q}^t)}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k)}\\ \text{s.t.} \quad & \mathbf{Q}_k \succeq \mathbf{0},\ \text{tr}(\mathbf{Q}_k) \leq P_k, \mathbf{Y}_k = \textstyle\sum_{j=1}^K \mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H, \forall k,\end{aligned} \tag{20}$$

and set the optimal solution of (20) as $(\mathbf{Q}^{t+1},\mathbf{Y}^{t+1})$ as $\mathbf{Q}^{t+1}$ is a better estimate of $\mathbf{Q}^\star$ than $\mathbf{Q}^t$:

$$\begin{aligned}\min_k \frac{r_k^+(\mathbf{Y}_k^{t+1}) - r_k^-(\mathbf{Q}^{t+1})}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k^{t+1})} &\geq \min_k \frac{\underline{r}_k^+(\mathbf{Y}^{t+1};\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q}^{t+1};\mathbf{Q}^t)}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k^{t+1})}\\ &\geq \min_k \frac{\underline{r}_k^+(\mathbf{Y}^t;\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q}^t;\mathbf{Q}^t)}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k^t)}\\ &= \min_k \frac{r_k^+(\mathbf{Y}_k^t) - r_k^-(\mathbf{Q}^t)}{P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k^t)},\end{aligned}$$

where the first inequality and the equality come from (19), and the second inequality from the optimality of $\mathbf{Q}^{t+1}$ for problem (20). Therefore, the achieved objective value of problem (16) is monotonically increasing. It turns out that the sequence $(\mathbf{Q}^t)_t$ generated by (20) has a limit point, and every limit point is a KKT point of problem (16) (and thus the original problem (4)). The proof follows standard arguments of successive lower bound maximization algorithms, see [2] for example, and is thus omitted here due to the page limit.

On solving the approximate problem (20). The approximate problem (20) is much easier to solve than the original problem (16), because it is a generalized fractional programming problem and can be solved by the generalized Dinkelbach's algorithm: given λ^τ (λ^0 can be set to 0) in iteration τ, we solve the following problem

$$\begin{aligned}\max_{\mathbf{Q},\mathbf{Y}} \quad & \min_k \quad \underline{r}_k^+(\mathbf{Y};\mathbf{Y}^t) - \overline{r}_k^-(\mathbf{Q};\mathbf{Q}^t) - \lambda^{t,\tau}(P_{0,k} + \rho_k\text{tr}(\mathbf{Q}_k))\\ \text{s.t.} \quad & \mathbf{Q}_k \succeq \mathbf{0},\ \text{tr}(\mathbf{Q}_k) \leq P_k, \mathbf{Y}_k = \textstyle\sum_{j=1}^K \mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H, \forall k,\end{aligned} \tag{21}$$

and denote its optimal point as $(\mathbf{Q}^\star(\lambda^{t,\tau}), \mathbf{Y}^\star(\lambda^{t,\tau}))$. Then the variable $\lambda^{t,\tau}$ is updated as follows:

$$\lambda^{t,\tau+1} = \min_{k=1,\ldots,K} \frac{\underline{r}_k^{+}(\mathbf{Y}^\star(\lambda^{t,\tau}); \mathbf{Y}^t) - \overline{r}_k^{-}(\mathbf{Q}^\star(\lambda^{t,\tau}); \mathbf{Q}^t)}{P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k^\star(\lambda^{t,\tau}))}.$$

It follows from the properties of the Dinkelbach's algorithm that $\lim_{\tau\rightarrow\infty} \mathbf{Q}^\star(\lambda^{t,\tau}) = \mathbf{Q}^{t+1}$ at a superlinear convergence rate.

On solving problem (21). As the pointwise minimum of multiple concave functions, the objective function of (21), although concave, is nondifferentiable. We thus introduce an auxiliary variable μ and rewrite problem (21) as

$$\max_{\mathbf{Q},\mathbf{Y},\mu} \quad \mu - c\textstyle\sum_{k=1}^K \left(\left\|\mathbf{Q}_k - \mathbf{Q}_k^t\right\|^2 + \left\|\mathbf{Y}_k - \mathbf{Y}_k^t\right\|^2\right) \tag{22a}$$

$$\text{s.t.} \quad \mathbf{Q}_k \succeq \mathbf{0},\, \mathrm{tr}(\mathbf{Q}_k) \leq P_k, \mathbf{Y}_k = \textstyle\sum_{j=1}^K \mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H, \tag{22b}$$

$$\underline{r}_k^{+}(\mathbf{Y}_k) - \overline{r}_k^{-}(\mathbf{Q}^t) - \textstyle\sum_{j\neq k}(\mathbf{Q}_j - \mathbf{Q}_j^t) \bullet \nabla_{\mathbf{Q}_j^*} \overline{r}_k^{-}(\mathbf{Q}^t) - \lambda^{t,\tau}(P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k)) \geq \mu, \forall k. \tag{22c}$$

Problem (22) is convex and the constraint set has a nonempty interior, so strong duality holds. Besides this, the auxiliary variables $(\mathbf{Y}_k)_{k=1}^K$ make the coupling constraints in (22), namely, the equality constraints in (22b) and the inequality constraints (22c), separable among the variables [20], so the dual decomposition technique can be used.

Suppose the dual variable associated with the coupling constraint in (22b) and (22c) is $\mathbf{\Sigma}_k$ and η_k, respectively. The Lagrangian is:

$$L(\mathbf{Q}, \mu, \mathbf{Y}, \mathbf{\Sigma}, \boldsymbol{\eta}) = \left(1 - \textstyle\sum_{k=1}^K \eta_k\right)\mu + \textstyle\sum_{k=1}^K (L_{Y,k}(\mathbf{Y}_k, \mathbf{\Sigma}, \boldsymbol{\eta}) - L_{Q,k}(\mathbf{Q}_k, \mathbf{\Sigma}, \boldsymbol{\eta})),$$

where

$$L_{Q,k}(\mathbf{Q}_k, \mathbf{\Sigma}, \boldsymbol{\eta}) \triangleq c\left\|\mathbf{Q}_k - \mathbf{Q}_k^t\right\|^2 + \mathbf{Q}_k \bullet \Big(\eta_k \lambda^{t,\tau}\rho_k \mathbf{I} + \textstyle\sum_{j\neq k}\eta_j \nabla_{\mathbf{Q}_k^*} \overline{r}_j^{-}(\mathbf{Q}_{-j}^t) - \sum_{j=1}^K \mathbf{H}_{jk}^H \mathbf{\Sigma}_j \mathbf{H}_{jk}\Big),$$

$$L_{Y,k}(\mathbf{Y}_k, \mathbf{\Sigma}, \boldsymbol{\eta}) \triangleq \eta_k \log\det(\sigma_k^2 \mathbf{I} + \mathbf{Y}_k) - c\left\|\mathbf{Y}_k - \mathbf{Y}_k^t\right\|^2 - \mathbf{\Sigma}_k \bullet \mathbf{Y}_k.$$

The dual function of (22) is

$$d(\mathbf{\Sigma}, \boldsymbol{\eta}) \triangleq \max_{(\mathbf{Q}_k\succeq\mathbf{0}, \mathrm{tr}(\mathbf{Q}_k)\leq P_k, \mathbf{Y}_k\succeq\mathbf{0})_{k=1}^K} L(\mathbf{Q}, \mu, \mathbf{Y}, \mathbf{\Sigma}, \boldsymbol{\eta}). \tag{23}$$

The maximization problem in (23) can be decomposed into many subproblems which are then solved in parallel: for all $k = 1, \ldots, K$,

$$\mathbf{Q}_k^\star(\mathbf{\Sigma}, \boldsymbol{\eta}) \triangleq \underset{\mathbf{Q}_k\succeq\mathbf{0}, \mathrm{tr}(\mathbf{Q}_k)\leq P_k}{\arg\max} \; L_{Q,k}(\mathbf{Q}_k, \mathbf{\Sigma}, \boldsymbol{\eta}), \tag{24a}$$

$$\mathbf{Y}_k^\star(\mathbf{\Sigma}, \boldsymbol{\eta}) \triangleq \underset{\mathbf{Y}_k\succeq\mathbf{0}}{\arg\max} \; L_{Y,k}(\mathbf{Y}_k, \mathbf{\Sigma}, \boldsymbol{\eta}). \tag{24b}$$

Both $\mathbf{Q}_k(\mathbf{\Sigma}, \boldsymbol{\eta})$ and $\mathbf{Y}_k(\mathbf{\Sigma}, \boldsymbol{\eta})$ are unique and have a closed-form expression, see [18, Lemma 2] and [20, Lemma 7], respectively. We remark that the optimal dual variable $(\mathbf{\Sigma}, \boldsymbol{\eta})$ can be found by the subgradient method; this is a standard technique and we omit the details due to page limit, see [21, Sec. IV-B] for a similar discussion.

Similar to the iterative algorithm proposed in [2, Section V-B], the proposed algorithm is essentially a successive lower bound maximization method. Nevertheless, a quadratic proximal term is incorporated into the proposed approximate function, cf. (17) and (18). This subtle but important difference is beneficial from several aspects. Firstly, the sequence $\{\mathbf{Q}^\star(\mathbf{\Sigma}^{t,\tau,\upsilon})\}_\upsilon$, calculated from the dual domain, converges to the optimal point of the primal problem (22). Secondly, the dual problem is differentiable. Thirdly, the dual decomposition technique can be used to solve the approximate problem efficiently and all variable updates have a closed-form expression. The proposed algorithm thus has a guaranteed convergence and a much lower complexity than [2] where the approximate problem is solved by a general purpose convex optimization solver.

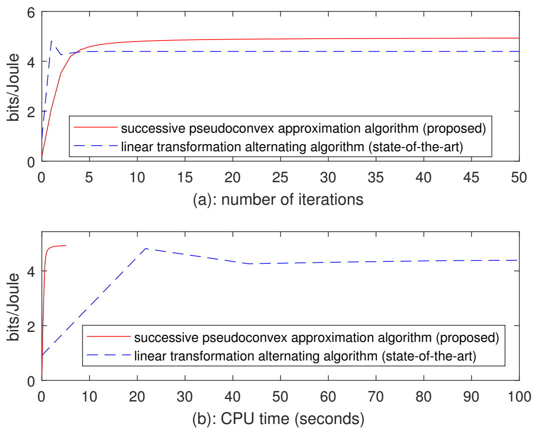

Figure 1. Achieved SEE versus the number of iterations and CPU time in seconds.

V. Simulations

We consider a cluster of $K = 7$ micro BSs with an inter-cell distance of 500m. The number of transmit antennas at each BS is $M_{T,k} = 4$ and the number of receive antennas at each user is $M_{R,k} = 4$. The power consumption at the zero RF output is $P_{0,k}$ =16 W, the power budget normalized by the number of transmit antennas is 36 dBm, i.e., $P_k/M_{T,k}$ =36 dBm, and the slope of power consumption ρ is 2.6; these parameters are mainly adopted from [14]. For each realization, all K users are randomly located in the multi-cell space where each user falls into the respective hexagonal cell, and the simulation results are averaged over 100 realizations.

Due to the page limit, we test the algorithm proposed in Sec. III for the SEE maximization problem only. Among the algorithms proposed in [9, 11, 12], we adopt the Linear Transformation Alternating (LTA) algorithm in [9] as the benchmark because other algorithms are applicable for MISO systems only. In Figure 1, we show the achieved SEE versus the number of iterations and the CPU time in seconds. It is easy to see from Figure 1 (a) that both the proposed algorithm and the LTA algorithm converges in the same number of iterations, but the proposed algorithm converges to a better solution. Besides this, we see from Figure 1 (b) that the proposed algorithm converge in less than 2 seconds, while the LTA algorithm converges in 100 seconds. This is because each iteration of the LTA algorithm consists of a BCD type algorithm, where the variables to be updated are matrices and the typical convergence speed is linear [22], while each iteration of the proposed algorithm consists of the Dinkelbach's algorithm, where the variable is a scalar and the convergence speed is superlinear [4]. Therefore, the proposed algorithm has a much lower complexity than the benchmark algorithm, namely, the LTA algorithm in [9].

VI. Conclusions

In this paper, we have studied the energy efficiency optimization problem in MIMO interference channels. Given the problem's multi-objective nature, we have considered two of the most popular design criteria: socially optimality and max-min fairness. For the socially optimal design, we have proposed for the first time a fast convergent and easily implementable parallel algorithm that is of a best-response type. For the max-min fairness design, we have also proposed a convergent algorithm, which exhibits a much lower complexity than state-of-the-art benchmarks, because, based on dual decomposition, the approximate problem can be decomposed into independent subproblems which can then be optimized in parallel and whose optimal solution have a closed-form expression.

REFERENCES

[1] 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, feb 2008.

[2] A. Zappone, E. Björnson, L. Sanguinetti, and E. Jorswieck, "Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2844–2859, 2017.

[3] O. Tervo, L. N. Tran, and M. Juntti, "Optimal Energy-Efficient Transmit Beamforming for Multi-User MISO Downlink," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5574–5588, 2015.

[4] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-Efficient Power Control: A Look at 5G Wireless Technologies," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1668–1683, apr 2016.

[5] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, 2016.

[6] Y. Yang and M. Pesavento, "A parallel algorithm for energy efficiency maximization in massive MIMO networks," *2016 IEEE Global Communications Conference, GLOBECOM 2016 - Proceedings*, 2016.

[7] ——, "A Unified Successive Pseudoconvex Approximation Framework," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3313–3328, 2017.

[8] Y. Yang and D. Sabella, "Mobile network energy efficiency optimization in MIMO multi-cell systems," in *2017 Fifth International Workshop on Cloud Technologies and Energy Efficiency in Mobile Communication Networks (CLEEN)*. IEEE, jun 2017, pp. 1–5.

[9] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 741–751, 2014.

[10] Q. Wu, W. Chen, D. W. Kwan Ng, J. Li, and R. Schober, "User-Centric Energy Efficiency Maximization for Wireless Powered Communications," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6898–6912, 2016.

[11] C. Pan, W. Xu, J. Wang, H. Ren, W. Zhang, N. Huang, and M. Chen, "Pricing-Based Distributed Energy-Efficient Beamforming for MISO Interference Channels," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 710–722, 2016.

[12] O. Tervo, A. Tolli, M. Juntti, and L.-N. Tran, "Energy-Efficient Beam Coordination Strategies With Rate-Dependent Processing Power," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 6097–6112, nov 2017.

[13] K. G. Nguyen, L. N. Tran, O. Tervo, Q. D. Vu, and M. Juntti, "Achieving energy efficiency fairness in multicell MISO downlink," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1426–1429, 2015.

[14] "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," INFSO-ICT-247733 EARTH D2.3. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/d71252/EARTH_WP2_D2.3_v2.pdf

[15] O. L. Mangasarian, *Nonlinear programming*. McGraw-Hill, 1969.

[16] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

[17] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by Partial Linearization: Parallel Optimization of Multi-Agent Systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641–656, feb 2014.

[18] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, "A Parallel Decomposition Method for Nonconvex Stochastic Multi-Agent Optimization Problems," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2949–2964, jun 2016.

[19] M. Razaviyayn, M. Hong, and Z.-q. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, jan 2013.

[20] G. Scutari, F. Facchinei, L. Lampariello, P. Song, and S. Sardellitti, "Parallel and Distributed Methods for Nonconvex Optimization–Part II: Applications," vol. 65, no. 8, pp. 1945–1960, 2016.

[21] Y. Yang, G. Scutari, P. Song, and D. P. Palomar, "Robust MIMO Cognitive Radio Systems Under Interference Temperature Constraints," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2465–2482, nov 2013.

[22] A. Beck and L. Tetruashvili, "On the Convergence of Block Coordinate Descent Type Methods," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2037–2060, jan 2013.