Maximizing Sum Rates in Cognitive Radio Networks: Convex Relaxation and Global Optimization Algorithms

Liang Zheng and Chee Wei Tan, *Senior Member, IEEE*

***Abstract*—A key challenge in wireless cognitive radio networks is to maximize the total throughput also known as the sum rates of all the users while avoiding the interference of unlicensed band secondary users from overwhelming the licensed band primary users. We study the weighted sum rate maximization problem with both power budget and interference temperature constraints in a cognitive radio network. This problem is nonconvex and generally hard to solve. We propose a reformulation-relaxation technique that leverages nonnegative matrix theory to first obtain a relaxed problem with nonnegative matrix spectral radius constraints. A useful upper bound on the sum rates is then obtained by solving a convex optimization problem over a closed bounded convex set. It also enables the sum-rate optimality to be quantified analytically through the spectrum of specially-crafted nonnegative matrices. Furthermore, we obtain polynomial-time verifiable sufficient conditions that can identify polynomial-time solvable problem instances, which can be solved by a fixed-point algorithm. As a by-product, an interesting optimality equivalence between the nonconvex sum rate problem and the convex max-min rate problem is established. In the general case, we propose a global optimization algorithm by utilizing our convex relaxation and branch-and-bound to compute an ϵ-optimal solution. Our technique exploits the nonnegativity of the physical quantities, e.g., channel parameters, powers and rates, that enables key tools in nonnegative matrix theory such as the (linear and nonlinear) Perron-Frobenius theorem, quasi-invertibility, Friedland-Karlin inequalities to be employed naturally. Numerical results are presented to show that our proposed algorithms are theoretically sound and have relatively fast convergence time even for large-scale problems.**

***Index Terms*—Optimization, convex relaxation, cognitive radio networks, nonnegative matrix theory.**

I. Introduction

A key challenge in cognitive radio networks is to enable efficient wireless resource allocation while ensuring the unlicensed band users (the secondary users) that share the spectrum have minimal impact on the licensed band users (the primary users). Due to the broadcast nature of the wireless medium, the data rates are affected by interference when all the users transmit simultaneously over the same spectrum. Thus, unlicensed radio transmission in cognitive radio networks can lead to overwhelming interference that adversely affects the overall performance of the network [1]–[3]. In this regard, the interference temperature, which is the total interference plus noise, is an important feature for interference management. Interference in cognitive radio networks can be suitably controlled by imposing interference temperature constraints on the secondary users (to limit the interference temperature below a predefined threshold) so that the interference experienced by the primary users are capped at a limit regardless of the number of secondary users [4]. Dynamic spectrum or power allocation to maximize the total throughput also known as the sum rates in response to varying quality of service demands is an important problem. In this paper, we focus on maximizing the overall weighted sum rates with both power budget constraints and interference temperature constraints in cognitive radio networks.

Throughput maximization for wireless cognitive radio networks is relatively new and challenging. The authors in [5] considered throughput maximization for secondary users over non-overlapping channels. Throughput maximization in interference channels is considerably harder, as it is a nonconvex problem. In [6], the two-user sum rate maximization special case with equal weights was studied. The authors in [7] proved the NP-hardness of the problem. An often used technique to tackle the nonconvexity is the standard Lagrange dual relaxation [7], [8]. The authors in [7] used the Lyapunov theorem in functional analysis to establish a zero duality gap property when the number of frequencies grows infinitely large. In [9], Lagrangian dual relaxation was used for estimating the duality gap in the general case. The authors in [10]–[13] developed successive convex approximation using geometric programming that only solves the problem suboptimally.

Global optimization algorithms that can compute the global optimal solution, albeit with exponential-time complexity, have been proposed in the literature recently. The authors in [14] proposed global optimization algorithms using DC (difference of convex functions) programming. In [15], [16], the problem was reformulated into a convex maximization problem over an unbounded convex constraint set. Based on this reformulation, an outer approximation algorithm and a branch-and-bound method were proposed in [15] and [16] respectively. However, the unbounded constraint set is susceptible to numerical artifacts in algorithm design, i.e., numerical computation is not exact. The authors in [17] obtained suffi-

Manuscript received November 13, 2012; revised April 3, 2013 and June 16, 2013. The work in this paper was partially supported by grants from the Research Grants Council of Hong Kong Project No. RGC CityU 125212, Qualcomm Inc., the Science, Technology and Innovation Commission of Shenzhen Municipality, Project No. JCYJ20120829161727318 on Green Communications in Small-cell Mobile Networks, and Project No. JCYJ20130401145617277 on Adaptive Spectrum Access Resource Allocation in Cognitive Radio Networks.

L. Zheng and C. W. Tan are with the Department of Computer Science, City University of Hong Kong, Tat Chee Ave., Hong Kong (e-mail: liangzheng.hkcityu@gmail.com, cheewtan@cityu.edu.hk).

Digital Object Identifier 10.1109/JSAC.2014.140324

cient conditions to solve the problem optimally in polynomial-time when there is only a single total power constraint, and to provide performance bounds for the problem with individual power constraints in [18].

However, techniques that shed light on achieving optimality (in polynomial time) or giving convex relaxation have come few and far. Most existing algorithms have drawbacks from a theoretical viewpoint as they cannot guarantee the optimality of the numerical solution. For example, algorithms based on Lagrange duality suffer from the positive duality gap and cannot guarantee finding a global optimal solution. Also, previously proposed global optimization algorithms do not consider interference temperature constraints, which are a unique feature of cognitive radio networks. To the best of our knowledge, there has been neither work that characterizes the quality of suboptimal solution analytically nor a (polynomial-time) optimal algorithm. Quantifying optimality in polynomial-time (whenever that is possible) and generating convex relaxation can provide new perspectives to this (still open) NP-hard problem, especially in the *large-scale* setting.

The contributions of this paper are as follows:

1) We propose a reformulation-relaxation technique based on nonnegative matrix theory that uses convex relaxation to bound the nonconvex achievable rate regions for the weighted sum rate maximization problem with power budget and interference temperature constraints.
2) Our technique can provide analytically sufficient conditions to identify problem instances that are polynomial-time optimally solvable, and a fixed-point algorithm (i.e., no parameter tuning required) is proposed to compute the global optimal solution.
3) In the general case, we propose a global optimization algorithm that leverages the convex relaxation and the branch-and-bound method to yield successively tighter bounds to the global optimal value (quantified by ϵ-optimality). Our optimization software is available for download at [19].

Using our convex relaxation, new insights can even be drawn, e.g., an interesting optimality equivalence between the weighted sum rate problem and the weighted max-min rate problem (that can be solved in polynomial time). These are consequences from exploiting the inherent nonnegativity of the physical quantities, e.g., channel gain, powers and rates, in the sum rate problem that enables key tools in nonnegative matrix theory and nonlinear Perron-Frobenius theory to be employed naturally and successfully.[1]

This paper is organized as follows. We present the system model in Section II. In Section III, we introduce the weighted sum rate maximization problem (in the transmit power and interference temperature domain) and propose an equivalent reformulation (in the achievable rate domain). In Section IV, we propose a convex relaxation technique to convexify the sum-rate reformulation and identify problem instances that can be solved optimally in polynomial-time. In Section V, we propose a global optimization algorithm to solve the problem in the general case. We highlight the performance of our algorithms using numerical examples in Section VI. We conclude with a summary in Section VII. All the proofs can be found in the Appendix.

[1]This idea relates to a repertory of theoretical and algorithmic developments for wireless utility maximization that offer useful insights based on the nonnegative matrix theory and the nonlinear Perron-Frobenius theory, see [15]–[18], [20]–[23] for related work.

The following notation is used in our paper. Column vectors and matrices are denoted by boldfaced lowercase and uppercase respectively. Let $\rho(\mathbf{F})$ denote the Perron-Frobenius eigenvalue of a nonnegative matrix $\mathbf{F}$, and $\mathbf{x}(\mathbf{F})$ and $\mathbf{y}(\mathbf{F})$ denote the Perron right and left eigenvectors of $\mathbf{F}$ associated with $\rho(\mathbf{F})$. We let $\mathbf{e}_l$ denote the lth unit coordinate vector, $\mathbf{I}$ denote the identity matrix and $\mathbf{1} = [1, \ldots, 1]^\top$. The superscripts $(\cdot)^\top$ denotes transpose, and $\|\cdot\|_F$ denotes the Frobenius norm. We denote $\mathbf{x} \circ \mathbf{y}$ as a Schur product of $\mathbf{x}$ and $\mathbf{y}$, i.e., $\mathbf{x} \circ \mathbf{y} = [x_1 y_1, \ldots, x_L y_L]^\top$. For a vector $\mathbf{x} = [x_1, \ldots, x_L]^\top$, $\mathrm{diag}(\mathbf{x})$ is its diagonal matrix $\mathrm{diag}(x_1, \ldots, x_L)$. Let $e^{\mathbf{x}}$ denote $e^{\mathbf{x}} = (e^{x_1}, \ldots, e^{x_L})^\top$, and $\log \mathbf{x}$ denote $\log \mathbf{x} = (\log x_1, \ldots, \log x_L)$. An equality involving eigenvectors is true up to a scaling constant.

II. System Model

We consider a cognitive radio network with a collection of primary users and secondary users transmitting simultaneously on a single-input-single-output (SISO) channel. There are L users (transmitter/receiver pairs) sharing a common frequency-flat fading channel, i.e., the total number of primary users and secondary users is L. Assume that there are S primary users in the cognitive radio network ($S < L$), then the number of secondary users is $L - S$. All users are indexed by l, where the lth user is a primary user when $l \leq S$, otherwise it is a secondary user. Each user employs a single-user decoder, i.e., treating the interference as additive Gaussian noise, and has perfect channel state information at the receiver. Let $\mathbf{G} = [G_{lj}]_{l,j=1}^{L} > 0_{L \times L}$ represent the channel gain, where G_{lj} is the channel gain from the jth transmitter to the lth receiver, and $\mathbf{n} = (n_1, \ldots, n_L)^\top > \mathbf{0}$, where n_l is the noise power at the lth user. The vector $\mathbf{p} = (p_1, \ldots, p_L)^\top$ is the transmit power vector.

The Signal-to-Interference-and-Noise Ratio (SINR) for the lth receiver is defined as the ratio of the received signal power to the sum of interference signal power and additive noise power. Now, the SINR of the lth user can be given in terms of $\mathbf{p}$:

$$\mathsf{SINR}_l(\mathbf{p}) = \frac{p_l G_{ll}}{\sum_{j \neq l} p_j G_{lj} + n_l}. \tag{1}$$

We also define a nonnegative matrix $\mathbf{F}$ with entries:

$$F_{lj} = \begin{cases} 0, & \text{if } l = j \\ G_{lj}/G_{ll}, & \text{if } l \neq j \end{cases} \tag{2}$$

and the vector $\mathbf{v} = (\frac{n_1}{G_{11}}, \frac{n_2}{G_{22}}, \ldots, \frac{n_L}{G_{LL}})^\top$. Moreover, we assume that $\mathbf{F}$ is irreducible, i.e., each link has at least an interferer. For brevity, the vector $\boldsymbol{\gamma}$ denotes the SINR for all users. With these notations, the SINR can be expressed as:

$$\gamma_l = \frac{p_l}{(\mathbf{F}\mathbf{p} + \mathbf{v})_l}. \tag{3}$$

Next, let the vector $\mathbf{q}$ denote the normalized interference temperature for all users with entries given by:

$$q_l = \sum_{j=1}^{L} p_j F_{lj} + v_l, \tag{4}$$

i.e., $\mathbf{q} = \mathbf{F}\mathbf{p} + \mathbf{v}$. The achievable data rate r_l of the lth user (assuming the Shannon capacity formula) is given by:

$$r_l = \log(1 + \mathsf{SINR}_l(\mathbf{p})) \tag{5}$$

measured in nats per symbol. This can be expressed as $e^{r_l} - 1 = p_l/q_l$ for all l. Since $\mathbf{p} = \operatorname{diag}(e^{\mathbf{r}} - \mathbf{1})\mathbf{q}$ and $\mathbf{q} = \mathbf{F}\mathbf{p} + \mathbf{v}$, the power and interference temperature can be written, respectively, as $\mathbf{p} = \operatorname{diag}(e^{\mathbf{r}} - \mathbf{1})(\mathbf{F}\mathbf{p} + \mathbf{v})$ and $\mathbf{q} = \mathbf{F}\operatorname{diag}(e^{\mathbf{r}} - \mathbf{1})\mathbf{q} + \mathbf{v}$. We can thus obtain a one-to-one mapping between $\mathbf{r}$ and $\mathbf{p}$ and also one between $\mathbf{r}$ and $\mathbf{q}$, given by:

$$\mathbf{p}(\mathbf{r}) = \big(\mathbf{I} + \mathbf{F} - \operatorname{diag}(e^{\mathbf{r}})\mathbf{F}\big)^{-1}\big(\operatorname{diag}(e^{\mathbf{r}} - \mathbf{1})\big)\mathbf{v} \tag{6}$$

$$\text{and} \quad \mathbf{q}(\mathbf{r}) = \big(\mathbf{I} + \mathbf{F} - \mathbf{F}\operatorname{diag}(e^{\mathbf{r}})\big)^{-1}\mathbf{v}, \tag{7}$$

respectively. Hence, given a feasible rate $\mathbf{r}$, the corresponding power and interference temperature can be computed by (6) and (7) respectively.

Typically, constraints in a cognitive radio network can be classified into either resource budget constraints that limit power at the transmitters or interference management constraints that limit interference at the receivers. As secondary users are allowed to coexist with primary users [4], interference temperature constraints are used to manage the interference of the secondary users from overwhelming the primary users. The weighted power constraint set $\mathcal{P}$ and the weighted interference temperature constraint set $\mathcal{Q}$ are respectively given by:

$$\mathcal{P} = \{\mathbf{p} \,|\, \mathbf{a}_k^\top \mathbf{p} \le \bar{p}_k,\ k = 1, \ldots, K\}, \tag{8}$$

$$\text{and} \quad \mathcal{Q} = \{\mathbf{q} \,|\, \mathbf{b}_m^\top \mathbf{q} \le \bar{q}_m,\ m = 1, \ldots, M\}, \tag{9}$$

where $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_K)^\top \in \mathbb{R}^K$ and $\mathbf{a}_k \in \mathbb{R}^L$ $k = 1, \ldots, K$ are respectively the upper bound and positive weight vectors for the weighted power constraints, and $\bar{\mathbf{q}} = (\bar{q}_1, \ldots, \bar{q}_M)^\top \in \mathbb{R}^M$ and $\mathbf{b}_m \in \mathbb{R}^L$ $m = 1, \ldots, M$ are respectively the upper bound and positive weight vectors for the weighted interference temperature constraints. Note that we must have $\bar{q}_m \ge \mathbf{b}_m^\top \mathbf{v}$ for a nonempty constraint set $\mathcal{Q}$. By setting the weights appropriately, we are able to guarantee prioritized access of the spectrum. In particular, we can set $(\mathbf{a}_k)_l$ for $l \le S$ to be relatively smaller than those for $l > S$ and $(\mathbf{b}_m)_l$ for $l \le S$ to be relatively larger than those for $l > S$, i.e., more restrictive constraints on secondary users' power and primary users' interference temperature. As special cases, when $\mathbf{a}_k = \mathbf{e}_k$, we have the individual power constraints $p_k \le \bar{p}_k$ for all k, and when $\mathbf{a}_k = \mathbf{1}$ and $K = 1$, we have a single total power constraint. Likewise, the weighted interference temperature constraints are general enough for modeling. For example, in the case of the individual interference temperature constraints, i.e., $\mathbf{b}_m = \mathbf{e}_m$, we can set $\bar{q}_m$ to some finite value for a primary user (when $m \le S$), while we can set $\bar{q}_m$ to infinity for a secondary user (when $m > S$).

Now, given the above power constraints, one can straightforwardly establish (loose) upper bounds to the achievable rates in (5). In particular, since $\gamma_l = \dfrac{p_l}{(\mathbf{F}\mathbf{p} + \mathbf{v})_l} \le \dfrac{p_l}{v_l} \le \max\limits_{k=1,\ldots,K} \dfrac{\bar{p}_k}{(\mathbf{a}_k)_l v_l}$, any $\mathbf{r} \in \mathbb{R}_+^L$ satisfies the inequalities

$$\mathbf{r} \le \bar{\mathbf{r}} = (\bar{r}_1, \ldots, \bar{r}_L)^\top, \tag{10}$$

where $\bar{r}_l = \log\left(1 + \max\limits_{k=1,\ldots,K} \dfrac{\bar{p}_k}{(\mathbf{a}_k)_l v_l}\right), \quad l = 1, \ldots, L.$

Thus, we have $\mathbf{r} \le \bar{\mathbf{r}}$ as an implicit box constraint set, albeit a loose one, that contains the nonconvex achievable rate region. In the following, we describe how to obtain better (tighter) rate constraint set using convex relaxation.

III. Problem Statement and Reformulation

The weighted sum rate maximization problem in a cognitive radio network is given by:

$$\begin{array}{rl} \text{maximize} & \displaystyle\sum_{l=1}^{L} w_l \log\big(1 + \mathsf{SINR}_l(\mathbf{p})\big) \\ \text{subject to} & \mathbf{p} \in \mathcal{P}, \\ & \mathbf{q} \in \mathcal{Q}, \\ \text{variable:} & \mathbf{p}, \mathbf{q}, \end{array} \tag{11}$$

where $\mathbf{w} = [w_1, \ldots, w_L]^\top > \mathbf{0}$ is a given probability vector, i.e., $\mathbf{1}^\top \mathbf{w} = 1$, and w_l is a weight assigned to the lth link to reflect priority (a larger weight reflects a higher priority). We denote the optimal solution of (11) by $\mathbf{p}^\star = (p_1^\star, \ldots, p_L^\star)^\top$ and $\mathbf{q}^\star = (q_1^\star, \ldots, q_L^\star)^\top$. In general, (11) is nonconvex (due to the nonconvex objective function) and thus is hard to solve.

We next study a reparameterization of (11) from the power $\mathbf{p}$ and $\mathbf{q}$ to an equivalent optimization problem that only has the achievable rate $\mathbf{r}$ and $\mathbf{q}$ as the explicit optimization variables. The advantage of the reformulation is an efficient technique to bound the nonconvex pareto rate region using convex relaxation. Let us first define the following set of nonnegative matrices:

$$\mathbf{B}_k = \mathbf{F} + \frac{1}{\bar{p}_k}\mathbf{v}\mathbf{a}_k^\top, \quad k = 1, \ldots, K, \tag{12}$$

$$\mathbf{D}_m = \Big(\mathbf{I} + \frac{1}{\bar{q}_m - \mathbf{b}_m^\top \mathbf{v}}\mathbf{v}\mathbf{b}_m^\top\Big)\mathbf{F}, \quad m = 1, \ldots, M. \tag{13}$$

Note that $\mathbf{B}_k$ and $\mathbf{D}_m$ contain the problem parameters associated with the kth power budget constraint in $\mathcal{P}$ and the mth interference temperature constraint in $\mathcal{Q}$ respectively. We now introduce a reformulation of (11) in the following.

Theorem 1: The optimal value in (11) is equal to the optimal value of the following problem:

$$\begin{array}{rl} \text{maximize} & \displaystyle\sum_{l=1}^{L} w_l r_l \\ \text{subject to} & \mathbf{B}_k \operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \le (\mathbf{I} + \mathbf{B}_k)\mathbf{q}, \quad k = 1, \ldots, K, \\ & \mathbf{D}_m \operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \le (\mathbf{I} + \mathbf{D}_m)\mathbf{q}, \quad m = 1, \ldots, M, \\ \text{variables:} & \mathbf{r}, \mathbf{q}. \end{array} \tag{14}$$

Denote the optimal $\mathbf{r}$ and $\mathbf{q}$ in (14) by $\mathbf{r}^\star$ and $\mathbf{q}^\star$ respectively. Using (4) and (5), we have $\mathbf{p}^\star$ in (11) given by $\mathbf{p}^\star = \operatorname{diag}(e^{\mathbf{r}^\star} - \mathbf{1})\mathbf{q}^\star$.

Though (14) has a linear objective function, (14) is still nonconvex due to the nonconvex constraint set. However, the transformation from (11) to (14) plays an instrumental role that facilitates a novel use of nonnegative matrix theory for convex relaxation and algorithm design. Fig. 1 gives an overview of the development in the paper, particularly the convex relaxation, the algorithms and the connections between key optimization problems.

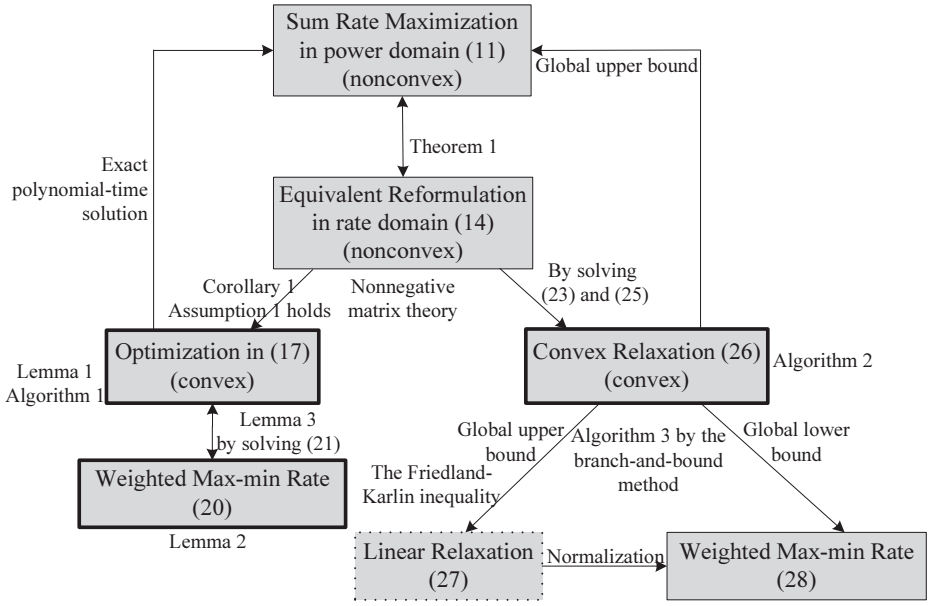


Fig. 1. Overview of our convex relaxation approach to the sum rate maximization problem. We obtain both a polynomial-time algorithm (for special cases) and a global optimization algorithm (for the general case). The interesting reformulations of the constraint sets among problems studied in this paper are illustrated in different boxes: the ones with solid boundaries are subject to original weighted power and interference temperature constraints, the ones with bold boundaries are subject to spectral radius constraints, and the one with dotted boundary is subject to linear constraints. Notably, (14) maximizes a linear objective function subject to a nonconvex constraint set, i,e, it is still nonconvex, but (14) enables the design of a polynomial-time algorithm (for special cases) and a global optimization algorithm (for the general cases).

IV. Convex Relaxation and Polynomial-time Algorithms

In this section, we first show that the constraints in (14) can be equivalently rewritten as convex spectral radius constraints for those special cases in Section IV-A that enable (14) to be solved in polynomial time. For the general case, we propose a convex relaxation approach in Section IV-B and a branch-and-bound method in Section V to compute the global optimal solution of (14), equivalently that of (11).

A. Polynomial-time Solvable Special Cases

The reformulation introduced in Theorem 1 allows us to decompose the weighted sum rate maximization problem in (11) into first optimizing $\mathbf{r}$ and $\mathbf{q}$ and then projecting the solution (through a one-to-one mapping) back to $\mathbf{p}$. Although (14) is nonconvex, the Perron-Frobenius theorem [24] enables us to rewrite (14) into one that has the rates characterized by a set of spectral radius constraints (i.e., $\mathbf{r}$ is the only variable) when the following assumption is satisfied.

Assumption 1: Let $\mathbf{B}_k$ and $\mathbf{D}_m$ be given in (12) and (13) respectively, the following is satisfied:

$$\widetilde{\mathbf{B}}_k = (\mathbf{I}+\mathbf{B}_k)^{-1}\mathbf{B}_k \geq \mathbf{0}, \quad k=1,\ldots,K, \tag{15}$$

$$\widetilde{\mathbf{D}}_m = (\mathbf{I}+\mathbf{D}_m)^{-1}\mathbf{D}_m \geq \mathbf{0}, \quad m=1,\ldots,M, \tag{16}$$

i.e., $\widetilde{\mathbf{B}}_k$ and $\widetilde{\mathbf{D}}_m$ are irreducible nonnegative matrices for all k and m.

When there is no interference, i.e., $G_{lj}=0,\ l\neq j$ and $\mathbf{F}$ in (2) is an all-zero matrix, we have $\mathbf{B}_k = \frac{1}{\bar{p}_k}\mathbf{v}\mathbf{a}_k^\top$, which implies that $\widetilde{\mathbf{B}}_k = \frac{1}{\bar{p}_k+\mathbf{a}_k^\top\mathbf{v}}\mathbf{v}\mathbf{a}_k^\top$ for all k. Thus, Assumption 1 can be satisfied under certain SNR and interference conditions. The conditions where Assumption 1 is more likely to hold are examined in Section VI.

Corollary 1: If Assumption 1 is satisfied, then (11) can be solved by the following convex optimization problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{l=1}^{L} w_l r_l \\ \text{subject to} \quad & \log\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0, \quad k=1,\ldots,K, \\ & \log\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0, \quad m=1,\ldots,M, \\ \text{variables:} \quad & \mathbf{r}, \end{aligned} \tag{17}$$

where $\widetilde{\mathbf{B}}_k$ and $\widetilde{\mathbf{D}}_m$ are given in (15) and (16) respectively.

Due to the log-convexity property of the Perron-Frobenius eigenvalue [25], $\log\rho\big(\widetilde{\mathbf{B}}_l \operatorname{diag}(e^{\mathbf{r}})\big)$ and $\log\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$ are convex functions in $\mathbf{r}$ for the irreducible nonnegative matrices $\widetilde{\mathbf{B}}_k$ and $\widetilde{\mathbf{D}}_m$ respectively. Hence, (17) is a convex optimization problem (implying that (11) can be solved in polynomial time). In fact, we can obtain a closed form solution to (11) whenever Assumption 1 holds.

Lemma 1: If Assumption 1 is satisfied, then we have

$$\begin{aligned} &\rho\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big)=1,\ \ \mathbf{p}^\star = \operatorname{diag}(e^{\mathbf{r}^\star}-\mathbf{1})\mathbf{x}\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big), \\ &\text{and}\ \ \mathbf{q}^\star = \mathbf{x}\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big), \\ &\text{where}\quad i = \arg\max_{k=1,\ldots,K} \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}^\star})\big), \end{aligned} \tag{18}$$

or

$$\begin{aligned} &\rho\big(\widetilde{\mathbf{D}}_\iota \operatorname{diag}(e^{\mathbf{r}^\star})\big)=1,\ \ \mathbf{p}^\star = \operatorname{diag}(e^{\mathbf{r}^\star}-\mathbf{1})\mathbf{x}\big(\widetilde{\mathbf{D}}_\iota \operatorname{diag}(e^{\mathbf{r}^\star})\big), \\ &\text{and}\ \ \mathbf{q}^\star = \mathbf{x}\big(\widetilde{\mathbf{D}}_\iota \operatorname{diag}(e^{\mathbf{r}^\star})\big), \\ &\text{where}\quad \iota = \arg\max_{m=1,\ldots,M} \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}^\star})\big). \end{aligned} \tag{19}$$

Next, we study the weighted max-min rate fairness problem whose solution is to provision rate fairness. How to adapt the weights in the weighted max-min rate fairness problem (a convex problem in general) to solve the sum rate maximization (a nonconvex problem in general)? This interesting and important connection between these two problems will be made precise in the following. Next, instead of solving (17) directly, let us consider the following weighted max-min rate problem with the same constraint set in (17):

$$\begin{array}{ll} \text{maximize} & \min\limits_{l=1,\ldots,L} \dfrac{r_l}{\beta_l} \\ \text{subject to} & \log\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \le 0, \quad k=1,\ldots,K, \\ & \log\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \le 0, \quad m=1,\ldots,M. \\ \text{variables:} & \mathbf{r}, \end{array} \tag{20}$$

where $\boldsymbol{\beta} \in \mathbb{R}_+^L$ are weights assigned to reflect priority among users. This means that a user requiring a higher quality-of-service has a larger weight, i.e., β_l is larger. In the following, we apply nonnegative matrix theory, particularly the Perron-Frobenius theorem, to characterize the optimal solution and the optimal value of (20).

Lemma 2: The optimal solution of (20) denoted by $\mathbf{r}^\star$ is a vector with $r_l^\star/\beta_l$ equal to a common value $\delta^\star$, i.e., $\mathbf{r}^\star = \delta^\star\boldsymbol{\beta}$, where $\delta^\star$ is the optimal value of (20).

We now explain how solving (20) can lead to a polynomial-time algorithm to compute the optimal solution of (17) that requires solving (20) as an intermediate step. In particular, with $\boldsymbol{\beta}$ appropriately chosen in (20), both (17) and (20) can have the same optimal solution.

Lemma 3: Both (17) and (20) have the same optimal solution with $\boldsymbol{\beta} = \big(\boldsymbol{\xi}^\star - (\log\varrho^\star)\mathbf{1}\big)/(\mathbf{1}^\top\boldsymbol{\xi}^\star - L\log\varrho^\star)$, where $\boldsymbol{\xi}^\star$ and $\log\varrho^\star$ are respectively the optimal solution and the optimal value of the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \max\Big\{ \max\limits_{k=1,\ldots,K} \Big\{\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}})\big)\Big\}, \\ & \qquad \max\limits_{m=1,\ldots,M} \Big\{\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big)\Big\}\Big\} \\ \text{subject to} & \sum\limits_{l=1}^{L} w_l\xi_l \ge 1, \\ \text{variables:} & \boldsymbol{\xi}. \end{array} \tag{21}$$

Remark 1: Using Lemma 3, an important special case will now be explained. Solving (20) with $\boldsymbol{\beta} = \mathbf{1}$ is equivalent to solving (17) with $\mathbf{w}$ given by $\mathbf{w} = \mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \arg_{\boldsymbol{\Omega}\in\{\widetilde{\mathbf{B}}_k,\widetilde{\mathbf{D}}_m\}} \max\big\{ \max\limits_{k=1,\ldots,K}\{\rho(\widetilde{\mathbf{B}}_k)\}, \max\limits_{m=1,\ldots,M}\{\rho(\widetilde{\mathbf{D}}_m)\}\big\}$. This interesting connection between (17) (a generally hard problem) and (20) (a polynomial-time solvable problem) is obtained using the Friedland-Karlin inequality (cf. Theorem 3.1 in [26]), which is a fundamental inequality in nonnegative matrix theory. In addition, this connection will be used in the convergence proofs of Algorithm 1 given below and the global optimization algorithm in Section V.

Suppose Assumption 1 holds, then the following two-time scale algorithm computes the optimal solution $\mathbf{r}^\star$ of (14): the outer loop iterates to solve (21) and to update the weight $\boldsymbol{\beta}(t)$ used in (20), which is solved by the inner loop.

Algorithm 1 (Polynomial-time Algorithm):

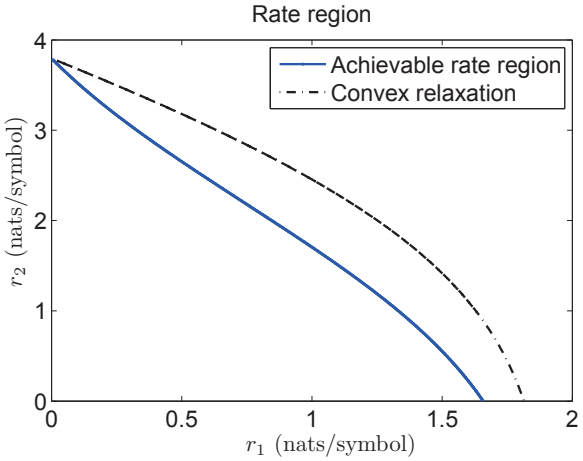
Initialize $\boldsymbol{\xi}(0)$.

1) Update $\boldsymbol{\xi}(t+1)$ as follows:
   if $\rho\big(\widetilde{\mathbf{B}}_{i_t} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big) \ge \rho\big(\widetilde{\mathbf{D}}_{\iota_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big)$
   $\xi_l(t+1) = \max\Big\{\xi_l(t) + w_l - \Big(\mathbf{x}\big(\widetilde{\mathbf{B}}_{i_t} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_{i_t} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big)\Big)_l, 0\Big\}$,
   else
   $\xi_l(t+1) = \max\Big\{\xi_l(t) + w_l - \Big(\mathbf{x}\big(\widetilde{\mathbf{D}}_{\iota_t} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_{\iota_t} \operatorname{diag}(e^{\boldsymbol{\xi}(t)})\big)\Big)_l, 0\Big\}$.
2) Normalize $\boldsymbol{\xi}(t+1)$:
   Compute $i_{t+1} = \arg\max\limits_{k=1,\ldots,K} \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)$ and $\iota_{t+1} = \arg\max\limits_{m=1,\ldots,M} \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)$.
   if $\rho\big(\widetilde{\mathbf{B}}_{i_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big) \ge \rho\big(\widetilde{\mathbf{D}}_{\iota_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)$
   $\boldsymbol{\xi}(t+1) \leftarrow \log\Big(e^{\boldsymbol{\xi}(t+1)}/\rho\big(\widetilde{\mathbf{B}}_{i_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)\Big)$,
   else
   $\boldsymbol{\xi}(t+1) \leftarrow \log\Big(e^{\boldsymbol{\xi}(t+1)}/\rho\big(\widetilde{\mathbf{D}}_{\iota_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)\Big)$.
3) Update $\boldsymbol{\beta}(t+1)$ as follows:
   $\varrho(k+1) = \max\{\rho\big(\widetilde{\mathbf{B}}_{i_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big), \rho\big(\widetilde{\mathbf{D}}_{\iota_{t+1}} \operatorname{diag}(e^{\boldsymbol{\xi}(t+1)})\big)\}$,
   $\boldsymbol{\beta}(k+1) = \dfrac{\boldsymbol{\xi}(t+1) - \big(\log\varrho(t+1)\big)\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\xi}(k+1) - L\log\varrho(t+1)}$.
4) Compute $\mathbf{p}(t+1)$ and $\mathbf{q}(t+1)$ by solving (20) with the weight replaced by $\boldsymbol{\beta}(t+1)$:
   while $\|\mathbf{p}(\tau) - \mathbf{p}(\tau-1)\| \ge \epsilon$
   $p_l(\tau+1) = \Big(\dfrac{\beta_l(t+1)}{\log(1+\mathsf{SINR}_l(\mathbf{p}(\tau)))}\Big) p_l(\tau)$,
   $\mathbf{p}(\tau+1) \leftarrow \dfrac{\mathbf{p}(\tau+1)}{\max\Big\{\max\limits_{k=1,\ldots,K}\{\frac{\mathbf{a}_k^\top\mathbf{p}(\tau+1)}{\bar{p}_k}\}, \max\limits_{m=1,\ldots,M}\{\frac{\mathbf{b}_m^\top\mathbf{F}\mathbf{p}(\tau+1)}{(\bar{q}_m - \mathbf{b}_m^\top\mathbf{v})}\}\Big\}}$,
   $r_l(\tau+1) = \log(1+\mathsf{SINR}_l(\mathbf{p}(\tau+1)))$,
   $q_l(\tau+1) = p_l(\tau+1)/(e^{r_l}-1)$.
   $\mathbf{p}(t+1) = \mathbf{p}(\tau)$, $\mathbf{q}(t+1) = \mathbf{q}(\tau)$, and $\mathbf{r}(t+1) = \mathbf{r}(\tau)$.
   when $\mathbf{p}(\tau)$ converges in the inner loop, and go to Step 1.

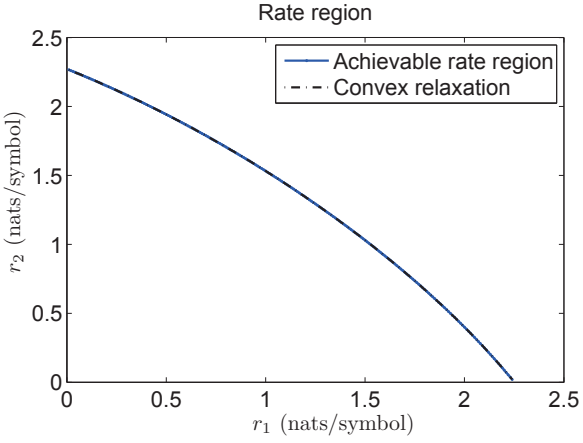
Lemma 4: Suppose Assumption 1 holds, starting from any initial point $\boldsymbol{\xi}(0)$, $\boldsymbol{\xi}(t)$ converges to the optimal solution of (21) and $\boldsymbol{\beta}(t)$ calculated in terms of $\boldsymbol{\xi}(t)$ is of the same scaling of the optimal solution $\mathbf{r}^\star$ of (17). In addition, $\mathbf{r}(t)$ converges to the optimal solution of (17) with $\mathbf{p}(t)$ and $\mathbf{q}(t)$ being the corresponding optimal power and interference temperature.

B. Convex Relaxation

In this section, we propose a novel convex relaxation to yield useful upper bounds to (11). Our convex relaxation stems from the quasi-invertibility introduced in Section IV-A. Whenever the quasi-inverse condition holds, the convex relaxation is in fact tight, otherwise the convex relaxation provides a useful bound to the global optimal value. In particular, we exploit the spectrum of specially-constructed nonnegative matrices ($\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$ introduced below) that are obtained from $\mathbf{B}_k$ and $\mathbf{D}_m$, thereby leveraging the techniques in Section IV-A.



(a)



(b)

Fig. 2. Achievable rate region and relaxed rate region for a 2-user case. The channel gains are given by $G_{11} = 0.85$, $G_{12} = 0.13$, $G_{21} = 0.14$ and $G_{22} = 0.87$. The noise power for both users are 0.1W. The other problem parameters are given as follows: (a) $\bar{\mathbf{p}} = (5,5)^\top$, $\mathbf{a}_1 = (10,1)^\top$, $\mathbf{a}_2 = (1,1)^\top$; (b) $\bar{\mathbf{p}} = (1,100)^\top$, $\mathbf{a}_1 = (1,1)^\top$, $\mathbf{a}_2 = (1,1)^\top$.

Let us introduce the following set of nonnegative matrices (without making Assumption 1):

$$\widehat{\mathbf{B}}_k = \big(\mathbf{I}+\mathbf{B}_k+\mathrm{diag}(\boldsymbol{\varepsilon})\big)^{-1}(\mathbf{B}_k-\widetilde{\mathbf{X}}_k^\star), \quad k=1,\ldots,K, \tag{22}$$

where $\widetilde{\mathbf{X}}_k^\star$ is obtained by solving

$$\begin{array}{ll} \text{minimize} & \|\widetilde{\mathbf{X}}_k\|_F \\ \text{subject to} & \big(\mathbf{I}+\mathbf{B}_k+\mathrm{diag}(\boldsymbol{\varepsilon})\big)^{-1}(\mathbf{B}_k-\widetilde{\mathbf{X}}_k)\geq 0, \\ & \widetilde{\mathbf{X}}_k \geq 0, \\ \text{variables:} & \widetilde{\mathbf{X}}_k, \end{array} \tag{23}$$

and

$$\widehat{\mathbf{D}}_m = \big(\mathbf{I}+\mathbf{D}_m+\mathrm{diag}(\boldsymbol{\varepsilon})\big)^{-1}(\mathbf{D}_m-\widetilde{\mathbf{Y}}_m^\star), \quad m=1,\ldots,M, \tag{24}$$

where $\widetilde{\mathbf{Y}}_m^\star$ is obtained by solving

$$\begin{array}{ll} \text{minimize} & \|\widetilde{\mathbf{Y}}_m\|_F \\ \text{subject to} & \big(\mathbf{I}+\mathbf{D}_m+\mathrm{diag}(\boldsymbol{\varepsilon})\big)^{-1}(\mathbf{D}_m-\widetilde{\mathbf{Y}}_m)\geq 0, \\ & \widetilde{\mathbf{Y}}_m \geq 0, \\ \text{variables:} & \widetilde{\mathbf{Y}}_m. \end{array} \tag{25}$$

Note that $\boldsymbol{\varepsilon}$ is a vector with each entry being a given small positive scalar (required in case $(\mathbf{I}+\mathbf{B}_k)$ or $(\mathbf{I}+\mathbf{D}_m)$ is not invertible); otherwise, $\boldsymbol{\varepsilon}$ can be an all-zero vector. Notably, (23) and (25) are convex optimization problems that can be solved numerically using interior-point solvers, e.g., the `cvx` software package [27]. Furthermore, $\widetilde{\mathbf{X}}_k^\star$ or $\widetilde{\mathbf{Y}}_m^\star$ are all-zero matrices whenever $\widetilde{\mathbf{B}}_k$ or $\widetilde{\mathbf{D}}_m$ are nonnegative, i.e., the quasi-inverse of $\widetilde{\mathbf{B}}_k$ or $\widetilde{\mathbf{D}}_m$ exists ($\widehat{\mathbf{B}}_k = \widetilde{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m = \widetilde{\mathbf{D}}_m$); otherwise $\widetilde{\mathbf{X}}_k^\star$ and $\widetilde{\mathbf{Y}}_m^\star$ are relatively small matrices (as compared to $\widetilde{\mathbf{B}}_k$ and $\widetilde{\mathbf{D}}_m$) with most of their entries being zeros.

We replace $\mathbf{B}_k$ and $\mathbf{D}_m$ on the righthand-side of the constraints in (14) such that we have, respectively, $(\mathbf{B}_k - \widetilde{\mathbf{X}}_k^\star)\,\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} \leq (\mathbf{I}+\mathbf{B}_k)\mathbf{q}$ and $(\mathbf{D}_m - \widetilde{\mathbf{Y}}_m^\star)\,\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} \leq (\mathbf{I}+\mathbf{D}_m)\mathbf{q}$. We can then compute the nonnegative $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$, and rewrite the constraints as $\widehat{\mathbf{B}}_k\,\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}$ and $\widehat{\mathbf{D}}_m\,\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}$. By using the Perron-Frobenius theorem, we consider the following convex optimization problem:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{l=1}^{L} w_l r_l \\ \text{subject to} & \log\rho\big(\widehat{\mathbf{B}}_k\,\mathrm{diag}(e^{\mathbf{r}})\big)\leq 0, \quad k=1,\ldots,K, \\ & \log\rho\big(\widehat{\mathbf{D}}_m\,\mathrm{diag}(e^{\mathbf{r}})\big)\leq 0, \quad m=1,\ldots,M, \\ \text{variables:} & \mathbf{r}. \end{array} \tag{26}$$

This relaxation given by (26) convexifies the nonconvex feasible region in (14) to one that is convex. Fig. 2 illustrates how the convex relaxation obtained by (26) produces a convex set for two different sets of parameter setting. The blue curves are the boundaries of the original pareto rate regions, while the dashed convex curves are the boundaries of the relaxed convex sets. As observed, Fig. 2(a) illustrates that the relaxation gap between the original rate region and the relaxed rate region can be remarkably small. Furthermore, when the (in general nonconvex) original rate region turns out to be convex in shape, the relaxed rate region coincides exactly with the original rate region, as illustrated in Fig. 2(b).

Observe that (26) has a problem structure (linear objective and spectral radius constraint set) similar to that of (17). This allows us to leverage Algorithm 1 to solve (26) as follows.

Algorithm 2 (Convex Relaxation Algorithm):

1) Solve (23) and (25) to obtain $\widetilde{\mathbf{X}}_k$ and $\widetilde{\mathbf{Y}}_m$ respectively. Then compute $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$ by (22) and (24) respectively.
2) Run Algorithm 1 with $\widetilde{\mathbf{B}}_k$ and $\widetilde{\mathbf{D}}_m$ replaced by $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$ respectively.

Although an upper bound of the optimal value of (11) can be computed by Algorithm 2 (which solves (26)), it is desirable that this bound can be further tightened and, more importantly, eventually yields the global optimal solution to (11). We pursue this by leveraging the relaxation in (26) and a branch-and-bound method [28], [29] in the next section.

V. Global Optimization Algorithm

In this section, we exploit the convex relaxation (26) together with the branch-and-bound method to iteratively

compute the global optimal solution of (11). Recall that we have an implicit box constraint set in (10), and this will be used as the initial set $\mathcal{B}_{\text{init}} = \{\mathbf{0} \le \mathbf{r} \le \bar{\mathbf{r}}\}$ in branch-and-bound and to be subdivided iteratively into smaller subsets for searching. In branch-and-bound, searching is organized using a binary tree, where the union of the sets represented by the leaf nodes is $\mathcal{B}_{\text{init}}$. At each leaf node, we can obtain a lower bound and an upper bound to (11) over the subset on that leaf node. Suppose the gap between the upper bound and the lower bound is some positive ϵ, which is nonincreasing with the search iterates. Then, branch-and-bound can locate an ϵ-optimal neighborhood of $\mathbf{r}^\star$. In the following, we give the details of this global optimization algorithm to compute the global optimal solution of (14), equivalently that of (11).

By applying the Friedland-Karlin inequality (cf. Theorem 3.1 in [26]) to the spectral radius constraints in (26), we can further relax (26) into a linear program (LP). In particular, the optimal value of the following LP is an upper bound to (26):

$$\begin{aligned}
\text{maximize} \quad & \sum_{l=1}^{L} w_l r_l \\
\text{subject to} \quad & \sum_{l=1}^{L} \big(\mathbf{x}(\widehat{\mathbf{B}}_k) \circ \mathbf{y}(\widehat{\mathbf{B}}_k)\big)_l r_l \le -\log \rho(\widehat{\mathbf{B}}_k), \\
& \qquad\qquad k = 1, \ldots, K, \\
& \sum_{l=1}^{L} \big(\mathbf{x}(\widehat{\mathbf{D}}_m) \circ \mathbf{y}(\widehat{\mathbf{D}}_m)\big)_l r_l \le -\log \rho(\widehat{\mathbf{D}}_m), \\
& \qquad\qquad m = 1, \ldots, M, \\
& \mathbf{0} \le \mathbf{r} \le \bar{\mathbf{r}}, \\
\text{variables:} \quad & \mathbf{r},
\end{aligned} \tag{27}$$

where $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$ are computed using (22) and (24) respectively, and $\mathbf{x}(\widehat{\mathbf{B}}_k) \circ \mathbf{y}(\widehat{\mathbf{B}}_k)$ and $\mathbf{x}(\widehat{\mathbf{D}}_m) \circ \mathbf{y}(\widehat{\mathbf{D}}_m)$ are probability vectors. Let $\acute{\mathbf{r}}$ denote the optimal solution of (27).

Observe that we have added an extra box constraint to (27) which will be subdivided iteratively. As mentioned, the LP in (27) is a relaxation of (26), which in turn is a relaxation of (11). The rate iterates obtained by solving this relaxed problem may not be feasible for (11). We propose a projection using the max-min rate fairness problem. In particular, with the weight of the objective function in (20) replaced by the optimal rate computed using the relaxed problem, and then solving (20) (which has the same feasible set as (11)), a projection onto the feasible rate domain of (11) is made at each iteration (thus yielding a feasible solution at each iteration as well as a lower bound to (11)). Replacing the weight in (20) by $\acute{\mathbf{r}}$ subject to the constraint set in (11), we thus have:

$$\begin{aligned}
\text{maximize} \quad & \min_{l=1,\ldots,L} \frac{\log(1 + \mathsf{SINR}_l(\mathbf{p}))}{\acute{r}_l} \\
\text{subject to} \quad & \mathbf{a}_k^\top \mathbf{p} \le \bar{p}_k, \quad k = 1, \ldots, K, \\
& \mathbf{b}_m^\top \mathbf{q} \le \bar{q}_m, \quad m = 1, \ldots, M, \\
\text{variables:} \quad & \mathbf{p}, \mathbf{q},
\end{aligned} \tag{28}$$

whose optimal solution is denoted by $\grave{\mathbf{r}}$. Note that $\grave{\mathbf{r}}$ is a suboptimal solution for (11). Thus, $\sum_{l=1}^{L} w_l \grave{r}_l$ is a lower bound to the optimal value of (11).

In summary, the computational methodology is as follows. Let $\mathcal{R}(t) = \{\mathcal{B}_1(t), \ldots, \mathcal{B}_J(t)\}$ denote the collection of box subsets $\mathcal{B}_j(t) = \{\underline{\mathbf{r}}^j(t) \le \mathbf{r} \le \bar{\mathbf{r}}^j(t)\}$ for all j at tth iteration, where $J(t) = |\mathcal{R}(t)|$ is the number of subsets in $\mathcal{R}(t)$. $\mathcal{B}_{\text{init}}$ is the initial rectangular constraint $\{\mathbf{0} \le \bar{\mathbf{r}}\}$ on the root node of the binary tree. In the tth iteration, we solve (27) over each subset $\mathcal{B}_j(t)$ to obtain the optimal solution $\acute{\mathbf{r}}^j(t)$. A lower bound $\mathbf{w}^\top \grave{\mathbf{r}}^j(t)$ is then obtained by solving (28) with the weight replaced by $\acute{\mathbf{r}}^j(t)$, i.e., taking the maximum over all the lower bound at each leaf nodes across all the levels in the binary tree. Next, we choose the maximum over all upper bounds as $U(t)$ and choose the maximum over all lower bounds as $L(t)$. We split $\mathcal{B}_j$, whose corresponding $\mathbf{w}^\top \acute{\mathbf{r}}^j(t)$ is maximal, into two subsets along one of its longest edges, removing it and adding the two new subsets to $\mathcal{R}(t)$. This method generates a nested sequence of box subsets from $\mathcal{B}_{\text{init}}$. The details are given in the following algorithm that computes the global optimal solution of (11).

Algorithm 3 (Global Optimization Algorithm):

Initialize $\mathcal{R}(0) = \{\mathcal{B}_{\text{init}}\}$.

1) Compute $J(t) = |\mathcal{R}(t)|$.
2) For each $\mathcal{B}_j(t) \in \mathcal{R}(t)$, solving the following LP:

$$\begin{aligned}
\text{maximize} \quad & \sum_{l=1}^{L} w_l r_l \\
\text{subject to} \quad & \sum_{l=1}^{L} \big(\mathbf{x}(\widehat{\mathbf{B}}_k) \circ \mathbf{y}(\widehat{\mathbf{B}}_k)\big)_l r_l \le -\log \rho(\widehat{\mathbf{B}}_k), \\
& \qquad\qquad k = 1, \ldots, K, \\
& \sum_{l=1}^{L} \big(\mathbf{x}(\widehat{\mathbf{D}}_m) \circ \mathbf{y}(\widehat{\mathbf{D}}_m)\big)_l r_l \le -\log \rho(\widehat{\mathbf{D}}_m), \\
& \qquad\qquad m = 1, \ldots, M, \\
& \mathbf{r} \in \mathcal{B}_j(t), \\
\text{variables:} \quad & \mathbf{r},
\end{aligned} \tag{29}$$

where $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{D}}_m$ are computed using (22) and (24) respectively, and we denote the optimal solution of (29) by $\acute{\mathbf{r}}^j(t)$.

Compute the optimal solution $\grave{\mathbf{r}}^j(t)$ of (28) with $\boldsymbol{\beta} = \acute{\mathbf{r}}^j(t)$ by the following computation which iterates as an inner loop:

while $\|\mathbf{p}'(\tau) - \mathbf{p}'(\tau - 1)\| \ge \epsilon$

$$p_l'(\tau + 1) = \left(\frac{\acute{r}_l^j(t)}{\log(1 + \mathsf{SINR}_l(\mathbf{p}'(\tau)))}\right) p_l'(\tau),$$

$$\mathbf{p}'(\tau + 1) \leftarrow \frac{\mathbf{p}'(\tau + 1)}{\max\Big\{\max\limits_{k=1,\ldots,K}\{\frac{\mathbf{a}_k^\top \mathbf{p}'(\tau+1)}{\bar{p}_k}\}, \max\limits_{m=1,\ldots,M}\{\frac{\mathbf{b}_m^\top \mathbf{F} \mathbf{p}'(\tau+1)}{(\bar{q}_m - \mathbf{b}_m^\top \mathbf{v})}\}\Big\}},$$

$$r_l'(\tau + 1) = \log(1 + \mathsf{SINR}_l(\mathbf{p}'(\tau + 1))),$$

$$q_l'(\tau + 1) = p_l'(\tau + 1)/(e^{r_l'} - 1).$$

$\grave{\mathbf{r}}^j(t) = \mathbf{r}'(\tau)$ when this inner loop converges.

3) Update the upper bound $L(t)$ and lower bound $U(t)$ as follow:

$$\mathcal{J} = \arg \max_{j=1,\ldots,J(t)} \mathbf{w}^\top \acute{\mathbf{r}}^j(t),$$

$$L(t) = \max_{j=1,\ldots,J(t)} \mathbf{w}^\top \grave{\mathbf{r}}^j(t),$$

$$U(t) = \max_{j=1,\ldots,J(t)} \mathbf{w}^\top \acute{\mathbf{r}}^j(t).$$

4) Update the collection of box subsets:

if $U(t) - L(t) \leq \epsilon$
Algorithm 3 terminates.
else
For any $\mathbf{w}^\top \grave{\mathbf{r}}^j(t) \leq U(t)$,
$\mathcal{R}(t+1) = \mathcal{R}(t) - \{\mathcal{B}_j(t)\}$.
Split $\mathcal{B}_{\mathcal{J}}(t)$ along one of its longest edges into $\mathcal{B}_I$ and $\mathcal{B}_{II}$,
$\mathcal{R}(t+1) = \big(\mathcal{R}(t) - \{\mathcal{B}_{\mathcal{J}}(t)\}\big) \bigcup \{\mathcal{B}_I, \mathcal{B}_{II}\}$.
Set $t \leftarrow t+1$, and go to Step 1.

Theorem 2: The iterates $\grave{\mathbf{r}}^{\mathcal{J}}(t)$ in Algorithm 3 converge to the ϵ-optimal solution of (14).

We add the following remarks regarding the implementation issues and the optimality of Algorithm 3.

Remark 2: The limit point of $\grave{\mathbf{r}}^{\mathcal{J}}(t)$ solves (11) by the following equations in (30) and (31) (see next page).

Remark 3: In Algorithm 3, we can delete the subsets that satisfy $\mathbf{w}^\top \grave{\mathbf{r}}^j(t) \leq U(t)$ by eliminating them from $\mathcal{R}(t)$ at Step 4, since these subsets do not contain the optimal solution and need not be considered in subsequent iterations. This reduces overall searching time of Algorithm 3.

Remark 4: At the tth iteration of Algorithm 3, $L(t) = \mathbf{w}^\top \acute{\mathbf{r}}^{\mathcal{J}}(t)$ is the maximum over all the upper bounds at each leaf nodes in the binary tree, which gives a global upper bound on the optimal value of (11). Practical stopping criterions for Algorithm 3 can be $U(t) - L(t) \leq \epsilon$ for a given small ϵ. This means that we have $\mathbf{w}^\top \mathbf{r}^\star \leq \mathbf{w}^\top \grave{\mathbf{r}}^{\mathcal{J}}(t) + \epsilon$. Therefore, whenever Algorithm 3 terminates with an error tolerance number ϵ, $\grave{\mathbf{r}}^{\mathcal{J}}(t)$ is an ϵ-optimal solution of (11).

In general, the worst case computational complexity of the branch-and-bound method is exponential. However, the average complexity of Algorithm 3 can be significantly lower: First, the initial relaxed set that leverages the Friedland-Karlin inequality in nonnegative matrix theory is a clever choice. Second, the number of iterations in branch and bound increases with the problem size. Thus, a stopping criterion ϵ can be used in Algorithm 3 to balance the solution accuracy and the number of iterations. Third, Algorithm 3 updates the global upper and lower bounds by solving a LP. Being able to solve a sequence of (polynomial-time solvable) LPs in each iteration enables a reasonably accurate solution to be computed with low complexity.

VI. Numerical Examples

In this section, we evaluate the performance of our algorithms to solve (11). Our optimization software implementation (that leverages the `cvx` software [27]) is available for download at [19]. We first evaluate the performance of Algorithm 1 using numerical examples in which Assumption 1 holds. In the general case (when Assumption 1 does not hold), we evaluate the performance of Algorithm 3.

To illustrate the validity of Assumption 1, we find the average number of instances when Assumption 1 holds for a 10-user network in Table I. The percentages $\mathbb{1}\{\widetilde{\mathbf{B}}_k \geq 0\}$ represents a statistical average (normalized to 100%) of when the nonnegative $\widetilde{\mathbf{B}}_k$ for all k exist, and $\mathbb{1}\{\widetilde{\mathbf{D}}_m \geq 0\}$ represents a statistical average (normalized to 100%) of when the nonnegative $\widetilde{\mathbf{D}}_m$ for all m exist. Roughly speaking, this indicates on average how often Assumption 1 holds. We generate each parameter randomly in a given range for fifty thousand times, and record the number of times that $\widetilde{\mathbf{B}}_k$ for all k and $\widetilde{\mathbf{D}}_m$ for all m are nonnegative. From Table I, we observe that both $\widetilde{\mathbf{B}}_k$ for all k and $\widetilde{\mathbf{D}}_m$ for all m are nonnegative when the cross-channel interference is relatively low. In other words, Assumption 1 is more likely to hold with not only smaller $G_{lj}(l \neq j)$ but also larger $\bar{\mathbf{q}}$. In addition, a smaller number of constraints and a smaller problem size make Assumption 1 more likely to hold.

Consider the following channel gain matrix for three users given by

$$\mathbf{G} = \begin{bmatrix} 1.55 & 0.02+\sigma & 0.03+\sigma \\ 0.05+\sigma & 1.50 & 0.08+\sigma \\ 0.05+\sigma & 0.02+\sigma & 1.62 \end{bmatrix}, \qquad (32)$$

where σ is a constant added to the non-diagonal entries of the channel gain matrix (which increase the interference in the network). To compute the global optimal sum rates, depending on σ, we evaluate the performance of Algorithm 1 whenever Assumption 1 holds, otherwise we evaluate the performance of Algorithm 3.

We assume that there are two power constraints and an interference temperature constraint ($K = 2$ and $M = 1$). The weights for the power and interference temperature constraints are: $\mathbf{a}_1 = (0.27, 0.51\,0.22)^\top$, $\mathbf{a}_2 = (0.32, 0.74, 0.61)^\top$ and $\mathbf{b}_1 = (0.63, 0.36, 0.56)^\top$. We also set $\bar{\mathbf{p}} = (1.0, 1.5)^\top$ and $\bar{q} = 1.2$. The noise power of each user is 1 W, and the weight for the sum rate function is $\mathbf{w} = (0.30, 0.37, 0.24)^\top$. First, we let $\sigma = 0$ in (32). Under this setting, $\widetilde{\mathbf{B}}_1$, $\widetilde{\mathbf{B}}_2$ and $\widetilde{\mathbf{D}}_1$ all exist, thus (11) can be solved optimally by Algorithm 1. Fig. 3 plots the evolution of rates for the three users that run Algorithm 1. We set the initial rate vector to $\mathbf{r}(0) = (2.11, 2.25, 2.52)^\top$ nats/symbol, and run Algorithm 1 for 10 iterations before it terminates. Fig. 3(a) shows that Algorithm 1 converges to the optimal solution (verifying Lemma 4). The optimal rate $\mathbf{r}^\star$ is $(2.75, 2.10, 2.26)^\top$ nats/symbol, which satisfies $\rho\big(\widetilde{\mathbf{B}}_2 \operatorname{diag}(e^{\mathbf{r}^\star})\big) = 1$ (verifying Lemma 1). We also run Algorithm 1 with a suitably large number of primary and secondary users. Fig. 3(b) shows the rate evolution for five users out of one hundred users, demonstrating that Algorithm 1 converges very fast even when the number of users is large.

Fig. 4 plots the evolution of the lower bound and upper bound to the optimal value of (11) for two users that run Algorithm 3. The parameter settings in Figs. 4(a) and (b) are the same as those in Figs. 2(a) and (b) respectively. The numerical example in Fig. 4(a) that runs Algorithm 3 converges at the 76th iteration with $\epsilon = 10^{-3}$, and the numerical example in Fig. 4(b) that runs Algorithm 3 converges at the 130th iteration with $\epsilon = 10^{-3}$ (verifying Theorem 2). The optimal $\mathbf{r}^\star$ in Fig. 4(a) is $(0, 1.68)^\top$ nats/symbol, and the optimal $\mathbf{r}^\star$ in Fig. 4(b) is $(2.23, 0.04)^\top$ nats/symbol.

Next, we compare Algorithm 3 with other previously proposed global optimization algorithms for solving (11), i.e., the branch-and-bound method in [16] and the outer approximation method in [15]. Notice that both [16] and [15] solve (11) using a reformulation in the $\tilde{\boldsymbol{\gamma}} = \log \mathsf{SINR}(\mathbf{p})$ domain (which is an unbounded constraint set). The unboundedness feature introduces numerical artifacts in the global optimization algorithms in [16] and [15].

TABLE I
TYPICAL NUMERICAL EXAMPLE IN A 10-USER NETWORK WITH DIFFERENT PARAMETER SETTINGS. ALL THE SETTINGS USE $\mathbf{a}_k \in [\mathbf{0}, \mathbf{1}]$, $\mathbf{b}_m \in [\mathbf{0}, \mathbf{1}]$ AND $\mathbf{n} = \mathbf{1}$. THE STATISTICAL PERCENTAGES OF INSTANCES WHERE $\widetilde{\mathbf{B}}_k$ EXISTS AND $\widetilde{\mathbf{D}}_m$ EXISTS ARE BOTH RECORDED.

L	$\mathbf{G}$	K	$\bar{\mathbf{p}}$	$\mathbb{1}\{\widetilde{\mathbf{B}}_k \geq \mathbf{0}\}$	M	$\bar{\mathbf{q}}$	$\mathbb{1}\{\widetilde{\mathbf{D}}_m \geq \mathbf{0}\}$
5	$G_{lj} \in [0.01, 0.04]$, $l \neq j$ $G_{lj} \in [1.50, 2.00]$, $l = j$	5	$\bar{\mathbf{p}} \in [\mathbf{1.5}, \mathbf{2.0}]$	44.14%	5	$\bar{\mathbf{q}} \in [\mathbf{2.5}, \mathbf{3.0}]$	99.19%
8	$G_{lj} \in [0.005, 0.01]$, $l \neq j$ $G_{lj} \in [2.50, 3.00]$, $l = j$	4	$\bar{\mathbf{p}} \in [\mathbf{1.0}, \mathbf{1.2}]$	72.68%	6	$\bar{\mathbf{q}} \in [\mathbf{2.0}, \mathbf{2.2}]$	93.00%
10	$G_{lj} \in [0.005, 0.01]$, $l \neq j$ $G_{lj} \in [2.50, 3.00]$, $l = j$	5	$\bar{\mathbf{p}} \in [\mathbf{0.4}, \mathbf{0.6}]$	41.56%	5	$\bar{\mathbf{q}} \in [\mathbf{2.0}, \mathbf{2.5}]$	37.68%
10	$G_{lj} \in [0.005, 0.01]$, $l \neq j$ $G_{lj} \in [2.50, 3.00]$, $l = j$	5	$\bar{\mathbf{p}} \in [\mathbf{1.5}, \mathbf{2.0}]$	52.63%	5	$\bar{\mathbf{q}} \in [\mathbf{2.5}, \mathbf{3.0}]$	98.63%
10	$G_{lj} \in [0.01, 0.04]$, $l \neq j$ $G_{lj} \in [1.50, 2.00]$, $l = j$	5	$\bar{\mathbf{p}} \in [\mathbf{1.5}, \mathbf{2.0}]$	53.18%	5	$\bar{\mathbf{q}} \in [\mathbf{2.5}, \mathbf{3.0}]$	60.76%
10	$G_{lj} \in [0.005, 0.01]$, $l \neq j$ $G_{lj} \in [2.50, 3.00]$, $l = j$	7	$\bar{\mathbf{p}} \in [\mathbf{1.5}, \mathbf{2.0}]$	40.29%	10	$\bar{\mathbf{q}} \in [\mathbf{2.5}, \mathbf{3.0}]$	97.14%

$$\mathbf{p}^\star = \left(\mathbf{I} + \mathbf{F} - \operatorname{diag}\left(e^{\lim_{t\to\infty} \grave{\mathbf{r}}^{\mathcal{J}}(t)}\right)\mathbf{F}\right)^{-1}\left(\operatorname{diag}\left(e^{\lim_{t\to\infty} \grave{\mathbf{r}}^{\mathcal{J}}(t)} - \mathbf{1}\right)\right)\mathbf{v}, \tag{30}$$

$$\mathbf{q}^\star = \left(\mathbf{I} + \mathbf{F} - \mathbf{F}\operatorname{diag}\left(e^{\lim_{t\to\infty} \grave{\mathbf{r}}^{\mathcal{J}}(t)}\right)\right)^{-1}\mathbf{v} \tag{31}$$

Let us consider $\sigma = 0.3$ in (32) and the other parameters are the same as that used in the example of Fig. 3 (under this setting Assumption 1 does not hold). This 3-user numerical example in Fig. 5(a) runs Algorithm 3 in this paper and the initial phase of Algorithm 2 in [16] respectively; the gap between the upper and lower bounds of our Algorithm 3 is 0.016 after 160 seconds, while Algorithm 2 in [16] still has a gap of 0.339. We next illustrate a 100-user numerical example with all parameters randomly generated in Fig. 5 that (b) runs Algorithm 3 in this paper and the initial phase of Algorithm 2 in [16] respectively. As shown, the gap between the lower and upper bounds is 1.274 for Algorithm 3, while the gap is still 2.214 for Algorithm 2 in [16]. We can observe from Fig. 5 that the relaxation gap for Algorithm 3 outperforms that in [16].

Next, we compare the outer approximation method in [15] that generates a nested sequence of polyhedrons and requires searching the vertices of these polyhedrons. Using a point $\tilde{\boldsymbol{\gamma}}(t)$ at time t of running Algorithm 1 in [15], we can obtain an upper bound to the optimal value by $\sum_{l=1}^{L} w_l \log(1 + e^{\tilde{\gamma}_l(t)})$, and a lower bound to the optimal value by mapping $\mathbf{p}(\tilde{\boldsymbol{\gamma}}(t)) = \left(\mathbf{I} - \operatorname{diag}(e^{\tilde{\boldsymbol{\gamma}}(t)})\right)^{-1}\operatorname{diag}(e^{\tilde{\boldsymbol{\gamma}}(t)})\mathbf{v}$ to the boundary of the constraint set, i.e., $\mathbf{p}'(\tilde{\boldsymbol{\gamma}}(t)) = \frac{\mathbf{p}(\tilde{\boldsymbol{\gamma}}(t))}{\max\left\{\max_{k=1,\ldots,K}\{\frac{\mathbf{a}_k^\top \mathbf{p}(\tilde{\boldsymbol{\gamma}}(t))}{\bar{p}_k}\}, \max_{m=1,\ldots,M}\{\frac{\mathbf{b}_m \mathbf{F}^\top \mathbf{p}(\tilde{\boldsymbol{\gamma}}(t))}{(\bar{q}_m - \mathbf{b}_m^\top \mathbf{v})}\}\right\}}$, and then computing $\sum_{l=1}^{L} w_l \log\left(1 + p_l'(\tilde{\boldsymbol{\gamma}}(t))/(\mathbf{F}\mathbf{p}'(\tilde{\boldsymbol{\gamma}}(t)) + \mathbf{v})_l\right)$.

In addition, we consider an equivalent formulation of (26) with the constraints replaced by $\widehat{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}$, $k = 1, \ldots, K$ and $\widehat{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}$, $m = 1, \ldots, M$. Using the logarithmic transformation, i.e., $\mathbf{q} = e^{\tilde{\mathbf{q}}}$ and then taking the logarithm on both sides of the inequality constraints, (26) can be written as a geometric program (GP) in terms of $\mathbf{r}$ and $\tilde{\mathbf{q}}$. This relaxation for (11) is tighter than (27), but requires a GP solver that is more complex than a LP solver. We also evaluate the case when (29) in Algorithm 3 is replaced by the GP that is an equivalent formulation of (26) in the following.

We compare the running time of Algorithm 1 in [15], Algorithm 3 with the LP, and Algorithm 3 with (29) replaced by a GP on an Intel (R) Core(TM) i7-2620M CPU 2.70GHz processor with a memory RAM of 4.00 GB (3.89 GB usable) and a 64-bit Unix operating system. We use $\sigma = 0.3$ in (32) and set the other parameters to be the same as that used in the examples of Fig. 3 and Fig. 5(a). This 3-user numerical example in Fig. 6(a) runs Algorithm 1 in [15] and Algorithm 3 in this paper with the LP and the GP respectively; Algorithm 1 in [15] converges much faster than Algorithm 3 (with $\epsilon = 0.02$), i.e., 2.115 seconds for Algorithm 1 in [15], 114.983 and 358.802 seconds respectively for Algorithm 3 with LP and GP. When the number of users is increased, e.g., the 8-user numerical example in Fig. 6(b), Algorithm 3 with the LP has a faster running time, i.e., 2387.828 seconds compared with 3949.179 seconds for Algorithm 1 in [15] (with $\epsilon = 0.2$). From Fig. 6, we observe that the running time of Algorithm 3 with the LP can be much faster than the running time of Algorithm 3 with the GP.

VII. CONCLUSION

We studied the weighted sum rate maximization problem with transmit power and interference temperature constraints in a cognitive radio network. To tackle the nonconvexity, we proposed a reformulation-relaxation technique by leveraging the nonnegative matrix theory. The reformulation transformed the original problem in the power and interference temperature domain to one with the achievable rates as explicit optimization variables. A convex relaxation to the sum rates was obtained by solving a convex optimization problem over a closed bounded convex set in the rate domain. The global optimality can even be characterized by the spectra of specially-crafted nonnegative matrices. Polynomial-time verifiable sufficient conditions were established to identify problem instances that can be solved optimally in polynomial time. We also provided an optimality equivalence between the

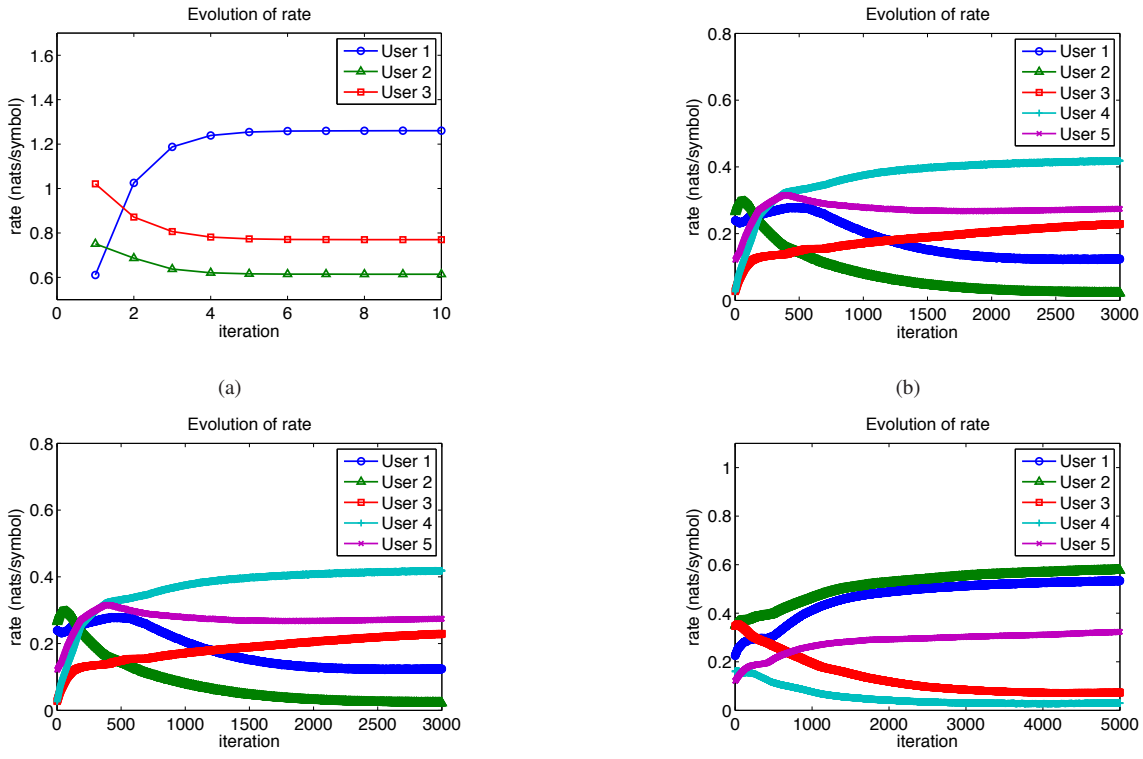


Fig. 3. Illustration of the convergence of Algorithm 1 for (a) 3 users, (b) 50 users, (c) 100 users , and (d) 200 users. For the numerical examples with large problem sizes, the problem parameters are generated randomly in a given range, and Assumption 1 holds for all of them. In (b), (c) and (d) of this figure, we show the rate evolution for only five users. We can observe from the figure that the convergence of Algorithm 1 is fast.

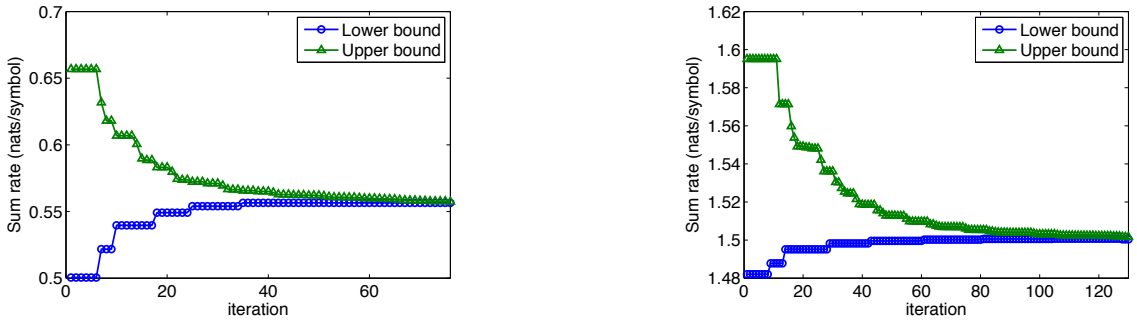


Fig. 4. Illustration of the convergence of Algorithm 3. The problem parameters are set the same as those in Fig. 2, and the weights for both of the numerical examples are set as $\mathbf{w} = (0.67,\ 0.33)^\top$.

nonconvex weighted sum rate maximization and the convex weighted max-min rate problem, and exploited it for algorithm design. In the general case, we proposed a global optimization algorithm to solve the sum rate maximization using the convex relaxation and branch-and-bound method. Numerical evaluations demonstrated that our proposed algorithms exhibited fast convergence behavior even for large-scale problems.

APPENDIX

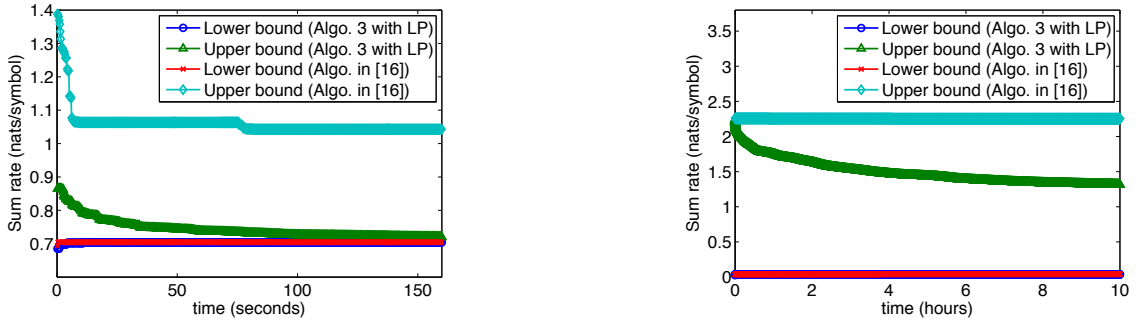
A. Proof of Theorem 1

For each lth user, by combining (3) and (5), its rate can be written as $e^{r_l} = p_l/(\mathbf{F}\mathbf{p}+\mathbf{v})_l + 1, \quad l = 1, \ldots, L$, which, in matrix form, can be given by $\big(\mathrm{diag}(e^{\mathbf{r}}) - \mathbf{I}\big)(\mathbf{F}\mathbf{p}+\mathbf{v}) = \mathbf{p}$. Substituting (4) into the above equation, it can be rewritten as $\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} = \mathbf{p} + \mathbf{q}$, whose righthand-side and lefthand-side are then both multiplied by $\mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top$ to obtain:

$$\begin{aligned} & \big(\mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top\big)\,\mathrm{diag}(e^{\mathbf{r}})\mathbf{q} \\ = \ & \big(\mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top\big)\mathbf{p} + \big(\mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top\big)\mathbf{q} \\ \overset{(a)}{\leq} \ & (\mathbf{F}\mathbf{p}+\mathbf{v}) + \big(\mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top\big)\mathbf{q} \\ \overset{(b)}{=} \ & \big(\mathbf{I} + \mathbf{F} + (1/\bar{p}_k)\mathbf{v}\mathbf{a}_k^\top\big)\mathbf{q}, \end{aligned} \tag{33}$$

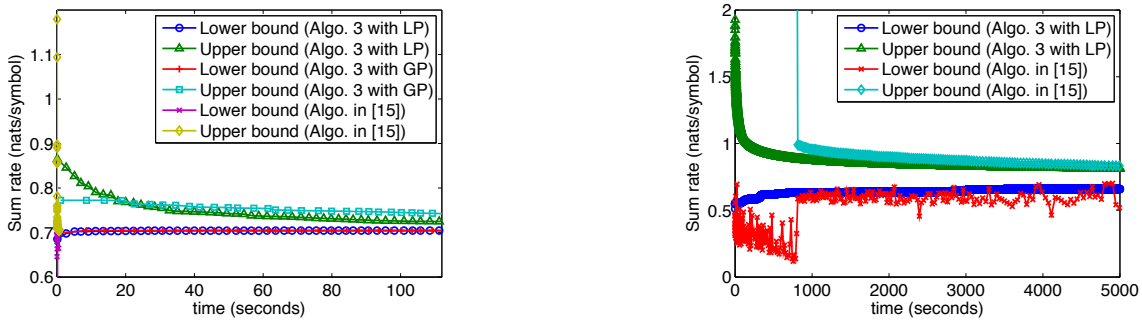
where (a) is due to the power constraint $(1/\bar{p}_k)\mathbf{a}_k^\top\mathbf{p} \leq 1$, and (b) is due to (4).

We first establish the following two matrix equations that



(a) (b)

Fig. 5. Comparison of the upper and lower bounds between Algorithm 3 in this paper and the global optimization algorithm proposed in [16]. (a) 3-user numerical example. We use $\sigma = 0.3$ in (32) and set the other parameters the same as that are used in Fig. 3; (b) 100-user numerical example. The channel gain is generated randomly in $G_{lj} \in [0.1,\ 0.2]$, $l \neq j$ and $G_{lj} \in [0.6,\ 0.9]$, $l = j$. We assume that the number of power and interference temperature constraints altogether are the same as the number of user with $\mathbf{a}_k \in [\mathbf{0},\ \mathbf{1}]$ and $\mathbf{b}_m \in [\mathbf{0},\ \mathbf{1}]$ randomly generated. We also set $\bar{\mathbf{p}} \in [\mathbf{10},\ \mathbf{20}]$ and $\bar{\mathbf{q}} \in [\mathbf{10},\ \mathbf{20}]$.



(a) (b)

Fig. 6. Comparison of the upper and lower bounds for the global optimization algorithm proposed in [15] and Algorithm 3 with LP and GP respectively. (a) 3-user numerical example. We use $\sigma = 0.3$ in (32) and set the other parameters the same as that are used in Fig. 3; (b) 8-user numerical example. The channel gain is generated randomly in $G_{lj} \in [0.1,\ 0.2]$, $l \neq j$ and $G_{lj} \in [0.6,\ 0.9]$, $l = j$. We assume that the number of power and interference temperature constraints altogether are the same as the number of user with $\mathbf{a}_k \in [\mathbf{0},\ \mathbf{1}]$ and $\mathbf{b}_m \in [\mathbf{0},\ \mathbf{1}]$ randomly generated. We also set $\bar{\mathbf{p}} \in [\mathbf{10},\ \mathbf{20}]$ and $\bar{\mathbf{q}} \in [\mathbf{10},\ \mathbf{20}]$.

are useful for a reformulation of the interference temperature constraints. The first one is:

$$\begin{aligned}
&\left(\mathbf{I} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right)^{-1} \overset{(c)}{=} \mathbf{I} + \lim_{N\to\infty}\sum_{n=1}^{N}\left(\frac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right)^n \\
&= \mathbf{I} + \lim_{N\to\infty}\sum_{n=1}^{N}\frac{1}{\bar{q}_m}\left(\frac{\mathbf{b}_m^\top\mathbf{v}}{\bar{q}_m}\right)^{n-1}\mathbf{v}\mathbf{b}_m^\top \\
&\overset{(d)}{=} \mathbf{I} + \lim_{N\to\infty}\frac{1}{\bar{q}_m}\left(\frac{1-(\mathbf{b}_m^\top\mathbf{v}/\bar{q}_m)^{N-1}}{1-\mathbf{b}_m^\top\mathbf{v}/\bar{q}_m}\right)\mathbf{v}\mathbf{b}_m^\top \\
&\overset{(e)}{=} \mathbf{I} + \big(1/(\bar{q}_m - \mathbf{b}_m^\top\mathbf{v})\big)\mathbf{v}\mathbf{b}_m^\top,
\end{aligned} \tag{34}$$

where (c) is due to the von Neumann's series expansion, (d) is due to the geometric series formula, and (e) is due to $\bar{q}_m \geq \mathbf{b}_m^\top\mathbf{v}$ for a nonempty constraint set so that $\lim_{N\to\infty}(\mathbf{b}_m^\top\mathbf{v}/\bar{q}_m)^{N-1} = 0$. Using (34), the second matrix equation is given as:

$$\begin{aligned}
&\mathbf{I} + \mathbf{F} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top \\
&= \left(\mathbf{I} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right) + \left(\mathbf{I} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right)\left(\mathbf{I} + \tfrac{1}{\bar{q}_m - \mathbf{b}_m^\top\mathbf{v}}\mathbf{v}\mathbf{b}_m^\top\right)\mathbf{F} \\
&= \left(\mathbf{I} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right)\left(\mathbf{I} + \mathbf{F} + \big(1/(\bar{q}_m - \mathbf{b}_m^\top\mathbf{v})\big)\mathbf{v}\mathbf{b}_m^\top\mathbf{F}\right).
\end{aligned} \tag{35}$$

Next, substituting $\mathbf{p} = \operatorname{diag}(e^{\mathbf{r}} - \mathbf{1})\mathbf{q}$ and $(1/\bar{q}_m)\mathbf{b}_m^\top\mathbf{q} \leq 1$ into (4), we have:

$$\begin{aligned}
&\mathbf{F}\big(\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} - \mathbf{q}\big) + (1/\bar{q}_m)\mathbf{v}\mathbf{b}_m^\top\mathbf{q} \leq \mathbf{q} \\
\Rightarrow\ &\mathbf{F}\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \overset{(f)}{\leq} \\
&\qquad \left(\mathbf{I} - \tfrac{1}{\bar{q}_m}\mathbf{v}\mathbf{b}_m^\top\right)\left(\mathbf{I} + \mathbf{F} + \tfrac{1}{\bar{q}_m - \mathbf{b}_m^\top\mathbf{v}}\mathbf{v}\mathbf{b}_m^\top\mathbf{F}\right)\mathbf{q} \\
\Rightarrow\ &\left(\mathbf{I} + \tfrac{1}{\bar{q}_m - \mathbf{b}_m^\top\mathbf{v}}\mathbf{v}\mathbf{b}_m^\top\right)\mathbf{F}\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \overset{(g)}{\leq} \\
&\qquad \left(\mathbf{I} + \mathbf{F} + \tfrac{1}{\bar{q}_m - \mathbf{b}_m^\top\mathbf{v}}\mathbf{v}\mathbf{b}_m^\top\mathbf{F}\right)\mathbf{q},
\end{aligned} \tag{36}$$

where (f) is due to (35) and (g) is due to (34). Thus, both (33) and (36) form the constraint set in (14). By substituting (5), we rewrite the objective function in (11) as $\sum_{l=1}^{L} w_l r_l$.

B. Proof of Corollary 1

Since the nonnegative matrix $\widetilde{\mathbf{B}}_k$ exists (cf. Assumption 1), the first constraint in (14) can be rewritten as:

$$\begin{aligned}
&\mathbf{B}_k\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq (\mathbf{I} + \mathbf{B}_k)\mathbf{q} \\
\Rightarrow\ &(\mathbf{I} + \mathbf{B}_k)^{-1}\mathbf{B}_k\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}, \\
\Rightarrow\ &\widetilde{\mathbf{B}}_k\operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q},
\end{aligned} \tag{37}$$

where $\mathbf{B}_k$ and $\widetilde{\mathbf{B}}_k$ are given in (12) and (15) respectively. Likewise, since the nonnegative matrix $\widetilde{\mathbf{D}}_m$ exists according to Assumption 1, the second constraint in (14) can be rewritten as $\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\mathbf{q} \leq \mathbf{q}, \quad m = 1, \ldots, M$, where $\mathbf{D}_m$ and $\widetilde{\mathbf{D}}_m$ are given in (13) and (16) respectively.

We now state the following lemma:

Lemma 5 (The Subinvariance Theorem [30]): Let $\mathbf{A}$ be a nonnegative irreducible matrix, Λ be a positive number, and $\boldsymbol{v}$ be a nonnegative vector satisfying $\mathbf{A}\boldsymbol{v} \leq \Lambda \boldsymbol{v}$. Then $\boldsymbol{v} > 0$ and $\Lambda > \rho(\mathbf{A})$. Moreover, $\Lambda = \rho(\mathbf{A})$ if and only if $\mathbf{A}\boldsymbol{v} = \Lambda \boldsymbol{v}$.

Letting $\Lambda = 1$, $\mathbf{A} = \widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})$ and $\Lambda = 1$, $\mathbf{A} = \widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})$ respectively in Lemma 5, the inequalities in (37) can be further written as $\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \leq 1$ and $\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \leq 1$ respectively. This algebraic manipulation transforms the constraint set in (14) to one that is sorely a spectral radius constraint set, which is convex (due to the log-convex property of the spectral radius function [25]).

C. Proof of Lemma 1

The first order derivative of $\log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)$ and $\log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$ at $\mathbf{r}$ are given by $\partial \log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)/\partial \mathbf{r} = \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)$ and $\partial \log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)/\partial \mathbf{r} = \mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$ respectively [16], [26]. Due to that $\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})$ and $\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})$ are nonnegative matrices, their Perron right eigenvector $\mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)$, $\mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$ and left eigenvector $\mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)$, $\mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$ are then all nonnegative [30], which implies that the spectral radius functions in the constraint set of (17) are monotonically increasing. Since the objective function in (17) is linear with positive weight and thus increasing in $\mathbf{r}$, maximizing the objective function in (17) leads to at least one of the constraints becoming tight. This also means that the corresponding constraint in (14) becomes tight. If i indexes the ith tight power constraint, we have $\rho\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big) = 1$ in (17) and $\mathbf{B}_i \operatorname{diag}(e^{\mathbf{r}^\star})\mathbf{q}^\star = (\mathbf{I} + \mathbf{B}_i)\mathbf{q}^\star$ in (14) respectively, which implies that $\mathbf{q}^\star = \mathbf{x}\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big)$. By combining $\mathbf{q}^\star = \mathbf{x}\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big)$ and $\mathbf{p} = \operatorname{diag}(e^{\mathbf{r}^\star} - \mathbf{1})\mathbf{q}$, we can obtain $\mathbf{p}^\star = \operatorname{diag}(e^{\mathbf{r}^\star} - \mathbf{1})\mathbf{x}\big(\widetilde{\mathbf{B}}_i \operatorname{diag}(e^{\mathbf{r}^\star})\big)$.

D. Proof of Lemma 2

We first write (20) in epigraph form as:

$$\begin{aligned} \text{maximize} \quad & \delta \\ \text{subject to} \quad & \delta \leq r_l/\beta_l, \quad l = 1, \ldots, L, \\ & \log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0, \quad k = 1, \ldots, K, \\ & \log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0, \quad m = 1, \ldots, M, \\ \text{variables:} \quad & \mathbf{r}, \delta. \end{aligned} \tag{38}$$

Next, we form the full Lagrangian for (38) by introducing the dual variable $\boldsymbol{\lambda} \in \mathbb{R}_+^L$ for the L inequality constraints $\delta \leq r_l/\beta_l \ l = 1, \ldots, L$, $\boldsymbol{\mu} \in \mathbb{R}_+^K$ for the K inequality constraints $\log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0 \ \ k = 1, \ldots, K$, and $\boldsymbol{\nu} \in \mathbb{R}_+^M$ for the M inequality constraints $\log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \leq 0 \ \ m = 1, \ldots, M$ to obtain:

$$\begin{aligned} \mathcal{L}(\delta, \mathbf{r}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= \delta - \boldsymbol{\lambda}^\top (\delta \mathbf{1} - \operatorname{diag}(\boldsymbol{\beta})^{-1}\mathbf{r}) \\ &- \sum_{k=1}^{K} \mu_k \log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) - \sum_{m=1}^{M} \nu_m \log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big). \end{aligned} \tag{39}$$

Taking the partial derivative of (39) with respect to $\mathbf{r}$, we have $\partial \mathcal{L}/\partial \mathbf{r} = \operatorname{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{\lambda} - \sum_{k=1}^{K} \mu_k \Big(\mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big)\Big) - \sum_{m=1}^{M} \nu_m \Big(\mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)\Big)$, and, setting it to zero, we have at optimality:

$$\begin{aligned} \boldsymbol{\lambda}^\star = \operatorname{diag}(\boldsymbol{\beta}) \Bigg(& \sum_{k=1}^{K} \mu_k^\star \Big(\mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}^\star})\big)\Big) \\ & + \sum_{m=1}^{M} \nu_m^\star \Big(\mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}^\star})\big)\Big)\Bigg) \end{aligned} \tag{40}$$

According to the Perron-Frobenius theorem, the Perron right and left eigenvectors of an irreducible nonnegative matrix are (entry-wise) nonnegative [30]. Since the dual variables are also nonnegative, we thus deduce from (40) that $\boldsymbol{\lambda}^\star$ is a positive vector. Combining the fact that $\boldsymbol{\lambda}^\star > 0$ with complementary slackness, we have $\delta^\star = r_l^\star/\beta_l$ for all l.

E. Proof of Lemma 3

We first form the Lagrangian for (17) by introducing the dual variable $\boldsymbol{\mu} \in \mathbb{R}_+^K$ for the K inequality constraints $\log \rho\big(\operatorname{diag}(e^{\mathbf{r}})\widetilde{\mathbf{B}}_k\big) \leq 0, \quad k = 1, \ldots, K$ and the dual variable $\boldsymbol{\nu} \in \mathbb{R}_+^M$ for the M inequality constraints $\log \rho\big(\operatorname{diag}(e^{\mathbf{r}})\widetilde{\mathbf{D}}_m\big) \leq 0, \quad m = 1, \ldots, M$ to obtain $\mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_{l=1}^{L} w_l r_l - \sum_{k=1}^{K} \mu_k \log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) - \sum_{m=1}^{M} \nu_m \log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$. Taking the first order derivative of the above Lagrangian with respect to $\mathbf{r}$, we have $\partial \mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\nu})/\partial \mathbf{r} = \mathbf{w} - \sum_{k=1}^{K} \mu_k \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big) - \sum_{m=1}^{M} \nu_m \mathbf{x}\big(\widetilde{\mathbf{B}}_m \operatorname{diag}(e^{\mathbf{r}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}})\big)$, and, setting it to zero, we have at optimality:

$$\begin{aligned} \mathbf{w} = & \sum_{k=1}^{K} \mu_k^\star \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\mathbf{r}^\star})\big) \\ & + \sum_{m=1}^{M} \nu_m^\star \mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\mathbf{r}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_m \operatorname{diag}(e^{\mathbf{r}^\star})\big). \end{aligned} \tag{41}$$

Combining (41) with (40), the equation $\boldsymbol{\lambda} = \operatorname{diag}(\boldsymbol{\beta})\mathbf{w}/\mathbf{1}^\top \mathbf{w}$ must hold if (17) and (20) have the same optimal solution. Moreover, by taking the partial derivative of (39) with respect to δ and then setting it to zero, we have $\mathbf{1}^\top \boldsymbol{\lambda} = 1$. Since $\mathbf{w}$ is a probability vector, we have $\sum_{l=1}^{L} \xi_l w_l \geq 1$ in the constraint set of (21). We then write (21) in epigraph form as

$$\begin{aligned} \text{minimize} \quad & \varrho \\ \text{subject to} \quad & \log \rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \leq \log \varrho, \quad k = 1, \ldots, K, \\ & \log \rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \leq \log \varrho, \quad m = 1, \ldots, M, \\ & \sum_{l=1}^{L} w_l \xi_l \geq 1, \\ \text{variables:} \quad & \varrho, \boldsymbol{\xi}. \end{aligned} \tag{42}$$

Next, we form the Lagrangian for (42) by introducing the dual variable $\hat{\boldsymbol{\mu}} \in \mathbb{R}_+^K$ for the K inequality constraints $\log\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \le \log\varrho,\ k=1,\ldots,K$, the dual variable $\hat{\boldsymbol{\nu}} \in \mathbb{R}_+^M$ for the M inequality constraints $\log\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \le \log\varrho,\ m=1,\ldots,M$, and the dual variable λ for the single inequality constraint $\sum_{l=1}^L w_l\xi_l \ge 1$ to obtain $\mathcal{L}(\varrho,\boldsymbol{\xi},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}},\lambda) = \varrho + \sum_{k=1}^K \hat{\mu}_k\Big(\log\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}})\big) - \log\varrho\Big) + \sum_{m=1}^M \hat{\nu}_m\Big(\log\rho\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) - \log\varrho\Big) - \lambda\big(\sum_{l=1}^L w_l\xi_l - 1\big)$. Taking the first order derivative of the above Lagrangian with respective to ϱ and $\boldsymbol{\xi}$ respectively, we have $\partial\mathcal{L}(\varrho,\boldsymbol{\xi},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}},\lambda)/\partial\varrho = 1 - \sum_{k=1}^K \hat{\mu}_k/\varrho - \sum_{m=1}^M \hat{\nu}_m/\varrho$ and $\partial\mathcal{L}(\varrho,\boldsymbol{\xi},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}},\lambda)/\partial\boldsymbol{\xi} = \sum_{k=1}^K \hat{\mu}_k \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) + \sum_{m=1}^M \hat{\nu}_m \mathbf{x}\big(\widetilde{\mathbf{B}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}})\big) - \lambda\mathbf{w}$, and, setting them to zero, we have at optimality:

$$\sum_{k=1}^K \hat{\mu}_k^\star + \sum_{m=1}^M \hat{\nu}_m^\star = \varrho^\star, \tag{43}$$

and

$$\sum_{k=1}^K \hat{\mu}_k^\star \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) + \sum_{m=1}^M \hat{\nu}_m^\star \mathbf{x}\big(\widetilde{\mathbf{B}}_m \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) = \lambda^\star\mathbf{w}, \tag{44}$$

which imply that $\lambda^\star = \varrho^\star$. Combining the fact that $\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(\frac{1}{\varrho}e^{\boldsymbol{\xi}^\star})\big) = \frac{1}{\varrho}\rho\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big)$, $\mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(\frac{1}{\varrho}e^{\boldsymbol{\xi}^\star})\big) = \mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big)$ and $\mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(\frac{1}{\varrho}e^{\boldsymbol{\xi}^\star})\big) = \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big)$ for any positive ϱ with (41), (43) and (44), we can deduce that the optimal solution $\mathbf{r}^\star$ of (17) and the optimal solution $\boldsymbol{\xi}^\star$ of (21) satisfy $e^{\mathbf{r}^\star} = \frac{1}{\varrho^\star}e^{\boldsymbol{\xi}^\star}$. Thus, if (17) and (20) have the same optimal solution, $\boldsymbol{\beta}$ is proportional to $\log(\frac{1}{\varrho^\star}e^{\boldsymbol{\xi}^\star})$.

Next, we discuss a special case and state the following inequality to connect (17) and (20) when $\boldsymbol{\beta} = \mathbf{1}$. Suppose $\boldsymbol{\Omega} \in \{\widetilde{\mathbf{B}}_1,\ldots,\widetilde{\mathbf{B}}_K,\widetilde{\mathbf{D}}_1,\ldots,\widetilde{\mathbf{D}}_M\}$. Then, taking the logarithm on the both sides of the Friedland-Karlin inequality (cf. Theorem 3.1 in [26]), we have $\log\rho(\boldsymbol{\Omega}) + \sum_{l=1}^L \big(\mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega})\big)_l r_l \le \log\rho\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\mathbf{r}})\big)$. Since $\log\rho\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\mathbf{r}})\big) \le 0$ in the constraint set in (17), we deduce that $\sum_{l=1}^L \big(\mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega})\big)_l r_l \le -\log\rho(\boldsymbol{\Omega}) \Rightarrow \big(\min_{l=1,\ldots,L}(\mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega}))_l/w_l\big)\sum_{l=1}^L w_l r_l \le \log\big(1/\rho(\boldsymbol{\Omega})\big) \Rightarrow \sum_{l=1}^L w_l r_l \overset{(h)}{\le} \big(\max_{l=1,\ldots,L} w_l/(\mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega}))_l\big)\log\big(1/\rho(\boldsymbol{\Omega})\big)$. Equality is achieved in (h) if and only if $\mathbf{w} = \mathbf{x}(\boldsymbol{\Omega}) \circ \mathbf{y}(\boldsymbol{\Omega})$ and $\rho(\boldsymbol{\Omega}) = \max\{\rho(\widetilde{\mathbf{B}}_1),\ldots,\rho(\widetilde{\mathbf{B}}_K),\rho(\widetilde{\mathbf{D}}_1),\ldots,\rho(\widetilde{\mathbf{D}}_M)\}$.

F. Proof of Lemma 4

We first prove the convergence of the outer loop by using the Friedland-Karlin inequality (cf. Theorem 3.1 in [26]). Suppose $\boldsymbol{\Omega} \in \{\widetilde{\mathbf{B}}_1,\ldots,\widetilde{\mathbf{B}}_K,\widetilde{\mathbf{D}}_1,\ldots,\widetilde{\mathbf{D}}_M\}$. Taking the logarithm of the both sides of the Friedland-Karlin inequality, we have

$$\log\rho\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}})\big) \le \sum_{l=1}^L \Big(\mathbf{x}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}})\big) \circ \mathbf{y}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}})\big)\Big)_l \xi_l + \log\rho\big(\boldsymbol{\Omega}\big). \tag{45}$$

Combining the complementary slackness with (44), we can deduce from $\lambda^\star > 0$ that $\sum_{l=1}^L w_l\xi_l^\star = 1$ at optimality of (21). By subtracting the value 1 and $\sum_{l=1}^L w_l\xi_l^\star$ respectively from left and right side of (45), we have, at optimality, the inequalities $\log\rho\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) - 1 \le \sum_{l=1}^L \Big(\mathbf{x}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big)\Big)_l \xi_l^\star + \log\rho\big(\boldsymbol{\Omega}\big) - \sum_{l=1}^L w_l\xi_l^\star \Rightarrow \log\rho\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \le \min_{\boldsymbol{\xi}}\Big\{\sum_{l=1}^L \big(\mathbf{x}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\boldsymbol{\Omega}\operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) - \mathbf{w}\big)_l\xi_l\Big\} + \log\rho\big(\boldsymbol{\Omega}\big) + 1$. Since $\log\rho\big(\widetilde{\mathbf{B}}_k\big) + 1$ is a constant, $\boldsymbol{\xi}^\star$ is solved by computing $\min_{\boldsymbol{\xi}}\Big\{\sum_{l=1}^L \big(\mathbf{x}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{B}}_k \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) - \mathbf{w}\big)_l\xi_l\Big\}$ or $\min_{\boldsymbol{\xi}}\Big\{\sum_{l=1}^L \big(\mathbf{x}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) \circ \mathbf{y}\big(\widetilde{\mathbf{D}}_m \operatorname{diag}(e^{\boldsymbol{\xi}^\star})\big) - \mathbf{w}\big)_l\xi_l\Big\}$.

Next, we prove the convergence of the inner loop at Step 4. From Lemma 2, we have $\delta^\star = r_l^\star/\beta_l \ \ l=1,\ldots,L$, i.e.,

$$\frac{1}{\delta^\star}\mathbf{p}^\star = f(\mathbf{p}^\star), \tag{46}$$

where $f:\mathbb{R}^L \to \mathbb{R}^L$ is given by:

$$f_l(\mathbf{p}^\star) = \frac{\beta_l}{\log(1+\mathsf{SINR}_l(\mathbf{p}^\star))}p_l^\star. \tag{47}$$

Now, we show that $\mathbf{p}^\star$ in (46) can be constrained by a monotone norm. Let us rewrite the interference temperature constraints in (11) as follows. By substituting $\mathbf{q} = \mathbf{F}\mathbf{p} + \mathbf{v}$ into $\mathbf{b}_m^\top\mathbf{q} \le \bar{q}_m$, we have $\mathbf{b}_m^\top\mathbf{F}\mathbf{p} \le \bar{q}_m - \mathbf{b}_m^\top\mathbf{v}$. Combining this with $\mathbf{a}_k^\top\mathbf{p} \le \bar{p}_k$, we see that $\mathbf{p}^\star$ must satisfy

$$\max\Big\{\max_{k=1,\ldots,K}\Big\{\frac{\mathbf{a}_k^\top\mathbf{p}^\star}{\bar{p}_k}\Big\}, \max_{m=1,\ldots,M}\Big\{\frac{\mathbf{b}_m^\top\mathbf{F}\mathbf{p}^\star}{\bar{q}_m - \mathbf{b}_m^\top\mathbf{v}}\Big\}\Big\} = 1, \tag{48}$$

which is a monotone norm in $\mathbf{p}^\star$. We now state the nonlinear Perron-Frobenius theory in [24].

Theorem 3 (Krause's theorem [24]): Let $\|\cdot\|$ be a monotone norm on $\mathbb{R}^L$. For a concave mapping $f:\mathbb{R}_+^L \to \mathbb{R}_+^L$ with $f(\mathbf{z}) > 0$ for $\mathbf{z} \ge \mathbf{0}$, the following statements hold. The conditional eigenvalue problem $f(\mathbf{z}) = \lambda\mathbf{z}$, $\lambda \in \mathbb{R}$, $\mathbf{z} \ge \mathbf{0}$, $\|\mathbf{z}\| = 1$ has a unique solution $(\lambda^*,\mathbf{z}^*)$, where $\lambda^* > 0$, $\mathbf{z}^* > \mathbf{0}$. Furthermore, $\lim_{k\to\infty}\tilde{f}(\mathbf{z}(k))$ converges geometrically fast to $\mathbf{z}^*$, where $\tilde{f}(\mathbf{z}) = f(\mathbf{z})/\|f(\mathbf{z})\|$.

Let $f(\mathbf{p})$ be given in (47), $\lambda = 1/\delta$, and we constrain $\mathbf{p}$ by the monotone norm $\|\cdot\|$ given in (48). The proof of $f(\mathbf{p})$ being concave can be found in [21]. By Theorem 3, $\mathbf{p}^\star$ converges to the fixed point $\mathbf{p}^\star = f(\mathbf{p}^\star)/\|f(\mathbf{p}^\star)\|$.

G. Proof of Theorem 2

Theorem 2 can be proved by combining the previous proofs in Section IV. The proof of convergence of $\grave{\mathbf{r}}^{\mathcal{J}}(t)$ to the optimal solution of (14) by the branch-and-bound method can be found in [28]. Since the maps in (6) and (7) are bijective, the limit point in (30) and (31) solves (11).

REFERENCES

[1] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 517–528, 2007.

[2] S. J. Shellhammer, A. K. Sadek, and W. Zhang, "Technical challenges for cognitive radio in the TV white space spectrum," in *Proc. Inf. Theory Appl. Workshop*, 2009.

[3] W. Ren, Q. Zhao, and A. Swami, "Power control in cognitive radio networks: How to cross a multi-lane highway," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1283–1296, 2009.

[4] J. S. Pang, G. Scutari, D. P. Palomar, and F. Facchinei, "Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3251–3271, 2007.

[5] S. Huang, X. Liu, and Z. Ding, "Decentralized cognitive radio control based on inference from primary link control information," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 394–406, 2011.

[6] M. Ebrahimi, M. A. Maddah-Ali, and A. K. Khandani, "Power allocation and asymptotic achievable sum-rates in single-hop wireless networks," in *Proc. IEEE 40th CISS*, 2006.

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

[8] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, 2006.

[9] S. Hayashi and Z.-Q. Luo, "Spectrum management for interference-limited multiuser communication systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1153–1175, 2009.

[10] C. W. Tan, D. P. Palomar, and M. Chiang, "Solving nonconvex power control problems in wireless networks: Low SIR regime and distributed algorithms," in *Proc. IEEE GLOBECOM*, 2005.

[11] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[12] T. Wang and L. Vandendorpe, "On the SCALE algorithm for multiuser multicarrier power spectrum management," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4992–4998, 2012.

[13] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, 2009.

[14] K. Eriksson, S. Shi, N. Vucic, M. Schubert, and E. G. Larsson, "Globally optimal resource allocation for achieving maximum weighted sum rate," in *Proc. IEEE GLOBECOM*, 2010.

[15] C. W. Tan, S. Friedland, and S. H. Low, "Spectrum management in multiuser cognitive wireless networks: Optimality and algorithm," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 421–430, 2011.

[16] C. W. Tan, S. Friedland, and S. H. Low, "Nonnegative matrix inequalities and their application to nonconvex power control optimization," *SIAM J. Matrix Analysis Appl.*, vol. 32, no. 3, pp. 1030–1055, 2011.

[17] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing MSE on multiuser downlink: Optimality, fast algorithms and equivalence via max-min SINR," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6127–6143, 2011.

[18] C. W. Tan, M. Chiang, and R. Srikant, "Fast algorithms and performance bounds for sum rate maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 706–719, 2013.

[19] L. Zheng and C. W. Tan, "Optimization software package for weighted sum rate maximization," 2013. [Online]. Available: https://sites.google.com/site/liangzhengcode/home/sumrate

[20] L. Zheng and C. W. Tan, "Cognitive radio network duality and algorithms for utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 500–513, 2013.

[21] D. W. H. Cai, C. W. Tan, and S. H. Low, "Optimal max-min fairness rate control in wireless networks: Perron-Frobenius characterization and algorithms," in *Proc. IEEE INFOCOM*, 2012.

[22] D. W. H. Cai, T. Quek, and C. W. Tan, "A unified analysis of max-min weighted SINR for MIMO downlink system," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3850–3862, 2011.

[23] C. W. Tan, "Optimal power control in Rayleigh-fading heterogeneous networks," in *Proc. IEEE INFOCOM*, 2011.

[24] U. Krause, "Concave Perron-Frobenius theory and applications," *Nonlinear Analysis*, vol. 47, pp. 1457–1466, 2001.

[25] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, 2004.

[26] S. Friedland and S. Karlin, "Some inequalities for the spectral radius of non-negative matrices and applications," *Duke Mathematical J.*, vol. 42, no. 3, pp. 459–490, 1975.

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

[28] V. Balakrishnan, S. Boyd, and S. Balem, "Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems," *Int. J. Robust Nonlinear Control*, vol. 1, no. 4, pp. 295–317, 1991.

[29] J. E. Falk, "Lagrange multipliers and nonconvex programs," *SIAM J. Control*, vol. 7, no. 4, pp. 534–545, 1969.

[30] E. Seneta, *Non-Negative Matrices and Markov Chains*, 2nd ed. New York: Springer-Verlag, 1981.

Liang Zheng received the bachelor's degree in software engineering from Sichuan University, Chengdu, China, in 2011. She is pursuing her Ph.D. degree at the City University of Hong Kong. Her research interests are in wireless networks, mobile computing, and nonlinear optimization and its applications. She was a Finalist of the Microsoft Research Asia Fellowship in 2013.

Chee Wei Tan (M'08-SM'12) received the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 2006 and 2008, respectively. He is an Assistant Professor at the City University of Hong Kong. Previously, he was a Postdoctoral Scholar at the California Institute of Technology (Caltech), Pasadena, CA. He was a Visiting Faculty member at Qualcomm R&D, San Diego, CA, in 2011. His research interests are in networks, inference in online large data analytics, and optimization theory and its applications. Dr. Tan currently serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was the recipient of the 2008 Princeton University Wu Prize for Excellence and was awarded the 2011 IEEE Communications Society AP Outstanding Young Researcher Award. He was a selected participant at the U.S. National Academy of Engineering China-America Frontiers of Engineering Symposium in 2013.