NONNEGATIVE MATRIX INEQUALITIES AND THEIR APPLICATION TO NONCONVEX POWER CONTROL OPTIMIZATION

CHEE WEI TAN*, SHMUEL FRIEDLAND†, AND STEVEN LOW‡

Abstract. Maximizing the sum rates in a multiuser Gaussian channel by power control is a nonconvex NP-hard problem that finds engineering application in Code Division Multiple Access (CDMA) wireless system. In this paper, we extend and apply several fundamental nonnegative matrix inequalities initiated by Friedland and Karlin in a 1975 paper to solve this nonconvex power control optimization problem. Leveraging nonnegative matrix theory such as the Subinvariance Theorem and the Perron-Frobenius Theorem, we (1) show that this problem in the power domain can be reformulated as an equivalent convex maximization problem over a closed unbounded convex set in the logarithmic signal-to-interference-noise ratio domain, (2) propose two relaxation technqiues that utilize the reformulation problem structure to compute progressively tight bounds, and (3) propose a global optimization algorithm with ϵ-suboptmality to compute the optimal power control allocation.

Keywords and phrases: Nonconvex optimization, convex relaxation, maximization of convex functions, nonnegative matrix theory, spectral radii of irreducible nonnegative matrices, wireless networks.
2000 Mathematics Subject Classification. 15A42, 15A48, 47N10, 49K35, 65K05, 94A40.

1. Introduction. We study the problem of total data throughput maximization using power control in Code Division Multiple Access (CDMA) wireless communication systems, where interference is a major source of performance impairment. Due to the broadcast nature of the wireless medium, the data rates in a wireless network are affected by interference when all the users transmit simultaneously over the same frequency band using CDMA. Power control is used to mitigate the effect of multiuser interference on performance and maximize the total data rates of all users [3]. The CDMA wireless network can be modeled by an information-theoretic interference channel that treats multiuser interference as additive noise [3]. Finding the optimal power allocation that maximizes sum rates over this multiuser Gaussian channel requires solving a nonconvex problem [18, 4, 3, 6, 14]. This nonconvex problem also finds applications in throughput maximization for digital subscriber line (DSL) wireline systems [13, 15, 16].[1]

The complexity of an exhaustive search is prohibitively expensive, since this problem is NP-hard, and may even be hard to approximate [15]. The authors in [4, 18] formulated the problem as a signomial program, and used a successive convex approximation method based on geometric programming. In [6], the solution to a two-user special case was analyzed. The authors in [15] showed the NP-hardness of the problem, and used the Lyapunov theorem in functional analysis to deduce a zero duality gap result between a related primal of a continuous problem formulation in the DSL setting and its dual. The authors in [16] estimated the size of this duality gap for the finite problem in the DSL setting using Lagrangian dual relaxation that is combined with a linear program. The authors in [19] proposed approximation algorithms to solve the problem with individual power constraints (CDMA uplink system). The authors in [20] solved the problem with a single total power constraint (CDMA downlink system) under low to medium interference conditions.

We now state briefly the sum rate maximization problem with individual power constraints. (See §2 for all definitions, notations and motivations.) Let $F = [f_{ij}]_{i=j=1}^{L}$ and

*Department of Computer Science, City University Hong Kong, Kowloon, Hong Kong (cheewtan@cityu.edu.hk).
†Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607-7045, USA, and Berlin Mathematical School, Berlin, Germany, (friedlan@uic.edu).
‡Department of Computer Science, California Institute of Technology, CA 91125, (slow@caltech.edu).

[1]The sum rate maximization problem we study in this paper differs from that in the DSL system in that each user allocates power only in a single frequency, whereas each user in a DSL system allocates their power over more than one frequency.

$\mathbf{v} = (v_1, \ldots, v_L)^\top$ be an $L \times L$ matrix with zero diagonal and positive off diagonal elements and a positive vector respectively. Let $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_L)^\top$ and $\mathbf{w} = (w_1, \ldots, w_L)^\top$ be a given positive vector and a given probability vector respectively. The sum rate maximization problem is given by

$$\max_{\mathbf{0} \leq p_l \leq \bar{p}_l \ \forall \, l} \quad \sum_{l=1}^{L} w_l \log \left(1 + \frac{p_l}{\sum_{j \neq l} f_{lj} p_j + v_l} \right). \tag{1.1}$$

This is a nonconvex optimization problem that has a nonlinear-fractional objective function of positive variables over a simple box constraint set. The exact solution to this problem is also known to be (strongly) NP-hard [15]. An often used technique to tackle nonconvexity is the standard Lagrange dual relaxation of (1.1) in the power domain. However, the shortcoming of this approach is that there can exist a positive duality gap between the global optimal primal and optimal dual value of (1.1) [15]. Also, finding an optimal primal solution given an optimal dual solution, or vice versa, is in general difficult.

We adopt a reformulation-relaxation approach to tackle (1.1). Our reformulation possesses certain desirable properties, which enables the application of nonnegative matrix theory, especially the Friedland-Karlin inequalities stated in [9], to find the global optimal solution and motivate efficient relaxation techniques. In particular, we utilize the problem structure to develop suitable fast computational procedures for solving and computing useful bounds to the sum rate maximization problem. Furthermore, analytical solution to both the sum rate maximization problem and its relaxation problems can also be characterized by the spectra of specially-crafted nonnegative matrices. A by-product of our analysis is a refinement of the Friedland-Karlin inequalities in [9] and its application to an inverse problem in nonnegative matrix theory. From an engineering perspective, our algorithms operate in the logarithmic signal-to-interference-noise ratio domain or, equivalently the dB domain that is lingua franca in existing wireless technology.

Overall, the contributions of the paper are as follows:

1. We study a reformulation of the sum rate maximization problem showing that it is equivalent to a convex maximization problem on a closed unbounded convex set.
2. Exploiting the reformulation problem structure, we propose two relaxation techniques that finds progressively tighter bounds on the global optimal value. The first one is a convex relaxation technique that uses Lagrange duality (and its connection to convex envelope relaxation) and the Friedland-Karlin inequalities (basic inequalities that characterize the spectral radius of a nonnegative matrix) to solve a sequence of linear programs. The second method exploits the spectra of specially-crafted nonnegative matrices in a successive convex approximation method.
3. Utilizing the relaxation techniques, we propose a global optimization algorithm (with ϵ-suboptmality) to solve the sum rate maximization problem.
4. We give new applications of the Friedland-Karlin inequalities to inverse problems in nonnegative matrix theory.
5. Numerical examples illustrate the performance of our techniques, and include a comparison between our relaxation techniques and the standard Lagrange dual relaxation.

This paper is organized as follows. In §2, we state definitions, notations and a short motivation. We give a characterization of the image of the multidimensional box $[\mathbf{0}, \bar{\mathbf{p}}] \subset \mathbb{R}^L_+$ by the map $\boldsymbol{\gamma}$, in terms of the spectral radii of a set of nonnegative matrices. In §3, we study the sum rate maximization problem for power control in wireless network. We give necessary and sufficient conditions on an extremal point $\mathbf{p} \in [\mathbf{0}, \bar{\mathbf{p}}]$ to be a local optimal value. In §4, we exploit the reformulation problem structure to study two relaxation techniques to find useful upper bounds to the global optimal value. In §5, we propose a global optimization algorithm to solve the sum rate maximization problem. In §6, we evaluate the performance of our algorithms. In §7, we conclude our paper. In §A, viewed as an appendix, we restate some useful results for [9] and give several applications and extensions, which are needed in this paper.

2. Notations and preliminary results. Throughout the paper, we use the following notations. Let $\mathbb{R}^{m\times n}\supset\mathbb{R}_+^{m\times n}$ denote the set of $m\times n$ matrices and its subset of nonnegative matrices. For $A,B\in\mathbb{R}^{m\times n}$, we denote $A\le B$ if $B-A\in\mathbb{R}_+^{m\times n}$. We denote $A\lneq B, A<B$ if $B-A$ is a nonzero nonnegative and positive matrix, respectively. We denote the entries of a matrix $A\in\mathbb{R}^{m\times n}$ by the small letters, i.e $A=[a_{ij}]_{i,j=1}^{m,n}$. Identify $\mathbb{R}^m=\mathbb{R}^{m\times 1},\mathbb{R}_+^m=\mathbb{R}_+^{m\times 1}$.

A column vector is denoted by the **bold** letter $\mathbf{x}=(x_1,\ldots,x_L)^\top\in\mathbb{R}^L$. We denote $e^{\mathbf{x}}:=(e^{x_1},\ldots,e^{x_m})^\top$. For $\mathbf{x}>\mathbf{0}$, we let $\mathbf{x}^{-1}:=(\frac{1}{x_1},\ldots,\frac{1}{x_m})^\top$ and $\log\mathbf{x}=(\log x_1,\ldots,\log x_L)^\top$. Let $\mathbf{x}\circ\mathbf{y}$ denote the Schur product of the vectors $\mathbf{x}$ and $\mathbf{y}$, i.e., $\mathbf{x}\circ\mathbf{y}=[x_1y_1,\ldots,x_Ly_L]^T$. Let $\mathbf{1}=(1,\ldots,1)^\top\in\mathbb{R}^L$. For $\underline{\mathbf{p}}\le\bar{\mathbf{p}}\in\mathbb{R}^L$, denote by $[\underline{\mathbf{p}},\bar{\mathbf{p}}]$ the set of all $\mathbf{x}\in\mathbb{R}^L$ satisfying $\underline{\mathbf{p}}\le\mathbf{x}\le\bar{\mathbf{p}}$. For a vector $\mathbf{y}=(y_1,\ldots,y_L)^\top$, denote by $\mathrm{diag}(\mathbf{y})$ the diagonal matrix $\mathrm{diag}(y_1,\ldots,y_L)$. We also let $(B\mathbf{y})_l$ denote the lth element of the vector $B\mathbf{y}$. The Perron-Frobenius eigenvalue of a nonnegative matrix F is denoted as $\rho(F)$, and the Perron (right) and left eigenvector of F associated with $\rho(F)$ are denoted by $\mathbf{x}(F)$ and $\mathbf{y}(F)$ (or simply $\mathbf{x}$ and $\mathbf{y}$ when the context is clear) respectively. Assume that F is an irreducible nonnegative matrix. Then $\rho(F)$ is simple and positive, and $\mathbf{x}(F),\mathbf{y}(F)>\mathbf{0}$ [2]. We will assume the normalization: $\mathbf{x}(F)\circ\mathbf{y}(F)$ is a probability vector. The super-script $(\cdot)^\top$ denotes transpose. For a positive integer n, denote by $\langle n\rangle$ the set $\{1,\ldots,n\}$. Let $P:X\to Y$ be a mapping from the space X to the space Y. For a subset $Z\subset X$, we denote by $P(Z)$ the image of the set Z.

Consider an interference channel with L logical transmitter/receiver pairs. The data transmission in this system with L users can be modeled as a Gaussian interference channel given by the following baseband signal model:

$$\boldsymbol{y}_l=\boldsymbol{h}_{ll}\boldsymbol{x}_l+\sum_{j\neq l}\boldsymbol{h}_{lj}\boldsymbol{x}_j+\boldsymbol{z}_l, \tag{2.1}$$

where $\boldsymbol{y}_l\in\mathbb{C}^{1\times 1}$ is the received signal of the lth user, $\boldsymbol{h}_{lj}\in\mathbb{C}^{1\times 1}$ is the channel coefficient between the transmitter of the jth user and the receiver of the lth user, $\boldsymbol{x}\in\mathbb{C}^{N\times 1}$ is the transmitted (information carrying) signal vector, and $\boldsymbol{z}_l$'s are the i.i.d. additive complex Gaussian noise coefficient with variance $n_l/2$ on each of its real and imaginary components. The first term on the right-hand side of (2.1) represents the desired signal, whereas the second term represents the interference signals from other users. At each transmitter, the signal is constrained by an average power constraint, i.e., $\mathbb{E}[|\boldsymbol{x}_l|^2]=p_l$, which we assume to be upper bounded by $\bar{p}_l$ for all l.

The vector $\mathbf{p}=(p_1,\ldots,p_L)^\top$ is the transmit power vector and is the optimization variable of interest in this paper. Let $G=[g_{lj}]_{l,j=1}^L>0_{L\times L}$ represent the channel gain, where $g_{lj}=|\boldsymbol{h}_{lj}|^2$ 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 receiver. Assuming a linear matched-filter receiver at each user (treating multiuser interference as additive Gaussian noise), the Signal-to-Interference-Noise Ratio (SIR) for the lth receiver is defined as the ratio of the received signal power $g_{ll}p_l$ to the sum of interference signal power and additive Gaussian noise power $\sum_{j\neq l}g_{lj}p_j+n_l$. We denote the SIR of the lth receiver by γ_l, and consider it as a scalar nonnegative function of $\mathbf{p}$ as follows. Let us first define

$$F=[f_{ij}]_{i,j=1}^L,\ \text{where } f_{ij}=\begin{cases}0, & \text{if } i=j\\ \frac{g_{ij}}{g_{ii}}, & \text{if } i\neq j\end{cases} \tag{2.2}$$

and

$$\mathbf{g} = (g_{11}, \ldots, g_{LL})^\top,\ \mathbf{n} = (n_1, \ldots, n_L)^\top, \mathbf{v} = \left(\frac{n_1}{g_{11}}, \frac{n_2}{g_{22}}, \ldots, \frac{n_L}{g_{LL}}\right)^\top. \tag{2.3}$$

For $\mathbf{p} = (p_1, \ldots, p_L)^\top \geq \mathbf{0}$, we define the following transformation: $\mathbf{p} \mapsto \boldsymbol{\gamma}(\mathbf{p})$, where

$$\gamma_l(\mathbf{p}) := \frac{g_{ll} p_l}{\sum_{j \neq l} g_{lj} p_j + n_l},\ l = 1, \ldots, L, \tag{2.4}$$

and we denote the vector $\boldsymbol{\gamma}(\mathbf{p}) = (\gamma_1(\mathbf{p}), \ldots, \gamma_L(\mathbf{p}))^\top = \mathbf{p} \circ (F\mathbf{p} + \mathbf{v})^{-1}$.

We state the following result on (2.4).

LEMMA 2.1 ([8]). *Let $\mathbf{p} \geq \mathbf{0}$ be a nonnegative vector. Assume that $\boldsymbol{\gamma}(\mathbf{p})$ is defined by (2.4). Then $\rho(\mathrm{diag}(\boldsymbol{\gamma}(\mathbf{p}))F) < 1$, where F is defined by (2.2). Hence, for $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{p})$,*

$$\mathbf{p} = P(\boldsymbol{\gamma}) := (I - \mathrm{diag}(\boldsymbol{\gamma})F)^{-1} \mathrm{diag}(\boldsymbol{\gamma})\mathbf{v}. \tag{2.5}$$

Vice versa, if $\boldsymbol{\gamma}$ is in the set

$$\boldsymbol{\Gamma} := \{\boldsymbol{\gamma} \geq \mathbf{0},\ \rho(\mathrm{diag}(\boldsymbol{\gamma})F) < 1\}, \tag{2.6}$$

then the vector $\mathbf{p}$ defined by (2.5) is nonnegative. Furthermore, $\boldsymbol{\gamma}(P(\mathbf{p})) = \boldsymbol{\gamma}$. That is, $\boldsymbol{\gamma} : \mathbb{R}_+^L \to \boldsymbol{\Gamma}$, and $P : \boldsymbol{\Gamma} \to \mathbb{R}_+^L$ are inverse mappings.

Proof. Observe that (2.4) is equivalent to the equality:

$$\mathbf{p} = \mathrm{diag}(\boldsymbol{\gamma})F\mathbf{p} + \mathrm{diag}(\boldsymbol{\gamma})\mathbf{v}. \tag{2.7}$$

First, let us assume that $\mathbf{p}$ is a positive vector, i.e., $\mathbf{p} > \mathbf{0}$. Hence, $\boldsymbol{\gamma}(\mathbf{p}) > \mathbf{0}$. Since all off-diagonal entries of $\mathbf{F}$ are positive it follows that the matrix $\mathrm{diag}(\boldsymbol{\gamma})F$ is irreducible. As $\mathbf{v} > \mathbf{0}$, we deduce that $\max_{l \in [1,n]} \frac{(\mathrm{diag}(\boldsymbol{\gamma})F\mathbf{p})_l}{p_l} < 1$. The min-max characterization of Wielandt of $\rho(\mathrm{diag}(\boldsymbol{\gamma})F)$, [2] and [10, (38), pp.64], implies that $\rho(\mathrm{diag}(\boldsymbol{\gamma})F) < 1$. Hence, $\boldsymbol{\gamma}(\mathbf{p}) \in \boldsymbol{\Gamma}$. Assume that $\mathbf{p} \geq \mathbf{0}$. Note that $p_l > 0 \iff \gamma_l(\mathbf{p}) > 0$. So $\mathbf{p} = \mathbf{0} \iff \gamma(\mathbf{p}) = 0$. Clearly, $\rho(\boldsymbol{\gamma}(\mathbf{0})F) = \rho(0_{L\times L}) = 0 < 1$. Assume that $\mathbf{p} \gneq \mathbf{0}$. Let $\mathcal{A} = \{l :\ p_l > 0\}$. Denote $\boldsymbol{\gamma}(\mathbf{p})(\mathcal{A})$ the vector composed of positive entries of $\boldsymbol{\gamma}(\mathbf{p})$. Let $F(\mathcal{A})$ be the principal submatrix of $\mathbf{F}$ with rows and columns in $\mathcal{A}$. It is straightforward to see that $\rho(\mathrm{diag}(\boldsymbol{\gamma}(\mathbf{p}))F) = \rho(\mathrm{diag}(\boldsymbol{\gamma}(p)(\mathcal{A})F(\mathcal{A}))$. The arguments above imply that

$$\rho(\mathrm{diag}(\boldsymbol{\gamma}(\mathbf{p}))F) = \rho(\mathrm{diag}(\boldsymbol{\gamma}(\mathbf{p})(\mathcal{A})\mathbf{F}(\mathcal{A})) < 1.$$

Assume that $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$. Then

$$(I - \mathrm{diag}(\boldsymbol{\gamma})F)^{-1} = \sum_{k=0}^{\infty} (\mathrm{diag}(\boldsymbol{\gamma})F)^k \geq 0_{L\times L}. \tag{2.8}$$

Hence, $P(\boldsymbol{\gamma}) \geq \mathbf{0}$. The definition of $P(\boldsymbol{\gamma})$ implies that $\boldsymbol{\gamma}(P(\boldsymbol{\gamma})) = \boldsymbol{\gamma}$. □

LEMMA 2.2. *The set $\boldsymbol{\Gamma} \subset \mathbb{R}_+^L$ is monotonic with respect to the order $\geq$. That is, if $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $\boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq \mathbf{0}$ then $\boldsymbol{\beta} \in \boldsymbol{\Gamma}$. Furthermore, the function $P(\boldsymbol{\gamma})$ is monotone on $\boldsymbol{\Gamma}$.*

$$P(\boldsymbol{\gamma}) \geq P(\boldsymbol{\beta}) \ \ \textit{if} \ \ \boldsymbol{\gamma} \in \boldsymbol{\Gamma} \ \ \textit{and} \ \ \boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq \mathbf{0}. \tag{2.9}$$

Equality holds if and only if $\boldsymbol{\gamma} = \boldsymbol{\beta}$.

Proof. Clearly, if $\boldsymbol{\gamma} \geq \boldsymbol{\beta} \geq \mathbf{0}$ then $\mathrm{diag}(\boldsymbol{\gamma})F \geq \mathrm{diag}(\boldsymbol{\beta})F$ which implies $\rho(\mathrm{diag}(\boldsymbol{\gamma})F) \geq \rho(\mathrm{diag}(\boldsymbol{\beta})F)$. Hence, $\boldsymbol{\Gamma}$ is monotonic. Next, we use the Neumann expansion (2.8) to deduce the monotonicity of P. The equality case is straightforward. $\square$

Note that $\boldsymbol{\gamma}(\mathbf{p})$ is not monotonic in $\mathbf{p}$. Indeed, if one increases only the lth coordinate of $\mathbf{p}$, then one increases the lth coordinate of $\boldsymbol{\gamma}(\mathbf{p})$ and decreases all other coordinates of $\boldsymbol{\gamma}(\mathbf{p})$. As usual, let $\mathbf{e}_l = (\delta_{l1}, \ldots, \delta_{lL})^\top$, $l = 1, \ldots, L$ be the standard basis in $\mathbb{R}^L$. In what follows, we need the following result.

Theorem 2.1. *Let $l \in [1, L]$ be an integer and $a > 0$. Denote $[0, a]_l \times \mathbb{R}_+^{L-1}$ the set of all $\mathbf{p} = (p_1, \ldots, p_L)^\top \in \mathbb{R}_+^L$ satisfying $p_l \le a$. Then the image of the set $[0, a]_l \times \mathbb{R}_+^{L-1}$ by the map $\boldsymbol{\gamma}$ (2.4), is given by*

$$\rho(\operatorname{diag}(\boldsymbol{\gamma})(F + (1/a)\mathbf{v}\mathbf{e}_l^\top)) \le 1, \ \mathbf{0} \le \boldsymbol{\gamma}. \tag{2.10}$$

Furthermore, $\mathbf{p} = (p_1, \ldots, p_L) \in \mathbb{R}_+^L$ satisfies the condition $p_l = a$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{p})$ satisfies

$$\rho(\operatorname{diag}(\boldsymbol{\gamma})(F + (1/a)\mathbf{v}\mathbf{e}_l^\top)) = 1. \tag{2.11}$$

Proof. Suppose that $\boldsymbol{\gamma}$ satisfies (2.10). We claim that $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$. Suppose first that $\boldsymbol{\gamma} > \mathbf{0}$. Then $\operatorname{diag}(\boldsymbol{\gamma})(F + t_1\mathbf{v}\mathbf{e}_l^\top) \lneq \operatorname{diag}(\boldsymbol{\gamma})(F + t_2\mathbf{v}\mathbf{e}_l^\top)$ for any $t_1 < t_2$. [10, Lemma 2, §2, Ch. XIII] yields

$$\begin{aligned}\rho(\operatorname{diag}(\boldsymbol{\gamma})F) < \rho(\operatorname{diag}(\boldsymbol{\gamma})(F + t_1\mathbf{v}\mathbf{e}_l^\top)) < \rho(\operatorname{diag}(\boldsymbol{\gamma})(F + t_2\mathbf{v}\mathbf{e}_l^\top)) < \\ \rho(\operatorname{diag}(\boldsymbol{\gamma})(F + (1/a)\mathbf{v}\mathbf{e}_l^\top)) \le 1 \text{ for } 0 < t_1 < t_2 < 1/a.\end{aligned} \tag{2.12}$$

Thus, $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$. Combine the above argument with the arguments of the proof of Lemma 2.1 to deduce that $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ for $\boldsymbol{\gamma} \ge \mathbf{0}$.

We now show that $P(\boldsymbol{\gamma})_l \le a$. The continuity of P implies that it suffices to consider the case $\boldsymbol{\gamma} > \mathbf{0}$. Combine the Perron-Frobenius theorem (see, e.g., [2]) with (2.12) to deduce

$$0 < \det(I - \operatorname{diag}(\boldsymbol{\gamma})(F + t\mathbf{v}\mathbf{e}_l^\top)) \text{ for } t \in [0, a^{-1}). \tag{2.13}$$

We now expand the right-hand side of the above inequality. Let $B = \mathbf{x}\mathbf{y}^\top \in \mathbb{R}^{L \times L}$ be a rank one matrix. Then B has $L-1$ zero eigenvalues and one eigenvalue equal to $\mathbf{y}^\top\mathbf{x}$. Hence, $I - \mathbf{x}\mathbf{y}^\top$ has $L-1$ eigenvalues equal to 1 and one eigenvalue is $(1 - \mathbf{y}^\top\mathbf{x})$. Therefore, $\det(I - \mathbf{x}\mathbf{y}^\top) = 1 - \mathbf{y}^\top\mathbf{x}$. Since $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, $(I - \operatorname{diag}(\boldsymbol{\gamma})F)$ is invertible. Thus, for any $t \in \mathbb{R}$,

$$\begin{aligned}&\det(I - \operatorname{diag}(\boldsymbol{\gamma})(F + t\mathbf{v}\mathbf{e}_l^\top)) = \\ &\det(I - \operatorname{diag}(\boldsymbol{\gamma})F)\det(I - t((I - \operatorname{diag}(\boldsymbol{\gamma})F)^{-1}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{v})\mathbf{e}_l^\top) \\ &\det(I - \operatorname{diag}(\boldsymbol{\gamma})F)(1 - t\mathbf{e}_l^\top(I - \operatorname{diag}(\boldsymbol{\gamma})F)^{-1}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{v}).\end{aligned} \tag{2.14}$$

Combine (2.13) with the above identity to deduce that

$$1 > t\mathbf{e}_l^\top(I - \operatorname{diag}(\boldsymbol{\gamma})F)^{-1}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{v} = tP(\boldsymbol{\gamma})_l \text{ for } t \in [0, a^{-1}). \tag{2.15}$$

Letting $t \nearrow a^{-1}$, we deduce that $P(\boldsymbol{\gamma})_l \le a$. Hence, the set of $\boldsymbol{\gamma}$ defined by (2.10) is a subset of $\boldsymbol{\gamma}([0, a]_l \times \mathbb{R}_+^{L-1})$.

Let $\mathbf{p} \in [0, a]_l \times \mathbb{R}_+^{L-1}$ and denote $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{p})$. We show that $\boldsymbol{\gamma}$ satisfies (2.10). Lemma 2.1 implies that $\rho(\operatorname{diag}(\boldsymbol{\gamma})F) < 1$. Since $\mathbf{p} = P(\boldsymbol{\gamma})$ and $p_l \le a$, we deduce (2.15). Use (2.14) to deduce (2.13). As $\rho(\operatorname{diag}(\boldsymbol{\gamma})F) < 1$, the inequality (2.13) implies that $\rho(\operatorname{diag}(\boldsymbol{\gamma})F + t\mathbf{v}^\top\mathbf{e}_l) < 1$ for $t \in (0, a^{-1})$. Hence, (2.10) holds.

It is left to show the condition (2.11) holds if and only if $P(\boldsymbol{\gamma})_l = a$. Assume that $\mathbf{p} = (p_1, \ldots, p_L)^\top \in \mathbb{R}_+^L$, $p_l = a$ and let $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{p})$. We claim that equality holds in (2.10).

Assume to the contrary that $\rho(\mathrm{diag}(\boldsymbol{\gamma})(F + (1/a)\mathbf{v}\mathbf{e}_l^\top)) < 1$. Then, there exists $\boldsymbol{\beta} > \boldsymbol{\gamma}$ such that $\rho(\mathrm{diag}(\boldsymbol{\beta})(F + (1/a)\mathbf{v}\mathbf{e}_l^\top)) < 1$. Since P is monotonic $P(\boldsymbol{\beta})_l > p_l = a$. On the other hand, since $\boldsymbol{\beta}$ satisfies (2.10), we deduce that $P(\boldsymbol{\beta})_l \le a$. This contradiction yields (2.11). Similarly, if $\boldsymbol{\gamma} \ge \mathbf{0}$ and (2.11), then $P(\boldsymbol{\gamma})_l = a$. □

Corollary 2.2. *Let $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_L)^\top > \mathbf{0}$ be a given positive vector. Then $\boldsymbol{\gamma}([\mathbf{0}, \bar{\mathbf{p}}])$, the image of the set $[\mathbf{0}, \bar{\mathbf{p}}]$ by the map $\boldsymbol{\gamma}$ (2.4), is given by*

$$\rho\left(\mathrm{diag}(\boldsymbol{\gamma})\left(F + (1/\bar{\mathbf{p}}_l)\mathbf{v}\mathbf{e}_l^\top\right)\right) \le 1, \text{ for } l = 1, \ldots, L, \text{ and } \gamma \in \mathbb{R}_+^L. \tag{2.16}$$

In particular, any $\boldsymbol{\gamma} \in \mathbb{R}_+^L$ satisfying the conditions (2.16) satisfies the inequalities

$$\boldsymbol{\gamma} \le \bar{\boldsymbol{\gamma}} = (\bar{\gamma}_1, \ldots, \bar{\gamma}_L)^\top, \text{ where } \bar{\gamma}_l = \frac{\bar{p}_l}{v_l}, \ i = 1, \ldots, L. \tag{2.17}$$

Proof. Theorem 2.1 yields that $\boldsymbol{\gamma}([\mathbf{0}, \bar{\mathbf{p}}])$ is given by (2.16). Using (2.4), we have

$$\gamma_l(\mathbf{p}) = \frac{p_l}{((F\mathbf{p})_l + v_l)} \le \frac{p_l}{v_l} \le \frac{\bar{p}_l}{v_l} \text{ for } \mathbf{p} \in [\mathbf{0}, \bar{\mathbf{p}}].$$

Note that equality holds for $\mathbf{p} = \bar{p}_l \mathbf{e}_l$. □

REMARK 1. *Corollary 2.2 shows that the (nonconvex) set (2.16) is contained in a rectangular set (2.17).*

3. The sum rate maximization problem.

We assume the use of singe-user decoder at each receiver, i.e., treating interference as additive Gaussian noise, and all users have perfect channel state information at the receiver. We also assume that the coherence time of the channel is less than the duration of the whole transmission by any user. This assumption is valid for example when fading occurs sufficiently slowly in the channel, i.e., flat-fading, so that the channel can be considered essentially fixed during transmission. We further assume that all users employ random Gaussian codes for transmission. In practice, Gaussian codes can be replaced by finite-order signal constellations such as the use of quadrature-amplitude modulation (QAM) or other practical (suboptimal) coding schemes. Assuming a fixed bit error rate (BER) at the receiver, the Shannon capacity formula can be used to deduce the achievable data rate (maximum information rate) of the lth user as [5]:

$$\log\left(1 + \frac{\gamma_l(\mathbf{p})}{\Gamma}\right) \quad \text{nats/symbol}, \tag{3.1}$$

where Γ is the SNR gap to capacity, which is always greater than 1. In this paper, we absorb $(1/\Gamma)$ into g_{ll} for all l, and instead write the achievable data rate as $\log(1 + \gamma_l(\mathbf{p}))$.

Let $\mathbf{w} = (w_1, \ldots, w_L)^\top \ge 0$ be a given probability vector, where w_l is a weight assigned to the lth link to reflect priority (a larger weight reflects a higher priority). The problem of maximizing the sum rate can be stated as the following optimization problem:

$$\begin{array}{rl} \text{maximize} & \Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p})) = \displaystyle\sum_{l=1}^{L} w_l \log(1 + \gamma_l(\mathbf{p})) \\ \text{subject to} & \mathbf{0} \le \mathbf{p} \le \bar{\mathbf{p}}, \\ \text{variables:} & \mathbf{p} = (p_1, \ldots, p_L)^\top \in \mathbb{R}_+^L. \end{array} \tag{3.2}$$

Let $\mathbf{p}^\star = (p_1^\star, \ldots, p_L^\star)^\top$ be a global optimal solution to (3.2). We first derive necessary conditions obtained by straightforward differentiation for an optimal solution $\mathbf{p}^\star$ of (3.2).

LEMMA 3.1. *Denote the gradient of $\Phi_{\mathbf{w}}$ by*

$$\nabla\Phi_{\mathbf{w}}(\boldsymbol{\gamma}) = \left(\frac{w_1}{1+\gamma_1}, \ldots, \frac{w_L}{1+\gamma_L}\right)^\top = \mathbf{w} \circ (\mathbf{1}+\boldsymbol{\gamma})^{-1}.$$

Let $\boldsymbol{\gamma}(\mathbf{p})$ be defined as in (2.4). Then, $\mathrm{H}(\mathbf{p}) = [\frac{\partial \gamma_l}{\partial p_j}]_{l=j=1}^L$, the Hessian matrix of $\boldsymbol{\gamma}(\mathbf{p})$, is given by

$$H(\mathbf{p}) = \operatorname{diag}((F\mathbf{p}+\mathbf{v})^{-1})(-\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))F + I).$$

In particular,

$$\nabla_{\mathbf{p}}\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p})) = \mathrm{H}(\mathbf{p})^\top \nabla\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p})).$$

Corollary 3.1. *Divide the set $\langle L\rangle = \{1, \ldots, L\}$ into the following three disjoint sets $\mathrm{S}_{\max}$, S_{in} and S_0:*

$$\mathrm{S}_{\max} = \{l \in \langle L\rangle,\ p_l^\star = \bar{p}_l\},\ \mathrm{S}_{\mathrm{in}} = \{l \in \langle L\rangle,\ p_l^\star \in (0, \bar{p}_l)\},\ \mathrm{S}_0 = \{l \in \langle L\rangle,\ p_l^\star = 0\}.$$

Then, the following conditions hold.

$$\begin{aligned}
&(\mathrm{H}(\mathbf{p}^\star)^\top \nabla\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}^\star)))_l \geq 0 \text{ for } l \in \mathrm{S}_{\max},\\
&(\mathrm{H}(\mathbf{p}^\star)^\top \nabla\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}^\star)))_l = 0 \text{ for } l \in \mathrm{S}_{\mathrm{in}},\\
&(\mathrm{H}(\mathbf{p}^\star)^\top \nabla\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}^\star)))_l \leq 0 \text{ for } l \in \mathrm{S}_0.
\end{aligned} \tag{3.3}$$

Proof. Assume that $p_l^\star = \bar{p}_l$. Then $\frac{\partial}{\partial p_l}\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}))(\mathbf{p}^\star) \geq 0$. Assume that $0 < p_l^\star < \bar{p}_l$. Then $\frac{\partial}{\partial p_l}\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}))(\mathbf{p}^\star) = 0$. Assume that $p_l^\star = 0$. Then $\frac{\partial}{\partial p_l}\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}))(\mathbf{p}^\star) \leq 0$. □

Instead of solving (3.2) directly, we now turn to a reformulation-relaxation approach that solves and provides useful bounds to (3.2) indirectly. We first need the following lemma.

LEMMA 3.2. *Let $\mathbf{w}$ be a probability vector, and assume that $\mathbf{p}^\star = (p_1^\star, \ldots, p_L^\star)^\top$ is an optimal solution to (3.2). Then $p_l^\star = \bar{p}_l$ for some l. Furthermore if $w_j = 0$ then $p_j^\star = 0$.*

Proof. Assume to the contrary that $\mathbf{p}^\star < \bar{\mathbf{p}}$. Let $\boldsymbol{\gamma}^\star = \boldsymbol{\gamma}(\mathbf{p}^\star)$. Since P is continuous on $\boldsymbol{\Gamma}$, there exists $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ such that $\boldsymbol{\gamma} > \boldsymbol{\gamma}^\star$ such that $P(\boldsymbol{\gamma}) < \bar{\mathbf{p}}$. Clearly, $\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}^\star)) < \Phi_{\mathbf{w}}(\boldsymbol{\gamma})$. As $\boldsymbol{\gamma} = \boldsymbol{\gamma}(P(\boldsymbol{\gamma}))$, we deduce that $\mathbf{p}^\star$ is not an optimal solution to (3.2), contrary to our assumptions.

Suppose that $w_j = 0$. For $\mathbf{p} = (p_1, \ldots, p_L)^\top$, let $\mathbf{p}_j$ be obtained from $\mathbf{p}$ by replacing the jth coordinate in $\mathbf{p}$ by 0. Assume that $p_j > 0$. Then $\gamma_l(\mathbf{p}) < \gamma_l(\mathbf{p}_j)$ for $l \neq j$. Since $w_j = 0$, it follows that $\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p})) < \Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}_j))$. □

We combine the above lemma with Theorem 2.1 and Corollary 2.2 to deduce an alternative formulation of (3.2).

Theorem 3.2. *Problem (3.2) is equivalent to the following optimization problem:*

$$\begin{array}{ll}
\textit{maximize} & \Phi_{\mathbf{w}}(\boldsymbol{\gamma})\\
\textit{subject to} & \rho(\operatorname{diag}(\boldsymbol{\gamma})(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top)) \leq 1 \quad \forall\, l,\\
\textit{variables:} & \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)^\top \in \mathbb{R}_+^L.
\end{array} \tag{3.4}$$

where $\boldsymbol{\gamma}^\star$ is an optimal solution of the above problem if and only if $P(\boldsymbol{\gamma}^\star)$ is an optimal solution $\mathbf{p}^\star$ of the problem (3.2). In particular, any optimal solution $\boldsymbol{\gamma}^\star$ satisfies the equality (2.16) for some integer $l \in [1, L]$.

REMARK 2. *Note that (3.4) is a nonconvex problem having a strictly concave objective function and a set of nonconvex spectral radius constraints.*

We now show that the optimization problem (3.4) can be restated as an optimization problem with a convex objective function on a closed unbounded convex domain. For $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)^\top > 0$, we define the logarithmic mapping:

$$\tilde{\boldsymbol{\gamma}} = \log \boldsymbol{\gamma}, \tag{3.5}$$

i.e., $\boldsymbol{\gamma} = e^{\tilde{\boldsymbol{\gamma}}}$. Recall that for an irreducible nonnegative matrix $B \in \mathbb{R}_+^{L\times L}$ $\log \rho(e^{\mathbf{x}} B)$ is a convex function [12]. Furthermore, $\log(1 + e^t)$ is a strictly convex function in $t \in \mathbb{R}$. Hence, the optimization problem in (3.4) is equivalent to the problem:

$$\begin{array}{ll} \text{maximize} & \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}}) \\ \text{subject to} & \log \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}})(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top)) \le 0 \quad \forall\, l, \\ \text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L. \end{array} \tag{3.6}$$

The unboundedness of the convex set in (3.6) is due to the identity $0 = e^{-\infty}$. In view of Lemma 3.2, it suffices to consider the optimization problem (3.2) in the case where $\mathbf{w} > \mathbf{0}$. Using the reformulation in (3.6), we deduce the following result that any solution satisfying (3.3) in (3.1) is also local optimal to (3.2).

Theorem 3.3. *Consider the optimization problem (3.2). Then any point $\mathbf{0} \le \mathbf{p}^\star \le \bar{\mathbf{p}}$ satisfying the conditions (3.3) is a local optimal solution.*

Proof. Since $\mathbf{w} > \mathbf{0}$, $\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}})$ is a strictly convex function in $\tilde{\boldsymbol{\gamma}} \in \mathbb{R}^L$. Hence, the optimal value of (3.6) is achieved exactly on the extreme points of the closed unbounded set specified in (3.6). (It may happen that some coordinates of the extreme point are $-\infty$.) Translating this observation to the optimization problem (3.2), we deduce the theorem. □

Since the reformulation in (3.6) is a convex maximization problem over a closed unbounded convex set, we choose not to rehash the standard global optimization methods for solving a standard convex maximization problem (cf. [21, 11]). Rather, we choose to exploit the problem structure of (3.6) to first compute good bounds to (3.6) (cf. Section 4) and then to propose a global optimization algorithm (with ϵ-suboptimality) to solve (3.6) (cf. Section 5). The global optimization algorithm is motivated by the relaxation techniques and the problem structure (log-convexity of the spectral radius and separability in the objective function), and differs from the standard global optimization technique found in the literature, e.g., [21, 11].

We now give simple lower and upper bounds on the value of (3.2).

LEMMA 3.3. *Consider the optimization problem (3.2). Let $B_l = (F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top))$ for $l = 1, \ldots, L$. Denote $R = \max_{l\in\langle L\rangle} \rho(B_l)$. Let $\bar{\boldsymbol{\gamma}}$ be defined by (2.17). Then*

$$\Phi_{\mathbf{w}}((1/R)\mathbf{1}) \le \max_{\mathbf{p}\in[\mathbf{0},\bar{\mathbf{p}}]} \Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p})) \le \Phi_{\mathbf{w}}(\bar{\boldsymbol{\gamma}}).$$

Proof. By Corollary 2.2, $\boldsymbol{\gamma}(\mathbf{p}) \le \bar{\boldsymbol{\gamma}}$ for $\mathbf{p} \in [\mathbf{0}, \bar{\mathbf{p}}]$. Hence, the upper bounds holds. Clearly, for $\boldsymbol{\gamma} = (1/R)\mathbf{1}$, we have that $\rho(\mathrm{diag}(\boldsymbol{\gamma})B_l) \le 1$ for $l \in \langle L\rangle$. Then, from Theorem 3.2, $\Phi_{\mathbf{w}}((1/R)\mathbf{1})$ yields the lower bound. Equality is achieved in the lower bound when $\mathbf{p}^\star = t\mathbf{x}(B_i)$, where $i = \arg\max_{l\in\langle L\rangle} \rho(B_l)$, for some $t > 0$. □

The upper bound in Lemma 3.3 is trivial and can be too loose to be useful (as it disregards the interference power and the number of interferers at each receiver). The lower bound is obtained when all the users have a common SIR value. Necessary conditions under which this lower bound is tight is given later (cf. Corollary 5.2). We will examine how to exploit the problem structure of (3.6) to obtain progressively tighter bounds in Section 4.

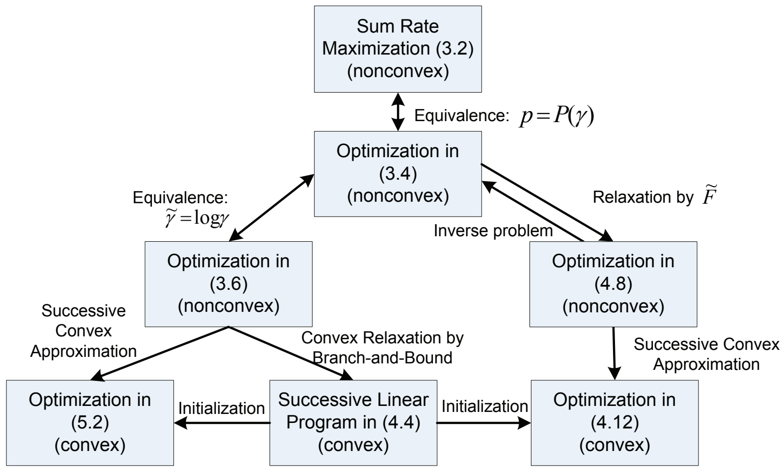


Fig. 4.1. *Overview of the two relaxation techniques and a global optimization technique used on the sum rate maximization problem: The relaxation techniques are 1) A convex relaxation with branch-and-bound method, and 2) a relaxation by three different versions of a matrix $\tilde{F}$ with successive convex approximation method and its connection to inverse problem given in the appendix (see Theorems 4.2 and 4.3 in the appendix). The global optimization technique relies on the first relaxation technique to find a good initial point (within ϵ-suboptimality) that is then combined with the successive convex approximation method to solve the sum rate maximization problem. The key optimization problems (whether convex or nonconvex) and their relationships are also highlighted.*

4. Relaxations and convex approximations. In this section, we use two different approaches that exploits the problem structure of (3.4) or equivalently (3.6) to construct several relaxed versions of (3.4), which can compute useful upper bounds to (3.4). The first relaxation approach leverages the Friedland-Karlin inequalities, the separability of the objective function and its convex envelope over box constraints (a sum of readily computed functions) to construct a linear program whose optimal value upper bounds that of (3.4). Progressively tigher bounds are then obtained by successive partitioning of the box constraints. The second relaxation approach replaces the L spectral constraints in (3.4) by a single one, which has three different versions depending on the choice of a nonnegative matrix. Each relaxed version is still nonconvex, but necessary conditions under which the relaxations are tight are given. Further, simpler algorithms derived from this approach are shown numerically to solve optimally these relaxation, thus providing useful upper bounds to (3.2). Figure 4.1 gives an overview of the development of these two relaxation approaches as well as a global optimization approach based on these relaxation techniques (see Section 5 later).

4.1. Convex Relaxation. We replace the (convex) spectral radius constraint set in (3.6) with a larger set by exploiting the Friedland-Karlin inequalities. We thus consider the following optimization problem:

$$
\begin{array}{ll}
\text{maximize} & \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}}) \\
\text{subject to} & \displaystyle\sum_j (\mathbf{x}(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \circ \mathbf{y}(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top))_j \tilde{\gamma}_j \leq -\log \rho(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \quad \forall\, l, \\
& -K \leq \tilde{\gamma}_l \leq \log \bar{\gamma}_l, \quad \forall\, l, \\
\text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L.
\end{array}
\tag{4.1}
$$

Note that (4.1) is a convex maximization problem with a polyhedron constraint set (still a nonconvex and NP-hard problem). Observe that the constraint set of (4.1) consists of a polyhedron and a box constraint set. In particular, this constraint structure allows us to compute useful upper bounds to (4.1) by exploiting several results in [7] that connects the relationship between relaxation via convexification and the Lagrange dual relaxation. More precisely, the optimal Lagrange dual of (4.1) (which upper bounds (4.1)) can be computed by considering the convex envelope of (4.1).

To compute the convex envelope of a separable function over a box constraint set, it is sufficient to compute the convex envelope of the individual summand of the function over their respective domains (cf. Theorem 2.3 in [7]). For any l, the convex envelope of $w_l \log(1 + e^{\tilde{\gamma}_l})$ over a constraint set $\grave{\tilde{\gamma}}_l \leq \tilde{\gamma}_l \leq \acute{\tilde{\gamma}}_l$ is a linear function in $\tilde{\gamma}_l$ given by

$$
w_l \left(\frac{\log(1 + e^{\acute{\tilde{\gamma}}_l}) - \log(1 + e^{\grave{\tilde{\gamma}}_l})}{\acute{\tilde{\gamma}}_l - \grave{\tilde{\gamma}}_l} (\tilde{\gamma}_l - \grave{\tilde{\gamma}}_l) + \log(1 + e^{\grave{\tilde{\gamma}}_l}) \right). \tag{4.2}
$$

Now, using (4.2) and let $\grave{\tilde{\gamma}}_l = -K, \acute{\tilde{\gamma}}_l = \log \bar{\boldsymbol{\gamma}}_l$ for all l, we replace the objective function of (4.1) by its convex envelope over the box constraint set $\{-K \leq \tilde{\gamma}_l \leq \log \bar{\gamma}_l, \ \forall\, l\}$ to obtain the following linear program:[2]

$$
\begin{array}{ll}
\text{maximize} & \Phi^c_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}}) := \displaystyle\sum_l w_l \left(\frac{\log(1 + \bar{\gamma}_l) - \log(1 + e^{-K})}{\log \bar{\gamma}_l + K} (\tilde{\gamma}_l + K) + \log(1 + e^{-K}) \right) \\
\text{subject to} & \displaystyle\sum_j (\mathbf{x}(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \circ \mathbf{y}(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top))_j \tilde{\gamma}_j \leq -\log \rho(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \quad \forall\, l, \\
& -K \leq \tilde{\gamma}_l \leq \log \bar{\gamma}_l, \quad \forall\, l, \\
\text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L.
\end{array}
\tag{4.3}
$$

Interestingly, the optimal Lagrange dual of (4.3) (equivalent to the optimal value of (4.3) and also obtained by a dual linear program) is equal to the optimal Lagrange dual of (4.1) [7]. In other words, (4.3) is the *"dual of the dual"* of (4.1). For all K (even for $K \to \infty$), the optimal Lagrange dual of (4.3) upper bounds the optimal Lagrange dual of (3.6). Thus, the optimal value of (4.3) gives an upper bound to (3.6).

Although the upper bound obtained by solving (4.3) may be loose, tighter bounds to (3.6) can be obtained iteratively by combining this convex relaxation approach with a branch-and-bound method in [1, 7, 11] that subdivides the set $\{-K \leq \tilde{\gamma}_l \leq \log \bar{\gamma}_l, \ \forall\, l\}$ into successively smaller subsets (the rectangular method, see, e.g., Chapter 7 in [11]). The search for the global optimal solution is performed over the subdivided sets organized in a binary tree data structure. More precisely, using the branch-and-bound method, (4.3) is solved in the first iteration (at the root of the binary tree). In subsequent iterations (lower levels of the binary tree), the set of lower and upper bounds on $\tilde{\boldsymbol{\gamma}}$ in (4.3) is replaced by a subdivided set, and the objective function of (4.3) is then replaced with a reweighted function

[2]Due to separability, the convex envelope of $\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}})$, denoted by $\Phi^c_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}})$, is the sum of the convex envelope of its constituents.

using (4.2), i.e., the convex envelope of $\Phi_{\mathbf{w}}(e^{\tilde{\gamma}})$ over the subdivided set. In particular, at the kth iteration of the branch-and-bound algorithm, we consider a subdivided box $\{\tilde{\gamma}_l \mid \grave{\tilde{\gamma}}_l(k) \le \tilde{\gamma}_l \le \acute{\tilde{\gamma}}_l(k),\ \ \forall\, l\}$ and solve

$$\begin{array}{ll}
\text{maximize} & \displaystyle\sum_l w_l \left(\frac{\log(1+e^{\acute{\tilde{\gamma}}_l(k)}) - \log(1+e^{\grave{\tilde{\gamma}}_l(k)})}{\acute{\tilde{\gamma}}_l(k) - \grave{\tilde{\gamma}}_l(k)} (\tilde{\gamma}_l - \grave{\tilde{\gamma}}_l(k)) + \log(1+e^{\grave{\tilde{\gamma}}_l(k)}) \right) \\
\text{subject to} & \displaystyle\sum_j (\mathbf{x}(F+(1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \circ \mathbf{y}(F+(1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top))_j \tilde{\gamma}_j \le -\log \rho(F+(1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \ \ \forall\, l, \\
& \grave{\tilde{\gamma}}_l(k) \le \tilde{\gamma}_l \le \acute{\tilde{\gamma}}_l(k),\ \ \forall\, l, \\
\text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L,
\end{array} \tag{4.4}$$

at one of the two nodes at the kth iteration of the branch-and-bound algorithm. We denote $\tilde{\boldsymbol{\gamma}}^{\mathrm{LP}_k}$ as the optimal solution to (4.4). A feasible power vector can be obtained from $\mathbf{p}^{\mathrm{LP}_k} = \min\{P(e^{\tilde{\boldsymbol{\gamma}}^{\mathrm{LP}_k}}), \bar{p}\}$, and a lower bound to (3.2) is then given by $\Phi_{\mathbf{w}}(\boldsymbol{\gamma}(\mathbf{p}^{\mathrm{LP}_k}))$. Rules to subdivide a rectangular box constraint and branch into a selected subdivided set as well as the convergence of this branch-and-bound method can be found in [1].

In this method, at the kth iteration of the branch-and-bound algorithm, taking the maximum over all the lower bound at each child node across all the levels in the binary tree (denote this maximum value as L_{bb}^k and its corresponding solution as $\tilde{\boldsymbol{\gamma}}^{\mathrm{BB}_k}$) gives a global lower bound on the optimal value of (3.2). Likewise, taking the maximum over all the upper bound at each child node across all the levels in the binary tree (denote this maximum value as U_{bb}^k) gives a global upper bound on the optimal value of (3.2). The difference between these two bounds is nonincreasing with k. Suppose $U_{bb}^k - L_{bb}^k \le \epsilon$ for some positive ϵ, then we have $\Phi_{\mathbf{w}}(\boldsymbol{\gamma}^\star) \le \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}^{\mathrm{BB}_k}}) + \epsilon$. From a geometrical perspective, this relaxation method systematically narrows down the SIR region that contains the global optimal solution of (3.6), i.e., locate a ϵ-suboptimal neighborhood of $\tilde{\boldsymbol{\gamma}}^\star$.

4.2. Relaxation by Interference Matrices. In this section, we study the second relaxation method that uses specially constructed nonnegative matrices to find useful upper bounds to (3.2). Conditions under which the relaxations are tight are stated, and simpler (lower complexity and faster) algorithms are proposed to compute the upper bounds.

Now, we consider a general matrix $\tilde{F}$ that is used to denote one of the following three matrices:

$$\begin{array}{ll}
1) & F + \operatorname{diag}(\bar{\boldsymbol{\gamma}})^{-1}, \\
2) & F + \left(1/\mathbf{1}^\top \bar{\mathbf{p}}\right) \mathbf{v}\mathbf{1}^\top, \\
3) & F + (1/\bar{p}_i)\, \mathbf{v}\mathbf{e}_i^\top, \ \ i = \arg\max\limits_l \rho(F + (1/\bar{p}_l)\, \mathbf{v}\mathbf{e}_l^\top).
\end{array} \tag{4.5}$$

Observe that the entries of $\tilde{F}$ are functions of all the problem parameters of (3.2). In the following, we consider a relaxation problem to (3.4) that has only a single spectral radius constraint in $\tilde{F}$. We then utilize the spectra of $\tilde{F}$ to find useful upper bounds to (3.4), which in turns upper bounds (3.2).

Lemma 4.1. *Let $\mathbf{0} \le \mathbf{p} \le \bar{\mathbf{p}}$. Assume that $\boldsymbol{\gamma}(\mathbf{p})$ is given by (2.4) and $\tilde{F}$ is given by any of the three matrices in (4.5). Then*

$$\mathbf{p} \ge \operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}\mathbf{p}, \tag{4.6}$$

and

$$\rho(\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}) \le 1. \tag{4.7}$$

Proof. The assumption that $0 \le p_l \le \bar{p}_l$ implies that $\frac{n_l}{g_{ll}} \ge \tilde{f}_{ll} p_l$. From the definition of $\boldsymbol{\gamma}(\mathbf{p})$, we deduce that $p_l = \gamma_l(\mathbf{p}) \left(\sum_{j \neq l} f_{lj} p_j + v_l \right)$, which together with the definition of $\tilde{F}$ and the above observation implies (4.6). The inequality (4.7) is a consequence of Wielandt's characterization of the spectral radius of an irreducible nonnegative matrix [10]. Indeed, if $\mathbf{p} > \mathbf{0}$, i.e. all the coordinates of $\mathbf{p}$ are positive, then $\boldsymbol{\gamma}(\mathbf{p}) > 0$. Hence, $\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}$ is a positive matrix. Then, using Wielandt's characterization, we have

$$\rho(\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}) \le \max_{l=1,\ldots,L} \frac{(\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}\mathbf{p})_l}{p_l} \le 1.$$

Observe next that if $p_l = 0$, then $(\boldsymbol{\gamma}(\mathbf{p}))_l = 0$. So if some of $p_l = 0$, then $\rho(\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F})$ is the spectral radius of the maximal positive submatrix of $\operatorname{diag}(\boldsymbol{\gamma}(\mathbf{p}))\tilde{F}$. By applying Wielandt's characterization to this positive submatrix, we deduce (4.7). $\square$

Lemma 4.1 shows that any feasible $\mathbf{p}$ satisfies (4.7). This leads to the following relaxation of (3.4) that has only a single constraint involving $\tilde{F}$ in (4.5).

LEMMA 4.2. *The optimal value of*

$$\begin{array}{ll} \textit{maximize} & \Phi_{\mathbf{w}}(\boldsymbol{\gamma}) \\ \textit{subject to} & \rho(\operatorname{diag}(\boldsymbol{\gamma})\tilde{F}) \le 1, \\ & \boldsymbol{\gamma} \le \bar{\boldsymbol{\gamma}}, \\ \textit{variables:} & \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)^\top \in \mathbb{R}^L_+ . \end{array} \tag{4.8}$$

is not less than the optimal value of (3.2). Further, using $P(\boldsymbol{\gamma})$, the optimal solution of (4.8) expressed in the power domain is given by $\mathbf{x}(\operatorname{diag}(\boldsymbol{\gamma}')\tilde{F})$, where $\boldsymbol{\gamma}'$ solves (4.8). In particular, $P(\boldsymbol{\gamma}) = \mathbf{p}^\star$ if $P(\boldsymbol{\gamma}')$ satisfies (3.3) in Corollary 3.1.

Proof. In view of (4.7), we see that the optimal value in (4.8) is achieved on a bigger set than the optimal value in (3.2). In view of (4.6), the optimal solution to (4.8) satisfies $\mathbf{p} = \operatorname{diag}(\boldsymbol{\gamma}')\tilde{F}\mathbf{p}$. Together with (4.7), this implies that $\mathbf{p} = \mathbf{x}(\operatorname{diag}(\boldsymbol{\gamma}')\tilde{F})$. $\square$

REMARK 3. *Note that the constraint $\boldsymbol{\gamma} \le \bar{\boldsymbol{\gamma}}$ has been included explicitly in (4.8), because the spectral radius constraint $\rho(\operatorname{diag}(\boldsymbol{\gamma})\tilde{F}) \le 1$ does not imply $\boldsymbol{\gamma} \le \bar{\boldsymbol{\gamma}}$ (cf. Corollary 2.2 and (3.4)).*

The second relaxation method is to solve (4.8) by considering all the three choices of $\tilde{F}$ in (4.5), and find the tightest relaxation to (3.2) among the three choices of $\tilde{F}$. Note that, in (4.5), the first two nonnegative matrices have positive diagonals, whereas the third nonnegative matrix has only a single positive diagonal element. This fact will be important in characterizing the optimal solution to the relaxation problems based on the inverse problem given in the appendix (see Theorems 4.2 and 4.3 later). From a computational viewpoint, solving (4.8) is also useful when L is large (as the computational time to solve (3.4) increases with L).

Corollary 4.1. *We have $\rho(\operatorname{diag}(\boldsymbol{\gamma}')\tilde{F}) = 1$ in (4.8), where $\boldsymbol{\gamma}'$ solves (4.8) optimally.*

Proof. Corollary 4.1 is easily proved by noting that both the objective function and the spectral radius function in (4.8) increase with $\boldsymbol{\gamma}$. $\square$

$\tilde{F}$	Necessary condition for $P(\boldsymbol{\gamma}') = \mathbf{p}^\star$
$F + \mathrm{diag}(\bar{\boldsymbol{\gamma}})^{-1}$	$p_l^\star = \{0, \bar{p}\}, \ \forall\, l$
$F + \left(1/\mathbf{1}^\top \bar{\mathbf{p}}\right) \mathbf{v}\mathbf{1}^\top$	$\mathbf{p}^\star = \bar{\mathbf{p}}$
$F + (1/\bar{p}_i)\, \mathbf{v}\mathbf{e}_i^\top$, $i = \arg\max_l \rho(F + (1/\bar{p}_l)\, \mathbf{v}\mathbf{e}_l^\top)$	$p_i^\star = \bar{p}_i$, $\mathbf{w} = \mathbf{x} \circ \mathbf{y}$ (cf. Corollary 5.2)

TABLE 4.1

A comparison of the different versions of $\tilde{F}$ in the second relaxation method, wherein the optimal solution in (4.8) is given by $\boldsymbol{\gamma}'$. Necessary conditions under which the relaxed problem (4.9) solves (3.6), equivalently (3.2), are given.

Using the logarithmic mapping in (3.5), solving (4.8) is thus equivalent to solving

$$\begin{array}{rl} \text{maximize} & \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}}) \\ \text{subject to} & \log \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}})\tilde{F}) \leq 0, \\ & \tilde{\boldsymbol{\gamma}} \leq \log \bar{\boldsymbol{\gamma}}, \\ \text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L. \end{array} \tag{4.9}$$

Still, (4.9) or equivalently (4.8) is nonconvex and hard to solve. In the following, we give conditions that relate the optimal power $\mathbf{p}^\star$ and the solution of (4.9) for different $\tilde{F}$. These conditions are also necessary when the solution of (4.9) solves (3.6), i.e., (4.9) is a tight relaxation of (3.6).

Corollary 4.1 implies that if the optimizer of (4.8) $\boldsymbol{\gamma}'$ satisfies $P(\boldsymbol{\gamma}') \preceq \bar{\mathbf{p}}$, then $P(\boldsymbol{\gamma}')$ is also the global optimizer of (3.2). Hence, $P(\boldsymbol{\gamma}') \preceq \bar{\mathbf{p}}$ is a necessary and sufficient condition for the relaxation to be tight. Weaker necessary conditions can however be obtained by checking that the following holds:

$$\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'})\tilde{F}\mathbf{p}^\star = \mathbf{p}^\star \tag{4.10}$$

for some $\mathbf{p}^\star$. A summary of the necessary conditions on $\mathbf{p}^\star$ satisfying (4.10) for the three different versions of $\tilde{F}$ is given in Table 4.1, whereby the relaxed problem (4.9) solves (3.6).

Now, we consider using a successive convex approximation method to solve (4.9) directly. This method is motivated by the inverse problem given in the appendix. This is given in the following algorithm to solve (4.9) and also yield a feasible solution to (3.2).

ALGORITHM 1 (Iteratively Reweighted Relaxation Algorithm).

1. *Compute the weight* $\mathbf{m}(k+1)$*:*

$$\mathbf{m}(k+1) = \frac{\nabla \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}(k)}) \circ e^{\tilde{\boldsymbol{\gamma}}(k)}}{\mathbf{1}^\top \left(\nabla \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}(k)}) \circ e^{\tilde{\boldsymbol{\gamma}}(k)}\right)}, \tag{4.11}$$

2. *Obtain* $\tilde{\boldsymbol{\gamma}}(k+1)$ *as the optimal solution to:*

$$\begin{array}{rl} \textit{maximize} & \sum_l m_l(k+1)\tilde{\gamma} \\ \textit{subject to} & \log \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}})\tilde{F}) \leq 0, \\ & \tilde{\gamma}_l \leq \log \bar{\gamma}_l, \ \ \forall\, l, \\ \textit{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L. \end{array} \tag{4.12}$$

3. Compute the power:

$$\mathbf{p}(k+1) = \min\left\{P(e^{\tilde{\gamma}(k+1)}),\ \bar{p}\right\}. \tag{4.13}$$

THEOREM 4.3. *For any $\tilde{\gamma}(0)$ in a sufficiently close neighborhood of $\tilde{\gamma}'$, $\tilde{\gamma}(k)$ in Algorithm 1 converges to the optimal solution of (4.8).*

Proof. We use the fact that in a sufficiently close neighborhood of $\tilde{\boldsymbol{\gamma}}'$, the domain set is convex, and the objective function $\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}})$ is twice continuously differentiable. We then use a successive convex approximation technique to compute $\tilde{\boldsymbol{\gamma}}'$ assuming that the initial point is sufficiently close to $\tilde{\boldsymbol{\gamma}}'$. The convergence conditions for such a technique are given in [17, 4]. Instead of solving (4.8) directly, we replace the objective function of (4.8) in a neighborhood of a feasible point $\tilde{\boldsymbol{\gamma}}(0)$ by its Taylor series (up to the first order terms): $\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}}) \circ e^{\tilde{\gamma}} \approx \Phi_{\mathbf{w}}(e^{\tilde{\gamma}(0)}) + (\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}(0)}) \circ e^{\tilde{\gamma}(0)})^{\top}(\tilde{\gamma} - \tilde{\gamma}(0))$. Assume a feasible $\tilde{\gamma}(0)$ that is close to $\tilde{\boldsymbol{\gamma}}'$. We then compute a feasible $\tilde{\boldsymbol{\gamma}}(k+1)$ by solving the $(k+1)$th approximation problem:

$$\begin{array}{ll} \text{maximize} & \left(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}(k)}) \circ e^{\tilde{\gamma}(k)}/\mathbf{1}^{\top}(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}(k)}) \circ e^{\tilde{\gamma}(k)})\right)^{\top}(\tilde{\gamma} - \tilde{\gamma}(k)) \\ \text{subject to} & \log\rho(\mathrm{diag}(e^{\tilde{\gamma}})\tilde{F}) \leq 0, \\ & \tilde{\gamma}_l \leq \log\bar{\gamma}_l, \ \ \forall\, l, \\ \text{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_n)^{\top} \in \mathbb{R}^L, \end{array} \tag{4.14}$$

where $\tilde{\boldsymbol{\gamma}}(k)$ is the optimal solution of the kth approximation problem. This inner approximation technique converges to a local optimal solution [17, 4]. In addition, if $\tilde{\gamma}(0)$ is sufficiently close to $\tilde{\boldsymbol{\gamma}}'$, then $\lim_{k\to\infty}\tilde{\gamma}(k) = \tilde{\gamma}'$.

Next, we leverage Corollary A.6 in the appendix to solve (4.14). At the global optimality of (4.8), we have the necessary condition of (4.8): $\mathbf{x}(\mathrm{diag}(e^{\tilde{\gamma}'})\tilde{F}) \circ \mathbf{y}(\mathrm{diag}(e^{\tilde{\gamma}'})\tilde{F}) = \nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'}/\mathbf{1}^{\top}(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'})$. □

Interestingly, Algorithm 1 can be viewed as an iteratively reweighted method that produces better estimates as the optimization progresses. The weights in (4.12) are determined by the previous solution. Now, suppose the positive weight vector $\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'}/\mathbf{1}^{\top}(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'})$ is used as the weight input to (4.12). Then, (4.12) outputs $\tilde{\boldsymbol{\gamma}}'$. Intuitively speaking, if the weight vector is approximately proportional to $\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'}/\mathbf{1}^{\top}(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\gamma}'}) \circ e^{\tilde{\gamma}'})$, then Algorithm 1 should converge to a unique $\tilde{\boldsymbol{\gamma}}'$. Based on the inverse problem given in the appendix (see Theorems 4.2 and 4.3), we quantify this in the following for the different choices of $\tilde{F}$ (and the diagonals of $\tilde{F}$ matter in a unique $\tilde{\gamma}'$).

Theorem 4.2. *Suppose that $\tilde{F}$ is given by 1) $F + \mathrm{diag}(\bar{\gamma})^{-1}$ or 2) $F + \left(1/\mathbf{1}^{\top}\bar{\mathbf{p}}\right)\mathbf{v}\mathbf{1}^{\top}$. Let $\mathbf{m} = (m_1, \ldots, m_L)^{\top}$ be a positive probability vector. Then*

$$\max_{\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_L)^{\top}>\mathbf{0},\rho(\mathrm{diag}(\boldsymbol{\gamma})\tilde{F})\leq 1} \quad \sum_{l=1}^{L} m_l \log\gamma_l = \sum_{l=1}^{L} m_l \log\gamma_l', \tag{4.15}$$

where $\boldsymbol{\gamma}' = (\gamma_1', \ldots, \gamma_L')^{\top} > \mathbf{0}$ is the unique vector satisfying the following conditions: $\rho(\mathrm{diag}(\boldsymbol{\gamma}')\tilde{F}) = 1$ and $\mathbf{x}(\mathrm{diag}(\boldsymbol{\gamma}')\tilde{F}) \circ \mathbf{y}(\mathrm{diag}(\boldsymbol{\gamma}')\tilde{F}) = \mathbf{m}$.

Proof. We use Theorem A.3 and Corollary A.6 in the appendix to prove Theorem 4.2. □

Combining Theorem A.3 and Corollary A.10 in the appendix, we deduce the following result.

Theorem 4.3. *Suppose that $\tilde{F}$ is given by 3) $F + (1/\bar{p}_i)\,\mathbf{v}\mathbf{e}_i^\top$, $i = \arg\max_l \rho(F + (1/\bar{p}_l)\,\mathbf{v}\mathbf{e}_l^\top)$. Let $\mathbf{m} = (m_1, \ldots, m_L)^\top$ be a positive probability vector satisfying the condition:*

$$\sum_{\forall\, j \neq l} m_j > m_l \text{ for all } l \in \langle L \rangle. \tag{4.16}$$

Then

$$\max_{\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_L)^\top > \mathbf{0}, \rho(\mathrm{diag}(\boldsymbol{\gamma})\hat{F}) \leq 1} \quad \sum_{l=1}^{L} m_l \log \gamma_l = \sum_{l=1}^{L} m_l \log \gamma_l', \tag{4.17}$$

where $\boldsymbol{\gamma}' = (\gamma_1', \ldots, \gamma_L')^\top > \mathbf{0}$ is a vector satisfying the following conditions: $\rho(\mathrm{diag}(\boldsymbol{\gamma}')\hat{F}) = 1$ and $\mathbf{x}(\mathrm{diag}(\boldsymbol{\gamma}')\hat{F}) \circ \mathbf{y}(\mathrm{diag}(\boldsymbol{\gamma}')\hat{F}) = \mathbf{m}$.

The above last two theorems enable us to choose $\mathbf{m}$ for which we know the solution to the optimization problems (4.15) and (4.17). Namely, choose $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 > \mathbf{0}$ such that $A_1 = \mathrm{diag}(\boldsymbol{\beta}_1)\tilde{F}, A_2 = \mathrm{diag}(\boldsymbol{\beta}_2)\hat{F}$ have spectral radius one. Let $\mathbf{m}_i = \mathbf{x}(A_i) \circ \mathbf{y}(A_i)$ for $i = 1, 2$. Then for $\mathbf{m}_1$, (4.15) has the unique optimal solution $\boldsymbol{\gamma}^\star = \boldsymbol{\beta}_1$. For $\mathbf{m}_2$, (4.17) has an optimal solution $\boldsymbol{\gamma}^\star = \boldsymbol{\beta}_2$. In view of Theorem A.3, $\mathbf{m}_2$ does not have to satisfy the condition $\sum_{\forall\, j \neq l} m_j > m_l$ for all $l \in \langle L \rangle$.

4.3. Relaxation with Improved Initialization. Observe that it is viable to apply the first relaxation technique, i.e., the convex relaxation and branch-and-bound method, to (4.9), and obtain upper bounds to (4.9). The bounds obtained will be looser than that employed on (3.6). On the other hand, Algorithm 1 requires an intital point that is sufficiently close to the optimal solution. We now propose a natural procedure of finding such a good initial point. The basic idea is to employ the first relaxation technique, i.e., solve (4.4) iteratively by branch and bound to locate a point $\log \boldsymbol{\gamma}(\min\{P(e^{\tilde{\boldsymbol{\gamma}}^{\mathrm{LP}_k}}), \bar{p}\})$ close enough to the optimal solution $\tilde{\boldsymbol{\gamma}}^\star$ (, i.e., a point in a ϵ-suboptimal region), and then input it as the initial point in Algorithm 1.

5. Global Optimization Algorithm. We now state a global optimization algorithm that combines the relaxation techniques and the improved initialization in Section 4 to solve (3.6) and equivalently to yield an optimal solution to (3.2) to within a prescribed accuracy on the suboptimality.

ALGORITHM 2 (Iteratively Reweighted Optimal Algorithm).
Initial Phase*: Set a prescribed accuracy ϵ. Find a ϵ-suboptimal region of (3.6) using the branch-and-bound technique in Section 4 and output the solution that yields the tightest lowest bound L_{bb}: $\tilde{\boldsymbol{\gamma}}^{BB}$. Set $\tilde{\boldsymbol{\gamma}}(0) = \tilde{\boldsymbol{\gamma}}^{BB}$.*

1. Compute the weight $\mathbf{m}(k+1)$:

$$\mathbf{m}(k+1) = \frac{\nabla \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}(k)}) \circ e^{\tilde{\boldsymbol{\gamma}}(k)}}{\mathbf{1}^\top \left(\nabla \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}(k)}) \circ e^{\tilde{\boldsymbol{\gamma}}(k)}\right)}, \tag{5.1}$$

2. Obtain $\tilde{\boldsymbol{\gamma}}(k+1)$ as the optimal solution to:

$$\begin{array}{ll} \textit{maximize} & \sum_l m_l(k+1)\tilde{\gamma} \\ \textit{subject to} & \log \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}})(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top)) \leq 0 \quad \forall\, l, \\ \textit{variables:} & \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L)^\top \in \mathbb{R}^L. \end{array} \tag{5.2}$$

3. Compute the power:

$$\mathbf{p}(k+1) = \min\left\{P(e^{\tilde{\boldsymbol{\gamma}}(k+1)}),\ \bar{p}\right\}. \tag{5.3}$$

THEOREM 5.1. *Suppose we have $U_{bb} - L_{bb} \le \epsilon$ at the completion of the initial phase in Algorithm 2, and the initial phase output $\tilde{\boldsymbol{\gamma}}^{BB}$ is a feasible solution to (3.6). If $\tilde{\boldsymbol{\gamma}}(0) = \tilde{\boldsymbol{\gamma}}^{BB}$, then $\tilde{\boldsymbol{\gamma}}(k)$ in Algorithm 2 converges to a point $\tilde{\boldsymbol{\gamma}}_\epsilon$ in a ϵ-suboptimal neighborhood of the optimal solution of (3.6), i.e., $\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}^\star}) - \Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}_\epsilon}) \le \epsilon$.*

Furthermore, $\tilde{\boldsymbol{\gamma}}(k)$ in Algorithm 2 converges to $\tilde{\boldsymbol{\gamma}}^\star$ for a sufficiently small ϵ.

Proof. Theorem 5.1 can be proved by combining the previous proofs in Section 4. □

The following result demonstrates a special case in which the optimal solution to (3.2) is given analytically, and can be computed by Algorithm 2 using a simpler initial point and in only one iteration.

COROLLARY 5.2. *If $\tilde{\gamma}_l^\star$ is equal for all l, then $\mathbf{w} = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$, where $i = \arg\max_l \rho(F + (1/\bar{p}_l)\,\mathbf{v}\mathbf{e}_l^\top)$. In this special case, $\tilde{\boldsymbol{\gamma}}(k)$ in Algorithm 2 converges to $-\log\rho(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ in only one iteration from any initial point $\tilde{\boldsymbol{\gamma}}(0)$ such that $\tilde{\gamma}_l(0)$ are equal for all l. Moreover, $\mathbf{p}(k)$ in Algorithm 2 converges to the optimal solution of (3.2) given by $\mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ (up to a scaling factor).*

Proof. Suppose that $\tilde{\gamma}_l^\star$ is equal (to a value $\tilde{\gamma}^\star$) for all l. At optimality, the constraint set of (3.6) reduces to $\tilde{\gamma}^\star + \log\rho(F + (1/\bar{p}_l)\mathbf{v}\mathbf{e}_l^\top) \le 0$ for all l, and since at least one of the spectral radius constraints in (3.6) is tight, $\tilde{\gamma}^\star = -\log\rho(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$, where $i = \arg\max_l \rho(F + (1/\bar{p}_l)\,\mathbf{v}\mathbf{e}_l^\top)$ (cf. the third matrix in (4.5)). Now, from Corollary A.6, we also have the optimality condition:

$$\mathbf{x}(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})(F+(1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top))\circ\mathbf{y}(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})(F+(1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)) = \nabla\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}^\star})\circ e^{\tilde{\boldsymbol{\gamma}}^\star}/\mathbf{1}^\top(\nabla\Phi_{\mathbf{w}}(e^{\tilde{\boldsymbol{\gamma}}^\star})\circ e^{\tilde{\boldsymbol{\gamma}}^\star}). \tag{5.4}$$

Using the fact that $\tilde{\gamma}_l^\star$ is equal for all l, (5.4) reduces to

$$\mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) = \mathbf{w}.$$

Hence, $\mathbf{w} = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ only if $\tilde{\gamma}_l^\star$ is equal for all l.

To prove the convergence of $\tilde{\boldsymbol{\gamma}}(k)$, any initial point $\tilde{\boldsymbol{\gamma}}(0)$ such that $\tilde{\gamma}_l(0)$ are equal for all l yields $\mathbf{m}(k) = \mathbf{w}$ for all k. Hence, the solution to (5.2) is always $-\log\rho(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$. Since the optimality of (3.2) and (3.6) implies

$$\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)\mathbf{p}^\star = \mathbf{p}^\star = \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top))\mathbf{p}^\star, \tag{5.5}$$

$\mathbf{p}^\star$ can be interpreted as the right eigenvector of $\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$. Together with the assumption that $\tilde{\gamma}_l^\star$ is equal for all l, this implies that $\mathbf{p}^\star = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ (up to a scaling factor). This proves Corollary 5.2. □

An example with a geometrical illustration of Corollary 5.2 is given in the following.

Example 5.1. *We give a simple illustrative example for the two user case. The channel gains are given by $G_{11} = 0.73, G_{12} = 0.04, G_{21} = 0.03, G_{22} = 0.89$ and the AWGN for the first and second user are 0.1 and 0.3 respectively. The individual maxmimum power vector $\bar{\mathbf{p}}$ is $(1, 50)^\top$. We then set $\mathbf{w} = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$, where $i = 1$ in (3.2). The rate of the two users evaluated at the solution of (3.2) given by $\mathbf{p}^\star = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ (up to a scaling factor) is then plotted on the achievable rate region (showing that maximizing the minimum rate coincides with the weighted sum rate). Algorithm 2 converges to this point in one iteration starting from any positive initial point such that $\tilde{\boldsymbol{\gamma}}(k) = \mathbf{1}$ (up to a scaling factor).*

6. Numerical Examples. In this section, we evaluate the performance of our global optimization algorithm, the relaxation techniques, Algorithm 1 and Algorithm 2. In the branch-and-bound technique, we choose the rectangular set with the largest upper bound and use the rule that splits the rectangular set with the longest edge [1]. We use nats per symbol for the sum rate unit.

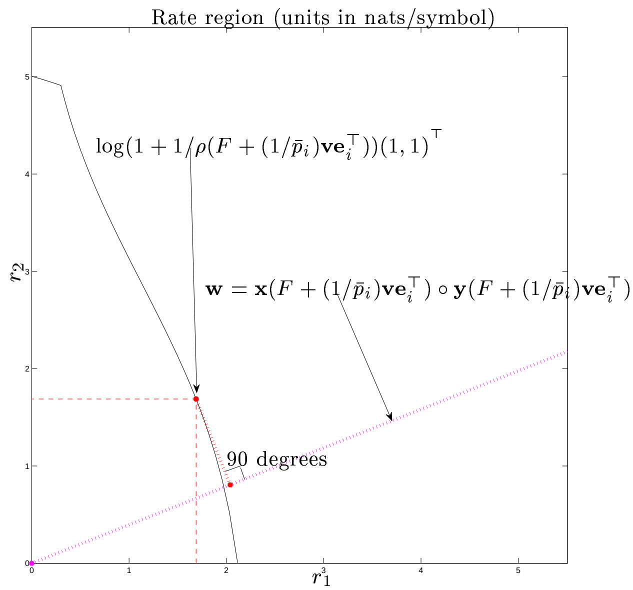


FIG. 5.1. *Achievable rate region for a 2-user interference channel. From a geometrical perspective, the weighted sum rate point (with the weight vector* $\mathbf{w} = \mathbf{x}(F+(1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F+(1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$*, where* $i=1$*, superimposed on the rate region) evaluated at the optimal solution* $\boldsymbol{\gamma}^\star = -\log\rho(F+(1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)(1,1)^\top$ *finds the largest hypercube that is contained inside the achievable rate region.*

6.1. Expt. 1 (Comparison with Lagrange dual relaxation). We compare the upper bounds obtained by our relaxation techniques and the Lagrange dual relaxation. We consider the two-user example in [15]:

$$\max_{\mathbf{0}\le p_l \le 2,\, l=1,2} \quad \frac{1}{2}\log\left(1+\frac{p_1}{p_2+1}\right)+\frac{1}{2}\log\left(1+\frac{p_2}{p_1+1}\right). \tag{6.1}$$

In [15], the optimal value of (6.1) and its optimal Lagrange dual value is computed explicitly as $(\log 3)/2$ and $(\log 5)/2$ respectively, i.e., a positive duality gap value of $\log(5/3)/2$. The optimal solution $\mathbf{p}^\star$ is either $(2,0)^\top$ or $(0,2)^\top$ (i.e., $\boldsymbol{\gamma}^\star$ is $(2,e^{-K})^\top$ or $(e^{-K},2)^\top$ in (3.4) respectively) as, for the two-user case with $w_1 = w_2$, it suffices to check the exteme points of the feasible set of (6.1) to solve (6.1) [6].

Figure 6.1 illustrates how the first method by convex relaxation computes an upper bound to (6.1). In this example, an upper bound very close to the optimal value of $(\log 3)/2$ within an acceptable accuracy can be obtained after solving two linear programs. An upper bound better than the optimal Lagrange dual value $(\log 5)/2$ is obtained after solving ten linear programs, and it takes another twelve more linear programs to certify that $(\log 3)/2$ is the global optimal value. The binary tree in Figure 6.1 has a total of fourteen levels (only the root and the first two levels are shown in Figure 6.1. Figure 6.2 shows the upper and lower bounds computed by the branch-and-bound method, and the branch and bound method converges after twenty-two iterations ($\epsilon = 1.5\times 10^{-3}$).

Next, we apply the second relaxation method that uses Algorithm 1 on (6.1), and we use an initial vector $(-0.172, -1.061)^\top$ (not too close to $\tilde{\boldsymbol{\gamma}}^\star$). The optimal value computed by Algorithm 1 when we use: 1) $\tilde{F} = F + \mathrm{diag}(\bar{\gamma})$, 2) $\tilde{F} = F + (1/\mathbf{1}^\top\bar{\mathbf{p}})\mathbf{v}\mathbf{1}^\top$ and 3)

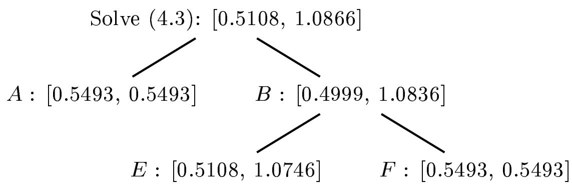


FIG. 6.1. *Solving (6.1) using the first relaxation technique: Successive linear program with a branch-and-bound algorithm (the rectangular method, see, e.g., Chapter 7 in [11]). We use a rectangular set* $[-K, \log(2)]^2$ *with* $K = 100$*. The lower and upper bounds are depicted in brackets next to the subdivided set. At the root of the tree, (4.3) is solved (original rectangle). We then have* $L_{bb} = 0.5108$ *and* $U_{bb} = 1.0866$*. In the second iteration, the rectangular set is partitioned into two: A and B (A is the set* $\{\tilde{\gamma}_1 \in [-K, \log(2)/2], \tilde{\gamma}_2 \in [-K, \log 2]\}$ *and B is the set* $\{\tilde{\gamma}_1 \in [\log(2)/2, \log 2], \tilde{\gamma}_2 \in [-K, \log 2]\}$*). We then have* $L_{bb} = 0.5493$ *and* $U_{bb} = 1.0836$*. At the third iteration, we partition the set B to obtain the bottom leaf children C and D (C is the set* $\{\tilde{\gamma}_1 \in [\log(2)/2, \log 2], \tilde{\gamma}_2 \in [\log(2)/2, \log 2]\}$ *and D is the set* $\{\tilde{\gamma}_1 \in [\log(2)/2, \log 2], \tilde{\gamma}_2 \in [-K, \log(2)/2]\}$*). We then have* $L_{bb} = 0.5493$ *and* $U_{bb} = 1.0746$*. Observe that a lower and upper bound of* 0.5493 *within acceptable accuracy to the optimal primal value* $(\log 3)/2$ *is obtained after solving two linear programs (at the node with set A containing the global optimal solution* $(-K, \log 2)^\top$*). Note that set F containing the other global optimal solution* $(\log 2, -K)^\top$ *also yields the same lower and upper bound of* 0.5493 *within acceptable accuracy after solving three linear programs.*

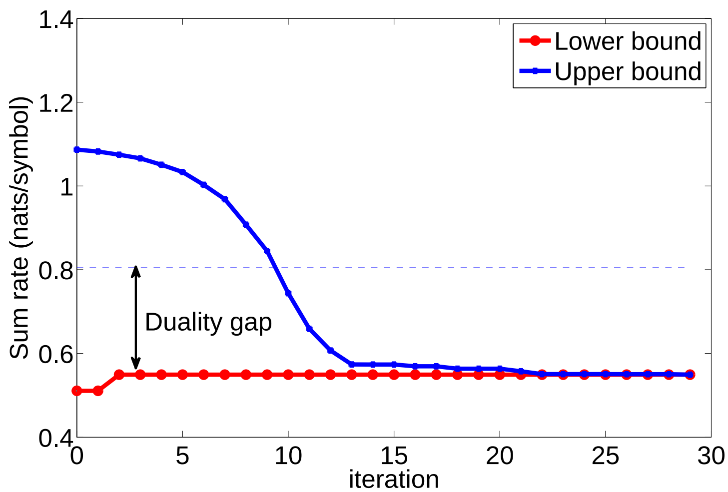
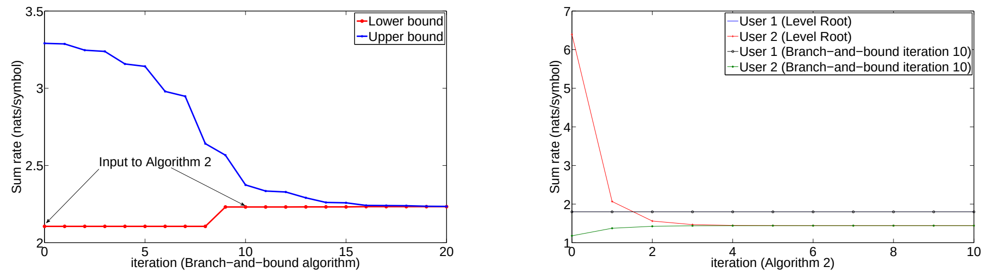


FIG. 6.2. *Upper and lower bounds on the sum rate computed by the first relaxation technique that uses successive linear programs and branch-and-bound in Expt. 1.*

$\tilde{F} = F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top$ are given by $(\log 3)/2$, $(\log 5)/2$ and $(\log 3)/2$, respectively. The first and third version of $\tilde{F}$ yield the global optimal primal value (as well as a feasible power vector $(2, 0)^\top$), whereas the second version yields the optimal dual value (with an infeasible power vector $(4, 0)^\top$). This shows that Algorithm 1 can find the global optimal power solution to (3.2) (cf. the necessary condition in the second row of Table 4.1).

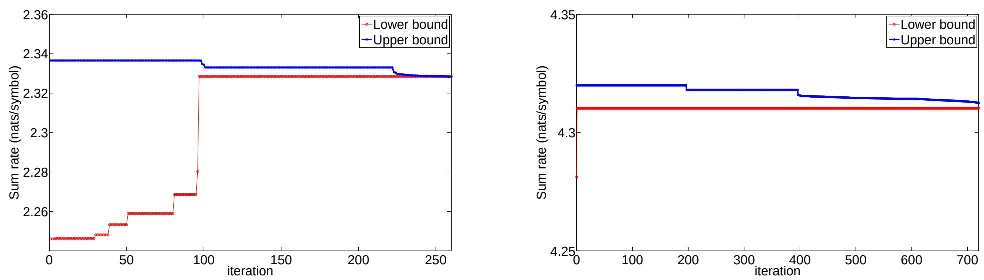
Lastly, using the feasible $\tilde{\boldsymbol{\gamma}}$ output by the first relaxation method: $(\log(2/3), \log(2/3))^\top$ and $(\log 2, -K)^\top$ that yields L_{bb} at the root and first level of the binary tree in Fig. 6.1, respectively, as initial points, Algorithm 2 converges to the global optimal solution $(\log 2, -K)^\top$.

6.2. Expt. 2 (Convergence of Algorithm 2). The channel gains are given by $G_{11} = 0.73, G_{12} = 0.04, G_{21} = 0.03, G_{22} = 0.89$ and the AWGN for the first and second user



(a) (b)

FIG. 6.3. *(a) Illustration of the convergence of the branch-and-bound relaxation with the feasible point of the tightest lower bound used as input to Algorithm 2 (b) Illustration of the convergence of Algorithm 2 with an initial point obtained at the root level and the* 10*th iteration of the branch-and-bound method.*



(a) (b)

FIG. 6.4. *Illustration of the convergence of the branch-and-bound relaxation for (a)* 100 *users and (b)* 200 *users in Expt. 3.*

are 0.1 and 0.1 respectively. The individual maxmimum power vector $\bar{\mathbf{p}}$ is $(1.8, 20.5)^\top$. We then set $\mathbf{w} = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top) \circ \mathbf{y}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$, where $i = 1$ in (3.2), and the optimal power is $\mathbf{p}^\star = \mathbf{x}(F + (1/\bar{p}_i)\mathbf{v}\mathbf{e}_i^\top)$ in (3.2) (hence the optimal solution to (3.4) is $8.334(1,1)^\top$.). Using the two different outputs from the branch-and-bound algorithm (at the root level and at the 10th iteration) shown in Figure 6.3(a) as initial points, the convergence of Algorithm 2 is shown in Figure 6.3(b). As observed in Figure 6.3(b), Algorithm 2 converges to the optimal solution with the two different initial points as input after four iterations.

6.3. Expt. 3 (Larger examples of branch-and-bound relaxation). Figure 6.4 illustrates the convergence of the branch-and-bound relaxation method for randomly generated problems with a larger number of users. As observed from Figure 6.4, when $L = 100$, the optimal solution can be computed after solving 190 linear programs, but it takes another 310 linear programs to certify that the optimal solution is within a (1×10^{-3})-suboptimal region. When $L = 200$, the optimal solution can be computed after solving 4 linear programs, but it takes approximately another 1400 linear programs to certify that the optimal solution is within a (1×10^{-3})-suboptimal region.

7. Conclusion. We studied the nonconvex sum rate maximization problem that finds applications in power control of CDMA wireless networks. Using tools from nonnegative matrix theory, in particular the Perron-Frobenius Theorem and the Friedland-Karlin inequalities, we showed that this problem can be reformulated as an equivalent convex maximization problem on a closed unbounded convex set. Utilizing the reformulation problem structure, we studied two relaxation techniques to compute progressively tight bounds to the nonconvex problem. One was based on convex relaxation by convex envelope and the other was based on successive convex approximation. We showed that the optimal solution to the sum rate maximization and its relaxation problems can be analytically characterized by the spectra of specially-crafted nonnegative matrices. Motivated by the relaxation techniques, we proposed a global optimization algorithm with ϵ-suboptmality to solve the sum rate maximization problem. We also gave new applications of the Friedland-Karlin inequalities to inverse problems in nonnegative matrix theory. As future work, we plan to extend our results in this paper to a multiple-frequency channel that has applications in a DSL multiuser system.

Appendix A. Results related to Friedland-Karlin inequalities. In this section, we recall some results from [9] and state the extensions of these results, and then illustrate their applications in this paper. We first state the following extension of [9, Theorem 3.1]:

Theorem A.1. *Let $A \in \mathbb{R}_+^{L\times L}$ be an irreducible nonnegative matrix. Assume that $\mathbf{x}(A) = (x_1(A), \ldots, x_L(A))^\top, \mathbf{y}(A) = (y_1(A), \ldots, y_L(A))^\top > \mathbf{0}$ are left and right Perron-Frobenius eigenvectors of A, normalized such that $\mathbf{x}(A) \circ \mathbf{y}(A)$ is a probability vector Suppose $\boldsymbol{\gamma}$ is a nonnegative vector. Then*

$$\rho(A) \prod_l \gamma_l^{(\mathbf{x}(A)\circ\mathbf{y}(A))_l} \leq \rho(\mathrm{diag}(\boldsymbol{\gamma})A). \tag{A.1}$$

If $\boldsymbol{\gamma}$ is a positive vector then equality holds if and only if all γ_l are equal. Furthermore, for any positive vector $\mathbf{z} = (z_1, \ldots, z_L)^\top$, the following inequality holds:

$$\rho(A) \leq \prod_{l=1}^{L} \left(\frac{(A\mathbf{z})_l}{z_l}\right)^{(\mathbf{x}(A)\circ\mathbf{y}(A))_l}. \tag{A.2}$$

If A is an irreducible nonnegative matrix with positive diagonal elements, then equality holds in (A.2) if and only if $\mathbf{z} = t\mathbf{x}(A)$ for some positive t.

Proof. Theorem 3.1 in [9] makes the following assumptions. First, in (A.1), it assumes that $\boldsymbol{\gamma} > \mathbf{0}$. Second, in (A.2), it assumes that $\rho(A) = 1$. Third, the equality case in (A.2) for $\mathbf{z} > \mathbf{0}$ is stated for a positive matrix A. We now show how to deduce the stronger version of Theorem 3.1 claimed here.

First, by using the continuity argument, we deduce the validity of (A.1) for any $\boldsymbol{\gamma} \geq \mathbf{0}$. Second, by replacing A by tA, where $t > 0$, we deduce that it suffices to show (A.2) in the case $\rho(A) = 1$.

Third, to deduce the equality case in (A.2) for $\mathbf{z} > \mathbf{0}$, we need to examine the proof of Lemma 3.2 in [9]. The proof of the Lemma 3.2 applies if the following condition holds. For any sequence of probability vectors $\mathbf{z}_i = (z_{1,i}, \ldots, z_{L,i})^\top, i = 1, \ldots$, which converges to a probability vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_L)^\top$, where $\boldsymbol{\zeta}$ has at least one zero coordinate, the function $\prod_{l=1}^{L} \left(\frac{(A\mathbf{z})_l}{z_l}\right)^{(\mathbf{x}(A)\circ\mathbf{y}(A))_l}$ tends to ∞ on the sequence $\mathbf{z}_i, i = 1, \ldots$. Assume that $\mathcal{A} = \{l \in \langle L\rangle, \zeta_l = 0\}$. Note that the complement of $\mathcal{A}$ in $\langle L\rangle$, denoted by $\mathcal{A}^c$ is nonempty.

Since $A = [a_{ij}]$ has positive diagonal entries, it follows that $\frac{(A\mathbf{z})_l}{z_l} \geq a_{ll} > 0$ for each $l \in \langle L\rangle$. Since A is irreducible, there exist $l \in \mathcal{A}$ and $m \in \mathcal{A}^c$ such that $a_{lm} > 0$. Hence, $\lim_{i\to\infty} \frac{(A\mathbf{z}_i)_l}{z_{l,i}} = \infty$. This shows that the unboundedness condition holds. $\square$

The following result gives an interpretation of (A.1) in terms of the supporting hyperplane of the convex function $\log \rho(e^{\boldsymbol{\xi}} B)$, where $B \in \mathbb{R}_+^{L\times L}$ is irreducible and $\boldsymbol{\xi} \in \mathbb{R}^L$.

Theorem A.2. *Let $B \in \mathbb{R}_+^{L\times L}$ be an irreducible nonnegative matrix. Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)^\top \in \mathbb{R}^L$ satisfy the condition $\rho(e^{\boldsymbol{\eta}} B) = 1$. Denote $A = e^{\boldsymbol{\eta}} B$ and assume that*

$\mathbf{x}(A) = (x_1(A), \ldots, x_L(A))^\top, \mathbf{y}(A) = (y_1(A), \ldots, y_L(A))^\top > \mathbf{0}$ *are left and right Perron-Frobenius eigenvectors of* A*, normalized such that* $\mathbf{x}(A) \circ \mathbf{y}(A)$ *is a probability vector. Let*

$$\mathrm{H}(\boldsymbol{\xi}) = \sum_{l=1}^{L} x_l(A) y_l(A)(\xi_l - \eta_l). \tag{A.3}$$

Then $\mathrm{H}(\boldsymbol{\xi}) \le 0$ *is the unique supporting hyperplane to the convex set* $\log \rho(e^{\boldsymbol{\xi}} B) \le 0$ *at* $\boldsymbol{\xi} = \boldsymbol{\eta}$.

Proof. Let $\boldsymbol{\xi} \in \mathbb{R}^L$. Then $e^{\boldsymbol{\xi}} B = e^{\boldsymbol{\xi} - \boldsymbol{\eta}} A$. Theorem A.1 implies that $\mathrm{H}(\boldsymbol{\xi}) \le \log \rho(e^{\boldsymbol{\xi}} B)$. Thus, $\mathrm{H}(\boldsymbol{\xi}) \le 0$ if $\log \rho(e^{\boldsymbol{\xi}} B) \le 0$. Clearly, $\mathrm{H}(\boldsymbol{\eta}) = 0$. Hence, $\mathrm{H}(\boldsymbol{\xi}) \le 0$ is a supporting hyperplane of the convex set $\log \rho(e^{\boldsymbol{\xi}} B) \le 0$. Since the function $\log \rho(e^{\boldsymbol{\xi}} B)$ is a smooth function of $\boldsymbol{\xi}$, it follows that $\mathrm{H}(\boldsymbol{\xi}) \le 0$ is unique. □

We now give an application of (A.2) in Theorem A.1.

Theorem A.3. *Let* $B \in \mathbb{R}_+^{L \times L}$ *be an irreducible nonnegative matrix. Let* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)^\top \in \mathbb{R}^L$. *Let* $\mathbf{m} = \mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) \circ \mathbf{y}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) = (m_1, \ldots, m_L)^\top$ *be a probability vector. Then, for any positive vector* $\mathbf{z} = (z_1, \ldots, z_L)^\top$,

$$\sum_{l=1}^{L} m_l \log \frac{z_l}{(B\mathbf{z})_l} \le -\log \rho(\mathrm{diag}(e^{\boldsymbol{\eta}})B) + \sum_{l=1} m_l \eta_l. \tag{A.4}$$

If B *has a positive diagonal, then equality holds if and only if* $\mathbf{z} = t\mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}})B)$ *for some* $t > 0$.

Proof. Let $A = \mathrm{diag}(e^{\boldsymbol{\eta}})B$. Then

$$\sum_{l=1}^{L} m_l \log \frac{(B\mathbf{z})_l}{z_l} = \sum_{l=1}^{L} m_l \log \frac{(A\mathbf{z})_l}{z_l} - \sum_{l=1} m_l \eta_l.$$

Use (A.2) to deduce (A.4). The equality case follows from the equality case in (A.2). □

We now turn to applying Theorem A.1 to solve the following inverse problem.

Problem A.4. *Let* $B \in \mathbb{R}_+^{L \times L}, \mathbf{m} \in \mathbb{R}_+^L$ *be given irreducible nonnegative matrix and positive probability vector, respectively. When does there exist* $\boldsymbol{\eta} \in \mathbb{R}^L$ *such that* $\mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) \circ \mathbf{y}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) = \mathbf{m}$*? If such* $\boldsymbol{\eta}$ *exists, when is it unique up to an addition* $t\mathbf{1}$*?*

To solve the inverse problem, we recall Theorem 3.2 in [9] (a consequence of Theorem A.1, i.e., Theorem 3.1 in [9]) that is reproduced in the following.

Theorem A.5. *Let* $A \in \mathbb{R}_+^{L \times L}, \mathbf{u}, \mathbf{v} \in \mathbb{R}_+^L$ *be given, where* A *is irreducible with positive diagonal elements and* $\mathbf{u}, \mathbf{v}$ *are positive. Then, there exists* $D_1, D_2 \in \mathbb{R}_+^{L \times L}$ *such that*

$$D_1 A D_2 \mathbf{u} = \mathbf{u},\ \mathbf{v}^\top D_1 A D_2 = \mathbf{v}^\top,\ D_1 = \mathrm{diag}(\mathbf{f}),\ D_2 = \mathrm{diag}(\mathbf{g}) \text{ and } \mathbf{f}, \mathbf{g} > \mathbf{0}. \tag{A.5}$$

The pair (D_1, D_2) *is unique to the change* $(tD_1, t^{-1}D_2)$ *for any* $t > 0$. *There exist* $\boldsymbol{\eta} \in \mathbb{R}^L$ *such that* $\mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}}B)) \circ \mathbf{y}(\mathrm{diag}(e^{\boldsymbol{\eta}}B)) = \mathbf{m}$. *Furthermore,* $\boldsymbol{\eta}$ *is unique up to an addition* $t\mathbf{1}$.

Corollary A.6. *Let* $B \in \mathbb{R}_+^{L \times L}, \mathbf{m} \in \mathbb{R}_+^L$ *be a given irreducible nonnegative matrix with positive diagonal elements and a positive probability vector, respectively. Then, there exists* $\boldsymbol{\eta} \in \mathbb{R}^L$ *such that* $\mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) \circ \mathbf{y}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) = \mathbf{m}$. *Furthermore,* $\boldsymbol{\eta}$ *is unique up* *to an addition of $t\mathbf{1}$. In particular, this $\boldsymbol{\eta}$ can be computed by solving the following convex optimization problem:*

$$\begin{array}{ll} \textit{maximize} & \mathbf{m}^\top \boldsymbol{\eta} \\ \textit{subject to} & \log \rho(\operatorname{diag}(e^{\boldsymbol{\eta}})B) \le 0, \\ \textit{variables:} & \boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)^\top \in \mathbb{R}^L. \end{array} \tag{A.6}$$

Proof. Let $\mathbf{u} = \mathbf{1}, \mathbf{v} = \mathbf{m}$. Then, from Theorem A.5, there exists D_1, D_2 two diagonal matrices with positive diagonal entries such that $D_1 B D_2 \mathbf{1} = \mathbf{1}, \mathbf{m}^\top D_1 B D_2 = \mathbf{m}^\top$. Consider the matrix $D_2 D_1 B = D_2(D_1 B D_2) D_2^{-1}$. It is straightforward to see that $\mathbf{x}(D_2 D_1 B) \circ \mathbf{y}(D_2 D_1 B) = \mathbf{m}$. Hence, $\boldsymbol{\eta}$ is the unique solution of $\operatorname{diag}(e^{\boldsymbol{\eta}}) = D_2 D_1$.

Assume that $\boldsymbol{\zeta} \in \mathbb{R}^L$ satisfies $\mathbf{x}(\operatorname{diag}(e^{\boldsymbol{\zeta}})B) \circ \mathbf{y}(\operatorname{diag}(e^{\boldsymbol{\zeta}})B) = \mathbf{m}$. By considering $\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta} + t\mathbf{1}$, we may assume that $\rho(\operatorname{diag}(e^{\boldsymbol{\zeta}})B) = 1$. Let $D_4 = \operatorname{diag}(\mathbf{x}(\operatorname{diag}(e^{\boldsymbol{\zeta}})B)$. Then $(D_4^{-1} \operatorname{diag}(e^{\boldsymbol{\zeta}}) B D_4)\mathbf{1} = \mathbf{1}$. Let $D_3 = D_4^{-1} \operatorname{diag}(e^{\boldsymbol{\zeta}})$. Hence, $\mathbf{y}(D_3 B D_4) = \mathbf{m}$. In view of Theorem A.5, $\operatorname{diag}(e^{\boldsymbol{\zeta}}) = D_4 D_3 = D_2 D_1 = \operatorname{diag}(e^{\boldsymbol{\eta}})$.

Next, we show that (A.6) computes the required $\boldsymbol{\eta}$. Since (A.6) is convex, we apply the Karush-Kuhn-Tucker conditions to (A.6). The stationarity of the Lagrangian yields $\mathbf{x}(\operatorname{diag}(e^{\boldsymbol{\eta}})B) \circ \mathbf{y}(\operatorname{diag}(e^{\boldsymbol{\eta}})B) = \mathbf{m}$, thus proving the corollary. $\square$

We illustrate the necessity of the irreducible nonnegative matrix having positive diagonal elements in Corollary A.6 by the following example.

Example A.7. *Let us look at the matrix F defined in (2.2) having zero diagonal entries and positive off-diagonal entries. For $L = 2$, it is easy to show that $\mathbf{x}(F) \circ \mathbf{y}(F) = (\frac{1}{2}, \frac{1}{2})^\top$. In particular, for $L = 2$, Problem A.4 is not solvable for $\mathbf{m} \neq (\frac{1}{2}, \frac{1}{2})^\top$. Similarly, given positive $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ such that $\mathbf{u} \circ \mathbf{v} \neq t(1,1)$ for any positive t, (A.5) does not hold for $A = F$. For $L \ge 3$, the situation is different, and is illustrated in the following result.*

Theorem A.8. *Let $L \ge 3, A \in \mathbb{R}_+^{L \times L}, \mathbf{u} = (u_1, \ldots, u_L)^\top, \mathbf{v} = (v_1, \ldots, v_L)^\top \in \mathbb{R}_+^L$ be given, where A is a matrix with zero diagonal entries and positive off-diagonal elements, and $\mathbf{u}, \mathbf{v}$ are positive. Assume that $\mathbf{m} = \mathbf{u} \circ \mathbf{v}$ is a probability vector satisfying the condition*

$$\sum_{\forall\, j \neq l} m_j > m_l \textit{ for all } l \in \langle L \rangle. \tag{A.7}$$

Then, there exists $D_1, D_2 \in \mathbb{R}_+^{L \times L}$ such that (A.5) holds.

Proof. Let $A_i = A + (1/i)I, i = 1, \ldots$, where I is the $L \times L$ identity matrix. Theorem A.5 implies

$$\begin{gathered} B_i = D_{1,i} A_i D_{2,i},\ D_{1,i} = \operatorname{diag}(\mathbf{f}_i),\ D_{2,i} = \operatorname{diag}(\mathbf{g}_i),\ B_i \mathbf{u} = \mathbf{u},\ \mathbf{v}^\top B_i = \mathbf{v}^\top, \\ \mathbf{f}_i = (f_{1,i}, \ldots, f_{L,i})^\top,\ \mathbf{w}_i = (g_{1,i}, \ldots, g_{L,i})^\top,\ s_i = \max_{j \in \langle L \rangle} f_{j,i} = \max_{j \in \langle L \rangle} g_{j,i},\ i = 1, \ldots. \end{gathered}$$

Note that each entry of B_i is bounded by $\frac{\max_j u_j}{\min_j u_j}$. By passing to the subsequence $B_{i_k}, \mathbf{f}_{i_k}, \mathbf{g}_{i_k}, 1 \le i_1 < i_2 < \ldots$, we can assume that the first subsequence converges to B, and the last two subsequences converge in generalized sense:

$$\begin{gathered} \lim_{k \to \infty} B_{i_k} = B = [b_{jl}] \in \mathbb{R}_+^{L \times L},\ \lim_{k \to \infty} \mathbf{f}_{i_k} = \mathbf{f} = (f_1, \ldots, f_L)^\top,\ \lim_{k \to \infty} \mathbf{g}_{i_j} = \mathbf{g} = (g_1, \ldots, g_L)^\top, \\ f_j, g_j \in [0, \infty], j = 1, \ldots, L,\ \lim_{k \to \infty} s_{i_k} = s = \max_{j \in \langle L \rangle} f_j = \max_{j \in \langle L \rangle} g_j \in [0, \infty]. \end{gathered}$$

Note that

$$B\mathbf{u} = \mathbf{u}, \quad \mathbf{v}^\top B = \mathbf{v}^\top. \tag{A.8}$$

Assume first that $s < \infty$. Then $B = \mathrm{diag}(\mathbf{f}) A\, \mathrm{diag}(\mathbf{g})$. In view of (A.8), $\mathbf{f} \circ \mathbf{g} > \mathbf{0}$. This proves the theorem in this case.

Assume now that $s = \infty$. Let

$$\mathcal{F}_\infty = \{j \in \langle L \rangle,\ f_j = \infty\},\ \mathcal{F}_+ = \{j \in \langle L \rangle,\ f_j \in (0, \infty),\}\ \mathcal{F}_0 = \{j \in \langle L \rangle,\ f_j = 0\},$$
$$\mathcal{G}_\infty = \{j \in \langle L \rangle,\ g_j = \infty\},\ \mathcal{G}_+ = \{j \in \langle L \rangle,\ g_j \in (0, \infty),\}\ \mathcal{G}_0 = \{j \in \langle L \rangle,\ g_j = 0\}.$$

Since off-diagonal entries of A are positive, and $B \in \mathbb{R}_+^{L \times L}$ it follows that $\mathcal{F}_\infty = \mathcal{G}_\infty = \{l\}$ for some $l \in \langle L \rangle$. Furthermore, $\mathcal{F}_+ = \mathcal{G}_+ = \emptyset$. So $\mathcal{F}_0 = \mathcal{G}_0 = \langle L \rangle \backslash \{l\}$. Assume first that $l = 1$. Then the principal submatrix $[b_{jl}]_{j=l=2}^L$ is zero. (A.8) yields that

$$b_{j1} = \frac{u_j}{u_1},\ b_{1j} = \frac{v_j}{v_1} \text{ for } j = 2, \ldots, L,\ b_{11} u_1 v_1 + \sum_{j=2}^L u_j v_j = u_1 v_1.$$

Since $b_{11} \geq 0$, the above last equality contradicts the condition (A.7) for $l = 1$. Similar argument implies the impossibility of $\mathcal{F}_\infty = \mathcal{G}_\infty = \{l\}$ for any $l \geq 2$. Hence, $s < \infty$ and we conclude the theorem. □

We do not know whether, under the conditions of Theorem A.8, the diagonal matrices (D_1, D_2) are unique up to the transformation $(tD_1, t^{-1} D_2)$. We now generalize the above theorem.

Theorem A.9. *Let*

$$L \geq 2,\ A = [a_{jl}]_{j=l=1}^L \in \mathbb{R}_+^{L \times L},\ \mathbf{0} < \mathbf{u} = (u_1, \ldots, u_L)^\top,\ \mathbf{v} = (v_1, \ldots, v_L)^\top \in \mathbb{R}_+^L$$

be given. Assume that A has positive off-diagonal elements, and $\mathbf{m} = \mathbf{u} \circ \mathbf{v}$ is a probability vector satisfying the condition

$$\sum_{\forall\, j \neq l} m_j > m_l \tag{A.9}$$

for each l such that $a_{ll} = 0$. Then, there exists $D_1, D_2 \in \mathbb{R}_+^{L \times L}$ such that (A.5) holds.

Proof. Assume first that $L \geq 3$. In view of Theorems A.5 and A.8, it suffices to assume that A has positive and zero diagonal entries. Apply the proof of Theorem A.8 and the following observation. If $\mathcal{F}_\infty = \mathcal{G}_\infty = \{l\}$ then $a_{ll} = 0$.

Assume now that $L = 2$. Note that if A has a zero diagonal then the condition (A.7) can not hold. Assume now that A has at least one positive diagonal element. Then the above arguments for $L \geq 3$ apply. □

Corollary A.10. *Let $B = [b_{jl}]_{j=l=1}^L \in \mathbb{R}_+^{L \times L}, \mathbf{m} \in \mathbb{R}_+^L$ be a given matrix with positive off-diagonal elements and a positive probability vector, respectively. Assume that $L \geq 2$ and $\mathbf{m}$ satisfies the conditions (A.9) for each l such that $b_{ll} = 0$. Then, there exists $\boldsymbol{\eta} \in \mathbb{R}^L$ such that $\mathbf{x}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) \circ \mathbf{y}(\mathrm{diag}(e^{\boldsymbol{\eta}})B) = \mathbf{m}$.*

REFERENCES

[1] V. Balakrishnan, S. Boyd and S. Balemi, Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems, *International Journal of Robust and Nonlinear Control*, 1 (4), 295-317, 1991.

[2] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, 1979.

[3] M. Chiang, P. Hande, T. Lan and C. W. Tan, Power Control in Wireless Cellular Networks, *Foundations and Trends in Networking*, 2 (4), 381-533, 2008.

[4] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill and D. Julian, Power Control by Geometric Programming, *IEEE Trans. on Wireless Communications*, 6 (7), 2640-2651, 2007.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.
[6] M. Ebrahimi and M. A. Maddah-Ali and A. K. Khandani, Power Allocation and Asymptotic Achievable Sum-Rates in Single-Hop Wireless Networks, *Proc. of IEEE 40th Annual Conference on Information Sciences and Systems*, 2006.
[7] J. E. Falk, Lagrange Multipliers and Nonconvex Programs, *SIAM J. Control*, 7 (4), 534-545, 1969.
[8] G. J. Foschini and Z. Miljanic, A Simple Distributed Autonomous Power Control Algorithm and its Convergence, *IEEE Trans. on Vehicular Technology*, 42 (4), 641-646, 1993.
[9] S. Friedland and S. Karlin, Some Inequalities for the Spectral Radius of Non-negative Matrices and Applications, *Duke Mathematical Journal* 42 (3), 459-490, 1975.
[10] F. R. Gantmacher, *The Theory of Matrices*, Vol. II, Chelsea Publ. Co., New York, 1974.
[11] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, 3rd Edition, Springer, New York, 1995.
[12] J. F. C. Kingman, A Convexity Property of Positive Matrices, *Quart. J. Math. Oxford Ser.* 12 (2), 283-284, 1961.
[13] R. Lui and W. Yu, Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems, *IEEE Trans. on Communications*, 54 (7), 1310-1322, 2006.
[14] Z.-Q. Luo and W. Yu, An Introduction to Convex Optimization for Communications and Signal Processing, (Tutorial Paper), *IEEE Journal on Selected Areas in Communications*, 24 (8), 1426-1438, 2006.
[15] Z.-Q. Luo and Z. Zhang, Dynamic Spectrum Management: Complexity and Duality, *IEEE Journal on Selected Areas in Signal Processing*, 2 (1), 57-73, 2008.
[16] Z.-Q. Luo and Z. Zhang, Duality Gap Estimation and Polynomial Time Approximation for Optimal Spectrum Management, *IEEE Trans. on Signal Processing*, 57 (7), 2675-2689, 2008.
[17] B. R. Marks and G. P. Wright, A General Inner Approximation Algorithm for Nonconvex Mathematical Programs, *Operations Research*, 26 (4), 681-683, 1978.
[18] C. W. Tan, D. P. Palomar and M. Chiang, Solving Nonconvex Power Control Problems in Wireless Networks: Low SIR Regime and Distributed Algorithms, *Proc. of IEEE Global Communications (Globecom)*, 2005.
[19] C. W. Tan, M. Chiang and R. Srikant, Fast Algorithms and Performance Bounds for Sum Rate Maximization in Wireless Networks, *Proc. of IEEE Conference on Computer Communications (Infocom)*, 2009.
[20] C. W. Tan, M. Chiang and R. Srikant, Maximizing Sum Rate and Minimizing MSE on Multiuser Downlink: Optimality, Fast Algorithm, and Equivalence via Max-min SIR, *Proc. of IEEE International Symposium on Information Theory (ISIT)*, 2009.
[21] H. Tuy, T. V. Thieu and N. Q. Thai, A Conical Algorithm for Globally Minimizing a Concave Function over a Closed Convex Set, *Mathematics of Operations Research*, 10 (3), 498-514, 1985.