Maximizing Sum Rate and Minimizing MSE on Multiuser Downlink: Optimality, Fast Algorithms and Equivalence via Max-min SIR

Chee Wei Tan,[1] Mung Chiang[2] and R. Srikant[3]
[1] City University of Hong Kong [2] Princeton University [3] University of Illinois at Urbana-Champaign

Abstract

Maximizing the minimum weighted SIR, minimizing the weighted sum MSE and maximizing the weighted sum rate in a multiuser downlink system are three important performance objectives in joint transceiver and power optimization, where all the users have a total power constraint. We show that, through connections with the nonlinear Perron-Frobenius theory, jointly optimizing power and beamformers in the max-min weighted SIR problem can be solved optimally in a distributed fashion. Then, connecting these three performance objectives through the arithmetic-geometric mean inequality and nonnegative matrix theory, we solve the weighted sum MSE minimization and the weighted sum rate maximization in the low to moderate interference regimes using fast algorithms. In the general case, we first establish the optimality conditions to the weighted sum MSE minimization and the weighted sum rate maximization problems and provide their further connection to the max-min weighted SIR problem. We then propose a distributed weighted proportional SIR algorithm that leverages our fast max-min weighted SIR algorithm to solve the two nonconvex problems, and give conditions under which global optimality is achieved. Numerical results are provided to complement the analysis.

Index Terms— Duality, Distributed optimization, Power control, Sum rate, MMSE, Beamforming, Interference channel.

I. Introduction

We consider the multiuser downlink transmission on a Multiple-Input-Single-Output (MISO) channel, where the transmitter (at the base station) is equipped with an antenna array and each user has a single receive antenna. Full channel information is available at both the transmitter and the receiver, and all the users share the same bandwidth under a total power constraint. The multiuser downlink system is modeled as an interference channel, where a minimal or low-complexity coordination among the users is desired for the purpose of decentralized implementation in a network. Under this setting, the antenna array provides an extra degree of freedom, in addition to power control, to optimize performance, e.g., increasing the total throughput (sum rates) or the total reliability in the system. Joint optimization of transmit power and beamformer involving the weighted sum rates and weighted sum mean squared error (MSE) as objectives are however challenging to solve, because these two problems are nonconvex. Further, the transmit beamformers are coupled across users, thereby making them hard to optimize in a distributed fashion.

Our approach to these two nonconvex optimization problems begins by first considering a joint optimization of power and transmit beamformer for the min-max (weighted) mean squared error (MSE) problem or, equivalently, the max-min (weighted) Signal-to-Interference Ratio (SIR) problem. While previous algorithms in the literature require centralized computation of the eigenvalue and eigenvector of an extended coupling matrix, we propose a fast distributed algorithm that computes the optimal power and transmit beamformer in the max-min weighted SIR problem with geometric convergence rate. This is achieved by applying the nonlinear Perron-Frobenius theory in [1]–[3] and the uplink-downlink duality in [4]–[10], wherein the uplink acts as an intermediate mechanism to optimize transmit beamformers in the downlink. We also show that the uplink-downlink duality is a special case of a nonlinear Perron-Frobenius minimax characterization.

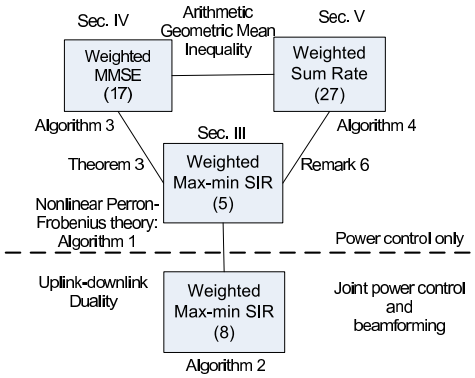


Fig. 1. Overview of the connection (solid lines) between the three optimization problems in the paper: i) Weighted sum MSE minimization in (27), ii) weighted sum rate maximization in (39), and iii) max-min weighted SIR in (8). The upper half of the dotted line considers power control only, while the lower half considers both power control and beamforming.

From an information theory viewpoint, treating interference as noise in the low interference regime has recently been justified in [11]. By considering simple and low-complexity receiver design, e.g., linear beamformer (effectively treating interference as noise), we study the nonconvex problems of, 1) minimizing the weighted sum MSE between the transmitted and estimated symbols, and, 2) maximizing the weighted sum rate. The max-min SIR problem is shown to be a special case of these two problems in the sense that optimal solutions are equivalent under special cases. Previous work in the literature, see e.g., [12], only solve these two nonconvex problems in a centralized manner and suboptimally. We develop fast algorithms (independent of stepsize and no configuration whatsoever) to solve these two nonconvex problems optimally under low to medium interference conditions. We leverage the standard interference function approach in [13] to show that our algorithms converge even under asynchronous updates.

We then turn to establishing the optimality conditions on the power and beamformer of these two nonconvex problems in the general case (any interference conditions). Using nonnegative matrix theory, the optimal beamformer will be shown to be the linear minimum mean squared error (LMMSE) filter. This relates to earlier work on the optimality of the LMMSE filter in the max-min SIR problem and a related total power minimization problem [4], [5], [8], [9]. Further theoretical and algorithmic connections between the two nonconvex problems and the max-min SIR problem are established using nonnegative matrix theory, and a fast algorithm (with minimal configuration) that leverages our fast max-min weighted SIR algorithm and the uplink-downlink duality is then proposed to solve these nonconvex problems in a distributed fashion.

This paper is organized as follows. We present the system model in Section II. In Section III, we look at the max-min weighted SIR power control problem and its extension to joint beamforming and power control, and we propose fast algorithms to solve them. In Section IV and Section V, we look at the MSE minimization and weighted sum rate maximization power control problems and solve them using fast algorithms in the low to medium interference regime. In Section VI, we establish the optimality conditions for these two problems and propose a weighted proportional SIR algorithm to solve them in the general case. We highlight the performance of our algorithms using numerical examples in Section VII. We conclude with a summary in Section VIII. All the proofs can be found in Appendix.

We refer the readers to Figure 1 for an overview of the connection between the three main optimization problems in the paper. The following notation is used. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, italics denote scalars, and $\mathbf{u} \geq \mathbf{v}$ ($\mathbf{B} \geq \mathbf{F}$) denotes componentwise inequality between nonnegative vectors $\mathbf{u}$ and $\mathbf{v}$ (nonnegative matrices $\mathbf{B}$ and $\mathbf{F}$). We let $(\mathbf{By})_l$ denote the lth element of $\mathbf{By}$. The Perron-Frobenius eigenvalue of a nonnegative matrix $\mathbf{F}$ is denoted as $\rho(\mathbf{F})$, and the Perron (right) and left eigenvectors of $\mathbf{F}$ associated with $\rho(\mathbf{F})$ are denoted by $\mathbf{x}(\mathbf{F})$ and $\mathbf{y}(\mathbf{F})$, respectively. The super-scripts $(\cdot)^\top$ and $(\cdot)^\dagger$ denote transpose and complex conjugate transpose respectively.

We let $\mathbf{e}_l$ denote the lth unit coordinate vector and $\mathbf{I}$ denote the identity matrix. Let $\mathbf{x} \circ \mathbf{y}$ denote $\mathbf{x} \circ \mathbf{y} = [x_1 y_1, \ldots, x_L y_L]^\top$ (Schur product). Let $e^{\mathbf{x}}$ denote $e^{\mathbf{x}} = [e^{x_1}, \ldots, e^{x_L}]^\top$.

II. System Model

We consider a single cell multiuser downlink system with N antennas at the base station and L decentralized users, each equipped with a single receive antenna, operating in a frequency-flat fading channel. The downlink channel can be modeled as a vector Gaussian broadcast channel:

$$\boldsymbol{y}_l = \boldsymbol{h}_l^\dagger \boldsymbol{x} + \boldsymbol{z}_l, \quad l = 1, \ldots, L, \tag{1}$$

where $\boldsymbol{y}_l \in \mathbb{C}^{1\times 1}$ is the received signal of the lth user, $\boldsymbol{h}_l \in \mathbb{C}^{N\times 1}$ is the channel matrix between the base station and the lth user, $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$ is the transmitted signal vector, and $\boldsymbol{z}_l$'s are the i.i.d. additive complex Gaussian noise vectors with variance $n_l/2$ on each of its real and imaginary components.

We assume that the multiuser system adopts a linear transmission and reception strategy. In transmit beamforming, the base station transmits a signal $\boldsymbol{x}$ in the form of $\boldsymbol{x} = \sum_{l=1}^{L} d_l \hat{\boldsymbol{w}}_l$, where $\hat{\boldsymbol{w}}_l \in \mathbb{C}^{N\times 1}$ is the transmit beamformer that carries the information signal d_l of the lth user. We assume a total power constraint at the transmit antennas, i.e., $\mathbb{E}[\boldsymbol{x}^\dagger \boldsymbol{x}] = \bar{P}$. From (1), the received signal for the lth user can be expressed as

$$\boldsymbol{y}_l = \left(\boldsymbol{h}_l^\dagger \hat{\boldsymbol{w}}_l\right) d_l + \sum_{j\neq l} \left(\boldsymbol{h}_l^\dagger \hat{\boldsymbol{w}}_j\right) d_j + \boldsymbol{z}_l. \tag{2}$$

Next, we write $\hat{\boldsymbol{w}}_l = \sqrt{p_l}\mathbf{u}_l$, where p_l is the downlink transmit power and $\mathbf{u}_l$ is the normalized transmit beamformer, i.e., $\mathbf{u}_l^\dagger \mathbf{u}_l = 1$, of the lth user. Now, the received SIR of the lth user in the downlink transmission can be given in terms of $\mathbf{p}$ and $\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_L]$:

$$\mathsf{SIR}_l(\mathbf{p}, \mathbf{U}) = \frac{p_l |\boldsymbol{h}_l^\dagger \mathbf{u}_l|^2}{\sum_{j\neq l} p_j |\boldsymbol{h}_l^\dagger \mathbf{u}_j|^2 + n_l}. \tag{3}$$

We define the matrix $\mathbf{G}$ with entries $G_{lj} = |\boldsymbol{h}_l^\dagger \mathbf{u}_j|^2$ in the downlink transmission. In terms of the beamforming matrix $\mathbf{U}$, we also define the (cross channel interference) matrix $\mathbf{F}(\mathbf{U})$ with entries:

$$F_{lj}(\mathbf{U}) = \begin{cases} 0, & \text{if } l = j \\ \frac{G_{lj}(\mathbf{U})}{G_{ll}(\mathbf{U})}, & \text{if } l \neq j \end{cases} \tag{4}$$

and

$$\mathbf{v}(\mathbf{U}) = \left(\frac{n_1}{G_{11}(\mathbf{U})}, \frac{n_2}{G_{22}(\mathbf{U})}, \ldots, \frac{n_L}{G_{LL}(\mathbf{U})}\right)^\top. \tag{5}$$

For brevity, we omit the dependency on $\mathbf{U}$ when we fix the beamformers and for the most part of the paper. This dependency is made explicit only in Section III-C.

Note that (3), as a linear fractional function of $\mathbf{p}$ and $\mathbf{U}$, depends only on the L pairs of parameters $\{\boldsymbol{h}_l, n_l\}$. An equivalent form of (3) is obtained if we rewrite (3) in terms of the normalized parameter pairs $\{\tilde{\boldsymbol{h}}_l, 1\}$, where $\tilde{\boldsymbol{h}}_l = \boldsymbol{h}_l/n_l$ [14]. Accordingly, the normalized $\mathbf{F}$ in (4) has entries $|\tilde{\boldsymbol{h}}_l^\dagger \mathbf{u}_j|^2$ for all $j \neq l$, and the normalized $\mathbf{v}$ in (5) becomes $\mathbf{1}$ subject to $|\tilde{\boldsymbol{h}}_l^\dagger \mathbf{u}_l| = 1$ for all l.

In the following, we study optimization problems having two performance metrics that are functions of $\mathsf{SIR}_l(\mathbf{p}, \mathbf{U})$, namely the MSE at the output of a LMMSE filter of each user [15], [16]:

$$\mathrm{MSE}_l(\mathbf{p}) = \frac{1}{1 + \mathsf{SIR}_l(\mathbf{p})} \tag{6}$$

and the throughput of each user (assuming the Shannon capacity formula) [10], [17]:

$$r_l(\mathbf{p}) = \log(1 + \mathsf{SIR}_l(\mathbf{p})). \tag{7}$$

III. MAX-MIN WEIGHTED SIR OPTIMIZATION

In this section, we first consider optimizing only power before we consider a joint optimization between power and transmit beamformers. Let $\boldsymbol{\beta}$ be a positive vector, where the lth entry β_l is assigned by the network to the lth link (to reflect some long-term priority). We first consider the following max-min weighted SIR problem:

$$\begin{array}{ll} \text{maximize} & \min\limits_{l} \ \dfrac{\mathsf{SIR}_l(\mathbf{p})}{\beta_l} \\ \text{subject to} & \mathbf{1}^{\top}\mathbf{p} \leq \bar{P}, \ \mathbf{p} \geq \mathbf{0}, \\ \text{variables:} & \mathbf{p}. \end{array} \tag{8}$$

Note that (8) is equivalent to the min-max weighted MSE problem:

$$\begin{array}{ll} \text{minimize} & \max\limits_{l} \ \beta_l \mathrm{MSE}_l(\mathbf{p}) = \dfrac{\beta_l}{(1+\mathsf{SIR}_l(\mathbf{p}))} \\ \text{subject to} & \mathbf{1}^{\top}\mathbf{p} \leq \bar{P}, \ \mathbf{p} \geq \mathbf{0}, \\ \text{variables:} & \mathbf{p}. \end{array} \tag{9}$$

Next, let us define the following nonnegative matrix

$$\mathbf{B} = \mathbf{F} + (1/\bar{P})\mathbf{v}\mathbf{1}^{\top}. \tag{10}$$

We will extensively exploit the spectra of $\mathbf{B}$ (particularly, its spectral radius, its corresponding eigenvectors, its quasi-inverse and other properties) in our problem formulation, their solution and algorithm design in this paper.

A. Optimal solution and algorithm

By exploiting a connection between the nonlinear Perron-Frobenius theory in [1], [2] and the algebraic structure of (8), we can give a closed form solution to (8).[1]

Lemma 1: The optimal objective and solution of (8) is given by $1/\rho(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B})$ and $(\bar{P}/\mathbf{1}^{\top}\mathbf{x}(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B}))\mathbf{x}(\mathrm{diag}(\boldsymbol{\beta}$ respectively.

Note that the optimal power in Lemma 1 can also be expressed as

$$\mathbf{p} = \left(\rho(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B})\mathbf{I} - \mathrm{diag}(\boldsymbol{\beta})\mathbf{F}\right)^{-1}\mathrm{diag}(\boldsymbol{\beta})\mathbf{v}. \tag{11}$$

The following algorithm computes the optimal power of (8) given in Lemma 1. We let k index discrete time slots.

Algorithm 1 (Max-min Weighted SIR):

1) Update power $\mathbf{p}(k+1)$:

$$p_l(k+1) = \left(\frac{\beta_l}{\mathsf{SIR}_l(\mathbf{p}(k))}\right) p_l(k) \quad \forall\, l. \tag{12}$$

2) Normalize $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P}/\mathbf{1}^{\top}\mathbf{p}(k+1). \tag{13}$$

Corollary 1: Starting from any initial point $\mathbf{p}(0)$, $\mathbf{p}(k)$ in Algorithm 1 converges geometrically fast to the optimal solution of (8), $(\bar{P}/\mathbf{1}^{\top}\mathbf{x}(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B}))\mathbf{x}(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B})$.

Remark 1: Interestingly, (12) in Algorithm 1 is simply the Distributed Power Control (DPC) algorithm in [18], where the lth user has a virtual SIR threshold of β_l in the downlink transmission. However, the

[1]A closed-form solution to (8) was first obtained in [8] using a nonnegative (increased dimension) matrix totally different from $\mathbf{B}$. As such, the algorithmic solution to (8) in [8] is different and is mainly centralized. On the other hand, our solution exploits the DPC algorithm in [18] and is distributed.

standard interference function approach in [13], a well-known method used to prove the convergence of the DPC algorithm, cannot prove the convergence of Algorithm 1 (due to (13) violating the standard interference function). Our convergence result in Corollary 1 follows from [1], [2], a special case of nonlinear Perron-Frobenius theory.

Remark 2: In principle, the normalization at Step 2 can be made distributed by using gossip algorithms to compute $\mathbf{1}^\top \mathbf{p}(k+1)$ at each user [19].

B. Nonlinear Perron-Frobenius minimax characterization

We first establish the following result based on the nonlinear Perron-Frobenius theory and the Friedland-Karlin inequality in [20], [21] (see (80) in Appendix and see [22] for its extension), and then discuss how it provides further insight into the analytical solution of (8). The result is also useful when we consider a different reformulation of (8) in Section VI-A.

Lemma 2: Let $\mathbf{A}$ be an irreducible nonnegative matrix, $\mathbf{b}$ a nonnegative vector and $\|\cdot\|$ a norm on $\mathbb{R}^L$ with a corresponding dual norm $\|\cdot\|_D$. Then,

$$\begin{aligned}\log\rho(\mathbf{A}+\mathbf{b}\mathbf{c}_*^\top) &= \max_{\|\mathbf{c}\|_D=1} \log\rho(\mathbf{A}+\mathbf{b}\mathbf{c}^\top) \\ &= \max_{\boldsymbol{\lambda}\geq\mathbf{0},\mathbf{1}^\top\boldsymbol{\lambda}=1} \min_{\|\mathbf{p}\|=1} \sum_l \lambda_l \log\frac{(\mathbf{A}\mathbf{p}+\mathbf{b})_l}{p_l} \qquad (14)\\ &= \min_{\|\mathbf{p}\|=1} \max_{\boldsymbol{\lambda}\geq\mathbf{0},\mathbf{1}^\top\boldsymbol{\lambda}=1} \sum_l \lambda_l \log\frac{(\mathbf{A}\mathbf{p}+\mathbf{b})_l}{p_l}, \qquad (15)\end{aligned}$$

where the optimal $\mathbf{p}$ in (14) and (15) are both given by $\mathbf{x}(\mathbf{A}+\mathbf{b}\mathbf{c}_*^\top)$, and the optimal $\boldsymbol{\lambda}$ in (14) and (15) are both given by $\mathbf{x}(\mathbf{A}+\mathbf{b}\mathbf{c}_*^\top)\circ\mathbf{y}(\mathbf{A}+\mathbf{b}\mathbf{c}_*^\top)$.

Furthermore, $\mathbf{p}=\mathbf{x}(\mathbf{A}+\mathbf{b}\mathbf{c}_*^\top)$ is the dual of $\mathbf{c}_*$ with respect to $\|\cdot\|_D$.[2]

Remark 3: Lemma 2 is a general version of the Friedland-Karlin spectral radius minimax characterization in [20], [21]. In particular, if $\mathbf{b}=\mathbf{0}$, we obtain Theorem 3.2 in [21].

Using Lemma 2 (let $\mathbf{A}=\mathrm{diag}(\boldsymbol{\beta})\mathbf{F}$, $\mathbf{b}=(1/\bar{P})\mathrm{diag}(\boldsymbol{\beta})\mathbf{v}$ and $\mathbf{c}_*=\mathbf{1}$ in Lemma 2), we deduce that the optimal SIR allocation in (8) is a weighted geometric mean of the optimal SIR, where the weights are the normalized[3] Schur product of the uplink power and the downlink power (Perron and left eigenvectors of $\mathrm{diag}(\boldsymbol{\beta})\mathbf{B}$, respectively):

$$\prod_l (\mathsf{SIR}_l(\mathbf{p})/\beta_l)^{p_l q_l/\beta_l} = 1/\rho(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B}). \qquad (16)$$

Further, both the optimal uplink and the downlink power form dual pairs with the vector $(1/\bar{P})\mathbf{1}$, i.e., $(\mathbf{p},(1/\bar{P})\mathbf{1})$ and $(\mathbf{q},(1/\bar{P})\mathbf{1})$ are dual pairs with respect to $\|\cdot\|_1$.

C. Uplink-downlink duality and joint optimization

Next, we consider the joint optimization of power and transmit beamformer in the following max-min weighted SIR problem:

$$\begin{array}{ll}\text{maximize} & \min_l \dfrac{\mathsf{SIR}_l(\mathbf{p},\mathbf{U})}{\beta_l} \\ \text{subject to} & \sum_{l=1}^L p_l \leq \bar{P},\ \ p_l\geq 0,\quad \mathbf{u}_l^\dagger\mathbf{u}_l=1\ \ \forall\, l, \\ \text{variables:} & \mathbf{U}=[\mathbf{u}_1\ldots\mathbf{u}_L],\quad \mathbf{p}.\end{array} \qquad (17)$$

We first review the notion of uplink-downlink duality. The duality theory states that, under a same total power constraint and additive white noise for all users, the achievable SIR region for a downlink

[2]A pair $(\mathbf{x},\mathbf{y})$ of vectors of $\mathbb{R}^L$ is said to be a dual pair with respect to $\|\cdot\|$ if $\|\mathbf{y}\|_D\|\mathbf{x}\|=1$.

[3]The normalization of the Schur product is such that $\sum_l p_l q_l/\beta_l = 1$.

transmission with joint transmit beamforming and power control optimization is equivalent to that of a *reciprocal* uplink transmission with joint receive beamforming and power control optimization. Further, the optimal *receive* beamforming vectors in the uplink is also the optimal *transmit* beamforming vectors in the downlink. Since joint power control and beamforming optimization in the uplink does not have the beamformer coupling difficulty associated with the downlink (hence easier to solve), the (dual) uplink problem can be first used to obtain the optimal transmit beamformers in the downlink. The optimal downlink transmit power is then computed by keeping the transmit beamformers fixed. In the case where the noise is different for each user, a virtual uplink transmission (assuming that all users have the same noise, i.e., $n_l = 1$ for all l) is constructed as an intermediary step to compute the optimal transmit beamforming vector.

Let the virtual uplink power be given by $\mathbf{q}$. Now, suppose there exists positive values γ (optimal max-min weighted SIR) and q_l for all l such that the virtual uplink $\hat{\mathsf{SIR}}_l$ satisfies

$$\hat{\mathsf{SIR}}_l(\mathbf{q}, \mathbf{U}) = \frac{q_l |\boldsymbol{h}_l^\dagger \mathbf{u}_l|^2}{\sum_{j \neq l} q_j |\boldsymbol{h}_j^\dagger \mathbf{u}_l|^2 + 1} \geq \beta_l \gamma \tag{18}$$

for all l. Since $\hat{\mathsf{SIR}}_l$ in (18) only depends on the beamforming vector $\mathbf{u}_l$, the receive beamforming optimization, with the power fixed at $\mathbf{q}$, is solved by

$$\mathbf{u}_l^* = \arg \min_{\mathbf{u}_l} \quad \sum_{j \neq l} \frac{G_{jl}(\mathbf{U})}{G_{ll}(\mathbf{U})} q_j + \frac{1}{G_{ll}(\mathbf{U})}, \tag{19}$$

whose solution is the LMMSE receiver given by (optimal up to a scaling factor):

$$\mathbf{u}_l^* = \left(\sum_{j \neq l} q_j \mathbf{h}_j \mathbf{h}_j^\dagger + \mathbf{I} \right)^+ \mathbf{h}_l, \tag{20}$$

where $(\cdot)^+$ denotes pseudo-inversion. Using this LMMSE receiver, the SIR constraint in (18) is always met with equality, i.e., $\hat{\mathsf{SIR}}_l(\mathbf{q}, \mathbf{U}) = \beta_l \gamma$. By the uplink-downlink duality, the LMMSE receiver is also the optimal transmit beamformer in the downlink max-min weighted SIR problem given by (17).

Now, we are ready to use Algorithm 1 to solve the joint power control and beamforming problem in (17) in a fast and distributed fashion. The following algorithm computes the optimal power and transmit beamformer in (17):

Algorithm 2 (Max-min Weighted SIR*–Joint Optimization):*

1) Update (virtual) uplink power $\mathbf{q}(k+1)$:

$$q_l(k+1) = \left(\frac{\beta_l}{\hat{\mathsf{SIR}}_l(\mathbf{q}(k), \mathbf{U}(k))} \right) q_l(k) \quad \forall\, l. \tag{21}$$

2) Normalize $\mathbf{q}(k+1)$:

$$\mathbf{q}(k+1) \leftarrow \mathbf{q}(k+1) \cdot \bar{P} / \mathbf{1}^\top \mathbf{q}(k+1). \tag{22}$$

3) Update transmit beamforming matrix $\mathbf{U}(k+1) = [\mathbf{u}_1(k+1) \ldots \mathbf{u}_L(k+1)]$:

$$\begin{aligned} \mathbf{u}_l(k+1) &= \left(\textstyle\sum_{j \neq l} q_j(k+1) \mathbf{h}_j \mathbf{h}_j^\dagger + \mathbf{I} \right)^+ \mathbf{h}_l \quad \forall\, l, \\ \mathbf{u}_l(k+1) &\leftarrow \mathbf{u}_l(k+1) / \|\mathbf{u}_l(k+1)\|_2 \quad \forall\, l. \end{aligned} \tag{23}$$

4) Update downlink power $\mathbf{p}(k+1)$:

$$p_l(k+1) = \left(\frac{\beta_l}{\mathsf{SIR}_l(\mathbf{p}(k), \mathbf{U}(k+1))} \right) p_l(k) \quad \forall\, l. \tag{24}$$

Uplink-downlink Duality Correspondence

Downlink		Uplink
$\mathbf{p} = \mathbf{x}(\text{diag}(\boldsymbol{\beta})\mathbf{B})$	$\leftrightarrow$	$\mathbf{q} = \mathbf{x}(\text{diag}(\boldsymbol{\beta})\mathbf{B}^{\top})$ $= \text{diag}(\boldsymbol{\beta})\mathbf{y}(\text{diag}(\boldsymbol{\beta})\mathbf{B})$
$\frac{\mathsf{SIR}_l(\mathbf{p})}{\beta_l} = \frac{1}{\rho(\text{diag}(\boldsymbol{\beta})\mathbf{B})}$	$\leftrightarrow$	$\frac{\hat{\mathsf{SIR}}_l(\mathbf{q})}{\beta_l} = \frac{1}{\rho(\text{diag}(\boldsymbol{\beta})\mathbf{B}^{\top})}$
$\boldsymbol{\lambda} = \mathbf{p} \circ \text{diag}(\boldsymbol{\beta})^{-1}\mathbf{q}$	$\leftrightarrow$	$\boldsymbol{\lambda} = \mathbf{q} \circ \text{diag}(\boldsymbol{\beta})^{-1}\mathbf{p}$
$(\mathbf{p}, \mathbf{c}_* = (1/\bar{P})\mathbf{1})$	$\leftrightarrow$	$(\mathbf{q}, \mathbf{c}_* = (1/\bar{P})\mathbf{1})$
LMMSE Transmit	$\leftrightarrow$	LMMSE Receive

Fig. 2. The uplink-downlink duality characterized through the nonlinear Perron-Frobenius theorem and the Friedland-Karlin spectral radius minimax theorem. The equality notation used in the equations denotes equality up to a scaling constant.

5) Normalize $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P}/\mathbf{1}^{\top}\mathbf{p}(k+1). \tag{25}$$

Theorem 1: Let the optimal power and beamforming matrix in (17) be $\mathbf{p}^{\star}$ and $\mathbf{U}^{\star}$ respectively. Then, starting from any initial point $\mathbf{q}(0)$ and $\mathbf{p}(0)$, $\mathbf{p}(k)$ in Algorithm 2 converges geometrically fast to $\mathbf{p}^{\star} = \mathbf{x}(\mathbf{B}(\mathbf{U}^{\star}))$ (unique up to a scaling constant).

Remark 4: In Algorithm 2, the uplink power $\mathbf{q}(k)$ converges geometrically fast to $(\bar{P}/\mathbf{1}^{\top}\mathbf{x}(\text{diag}(\boldsymbol{\beta})\mathbf{B}^{\top}))\mathbf{x}(\text{diag}(\boldsymbol{\beta})$ It is equal to the left eigenvector of $\text{diag}(\boldsymbol{\beta})\mathbf{B}$ if $\boldsymbol{\beta} = \mathbf{1}$.

Remark 5: Note that (21) and (24) of Algorithm 2 use the DPC algorithm in [18], where the lth user has a virtual SIR threshold of β_l in both the (virtual) uplink and downlink transmission. In the case where n_l's are equal for all l, $\mathbf{q}$ in (21) is the exact uplink transmit power, and only computing $\mathbf{1}^{\top}\mathbf{q}(k+1)$ in (22) requires a global coordination at the base station. Compared to previous centralized solution in [8], [9], our solution has less complexity and provable geometric convergence rate. In principle, $\mathbf{1}^{\top}\mathbf{q}(k+1)$ at Step 2 and $\mathbf{1}^{\top}\mathbf{p}(k+1)$ at Step 5 can also be computed in a distributed manner using gossip algorithms [19].

Finally, we summarize this interesting connection between the uplink-downlink duality with the nonlinear Perron-Frobenius theorem and the Friedland-Karlin spectral radius minimax characterization from Section III-B in Figure 2.

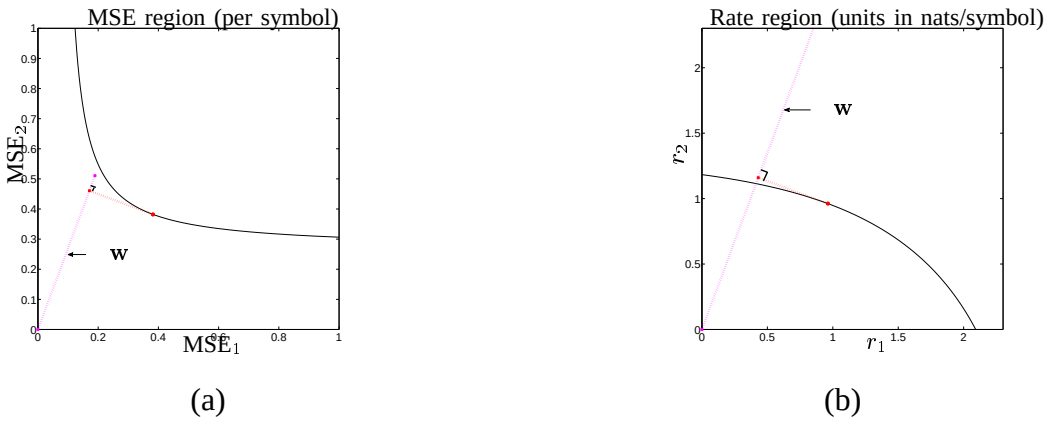
IV. Weighted Sum MSE Minimization

In this section, we state the optimization problem of minimizing the weighted sum of the MSE's of individual data streams under a sum power constraint. This problem is nonconvex, and we will approach this problem by two methods. In our first method (Section IV-B), we will show that it can be solved exactly as special cases when the problem parameters satisfy certain conditions (which will be associated with the interference and SNR level). In our second method (Section VI), for the general case, we will deduce optimality conditions and propose suboptimal algorithm that exploits the connection with max-min weighted SIR (and its associated fast algorithms in the previous section) to solve it.

A. Problem statement

We assume that all the receivers use the LMMSE filter for estimating the received symbols of all users. The weighted sum MSE at the output of the LMMSE receiver is given by [15]:

$$\sum_{l=1}^{L} w_l \text{MSE}_l(\mathbf{p}) = \sum_{l=1}^{L} w_l \frac{1}{1+\mathsf{SIR}_l(\mathbf{p})}, \tag{26}$$



(a) (b)

Fig. 3. (a) A geometrical illustration of the weighted sum MSE minimization problem. The probability vector $\mathbf{w}$ is superimposed on the MSE achievable region, and its perpendicular coincides with the optimal point in the MSE achievable region. (b) A geometrical illustration of the weighted sum rate maximization problem. The probability vector $\mathbf{w}$ is superimposed on the MSE achievable region, and its perpendicular coincides with the optimal point in the rate achievable region.

where w_l is some positive weight assigned by the network to the lth link (to reflect some long-term priority). Without loss of generality, we assume that $\mathbf{w}$ is a probability vector. The weighted sum MSE minimization problem is given by

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^{L} w_l \frac{1}{1+\mathsf{SIR}_l(\mathbf{p})} \\ \text{subject to} & \sum_{l=1}^{L} p_l \leq \bar{P}, \;\; p_l \geq 0 \;\; \forall\, l, \\ \text{variables:} & p_l \;\; \forall\, l. \end{array} \tag{27}$$

We denote the optimal power vector to (27) by $\mathbf{p}^\star$. By expressing the solution in the MSE variable, a geometrical illustration of (27) is given in Figure 3(a).

B. Exact solution to special cases

We can rewrite (27) as

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^{L} w_l \frac{\sum_{k\neq l} G_{lk}p_k + n_l}{\sum_{k} G_{lk}p_k + n_l} \\ \text{subject to} & \sum_{l=1}^{L} p_l \leq \bar{P}, \;\; p_l \geq 0 \;\; \forall\, l, \\ \text{variables:} & p_l \;\; \forall\, l. \end{array} \tag{28}$$

It can be shown that the total power constraint in (27) and (28) are tight at optimality (see Appendix IX-E), which we exploit to transform (28) in the variables $\mathbf{p}$ into another optimization problem that can be used to solve (28) optimally. To proceed further, we need to introduce the notion of quasi-invertibility of a nonnegative matrix in [23], which will be useful in solving (28) optimally.

Definition 1 (Quasi-invertibility): A square nonnegative matrix $\mathbf{B}$ is a quasi-inverse of a square nonnegative matrix $\tilde{\mathbf{B}}$ if $\mathbf{B} - \tilde{\mathbf{B}} = \mathbf{B}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}\mathbf{B}$. Furthermore, $(\mathbf{I} - \tilde{\mathbf{B}})^{-1} = \mathbf{I} + \mathbf{B}$ [23].

We will now apply the definition of quasi-invertibility to $\mathbf{B}$ in (10), and study the existence of $\tilde{\mathbf{B}}$, which can interestingly be associated with the SNR regime. In the case where the total power is very large, i.e., $\bar{P} \to \infty$ (high SNR regime) or when interference (off-diagonals of $\mathbf{F}$) is very large, it is deduced in the following that $\tilde{\mathbf{B}}$ does not exist.

Lemma 3: $\tilde{\mathbf{B}}$ does not exist when $\mathbf{B} = \mathbf{F}$, where $F_{lj} > 0$ for all l, j and $l \neq j$.

However, when $\mathbf{F} = \mathbf{0}$ (no interference) such that $\mathbf{B} = \mathbf{v}\mathbf{1}^\top/\bar{P}$ or when $\bar{P}$ is sufficiently small (low SNR regime) such that $\mathbf{B} \approx \mathbf{v}\mathbf{1}^\top/\bar{P}$, then $\tilde{\mathbf{B}}$ always exists, as shown by the following lemma.

Lemma 4: For any nonnegative vector $\mathbf{v}$, $\tilde{\mathbf{B}} = 1/(1 + \mathbf{1}^\top\mathbf{v})\mathbf{v}\mathbf{1}^\top$ when $\mathbf{B} = \mathbf{v}\mathbf{1}^\top$.

Example 1: In a numerical example for a ten-user IEEE 802.11b network, we experiment with a total power constraint of 33mW and 1W (the largest possible value allowed in IEEE 802.11b). Averaging over

$10,000$ random channel coefficient instances, the percentage of instances where $\tilde{\mathbf{B}}$ exists is 99% and 65% corresponding to the total power constraint of 33mW and 1W, respectively.

In the rest of Section IV-B, we focus on the case when $\tilde{\mathbf{B}}$ exists. We next solve (28) in the following. Let us define

$$\mathbf{z} = (\mathbf{I} + \mathbf{B})\mathbf{p}. \tag{29}$$

Note that $G_{ll}z_l$ is the total received (desired and interfering) signal power plus the additive white noise at the lth receiver.

Then, we can rewrite (28) in terms of $\mathbf{z}$ as

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^{L} w_l \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} \\ \text{subject to} & z_l \geq (\tilde{\mathbf{B}}\mathbf{z})_l, \quad l = 1, \ldots, L, \\ \text{variables:} & z_l \quad \forall\, l, \end{array} \tag{30}$$

where the constraints in (30) are due to the nonnegativity of $\mathbf{p}$, since, using Definition 1, $\mathbf{p} = (\mathbf{I}+\mathbf{B})^{-1}\mathbf{z} = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z} \geq \mathbf{0}$.

The following result provides a condition under which the optimal solution to (30), $\mathbf{z}^\star$, can be transformed to yield the optimal solution to (28) or equivalently (27).

Theorem 2: The optimal solution to (27) is given by $\mathbf{p}^\star = (\mathbf{I}+\mathbf{B})^{-1}\mathbf{z}^\star$, where $\mathbf{z}^\star$ is the optimal solution to (30), if $\mathbf{B}$ is the quasi-inverse of a nonnegative matrix $\tilde{\mathbf{B}}$, where $\rho(\tilde{\mathbf{B}}) < 1$.

Lemma 5: If $\tilde{\mathbf{B}}$ exists, then $\tilde{\mathbf{B}}$ has the spectral radius

$$\rho(\tilde{\mathbf{B}}) = \frac{\rho(\mathbf{B})}{1 + \rho(\mathbf{B})}, \tag{31}$$

with the corresponding left and Perron eigenvectors of $\mathbf{B}$.

In the following, we derive useful lower bounds to (27), investigate special cases, and finally characterize the exact solution to (27).

C. A lower bound to weighted sum MSE minimization

By exploiting the eigenspace of $\tilde{\mathbf{B}}$ and the Friedland-Karlin inequality, the following result gives a lower bound on the weighted sum MSE problem in (27).

Theorem 3: If $\tilde{\mathbf{B}}$ exists,

$$\sum_{l=1}^{L} w_l \frac{1}{1 + \mathsf{SIR}_l(\mathbf{p})} \geq \left(\frac{1}{1 + 1/\rho(\mathbf{B})}\right)^{\|\mathbf{w}\|_\infty^{\mathbf{x}(\mathbf{B})\circ\mathbf{y}(\mathbf{B})}} \tag{32}$$

for all feasible $\mathbf{p}$ in (27).

Equality is achieved if and only if $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$. Thus, $\mathsf{SIR}_l(\mathbf{p}^\star) = (1/\rho(\mathbf{B}))$ for all l. In this case, $\mathbf{p}^\star = \mathbf{x}(\mathbf{B})$ solves (27).

Interestingly, Theorem 3 shows that solving (8) with $\boldsymbol{\beta} = \mathbf{1}$ can be seen as an approximation method to solving (27) suboptimally, but with an approximation guarantee. In particular, by taking the logarithm of the objective function of (27), $\|\mathbf{w}\|_\infty^{\mathbf{x}(\mathbf{B})\circ\mathbf{y}(\tilde{\mathbf{B}})}$ can be interpreted as the approximation ratio of Algorithm 1 with $\boldsymbol{\beta} = \mathbf{1}$ in solving (27).

Remark 6: Based on Theorem 3, we obtain a connection between the min-max MSE problem ((9) with $\boldsymbol{\beta} = \mathbf{1}$) and the weighted sum MSE optimization in (27). Suppose we consider min-max MSE:

$$\min_{\mathbf{p}} \max_{l} 1/(1 + \mathsf{SIR}_l(\mathbf{p})). \tag{33}$$

Then, by the max-min characterization of $\rho(\tilde{\mathbf{B}})$:[4]

$$\max_{\mathbf{z}>\mathbf{0}} \min_{l} \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} = \min_{\mathbf{z}>\mathbf{0}} \max_{l} \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} = \rho(\tilde{\mathbf{B}}), \tag{34}$$

the optimal objective of (33) is simply $\rho(\tilde{\mathbf{B}}) = 1/(1 + 1/\rho(\mathbf{B}))$. It immediately follows from Theorem 3 that (8) with $\boldsymbol{\beta} = \mathbf{1}$ yields the equivalent power allocation as (33) and the optimal sum MSE with $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$.

We will establish further connection and equivalence results between the max-min weighted SIR problem and the two nonconvex problems (27) and (39) in Section V-C (still requiring the existence of $\tilde{\mathbf{B}}$) and in Section VI-A for the general case (without requiring the existence of $\tilde{\mathbf{B}}$).

D. Exact solution and fast algorithm

The existence of $\tilde{\mathbf{B}}$ allows us to delineate cases of (27) that can be solved optimally from the general problem. The following result gives the exact closed-form solution to (27), which motivates a fast algorithm (Algorithm 3 below) to compute the optimal solution.

Theorem 4: If $\tilde{\mathbf{B}}$ exists, then the optimal solution to (27) is given by $\mathbf{p}^\star = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^\star \geq \mathbf{0}$, where $\mathbf{z}^\star$ is given by

$$z_l^\star = \sqrt{\frac{w_l \sum_{j\neq l} \tilde{B}_{lj} z_j^\star}{\sum_{j\neq l} w_j \tilde{B}_{jl}/z_j^\star}} \tag{35}$$

for all l and satisfies $\mathbf{1}^\top \mathbf{z}^\star - \mathbf{1}^\top \tilde{\mathbf{B}}\mathbf{z}^\star = \bar{P}$.

Example 2: The solution for the two-user case assumes a simple form, because Theorem 4 simplifies to: If $F_{12}F_{21}\bar{P} + F_{12}v_2 + F_{21}v_1 \leq \min\{v_1, v_2\}$, then $\mathbf{p}^\star = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^\star$, where $z_1^\star = \sqrt{w_1\tilde{B}_{12}/(w_2\tilde{B}_{21})}z_2^\star$ and $z_2^\star = \bar{P}/(\sqrt{w_1\tilde{B}_{12}/(w_2\tilde{B}_{21})}(1 - \tilde{B}_{11} - \tilde{B}_{21}) + 1 - \tilde{B}_{12} - \tilde{B}_{22})$.

Now, (35) in Theorem 4 can be written in the form of $\mathbf{z} = I(\mathbf{z})$, where I is a homogeneous function. We will leverage the standard interference function results in [13] to propose the following (step size free) algorithm that computes $\mathbf{z}^\star$ in Theorem 4, and implicitly, the optimal transmit power of (27).

Algorithm 3 (Weighted Sum MSE Minimization):

1) Initialize an arbitrarily small $\epsilon > 0$.
2) Update auxiliary variable $\mathbf{z}(k+1)$:

$$z_l(k+1) = \sqrt{\frac{w_l \sum_{j\neq l} \tilde{B}_{lj} z_j(k)}{\sum_{j\neq l} w_j \tilde{B}_{jl}/z_j(k)}} + \epsilon \quad \forall\, l. \tag{36}$$

3) Update $\mathbf{p}(k+1)$:

$$p_l(k+1) = \frac{\mathsf{SIR}_l(\mathbf{p}(k))}{1 + \mathsf{SIR}_l(\mathbf{p}(k))} z_l(k+1) \quad \forall\, l. \tag{37}$$

4) Normalize $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P}/(\mathbf{1}^\top \mathbf{p}(k+1)). \tag{38}$$

The following theorem shows that Algorithm 3 converges to the optimal solution $\mathbf{z}^\star$ and $\mathbf{p}^\star$ in Theorem 4.

[4]The max-min characterization of the spectral radius of an irreducible nonnegative matrix $\mathbf{B}$ is also known as the Collatz-Wielandt characterization in nonnegative matrix theory, e.g., see [20], [24].

Theorem 5: If $\tilde{\mathbf{B}}$ exists, for arbitrarily small $\epsilon > 0$, Algorithm 3 converges to the unique fixed point $\mathbf{z}^\star$ and $\mathbf{p}^\star$ in Theorem 4 from any initial point $\mathbf{z}(0)$ under asynchronous update.

Remark 7: Algorithm 3 requires a complexity of O(L^3) to compute $\tilde{\mathbf{B}}$. Step 2 of Algorithm 3 can be implemented by distributed message passing. Transforming from $\mathbf{z}(k+1)$ to $\mathbf{p}(k+1)$ in (37) is performed locally by each user, and the normalization at Step 38 is performed at the base station.

Remark 8: The convergence of Algorithm 3 leverages the standard interference function approach in [13] by introducing ϵ in (36) of Algorithm 3.

V. Weighted Sum Rate Maximization

In this section, we consider the weighted sum rate of all users as a performance metric to be optimized. Similar to Section IV, we will first show that it can be solved exactly as special cases when the problem parameters satisfy certain conditions, and then look at the general case in Section VI. Further, we quantify the connection of the weighted sum rate maximization and the weighted sum MSE minimization.

A. Problem statement

The weighted sum rate maximization problem is given by

$$\begin{array}{ll} \text{maximize} & \sum_{l=1}^{L} w_l \log(1+\mathsf{SIR}_l(\mathbf{p})) \\ \text{subject to} & \sum_{l=1}^{L} p_l \le \bar{P}, \;\; p_l \ge 0 \;\; \forall\, l, \\ \text{variables:} & p_l \;\; \forall\, l. \end{array} \tag{39}$$

By expressing the solution in the rate variable, a geometrical illustration of (39) is given in Figure 3(b).

B. Exact solution and fast algorithm

We can rewrite (39) to be equivalent to

$$\begin{array}{ll} \text{minimize} & \prod_{l=1}^{L} \left(\frac{\sum_{k\neq l} G_{lk} p_k + 1}{\sum_{k} G_{lk} p_k + 1} \right)^{w_l} \\ \text{subject to} & \sum_{l=1}^{L} p_l \le \bar{P}, \;\; p_l \ge 0 \;\; \forall\, l, \\ \text{variables:} & p_l \;\; \forall\, l. \end{array} \tag{40}$$

Similar to Section IV, if $\mathbf{B}$ is the quasi-inverse of $\tilde{\mathbf{B}}$, we can rewrite (40) as

$$\begin{array}{ll} \text{minimize} & \prod_{l=1}^{L} \left(\frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} \right)^{w_l} \\ \text{subject to} & z_l \ge (\tilde{\mathbf{B}}\mathbf{z})_l, \;\; l = 1, \ldots, L, \\ \text{variables:} & z_l \;\; \forall\, l, \end{array} \tag{41}$$

where $\mathbf{z}$ is given by (29).

Similar to Theorem 2, the following Theorem 6 gives the condition under which (39) is solved optimally.

Theorem 6: If $\tilde{\mathbf{B}}$ exists, the optimal solution to (39) is given by $\mathbf{p}^\star = (\mathbf{I}+\mathbf{B})^{-1}\mathbf{z}^\star$, where $\mathbf{z}^\star$ is the optimal solution of (41).

Theorem 6 is used to give the following solution of (39).

Theorem 7: If $\tilde{\mathbf{B}}$ exists, then the optimal solution to (39) is given by $\mathbf{p}^\star = (\mathbf{I}-\tilde{\mathbf{B}})\mathbf{z}^\star \ge \mathbf{0}$, where $\mathbf{z}^\star$ is given by

$$z_l^\star = \frac{w_l}{\sum_j w_j \tilde{B}_{jl} / (\tilde{\mathbf{B}}\mathbf{z}^\star)_j} \tag{42}$$

for all l and satisfies $\mathbf{1}^\top \mathbf{z}^\star - \mathbf{1}^\top \tilde{\mathbf{B}}\mathbf{z}^\star = \bar{P}$.

Example 3: The solution for the two-user case assumes a simple form, because Theorem 7 simplifies to: If $F_{12}F_{21}\bar{P} + F_{12}v_2 + F_{21}v_1 \le \min\{v_1, v_2\}$, then $\mathbf{p}^\star = (\mathbf{I}-\tilde{\mathbf{B}})\mathbf{z}^\star$, where $z_1^\star$ and $z_2^\star$ are the solutions to $w_1\tilde{B}_{12}\tilde{B}_{22}(z_2^\star)^2 + (w_1-w_2)\tilde{B}_{12}\tilde{B}_{21}z_1^\star z_2^\star - w_2\tilde{B}_{11}\tilde{B}_{21}(z_1^\star)^2 = 0$ and $z_2^\star = (\bar{P} - (1-\tilde{B}_{11}-\tilde{B}_{21})z_1^\star)/(1-$

$\tilde{B}_{12} - \tilde{B}_{22}$). As a side remark, we note the close similarity in our sufficient condition (existence of $\tilde{\mathbf{B}}$ for two users) obtained here and the information-theoretic sufficient condition of sum-capacity optimality established in [11]. More precisely, our sufficient condition contains the one obtained in [11]. Could a fundamental relatiohsip be found theoretically? We leave this for a future work.

As in the previous, (42) in Theorem 7 can be expressed as $\mathbf{z} = I(\mathbf{z})$, where I is a homogeneous function. Using the standard interference function approach in [13], the following algorithm computes the optimal solution of (39).

Algorithm 4 (Weighted Sum Rate Maximization):

1) Initialize an arbitrarily small $\epsilon > 0$.
2) Update auxiliary variable $\mathbf{z}(k+1)$:

$$z_l(k+1) = \frac{w_l}{\sum_j w_j \tilde{B}_{jl}/(\tilde{\mathbf{B}}\mathbf{z}(k))_j} + \epsilon \quad \forall\, l. \tag{43}$$

3) Update $\mathbf{p}(k+1)$:

$$p_l(k+1) = \frac{\mathsf{SIR}_l(\mathbf{p}(k))}{1+\mathsf{SIR}_l(\mathbf{p}(k))} z_l(k+1) \quad \forall\, l. \tag{44}$$

4) Normalize $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P}/(\mathbf{1}^\top \mathbf{p}(k+1)). \tag{45}$$

The following result shows that Algorithm 4 converges to the optimal solution $\mathbf{z}^\star$ and $\mathbf{p}^\star$ in Theorem 7.

Theorem 8: If $\tilde{\mathbf{B}}$ exists, for arbitrarily small $\epsilon > 0$, Algorithm 4 converges to the unique fixed point $\mathbf{z}^\star$ and $\mathbf{p}^\star$ in Theorem 7 from any initial point $\mathbf{z}(0)$ under asynchronous update.

Next, we connect the three optimization problems given in (27), (39) and (8) with $\boldsymbol{\beta} = \mathbf{1}$.

C. Connection between weighted sum rate, sum MSE and max-min SIR

Applying the arithmetic-geometric mean inequality (see (82) in Appendix) to connect (30) and (41), we deduce that the weighted sum rate maximization has the same optimal power as the weighted sum MSE minimization when $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$. Furthermore, from Remark 6, this optimal power is also the solution to the max-min weighted SIR problem (with $\boldsymbol{\beta} = \mathbf{1}$).

Example 4: We give an 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 total power is $2W$. It can be easily checked that $\tilde{\mathbf{B}}$ exists. We solve (8) with $\boldsymbol{\beta} = \mathbf{1}$. We then set $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$ in both (27) and (39), and find their corresponding optimal solution by exhaustive search. These are then plotted on the respective achievable rate and MSE region. Figure 5 shows that the optimal weighted sum rate point is the same as the weighted sum rate point evaluated using the max-min SIR power control. Figure 4 shows that the optimal weighted sum MSE point is the same as the weighted sum MSE point evaluated using the max-min SIR power control.

In Section VI-B, we continue to highlight this connection between (8) with $\boldsymbol{\beta} = \mathbf{1}$ and the two problems (27) and (39) with $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$ in the general case, i.e., without the $\tilde{\mathbf{B}}$ existence condition (see Corollary 2).

VI. Global Optimization in SIR

In this section, we first study the optimality conditions of (27) and (39) in the general case. We then solve the problems using a centralized algorithm that can be made distributed using the gradient method and Algorithm 2. Conditions under which these algorithms solve them optimally will also be given.

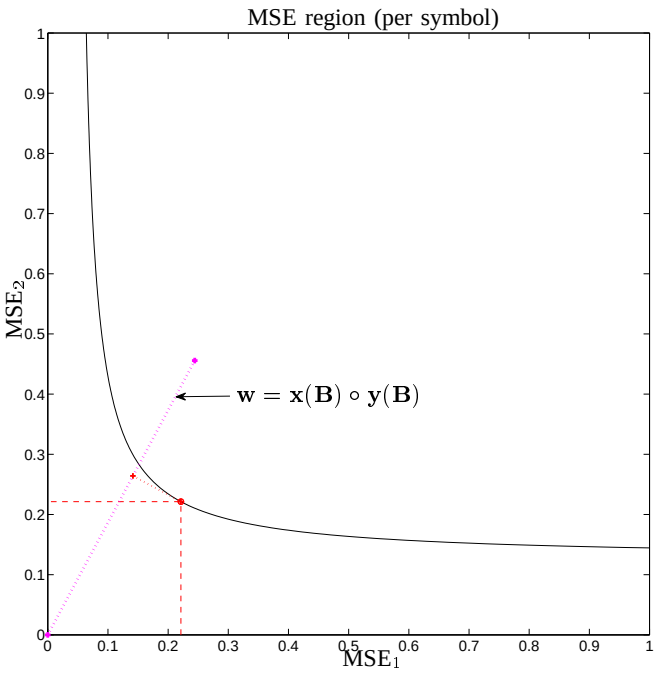


Fig. 4. Achievable MSE region for a 2-user interference channel. From a geometrical perspective, the weighted sum MSE point evaluated at the solution of the max-min SIR power control finds the largest hypercube that supports the pareto boundary of the achievable MSE region.

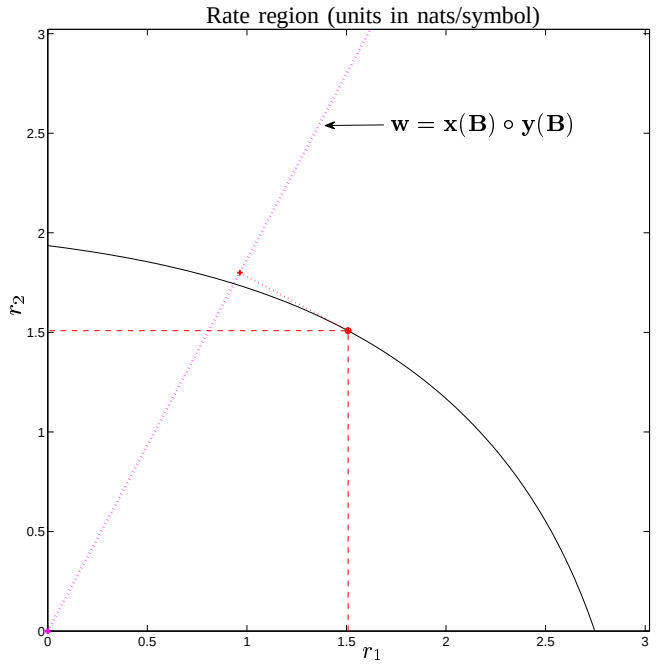


Fig. 5. Achievable rate region for a 2-user interference channel. From a geometrical perspective, the weighted sum rate point evaluated at the solution of the max-min SIR power control finds the largest hypercube that is contained inside the achievable rate region.

A. Optimality conditions in SIR

First, we reformulate both (39) and (27) as optimization problems having a new set of variables (in the SIR domain) and a spectral radius constraint involving $\mathbf{B}$ in (10). The new formulation permits us to derive optimality conditions, propose fast algorithms and further connect the max-min SIR problem to (27) and (39). In particular, by leveraging the beamforming result in Section III-C, we also address the optimal beamformer to (27) and (39).

First, let us consider the following optimization problem:

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{\gamma}) \\ \text{subject to} & \rho(\text{diag}(\boldsymbol{\gamma})\mathbf{B}) \leq 1, \\ \text{variables:} & \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \quad \boldsymbol{\gamma}. \end{array} \tag{46}$$

Let us denote the optimal solution of (51) by $\boldsymbol{\gamma}^\star$. Now, the optimal power to the weighted MSE minimization problem in (27) and the weighted sum rate maximization problem in (39) can be implicitly obtained by solving (46) using, respectively,

$$f(\boldsymbol{\gamma}) = -\sum_l w_l/(1+\gamma_l) \qquad \text{(weighted sum MSE)} \tag{47}$$

and

$$f(\boldsymbol{\gamma}) = \sum_l w_l \log(1+\gamma_l) \qquad \text{(weighted sum rate)} \tag{48}$$

in (46). We summarize this in the following result.

Lemma 6: For a feasible $\boldsymbol{\gamma}$ in (46), a feasible power in (27) and in (39) can be computed by

$$\mathbf{p} = (\mathbf{I} - \text{diag}(\boldsymbol{\gamma})\mathbf{F})^{-1}\text{diag}(\boldsymbol{\gamma})\mathbf{v}. \tag{49}$$

This means that if we first solve (46) to obtain $\boldsymbol{\gamma}^\star$, the optimal power in (27) and (39) can be obtained using (49) in Lemma 6.

We continue with a further change of variable technique in (46). For $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_L)^\top > 0$, let

$$\tilde{\gamma}_l = \log \gamma_l \quad \text{for all } \ l, \tag{50}$$

i.e., $\boldsymbol{\gamma} = e^{\tilde{\boldsymbol{\gamma}}}$. Then, (46) is equivalent to:

$$\begin{array}{ll} \text{maximize} & f(e^{\tilde{\boldsymbol{\gamma}}}) \\ \text{subject to} & \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \leq 0, \\ \text{variables:} & \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \quad \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_n)^\top \in \mathbb{R}^L. \end{array} \tag{51}$$

Note that the constraint set in (51) is an unbounded convex set. Let us denote the optimal solution of (51) by $\tilde{\boldsymbol{\gamma}}^\star$. Note that $\tilde{\gamma}_l^\star = \log \gamma_l^\star$ for all l.

Suppose we know the optimal $\mathbf{U}$ in (51). Define the first order derivative of the objective function in (51) with respect to $\tilde{\boldsymbol{\gamma}}$ by $f'(e^{\tilde{\boldsymbol{\gamma}}}) \in \mathbb{R}^L$. The following result characterizes the optimality condition of (51), and also applies to (46) after using the mapping in (50).

Theorem 9: The optimal solution of (51), $\tilde{\boldsymbol{\gamma}}^\star$, satisfies

$$\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) = \frac{f'(e^{\tilde{\boldsymbol{\gamma}}^\star})}{\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}^\star})} \tag{52}$$

and

$$\rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) = 1. \tag{53}$$

Furthermore, $\tilde{\boldsymbol{\gamma}}^\star$ is unique.

Our results in Sections IV and V established the connection between these three problems only in the low interference or low signal-to-noise regime. Now, we establish the connection for the general case.

From (53) in Theorem 9, we see that if the optimal SIR_l is equal for all l, then the optimal solution to (51) can be obtained by solving the max-min SIR problem in (17).

Now, by extending this observation to the constraint set in (46), we see that the uplink-downlink duality applies to the weighted sum rate maximization and weighted sum MSE minimization problems. In particular, the optimal $\mathbf{U}$ in (46) is given by the LMMSE filter in (20), where

$$\mathbf{q} = (\mathbf{I} - \mathrm{diag}(\boldsymbol{\gamma}^\star)\mathbf{F}^\top)^{-1}\mathrm{diag}(\boldsymbol{\gamma}^\star)\mathbf{v} \tag{54}$$

in (20).

Further, using (14) in Lemma 2, we see that a simple proof to the uplink-downlink duality is to observe that

$$\rho(\mathrm{diag}(\boldsymbol{\gamma})(\mathbf{F} + (1/\bar{P})\mathbf{1}\mathbf{1}^\top)) = \rho(\mathrm{diag}(\boldsymbol{\gamma})(\mathbf{F}^\top + (1/\bar{P})\mathbf{1}\mathbf{1}^\top)) = 1, \tag{55}$$

where we have used both the fact that $\rho(\mathbf{AC}) = \rho(\mathbf{CA})$ and $\rho(\mathbf{A}) = \rho(\mathbf{A}^\top)$ for any irreducible nonnegative matrices $\mathbf{A}$ and $\mathbf{C}$ [24]. The first and second spectral radius expression in (55) correspond to the uplink and the downlink systems, respectively. Note that (55) characterizes all *pareto efficient* achievable SIR in both the uplink and downlink systems (since the spectral radius increases monotonically in the entry of a matrix [24]).

Interestingly, using the weighted geometric mean equivalence result in (16), (17) is equivalent to an optimization problem given in the form of (46) by letting

$$f(\boldsymbol{\gamma}) = \textstyle\sum_l x_l(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B})y_l(\mathrm{diag}(\boldsymbol{\beta})\mathbf{B})\log\gamma_l, \quad \begin{array}{l}\text{(max-min}\\ \text{weighted } \mathsf{SIR})\end{array} \tag{56}$$

and its corresponding transformed problem in (51), which is convex in $\tilde{\boldsymbol{\gamma}}$, can be solved efficiently. We will now exploit this fact and our max-min weighted SIR algorithm in Section III-A to propose a fast algorithm that can be made distributed to solve (27) and (39) in the following.

B. A weighted proportional SIR *algorithm*

We first state our weighted proportional SIR algorithm, which uses Algorithm 1 as a sub-module and adapts the weight parameter $\boldsymbol{\beta}$ of Algorithm 1 iteratively to solve either (27) or (39).

Algorithm 5 (Weighted Proportional SIR*):*

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

$$\mathbf{m}(k+1) = \frac{f'(e^{\tilde{\boldsymbol{\gamma}}(k)})}{\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}(k)})}. \tag{57}$$

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

$$\begin{array}{ll}\text{maximize} & \sum_l m_l(k+1)\tilde{\gamma}\\ \text{subject to} & \log\rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \le 0,\\ \text{variables:} & \tilde{\boldsymbol{\gamma}}.\end{array} \tag{58}$$

3) Set the output of Algorithm 1 using the input weight parameter ($\boldsymbol{\beta} = e^{\tilde{\boldsymbol{\gamma}}(k+1)}$) as $\mathbf{p}(k+1)$.[5]

Depending on the $f'(\tilde{\boldsymbol{\gamma}}(k))$ used in (57), Algorithm 5 can compute the optimal solution of (27) or (39) when the initial point is sufficiently close to the optimal solution. This is stated in the following result.

[5]The output of Algorithm 1 is obtained after it converges to within a given tolerance. Algorithm 5 is essentially a two time-scale algorithm - the power and beamformer are updated at a much faster timescale than the variable $\tilde{\boldsymbol{\gamma}}$.

Theorem 10: For any $\tilde{\boldsymbol{\gamma}}(0)$ in a sufficiently close neighborhood of $\tilde{\boldsymbol{\gamma}}^\star$, $\mathbf{p}(k+1)$ in Algorithm 5 converges to the optimal solution of (27) and (39) for, respectively,

$$f'(e^{\tilde{\boldsymbol{\gamma}}(k)}) = \left(-w_1 \frac{e^{\tilde{\gamma}_1(k)}}{(1+e^{\tilde{\gamma}_1(k)})^2}, \ \ldots \ , -w_L \frac{e^{\tilde{\gamma}_L(k)}}{(1+e^{\tilde{\gamma}_L(k)})^2}\right)^\top \tag{59}$$

and

$$f'(e^{\tilde{\boldsymbol{\gamma}}(k)}) = \left(w_1 \frac{e^{\tilde{\gamma}_1(k)}}{1+e^{\tilde{\gamma}_1(k)}}, \ \ldots \ , w_L \frac{e^{\tilde{\gamma}_L(k)}}{1+e^{\tilde{\gamma}_L(k)}}\right)^\top . \tag{60}$$

Example 5: Algorithm 5 for the two-user case assumes a simple form, because the optimal solution to (58), $(e^{\tilde{\boldsymbol{\gamma}}(k+1)})_1$, is given by the positive root of the quadratic equation $abm_1(k)\gamma_1^2 + (bc(m_1(k)+m_2(k)) - a(m_1(k)-m_2(k)))\gamma_1 - m_1(k)c = 0$, and $(e^{\tilde{\boldsymbol{\gamma}}(k+1)})_2 = (bm_1(k)(e^{\tilde{\boldsymbol{\gamma}}(k+1)})_1 - (m_1(k)-m_2(k)))/(cm_2(k))$, where $a = F_{12}F_{21} + F_{21}v_1/\bar{P} + F_{12}v_2/\bar{P}$, $b = v_1/\bar{P}$ and $c = v_2/\bar{P}$.

The convergence in Theorem 10 is proved only for initial points in the neighbourhood of the optimal solution. However, we now state a result stronger than Theorem 10 (for a more relaxed initialization) that also highlights the connection between the max-min weighted SIR problem and the two nonconvex problems (27) and (39).

Corollary 2: If $\tilde{\gamma}_l^\star$ are equal for all l and $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$, then $\mathbf{p}(k+1)$ in Algorithm 5 converges to the optimal solution of (27) and (39) for (59) and (60), respectively, from any initial point $\tilde{\boldsymbol{\gamma}}(0)$ such that $\tilde{\gamma}_l(0)$ are equal for all l.

Remark 9: This result is similar in flavor to the result in Section V-C, but is more general (without the existence condition of $\tilde{\mathbf{B}}$).

C. Distributed optimization

Although Algorithm 5 is centralized (due to the need to solve (58)), it can be made distributed by leveraging the output of Algorithm 1. In particular, we will use the approximate projected gradient method to obtain a distributed solution to solve (58). Distributed technique involving the subgradient method to solve eigenvalue problems has recently also been used in randomized gossip algorithms [19].

Recall that the gradient $\mathbf{g} \in \mathbb{R}^L$ of $\log\rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$ at $\hat{\boldsymbol{\gamma}}$ satisfies

$$\log\rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \geq \log\rho(\text{diag}(e^{\hat{\boldsymbol{\gamma}}})\mathbf{B}) + \mathbf{g}^\top(\tilde{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}) \tag{61}$$

for any feasible $\tilde{\boldsymbol{\gamma}}$. In fact, $\mathbf{g}$ is given by [21]:

$$\mathbf{g} = \mathbf{x}(\text{diag}(e^{\hat{\boldsymbol{\gamma}}})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\hat{\boldsymbol{\gamma}}})\mathbf{B}), \tag{62}$$

normalized such that $\mathbf{1}^\top\mathbf{g}$=1. Computing the left and right Perron eigenvector requires centralized computation in general. However, observe that it is also the normalized Schur product of the optimal uplink and downlink power when $\boldsymbol{\beta} = e^{\hat{\boldsymbol{\gamma}}}$ in (8) (cf. Figure 2). Leveraging on this fact, we next apply the approximate projected gradient method to obtain a distributed version of Algorithm 5.

Algorithm 6 (Distributed Weighted Proportional SIR): Set the step size $\nu(0) \in (0,1)$.

1) Compute the weight $\mathbf{m}(k+1)$:
$$\mathbf{m}(k+1) = \frac{f'(e^{\tilde{\boldsymbol{\gamma}}(k)})}{\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}(k)})}. \tag{63}$$
2) Set the downlink power and uplink power output of Algorithm 2 using the input weight parameter ($\boldsymbol{\beta} = e^{\tilde{\boldsymbol{\gamma}}(k)}$) as $\mathbf{p}(k)$ and $\mathbf{q}(k)$, respectively.
3) Each lth user computes:
$$\begin{aligned}\tilde{\gamma}_l(k+1) &= \tilde{\gamma}_l(k) + \log \text{SIR}_l(\mathbf{p}(k)) \\ &\quad + \nu(k)\left(\frac{m_l(k+1)(e^{\tilde{\boldsymbol{\gamma}}(k)})_l \mathbf{p}(k)^\top \text{diag}(e^{-\tilde{\boldsymbol{\gamma}}(k)})\mathbf{q}(k)}{(\mathbf{p}(k)\circ\mathbf{q}(k))_l} - 1\right)\end{aligned} \tag{64}$$

for all l. Update $\nu(k+1)=\nu(k)/k$.

Theorem 11: Let $\mathbf{x}(k)=\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{B})$ and $\mathbf{y}(k)=\mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{B}))$ be the optimal solution of Algorithm 2 when $\boldsymbol{\beta}=e^{\tilde{\boldsymbol{\gamma}}(k)}$. Suppose the output of Algorithm 2 (at Step 2 of Algorithm 6) satisfies

$$\limsup_k \|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{x}(k)}\|\|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{y}(k)} \le M_1$$

and

$$\limsup_k \|\mathbf{x}(k)\|_\infty^{\mathbf{p}(k)}\|\|\mathbf{y}(k)\|_\infty^{\mathbf{q}(k)} \le M_2$$

for some positive M_1 and M_2 respectively. For any $\tilde{\boldsymbol{\gamma}}(0)$ in a sufficiently close neighborhood of $\tilde{\boldsymbol{\gamma}}^\star$ and a sequence of step sizes $\nu(k)>0$, $\{\nu(k)\}$, that satisfies

$$\sum_{k=0}^{\infty}\nu(k)=\infty,\ \ \sum_{k=0}^{\infty}(\nu(k))^2<\infty, \tag{65}$$

$\mathbf{p}(k+1)$ in Algorithm 6 converges to within a closed neighbourhood of the optimal solution of (27) and (39) for, respectively, (59) and (60).

Remark 10: We make the following remarks on Algorithm 6. At Step 2, the computation of $\mathbf{p}(k)$ and $\mathbf{q}(k)$ is approximately optimal in the sense of max-min weighted SIR as the time to run Algorithm 2 is finite. These approximation errors carry over to Step 3, thus leading to an approximate gradient with error. Theorem 11 states that these accumulated approximation errors however do not affect the overall convergence as long as the number of iterations to execute Algorithm 2 (at Step 2 of Algorithm 6) is sufficiently long and the step size is tuned appropriately at each iteration. At Step 3, the computation of both the normalized $\mathbf{m}(k)$ and $\mathbf{p}(k)^\top\text{diag}(e^{-\tilde{\boldsymbol{\gamma}}(k)})\mathbf{q}(k)$ can be obtained by a gossip algorithm.

VII. Numerical Examples

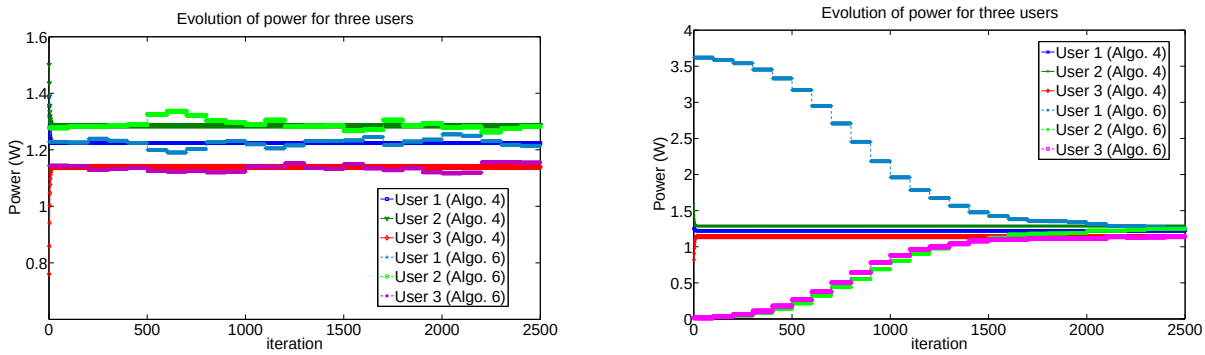
In this section, we evaluate the performance of 3, 4 and 6 numerically. We focus on the case when $\tilde{\mathbf{B}}$ exists, whereby Algorithm 3 and 4 can be used. We use the following channel gain matrix:

$$\mathbf{G}=\begin{bmatrix}0.73 & 0.14 & 0.13\\ 0.15 & 0.69 & 0.12\\ 0.15 & 0.12 & 0.79\end{bmatrix}. \tag{66}$$

We set the total power constraint as $\bar{\mathbf{p}}=3.65$W and the noise power of each user be 1W. The weight vector is given by $\mathbf{w}=\mathbf{x}(\mathbf{B})\circ\mathbf{y}(\mathbf{B})$. It is easily verified that $\tilde{\mathbf{B}}$ exists. Using exaustive search, both the optimal solution to (27) and (39) are attained at the equal SIR allocation of 0.673 for the three users (equal to the solution of (8) with $\boldsymbol{\beta}=\mathbf{1}$), where $\mathbf{p}^\star=\mathbf{x}(\mathbf{B})=[1.2238\ \ 1.2870\ \ 1.1392]^\top W$. Thus, the optimal sum rate is 0.5144 nats/symbol.

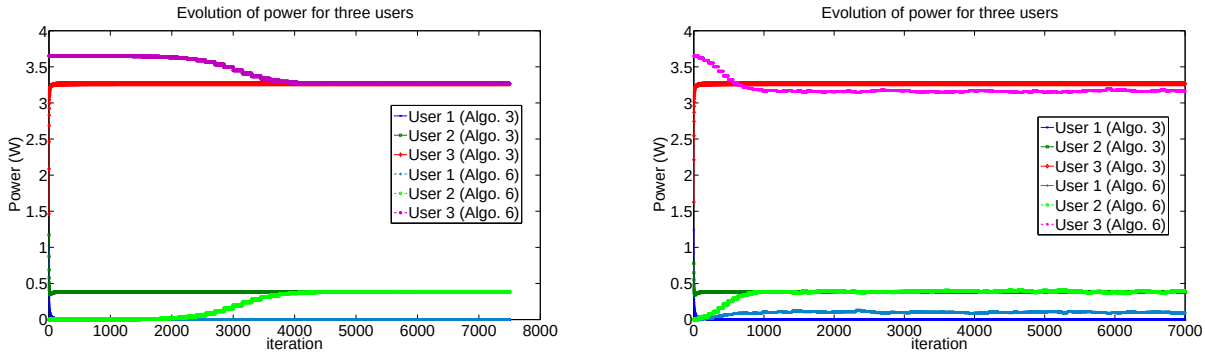
Figure 6 plots the evolution of the power for the three users that run Algorithm 4 and 6. At Step 2 of Algorithm 6, we run Algorithm 2 for 100 iterations before it terminates. We use a diminishing step size $\nu(k+1)=\nu(k)/k, k\ge 1$. In Figure 6(a) and (b), we set the initial SIR vector in Algorithm 6 to $[0.473\ \ 0.473\ \ 0.473]^\top$ (closer to the optimal solution) and $[\text{SNR}_1\ \ 0\ \ 0]^\top$ (further from the optimal solution) respectively. It is observed that Algorithm 4 converges geometrically fast to the optimal solution (verifying Theorem 8) under synchronous updates. In the case of Algorithm 6, convergence to the optimal solution is observed for initial vectors close to the optimal solution (verifying Theorem 11). Interestingly, we have never observed a case where convergence to the optimal solution fails even for the other initial vectors further away from the optimal solution that we have tested. We also observe that the convergence speed is faster when the initial SIR vector is closer to the optimal solution.

Next, we evaluate the performance of using a constant step size of 0.07 in Algorithm 6, and evaluate its performance to solve (27) for different number of inner loops executed at Step 2 of Algorithm 6. The



(a) (b)

Fig. 6. Illustration of the convergence of Algorithm 3 and Algorithm 6 with an initial SIR vectors that are closer to, in (a), and further from, in (b), the optimal solution.



(a) (b)

Fig. 7. Illustration of the convergence of Algorithm 4 and Algorithm 6 with different number of inner loops executed by Algorithm 2 of 150 (a), and 70 in (b).

weight vector is given by $\mathbf{w} = [0.1309 \;\; 0.1678 \;\; 0.7013]^\top$. Using exaustive search, the optimal solution to (27) is attained at the point $[0 \;\; 0.3884 \;\; 3.2616]^\top$. The initial SIR vectors in Algorithm 6 are set close to the optimal solution. Figure 7 plots the evolution of the power for the three users that run Algorithm 3 and Algorithm 6. Figure 7(a) and (b) illustrate the cases when we run Algorithm 2 for 150 and 70 iterations before it terminates, respectively. We observe that if the number of inner loops to execute Algorithm 2 is smaller, we obtain a faster convergence speed of Algorithm 6 to the neighborhood of the optimal solution, but it may not converge to the optimal solution due to the approximate gradient error (In Figure 7(b), a maximum of 9% deviation from the optimal solution is observed).

VIII. Conclusion

Maximizing the minimum weighted SIR, minimizing weighted sum MSE and maximizing weighted sum rate on a multiuser downlink system are three important goals of joint transceiver and power optimization. We established a theoretical connection between these three problems using nonnegative matrix theory, nonlinear Perron-Frobenius theory and the arithmetic-geometric mean inequality. Under sufficiently low to medium interference, we showed that the weighted sum MSE minimization and the sum rate maximization problems can be solved optimally using fast algorithms (without any configuration). In the general case, we established optimality conditions and also theoretical and algorithmic connections between the three problems. We then proposed a weighted proportional SIR algorithm that leveraged our fast max-min weighted SIR algorithm and the uplink-downlink duality to solve these two nonconvex problems in

a distributed manner. Our numerical analysis highlighted the robust performance of our algorithms in finding the global optimal solution of these nonconvex problems. As future work, it will be interesting to find a connection between the quasi-inverse existence and the information-theoretic sum-capacity optimal condition in [11], and to prove our empirical evidence on why Algorithm 5 (centralized version) and Algorithm 6 (distributed version) never fail to find the global optimal solution from any initial point.

ACKNOWLEDGEMENT

We acknowledge helpful discussions with Steven Low at Caltech, Shmuel Friedland at UIC and Kevin Tang at Cornell. This research has been supported in part by AFOSR FA9550-09-1-0643, ONR grant N00014-07-1-0864, NSF CNS 0720570 and ARO MURI Award W911NF-08-1-0233.

IX. APPENDIX

A. Proof of Lemma 1

Proof: Our proof is based on the nonlinear Perron-Frobenius theory [1], [2]. We let the optimal weighted max-min $\mathsf{SIR}(\mathbf{p}^*)$ in (8) be τ^*. A key observation is that all the SIR constraints are tight at optimality. This implies, at optimality of (8),

$$\frac{(p_l^*/\sum_l p_l^*)}{\sum_{j\neq l} F_{lj}(p_l^*/\sum_l p_l^*) + (v_l/\sum_l p_l^*)} = \beta_l \tau^* \tag{67}$$

for all l. Letting $\mathbf{s}^* = (1/\sum p_l^*)\mathbf{p}^*$, (67) can be rewritten as

$$(1/\tau^*)\mathbf{s}^* = \text{diag}(\boldsymbol{\beta})\mathbf{F}\mathbf{s}^* + (1/\sum_l p_l^*)\text{diag}(\boldsymbol{\beta})\mathbf{v}. \tag{68}$$

We first state the following conditional eigenvalue lemma:

Lemma 7 (Conditional eigenvalue [1], Corollary 13): Let $\mathbf{A}$ be a nonnegative matrix and $\mathbf{b}$ be a nonnegative vector. If $\rho(\mathbf{A}+\mathbf{b}\mathbf{1}^\top) > \rho(\mathbf{A})$, then the conditional eigenvalue problem

$$\lambda\mathbf{s} = \mathbf{A}\mathbf{s} + \mathbf{b}, \quad \lambda \in \mathbb{R}, \quad \mathbf{s} \geq \mathbf{0}, \quad \sum_l s_l = 1,$$

has a unique solution given by $\lambda = \rho(\mathbf{A}+\mathbf{b}\mathbf{1}^\top)$ and $\mathbf{s}$ being the unique normalized Perron eigenvector of $\mathbf{A}+\mathbf{b}\mathbf{1}^\top$.

Letting $\lambda = 1/\tau^*, \mathbf{A} = \text{diag}(\boldsymbol{\beta})\mathbf{F}, \mathbf{b} = (1/\bar{P})\text{diag}(\boldsymbol{\beta})\mathbf{v}$ in Lemma 7 and noting that $\sum_l s_l^* = 1$ shows that $\mathbf{p}^* = (\bar{s}_i/\bar{P})\mathbf{x}(\text{diag}(\boldsymbol{\beta})(\mathbf{F} + (1/\bar{P})\mathbf{v}\mathbf{1}^\top))$ is a fixed point of (68). ■

B. Proof of Corollary 1

Proof: The fixed point in Lemma 7 is also a unique fixed point of the following equation [1]:

$$\mathbf{s} = \frac{\mathbf{A}\mathbf{s}+\mathbf{b}}{\|\mathbf{A}\mathbf{s}+\mathbf{b}\|_1}. \tag{69}$$

Applying the power method in (69) to the system of equations in (68) yields the following iterative method:

1) Update auxiliary variable $\mathbf{s}(k+1)$:

$$\mathbf{s}(k+1) = \mathbf{F}\mathbf{s}(k) + (1/\bar{P})\mathbf{v}. \tag{70}$$

2) Normalize $\mathbf{s}(k+1)$:

$$\mathbf{s}(k+1) \leftarrow \mathbf{s}(k+1)/\sum_l s_l(k+1). \tag{71}$$

3) Compute power $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) = \mathbf{s}(k+1)\bar{P}. \tag{72}$$

Combining Step 1 and 3 in the above and after some rearrangement yields Algorithm 1.

We first state the following key theorem in [2] to establish the convergence rate of Algorithm 1.

Theorem 12 (Krause's theorem [2]): Let $\|\cdot\|$ be a monotone norm on $\mathbb{R}^L$. For a concave mapping $f : \mathbb{R}_+^L \to \mathbb{R}_+^L$ with $f(\mathbf{z}) > 0$ for $\mathbf{z} \geq \mathbf{0}$, the following statements hold. The conditional eigenvalue problem $f(\mathbf{z}) = \lambda \mathbf{z}$, $\lambda \in \mathbb{R}$, $\mathbf{z} \geq \mathbf{0}$, $\|\mathbf{z}\| = 1$ has a unique solution $(\lambda^*, \mathbf{z}^*)$, where $\lambda^* > 0$, $\mathbf{z}^* > \mathbf{0}$. Furthermore, $\lim_{k\to\infty} \tilde{f}^k(\mathbf{z})$ converges geometrically fast to $\mathbf{z}^*$, where $\tilde{f}(\mathbf{z}) = f(\mathbf{z})/\|(\mathbf{z})\|$.

It is shown in [1] that $f(\mathbf{z}) = \mathbf{A}\mathbf{z}+\mathbf{b}$ is a concave mapping, where $\mathbf{A} \geq \mathbf{0}$, $\mathbf{b} \geq \mathbf{0}$. Hence, the convergence of the iteration $(\mathbf{A}\mathbf{z}+\mathbf{b})/\|\mathbf{A}\mathbf{z}+\mathbf{b}\|_1$ to the fixed point is geometrically fast, regardless of the initial point. ■

C. Proof of Lemma 2

Proof: Let $\mathbf{A}$ be an irreducible nonnegative matrix, $\mathbf{b}$ a nonnegative vector and $\|\cdot\|$ a norm on $\mathbb{R}^L$ with a corresponding dual norm $\|\cdot\|_D$. Then, Proposition 5 in [1] establishes that

$$\log \rho(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top) = \max_{\|\mathbf{c}\|_D = 1} \log \rho(\mathbf{A} + \mathbf{b}\mathbf{c}^\top),$$

and the fact that $\mathbf{p} = \mathbf{x}(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top)$, which is the Perron eigenvector of $\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top$, is the dual of $\mathbf{c}_*$ with respect to $\|\cdot\|_D$. Further, this $\mathbf{p}$ is also the unique solution to the problem:

$$\tau\mathbf{p} = \mathbf{A}\mathbf{p} + \mathbf{b},\ \tau \in \mathbb{R},\ \mathbf{p} > \mathbf{0},\ \|\mathbf{p}\| = 1. \tag{73}$$

The above result is a special case of Krause's theorem [2] (see above proof of Corollary 1). Next, we give further characterization of the above result by applying the Friedland-Karlin inequality and minimax theorem in [20], [21] on $\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top$.

The equality between (14) and (15) in Lemma 2 can be proved in two different ways. We sketch the first proof (see [3] for more details). Observe that (15) is equivalent to

$$\min_{\|\mathbf{p}\|=1} \max_l \log \frac{(\mathbf{A}\mathbf{p}+\mathbf{b})_l}{p_l},$$

which can be solved optimally by first casting it in epigraph form (introduce an auxiliary variable $\tilde{\tau}$ and additional constraints $\log \frac{(\mathbf{A}\mathbf{p}+\mathbf{b})_l}{p_l} \leq \tilde{\tau}$ for all l), and then using the logarithmic change of variable technique on $\mathbf{p}$, i.e., $\mathbf{p} = e^{\tilde{\mathbf{p}}}$ to make the problem convex in $\tilde{\mathbf{p}}$ and $\tilde{\tau}$. Next, a partial Lagrangian from the epigraph form is considered by relaxing the constraints $\log \frac{(\mathbf{A}e^{\tilde{\mathbf{p}}}+\mathbf{b})_l}{e^{\tilde{p}_l}} \leq \tilde{\tau}$ for all l. Using the Lagrange duality, i.e., the Karush-Kuhn-Tucker (KKT) conditions (cf. [25]), to solve the primal and dual problems, the optimal $\boldsymbol{\lambda}$ in (15) turns out to be the optimal dual variable.

The next observation is to apply the Friedland-Karlin inequality to upper bound the dual problem, which turns out to be tight for a feasible $\tilde{\mathbf{p}}$. The optimal $\mathbf{p}$ and $\boldsymbol{\lambda}$ is thus given by $\mathbf{x}(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top)$ and $\mathbf{x}(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top) \circ \mathbf{y}(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top)$ respectively. In fact, the optimal $\tilde{\tau}$ in the epigraph form is related to the τ in (73): $\tilde{\tau} = \log \tau$. Once (15) is solved completely, the equality between (14) and (15) follows from strong duality. The last step is to observe that the Collatz-Wielandt characterization (cf. (34)) relates (15) to $\log \rho(\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top)$.

The second and simpler proof is a direct application of the Friedland-Karlin minimax theorem in [21] (See Theorem 3.2 in [21]) to the matrix $\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top$. ■

D. Proof of Theorem 1

Proof: Let the optimal power and beamforming matrix in (17) be $\mathbf{p}^\star$ and $\mathbf{U}^\star$ respectively.

First, the authors in [8] used the *extended coupling matrix* approach to show that the downlink max-min SIR problem in (17) has an optimal objective and an optimal power vector $\mathbf{p}^\star$, given respectively by $1/\rho(\mathbf{E}(\mathbf{U}^\star))$ and the vector that contains the first L elements of the eigenvector $\mathbf{x}(\mathbf{E}(\mathbf{U}^\star))$ (with $(\mathbf{x}(\mathbf{E}(\mathbf{U})))_{L+1} = 1$), where

$$\mathbf{U}^\star = \arg \min_{\mathbf{U}\in\mathcal{C}^{N\times L}, \mathbf{u}_l^\dagger \mathbf{u}_l = 1 \; \forall \, l} \rho(\mathbf{E}(\mathbf{U})) \tag{74}$$

and

$$\mathbf{E}(\mathbf{U}) = \begin{bmatrix} \mathbf{F}(\mathbf{U}) & \mathbf{v}(\mathbf{U}) \\ (1/\bar{P})\mathbf{1}^\top \mathbf{F}(\mathbf{U}) & (1/\bar{P})\mathbf{1}^\top \mathbf{v}(\mathbf{U}) \end{bmatrix}. \tag{75}$$

Now, using the fact that the optimal max-min SIR is unique and the above fact in [8] that $\begin{bmatrix}\mathbf{p}^\star & 1\end{bmatrix}^\top = \mathbf{x}(\mathbf{E}(\mathbf{U}^\star))$ and

$$\rho(\mathbf{E}(\mathbf{U}^\star)) = \frac{1}{\bar{P}}\mathbf{1}^\top \mathbf{F}(\mathbf{U}^\star)\mathbf{p}^\star + \frac{1}{\bar{P}}\mathbf{1}^\top \mathbf{v}(\mathbf{U}^\star), \tag{76}$$

we deduce that

$$\rho(\mathbf{E}(\mathbf{U}^\star)) = \rho(\mathbf{B}(\mathbf{U}^\star)) \tag{77}$$

and

$$\mathbf{x}(\mathbf{E}(\mathbf{U}^\star)) = \begin{bmatrix} \mathbf{x}(\mathbf{B}(\mathbf{U}^\star)) \\ 1 \end{bmatrix}, \tag{78}$$

where $\mathbf{1}^\top \mathbf{x}(\mathbf{B}(\mathbf{U}^\star)) = \bar{P}$.

Next, we focus on Algorithm 2 and show that it produces $\mathbf{p}^\star$ and $\mathbf{U}^\star$. It was shown in [9] that the uplink-downlink duality can be combined with the *extended coupling matrix* approach in [8] to solve (17). In particular, a virtual uplink transmission that considers the uplink receive SIR expression given by (18) is first constructed. By considering the following *virtual uplink coupling matrix*:

$$\hat{\mathbf{E}}(\mathbf{U}) = \begin{bmatrix} \mathbf{F}^\top(\mathbf{U}) & \mathbf{v}(\mathbf{U}) \\ (1/\bar{P})\mathbf{1}^\top \mathbf{F}^\top(\mathbf{U}) & (1/\bar{P})\mathbf{1}^\top \mathbf{v}(\mathbf{U}) \end{bmatrix}, \tag{79}$$

the authors in [9] showed that the first L elements of the eigenvector $\mathbf{x}(\hat{\mathbf{E}}(\mathbf{U}^\star))$ (with $(\mathbf{x}(\hat{\mathbf{E}}(\mathbf{U})))_{L+1} = 1$) yields the virtual uplink power vector $\mathbf{q}^\star$, which when plugged into (19) yields the optimal beamforming matrix $\mathbf{U}^\star$. This leads to an iterative method (see Table I in [9]) that computes the eigenvector and spectral radius of $\hat{\mathbf{E}}(\mathbf{U})$ in [9].

Building on the connection between the virtual uplink coupling matrix in [9] and the virtual uplink power $\mathbf{q}$, we can solve both the virtual uplink power $\mathbf{q}$ and the downlink power using Algorithm 1. This leads to Algorithm 2, where in Steps 1-2, $\mathbf{q}(k)$ is computed and is used to compute the transmit beamforming matrix $\mathbf{U}(k)$ in Step 3. Lastly, Steps 4-5 keep the transmit beamfomers fixed and compute the downlink power $\mathbf{p}(k)$. Using (77) and (78), the convergence of Algorithm 2 follows from the convergence result of the iterative method (Table I in [9]). This completes the proof of Theorem 1. ■

E. Total power constraint in (27) and (28) tight at optimality

Proof: Suppose $\mathbf{1}^\top \mathbf{p} < \mathbf{1}^\top \bar{\mathbf{p}}$ at optimality. The objective function in (28) can be strictly minimized by increasing the power of all users proportionally such that $\mathbf{1}^\top \mathbf{p} = \mathbf{1}^\top \bar{\mathbf{p}}$, since $\mathsf{SIR}_l(\mathbf{p})$ for all l increases monotonically. But this contradicts the assumption, thus the total power constraint in (27) and (28) are tight at optimality. ■

F. Proof of Lemma 3

Proof: Let $\mathbf{B} = \mathbf{F}$, where $F_{lj} > 0$ for all l, j and $l \neq j$. Suppose $\tilde{\mathbf{B}} \geq \mathbf{0}$ exists. Since $F_{ll} = 0$ for all l, by definition 1, $\mathbf{F} \geq \tilde{\mathbf{B}}$, thus $\tilde{B}_{ll} = 0$ for all l. We assume $\tilde{\mathbf{B}}$ with $\tilde{B}_{ll} = 0$ for all l exists. By definition 1, $\mathbf{F} - \tilde{\mathbf{B}} = \mathbf{F}\tilde{\mathbf{B}}$. Thus, $\mathrm{Tr}[\mathbf{F}\tilde{\mathbf{B}}] = \mathrm{Tr}[\tilde{\mathbf{B}}\mathbf{F}] = 0$, where $\mathrm{Tr}[\cdot]$ denotes the trace operator. But, this cannot happen unless $\mathbf{F} = \mathbf{0}$ or $\tilde{\mathbf{B}} = \mathbf{0}$, which is at once a contradiction that $\tilde{\mathbf{B}}$ exists. This proves Lemma 3. ■

G. Proof of Lemma 4

Proof: Suppose that $\tilde{\mathbf{B}} = a\mathbf{v}\mathbf{1}^\top$ for some positive a when $\mathbf{B} = \mathbf{v}\mathbf{1}^\top$. We shall show that $\tilde{\mathbf{B}}$ satisfies definition 1 for a unique $a < 1$. By definition 1, $(\mathbf{I} + \mathbf{v}\mathbf{1}^\top) = (\mathbf{I} - a\mathbf{v}\mathbf{1}^\top)^{-1}$, which can be written as $(\mathbf{I} + \mathbf{v}\mathbf{1}^\top) = \sum_{l=0}^{\infty}(a\mathbf{v}\mathbf{1}^\top)^l$ using the von Neumann's expansion [24]. This leads to $\sum_{l=1}^{\infty} a^l(\mathbf{v}^\top\mathbf{1})^{l-1} = 1$, which yields $a = 1/(1 + \mathbf{1}^\top\mathbf{v})$. It can be checked, by definition 1, that $\mathbf{B} - \tilde{\mathbf{B}} = (1 - 1/(1 + \mathbf{1}^\top\mathbf{v}))\mathbf{1}\mathbf{v}^\top = \mathbf{B}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}\mathbf{B} \geq \mathbf{0}$, thus proving Lemma 4. ■

H. Proof of Theorem 2

Proof: We first show that $\rho(\tilde{\mathbf{B}}) < 1$. Now, $\mathbf{I} + \mathbf{B} = (\mathbf{I} - \tilde{\mathbf{B}})^{-1} = \sum_{k=0}^{\infty}(\tilde{\mathbf{B}})^k$, which converges if and only if $\rho(\tilde{\mathbf{B}}) < 1$. Note $\rho(\tilde{\mathbf{B}}) = 1$ corresponds to the trivial solution that $\mathbf{p} = 0$, shown as follows. Let $\mathbf{z} = \mathbf{x}(\tilde{\mathbf{B}})$. Then, take the inner product on both sides of $\mathbf{x}(\tilde{\mathbf{B}}) \geq \tilde{\mathbf{B}}\mathbf{x}(\tilde{\mathbf{B}})$ in (30) by $\mathbf{y}(\tilde{\mathbf{B}})$, which yields $\mathbf{y}(\tilde{\mathbf{B}})^\top\tilde{\mathbf{B}}\mathbf{x}(\tilde{\mathbf{B}})/\mathbf{y}(\tilde{\mathbf{B}})^\top\mathbf{x}(\tilde{\mathbf{B}}) = \rho(\tilde{\mathbf{B}}) \leq 1$. If $\rho(\tilde{\mathbf{B}}) = 1$, then $\mathbf{z} = \mathbf{B}\mathbf{z}$, which implies $\mathbf{p} = \mathbf{0}$, which can be ignored since, at optimality, $p_l > 0$ for some l.

Suppose $\mathbf{B}$ is the quasi-inverse of $\tilde{\mathbf{B}}$. Then, $(\mathbf{I} - \tilde{\mathbf{B}})^{-1}$ is always nonnegative if $\rho(\tilde{\mathbf{B}}) < 1$ since $\mathbf{I} - \tilde{\mathbf{B}}$ is an M-matrix [24]. Thus, there is a unique mapping between all feasible nonnegative $\mathbf{z}$ in (30) and feasible $\mathbf{p}$ in (28), and the optimal solution to (27) is given by $\mathbf{p}^\star = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^\star$. Since $\rho(\tilde{\mathbf{B}}) < 1$, $\mathbf{p}^\star \geq \mathbf{0}$. ■

I. Proof of Lemma 5

Proof: We first state the following lemma.

Lemma 8 (Splitting lemma [24], Chapter 7, Theorem 5.2): Let $\mathbf{A} = \mathbf{M} - \mathbf{N}$ with $\mathbf{A}$ and $\mathbf{M}$ nonsingular. Suppose that $\mathbf{H} \geq \mathbf{0}$, where $\mathbf{H} = \mathbf{M}^{-1}\mathbf{N}$. Then $\rho(\mathbf{H}) = \rho(\mathbf{A}^{-1}\mathbf{N})/(1 + \rho(\mathbf{A}^{-1}\mathbf{N}))$ if and only if $\mathbf{A}^{-1}\mathbf{N} \geq \mathbf{0}$.

Letting $\mathbf{A} = (\mathbf{I}+\mathbf{B})^{-1}$, $\mathbf{M} = \mathbf{I}$ and $\mathbf{N} = \tilde{\mathbf{B}}$ in Lemma 8, we have $\rho(\tilde{\mathbf{B}}) = \rho((\mathbf{I}+\mathbf{B})\tilde{\mathbf{B}})/(1+\rho((\mathbf{I}+\mathbf{B})\tilde{\mathbf{B}}))$. But $(\mathbf{I} + \mathbf{B})\tilde{\mathbf{B}} = \mathbf{B}$, thus obtaining (31). Next, we multiply both sides of $\mathbf{B} - \tilde{\mathbf{B}} = \tilde{\mathbf{B}}\mathbf{B}$ with the Perron eigenvector $\mathbf{x}$ of $\mathbf{B}$. After rearranging, we obtain $\tilde{\mathbf{B}}\mathbf{x} = \rho(\tilde{\mathbf{B}})\mathbf{x}$. Thus, $\tilde{\mathbf{B}}$ and $\mathbf{B}$ have the same Perron eigenvector $\mathbf{x}$. A similar proof for the left eigenvector $\mathbf{y}$ can also be shown. ■

J. Proof of Theorem 3

Proof: We first recall the following result in [20].

Theorem 13 (Theorem 3.1 in [20]): For any irreducible nonnegative matrix $\mathbf{A}$,

$$\prod_l \left(\frac{(\mathbf{A}\mathbf{z})_l}{z_l}\right)^{(\mathbf{x}\circ\mathbf{y})_l} \geq \rho(\mathbf{A}) \tag{80}$$

for all strictly positive $\mathbf{z}$, where $\mathbf{x}$ and $\mathbf{y}$ are the Perron and left eigenvectors of $\mathbf{A}$ respectively.
We will call (80) the Friedland-Karlin inequality in this paper.

Furthermore, if $\mathbf{z} \geq \mathbf{A}\mathbf{z}$ (which implies $\rho(\mathbf{A}) \leq 1$), then for any positive vector $\mathbf{w}$, (80) can be extended to

$$\prod_l \left(\frac{(\mathbf{A}\mathbf{z})_l}{z_l}\right)^{w_l} \geq (\rho(\mathbf{A}))^{\|\mathbf{w}\|_\infty^{\mathbf{x}\circ\mathbf{y}}}. \tag{81}$$

Clearly, (81) reduces to (80) when $\mathbf{w} = \mathbf{x} \circ \mathbf{y}$.

Next, the arithmetic-geometric mean inequality states that

$$\sum_l \alpha_l v_l \geq \prod_l v_l^{\alpha_l}, \tag{82}$$

where $\mathbf{v} > \mathbf{0}$ and $\boldsymbol{\alpha} \geq \mathbf{0},\ \mathbf{1}^\top \boldsymbol{\alpha} = 1$. Equality is achieved in (82) if and only if $v_1 = v_2 = \cdots = v_L$.

Now, we are ready to prove (32). From the objective function of (30), we can establish the following chain of inequalities:

$$\begin{array}{rl} \sum_l w_l(\tilde{\mathbf{B}}\mathbf{z})_l/z_l & \overset{(a)}{\geq} \prod_l \left((\tilde{\mathbf{B}}\mathbf{z})_l/z_l\right)^{w_l} \\ & \overset{(b)}{\geq} (\rho(\tilde{\mathbf{B}}))^{\|\mathbf{w}\|_\infty^{\mathbf{x}(\tilde{\mathbf{B}})\circ\mathbf{y}(\tilde{\mathbf{B}})}} \\ & \overset{(c)}{=} \left(\frac{\rho(\mathbf{B})}{1+\rho(\mathbf{B})}\right)^{\|\mathbf{w}\|_\infty^{\mathbf{x}(\mathbf{B})\circ\mathbf{y}(\mathbf{B})}}, \end{array} \tag{83}$$

where (a) is due to letting $v_l = (\tilde{\mathbf{B}}\mathbf{z})_l/z_l$ for all l and $\boldsymbol{\alpha} = \mathbf{w}$ in (82), (b) is due to letting $\mathbf{A} = \tilde{\mathbf{B}}$ in (81) since $\mathbf{z} \geq \tilde{\mathbf{B}}\mathbf{z}$, and (c) is due to Lemma 5. But, using (29), $\sum_l w_l(\tilde{\mathbf{B}}\mathbf{z})_l/z_l = \sum_{l=1}^L w_l/(1+\gamma_l)$. Thus, we establish (32).

To prove the second part, we note that both inequalities (a) and (b) in (83) become tight if and only if $\gamma_1 = \gamma_2 = \cdots = \gamma_L$ and $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$ (required only for (b) to become tight). In particular, all users receive a common SIR given by $\gamma_l^\star = 1/\rho(\mathbf{B})$ for all l.

To establish the power vector corresponding to $\gamma_l^\star = 1/\rho(\mathbf{B})$, we further note that equality is achieved in (32) of Theorem 3 by the max-min SIR power allocation. More precisely, $\gamma_l^\star = 1/\rho(\mathbf{B})$ is in fact the total power constrained max-min SIR whose associated power vector is given in Lemma 1. This completes the proof of Theorem 3. ■

K. Proof of Theorem 4

Proof: First, we note that though (30) is seemingly nonconvex in $\mathbf{z}$, (30) can nevertheless be solved optimally by further making a change of variables $\tilde{z}_l = \log z_l$ for all l. This is allowed since $\mathbf{z} > \mathbf{0}$. We thus consider

$$\begin{array}{ll} \text{minimize} & \sum_{l=1}^L w_l(\tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}))_l/e^{\tilde{z}_l} \\ \text{subject to} & (\tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}))_l/e^{\tilde{z}_l} \leq 1,\ \ l = 1, \ldots, L, \\ \text{variables:} & \tilde{z}_l\ \ \forall\, l, \end{array} \tag{84}$$

which is clearly strictly convex in $\tilde{\mathbf{z}}$. Though the constraint set in (84) is unbounded, the optimal solution to (84) cannot have $\tilde{z}_l^\star = -\infty$ for some l since, at optimality, $\mathbf{z}^\star > \mathbf{0}$. Next, the Lagrangian of (84) can be written as

$$L(\tilde{\mathbf{z}}, \boldsymbol{\lambda}) = \sum_{l=1}^L w_l(\tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}))_l/e^{\tilde{z}_l} + \sum_{l=1}^L \lambda_l(\tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}))_l/e^{\tilde{z}_l}. \tag{85}$$

We can further simplify (85) by noting that, at optimality, $\mathbf{z}^\star > \tilde{\mathbf{B}}\mathbf{z}^\star$. This is proved easily by noting the following lemma.

Lemma 9 ([24], Chapter 2): If $\mathbf{z} > \mathbf{0}$, then $\mathbf{A}\mathbf{z} \leq \beta\mathbf{z}$ implies $\rho(\mathbf{A}) \leq \beta$ and $\mathbf{A}\mathbf{z} < \beta\mathbf{z}$ implies $\rho(\mathbf{A}) < \beta$.

Thus, $\exp(\tilde{\mathbf{z}}^\star) > \tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}^\star)$ implies $\rho(\tilde{\mathbf{B}}) < 1$, which satisfies Theorem 2. This also implies, using complementary slackness, that the optimal dual variables $\boldsymbol{\lambda}^\star = \mathbf{0}$. Hence, $\mathbf{p}^\star = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^\star \geq \mathbf{0}$, and (85) becomes

$$L(\tilde{\mathbf{z}}, \boldsymbol{\lambda}) = \sum_{l=1}^L w_l(\tilde{\mathbf{B}}\exp(\tilde{\mathbf{z}}))_l/e^{\tilde{z}_l}, \tag{86}$$

which is in fact an unconstrained version of (84). Taking the first order derivative of (86) with respect to $\tilde{z}_l$ for each l and setting it to zero, we have (35) after making the change of variables back to $\mathbf{z}$. Since $\mathbf{1}^\top\mathbf{p}^\star = \mathbf{1}^\top(\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^\star = \bar{P}$, this completes the proof of Theorem 4. ■

L. Proof of Theorem 5

Proof: The convergence proof of Algorithm 3 is based on the standard function approach [13], which is summarized as follows.

Definition 2: $I(\mathbf{p})$ is a standard function if it satisfies [13]:

1) (monotonicity) $I(\mathbf{p}') > I(\mathbf{p})$ if $\mathbf{p}' > \mathbf{p}$.
2) (scalability) Suppose $\beta > 1$. Then, $\beta I(\mathbf{p}) > I(\beta\mathbf{p})$.

Lemma 10 ([13]): If $I(\mathbf{p})$ is standard, then $\mathbf{p}(k+1) = I(\mathbf{p}(k))$ converges to the unique fixed point from any initial point $\mathbf{p}(0)$ under synchronous and asynchronous updates.

It can be verified that the iterative equation in (36) is standard. Thus, by Lemma 10, $\mathbf{z}(k+1)$ converges to the fixed point of (35) for arbitrarily small $\epsilon > 0$ under synchronous and asynchronous updates. Hence, $\mathbf{p}(k+1)$ also converges to $\mathbf{p}^\star$ under synchronous and asynchronous updates.

Now, in order to compute the power $\mathbf{p}(k+1)$ from $\mathbf{z}(k+1)$, we could use

$$\mathbf{p}(k+1) = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}(k+1), \tag{87}$$

which however has to be computed in a centralized manner. Instead of using (87) in Algorithm 3, we derive in the following a simpler mapping. From (29), observe that

$$z_l = \sum_{j \neq l} F_{lj} p_j + v_l + p_l \tag{88}$$

can be rewritten as

$$p_l = \frac{\mathsf{SIR}_l(\mathbf{p})}{1 + \mathsf{SIR}_l(\mathbf{p})} z_l. \tag{89}$$

Thus, we use (89) as the nonlinear map between $\mathbf{p}$ and $\mathbf{z}$, and write

$$p_l(k+1) = \frac{\mathsf{SIR}_l(\mathbf{p}(k))}{1 + \mathsf{SIR}_l(\mathbf{p}(k))} z_l(k+1) \quad \forall\, l, \tag{90}$$

in (37), which can be computed locally by the lth user. ■

M. Proof of Theorem 6

Proof: Theorem 6 can be proved in a manner similar to the proof of Theorem 4, and therefore is omitted. ■

N. Proof of Theorem 7

Proof: Similar to the proof of Theorem 4, we make a change of variables $\tilde{z}_l = \log z_l$ for all l in (41), and thus convexify (41) to yield:

$$\begin{array}{ll} \text{minimize} & \prod_{l=1}^{L} w_l (\tilde{\mathbf{B}} \exp(\tilde{\mathbf{z}}))_l / e^{\tilde{z}_l} \\ \text{subject to} & (\tilde{\mathbf{B}} \exp(\tilde{\mathbf{z}}))_l / e^{\tilde{z}_l} \leq 1 \;\; l = 1, \ldots, L, \\ \text{variables:} & \tilde{z}_l, \;\; \forall\, l, \end{array} \tag{91}$$

which is strictly convex in $\tilde{\mathbf{z}}$. Applying the KKT conditions to (91) and making a change of variables back to the $\mathbf{z}$ domain, we obtain (42). The optimal power vector $\mathbf{p}^\star$ in (39) is then recovered from $\mathbf{z}^\star$ in (42) using Theorem 6. ■

O. Proof of Theorem 8

Proof: Similar to the proof of Theorem 5, the convergence proof of Algorithm 3 is based on the standard function approach [13]. It can be verified that the equation $\mathbf{z} = I(\mathbf{z})$ in (42) is homogeneous of degree 1. As in the previous, we incorporate an arbitrarily small positive value ϵ to (42) to make it a standard interference function. We thus consider the iterative method $\mathbf{z}(k+1) = I(\mathbf{z}(k)) + \epsilon$. Using Lemma 10, $\mathbf{z}(k+1)$ converges to the fixed point of (42) for arbitrarily small $\epsilon > 0$ under synchronous and asynchronous updates. Hence, using the mapping in (90), $\mathbf{p}(k+1)$ also converges to $\mathbf{p}^\star$ of (39) under synchronous and asynchronous updates. ■

P. Proof of Lemma 6

Proof: It is easy to observe that for a given SIR value $\boldsymbol{\gamma}$, the power to achieve $\boldsymbol{\gamma}$ is given by (49) [10], [18]. Next, we will show that this power vector is feasible in (27) and (39), i.e., it satisfies the total power constraint. First, observe that $\text{diag}(\boldsymbol{\gamma})\mathbf{B}$ is an irreducible nonnegative matrix. Now, using (49), we have

$$\begin{aligned} \text{diag}(\boldsymbol{\gamma})\mathbf{B}\mathbf{p} &= \text{diag}(\boldsymbol{\gamma})\mathbf{F}\mathbf{p} + (\mathbf{1}^\top \mathbf{p}/\bar{P})\mathbf{v} \\ &\leq \text{diag}(\boldsymbol{\gamma})\mathbf{F}\mathbf{p} + \mathbf{v} = \mathbf{p}. \end{aligned} \tag{92}$$

Thus, $\text{diag}(\boldsymbol{\gamma})\mathbf{B}\mathbf{p} \leq \mathbf{p}$ for any $\mathbf{p}$ in (49). Using the Perron-Frobenius Theorem [24], this implies that $\rho(\text{diag}(\boldsymbol{\gamma})\mathbf{B}) \leq 1$. ■

Q. Proof of Theorem 9

Proof: Theorem 9 is proved using Lagrange duality, i.e., the KKT conditions, and the fact that the optimal dual variable is unique as follows. We introduce a dual variable to the constraint of (51) and write the Lagrangian:

$$L(\tilde{\boldsymbol{\gamma}}, \lambda) = f(e^{\tilde{\boldsymbol{\gamma}}}) - \lambda \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}).$$

Now, the gradient of $\log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$ with respect to $\tilde{\boldsymbol{\gamma}}$ is given by (unique up to a scaling constant) [21]:

$$\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}),$$

where $\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$ and $\mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$ are scaled such that $\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})^\top \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) = 1$. Using this fact, we can compute $\partial L(\tilde{\boldsymbol{\gamma}}, \lambda)/\partial \tilde{\gamma}_l$ for all l. In addition, we see that λ is unique and equals $\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}^\star})$ at optimality. Hence, $\tilde{\boldsymbol{\gamma}}^\star$ satisfies

$$\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) = \frac{f'(e^{\tilde{\boldsymbol{\gamma}}^\star})}{\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}^\star})}$$

and

$$\rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}^\star})\mathbf{B}) = 1.$$

Lastly, the uniqueness of $\tilde{\boldsymbol{\gamma}}'$ follows from the following result:

Lemma 11 ([26]): Let $\mathbf{B} \in \mathbb{R}_+^{L \times L}, \mathbf{w} \in \mathbb{R}_+^L$ be a given irreducible matrix with positive diagonal elements and positive probability vector, respectively. Then there exists $\boldsymbol{\eta} \in \mathbb{R}^L$ such that $\mathbf{x}(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{B}) = \mathbf{w}$. Furthermore, $\boldsymbol{\eta}$ is unique up to an addition of $t\mathbf{1}$, t being a scalar. In particular, this $\boldsymbol{\eta}$ can be computed by solving the following convex optimization problem:

$$\begin{array}{ll} \text{maximize} & \mathbf{w}^\top \boldsymbol{\eta} \\ \text{subject to} & \log \rho(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{B}) \leq 0, \\ \text{variables:} & \boldsymbol{\eta}. \end{array} \tag{93}$$

■

R. Proof of Theorem 10

Proof: We use the fact that in a sufficiently close neighborhood of $\tilde{\boldsymbol{\gamma}}^\star$, the domain set is convex, and the objective function $f(e^{\tilde{\boldsymbol{\gamma}}^\star})$ is twice continuously differentiable. We then use a successive convex approximation technique to compute $\tilde{\boldsymbol{\gamma}}^\star$ assuming that the initial point is sufficiently close to $\tilde{\boldsymbol{\gamma}}^\star$. The convergence conditions for such a technique are given in [27], [28]. Instead of solving (51) directly, we replace the objective function of (51) in a neighborhood of a feasible point $\tilde{\boldsymbol{\gamma}}(0)$ by its Taylor series (up to the first order terms):

$$f'(e^{\tilde{\boldsymbol{\gamma}}}) \approx f(e^{\tilde{\boldsymbol{\gamma}}(0)}) + f'(e^{\tilde{\boldsymbol{\gamma}}(0)})^\top (\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}(0)).$$

Assume a feasible $\tilde{\boldsymbol{\gamma}}(0)$ that is close to $\tilde{\boldsymbol{\gamma}}^\star$. We then compute a feasible $\tilde{\boldsymbol{\gamma}}(k+1)$ by solving the $(k+1)$th approximation problem:

$$\begin{array}{ll} \text{maximize} & f'(e^{\tilde{\boldsymbol{\gamma}}(k)})^\top (\tilde{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}(k)) \\ \text{subject to} & \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \leq 0, \\ \text{variables:} & \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \quad \tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_n)^\top \in \mathbb{R}^L, \end{array} \tag{94}$$

where $\tilde{\boldsymbol{\gamma}}(k)$ is the optimal solution of the kth approximation problem. This inner approximation technique converges to a local optimal solution [27], [28]. In addition, if $\tilde{\boldsymbol{\gamma}}(0)$ is sufficiently close to $\tilde{\boldsymbol{\gamma}}^\star$, then $\lim_{k\to\infty} \tilde{\boldsymbol{\gamma}}(k) = \tilde{\boldsymbol{\gamma}}^\star$.

Next, we leverage Lemma 11 in [26] and Algorithm 1 to solve (94). To be more precise, note that the solution to the max-min weighted SIR problem satisfies the following optimality condition:

$$\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) = \mathbf{x}(\text{diag}(\boldsymbol{\beta})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\boldsymbol{\beta})\mathbf{B}),$$

which can be made equal to $f'(e^{\tilde{\boldsymbol{\gamma}}(k)})/\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}(k)})$ by choosing $\boldsymbol{\beta}$ appropriately. In particular, by setting $\boldsymbol{\beta} = e^{\tilde{\boldsymbol{\gamma}}(k)}$, Algorithm 1 computes a feasible power $\mathbf{p}(k+1)$:

$$\mathbf{p}(k+1) = (\mathbf{I} - \text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{F})^{-1}\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{v},$$

which converges to the optimal solution of (27) and (39) for, respectively, $f'(e^{\tilde{\boldsymbol{\gamma}}(k)}) = \left(-w_1 \frac{e^{\tilde{\gamma}_1(k)}}{(1+e^{\tilde{\gamma}_1(k)})^2}, \dots, -w_L \frac{e^{\tilde{\gamma}_L}}{(1+e^{\tilde{\gamma}_L}}\right.$ and $f'(e^{\tilde{\boldsymbol{\gamma}}(k)}) = \left(w_1 \frac{e^{\tilde{\gamma}_1(k)}}{1+e^{\tilde{\gamma}_1(k)}}, \dots, w_L \frac{e^{\tilde{\gamma}_L(k)}}{1+e^{\tilde{\gamma}_L(k)}}\right)^\top$. ■

S. Proof of Corollary 2

Proof: If $\tilde{\boldsymbol{\gamma}}^\star = \mathbf{1}$ (up to a scaling constant), and $\mathbf{w} = \mathbf{x}(\mathbf{B}) \circ \mathbf{y}(\mathbf{B})$, then the optimality conditions in Theorem 9 are clearly satisfied. Any initial condition $\tilde{\boldsymbol{\gamma}}(0) = \mathbf{1}$ (up to a scaling constant), will satisfy (up to a scaling constant):

$$\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(0)})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(0)})\mathbf{B}) = \frac{f'(e^{\tilde{\boldsymbol{\gamma}}(0)})}{\mathbf{1}^\top f'(e^{\tilde{\boldsymbol{\gamma}}(0)})} = \mathbf{w}.$$

Thus, $\tilde{\boldsymbol{\gamma}}(k) = \mathbf{1}$ (up to a scaling constant) for all k. This proves Corollary 2. ■

T. Proof of Theorem 11

Proof: Theorem 11 is proved by using the projected gradient method with error [25] to solve (58) and then applying a result on the convergence of approximate gradient method in [29]. We first look at computing an approximate projected gradient for (58), and then show that the approximate projected gradient method converges to a neighbourhood of the optimal solution under certain conditions on the sequence of step sizes.

Define $\boldsymbol{\omega} \in \mathbb{R}^L$ as the vector with the lth entry given by $1/(\mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}))_l$. Now, based on the gradient of $\log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$, $\nabla \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B}) = (\partial \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})/\partial\tilde{\gamma}_1, \ldots, \partial \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})/\partial\tilde{\gamma}_L)^\top$, in (62), an approximate gradient to $\tilde{\boldsymbol{\gamma}}$ of (58) can be given by

$$g(\tilde{\boldsymbol{\gamma}}) = \text{diag}(\boldsymbol{\omega})(\mathbf{m} - \nabla \log \rho(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})).$$

Note that, at Step 2, the downlink power $\mathbf{p}(k)$ and uplink power $\mathbf{q}(k)$ output of Algorithm 2 using the input weight parameter ($\boldsymbol{\beta} = e^{\tilde{\boldsymbol{\gamma}}(k)}$) are approximately optimal (in the sense of solving the downlink and uplink max-min weighted SIR respectively), because the time to compute them is finite and this computation terminates within some tolerance.

Let $\mathbf{g}(k)$ be the vector with the lth entry:

$$\frac{m_l(k+1)(e^{\tilde{\boldsymbol{\gamma}}(k)})_l \mathbf{p}(k)^\top \text{diag}(e^{-\tilde{\boldsymbol{\gamma}}(k)})\mathbf{q}(k)}{(\mathbf{p}(k) \circ \mathbf{q}(k))_l} - 1.$$

Note that if $\mathbf{p}(k) = \mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$ and $\mathbf{q}(k) = \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}})\mathbf{B})$, then $\mathbf{g}(k) = g(\tilde{\boldsymbol{\gamma}}(k))$. However, $\mathbf{p}(k)$ and $\mathbf{q}(k)$ contains errors, as Algorithm 2 terminates in finite time.

We now show that the gradient method converges depite approximation errors made in the computation of the eigenvectors ($\mathbf{p}(k)$ and $\mathbf{q}(k)$) at Step 2, which spills into the gradient projection computation at Step 3. We will use the following result from [29] (Proposition 1 in [29], see also [25], Sec. 1.3, pp. 61):

Theorem 14: Let $\{\mathbf{s}(k+1)\}$ be a sequence generated by the gradient method with errors

$$\mathbf{s}(k+1) = \mathbf{s}(k) + \nu(k)(\mathbf{d}(k) + \mathbf{e}(k)),$$

where $\nabla \boldsymbol{f}$ satisfies the Lipschitz assumption, $\mathbf{d}(k)$ safisfies

$$c_1 \|\nabla \boldsymbol{f}(\mathbf{s}(k))\|^2 \leq \nabla \boldsymbol{f}(\mathbf{s}(k))^\top \mathbf{d}(k), \;\; \|\mathbf{d}(k)\| \leq c_2(1 + \|\nabla \boldsymbol{f}(\mathbf{s}(k))\|), \; \forall\, k,$$

where c_1 and c_2 are some scalars, the stepsizes $\{\nu(k)\}$ satisfy

$$\sum_{k=0}^{\infty} \nu(k) = \infty, \;\; \sum_{k=0}^{\infty} (\nu(k))^2 < \infty,$$

and the errors $\{\mathbf{e}(k)\}$ satisfy

$$\|\mathbf{e}(k)\| \leq \nu(k)(c_3 + c_4 \|\nabla \boldsymbol{f}(\mathbf{s}(k))\|), \; \forall\, k,$$

where c_3 and c_4 are some scalars. Then either $\boldsymbol{f}(\mathbf{s}(k)) \to -\infty$ or else $\{\boldsymbol{f}(\mathbf{s}(k))\}$ converges to a finite value and $\nabla \boldsymbol{f}(\mathbf{s}(k)) \to 0$. Furthermore, every limit point of $\{\mathbf{s}(k)\}$ is a stationary point of $\boldsymbol{f}$.

The idea of Theorem 14 is that as long as the descent direction is sufficiently aligned with the gradient and the errors are sufficiently bounded, then the gradient method converges to the optimal solution. Now, $(\mathbf{g}(k))_l$ can be written as (let $\mathbf{x}(k) = \mathbf{x}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{B})$ and $\mathbf{y}(k) = \mathbf{y}(\text{diag}(e^{\tilde{\boldsymbol{\gamma}}(k)})\mathbf{B})$):

$$(\mathbf{g}(k))_l = g_l(\tilde{\boldsymbol{\gamma}}(k))\alpha(k) + g_l(\tilde{\boldsymbol{\gamma}}(k)) - \left(\frac{m_l(k+1)}{(\mathbf{x}(k) \circ \mathbf{y}(k))_l} - \alpha(k)\right)$$

where $\alpha(k) = \frac{(\mathbf{x}(k) \circ \mathbf{y}(k))_l}{(\mathbf{p}(k) \circ \mathbf{q}(k))_l}$.

In Theorem 14, we let $\mathbf{s} = \tilde{\boldsymbol{\gamma}}$, $f = g(\tilde{\boldsymbol{\gamma}})$, $\mathbf{d}_l(k) = g_l(\tilde{\boldsymbol{\gamma}}(k))\alpha(k)$ and $\mathbf{e}_l(k) = g_l(\tilde{\boldsymbol{\gamma}}(k)) - \left(\frac{m_l(k+1)}{(\mathbf{x}(k) \circ \mathbf{y}(k))_l} - \alpha(k)\right)$. It can be shown that

$$\mid \mathbf{d}_l(k)^\top g_l(\tilde{\boldsymbol{\gamma}}(k)) \mid \leq \|g_l(\tilde{\boldsymbol{\gamma}}(k))\|^2 \|\mathbf{x}(k)\|_\infty^{\mathbf{p}(k)} \|\mathbf{y}(k)\|_\infty^{\mathbf{q}(k)}$$

and

$$\begin{aligned} \|\mathbf{e}(k)\| \;\; &\leq \|g_l(\tilde{\boldsymbol{\gamma}}(k))\| + \|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{x}(k)} \|\|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{y}(k)} \\ &\quad + \|\mathbf{x}(k)\|_\infty^{\mathbf{p}(k)} \|\|\mathbf{y}(k)\|_\infty^{\mathbf{q}(k)}, \end{aligned}$$

where $\sqrt{\mathbf{m}(k+1)} = (\sqrt{m_1(k+1)}, \ldots, \sqrt{m_L(k+1)})^\top$, and we have used the fact that $\|\mathbf{x} \circ \mathbf{y}\|_\infty^{\mathbf{p}\circ\mathbf{q}} \leq \|\mathbf{x}\|_\infty^{\mathbf{p}} \|\mathbf{y}\|_\infty^{\mathbf{q}}$ for any positive vector $\mathbf{x}$, $\mathbf{y}$, $\mathbf{p}$ and $\mathbf{q}$.

Assume that

$$\limsup_k \|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{x}(k)} \|\|\sqrt{\mathbf{m}(k+1)}\|_\infty^{\mathbf{y}(k)} \leq M_1$$

and

$$\limsup_k \|\mathbf{x}(k)\|_\infty^{\mathbf{p}(k)} \|\|\mathbf{y}(k)\|_\infty^{\mathbf{q}(k)} \leq M_2$$

for some positive M_1 and M_2 respectively, we can select $\nu(k)$ such that

$$\|\mathbf{e}(k)\| \leq \nu(k)(\limsup_k (1/\nu(k))(M_1 + M_2) + \limsup_k (1/\nu(k)) \|g_l(\tilde{\boldsymbol{\gamma}}(k))\|), \ \forall\, k,$$

in the following gradient update:

$$\begin{aligned}\tilde{\gamma}_l'(k+1) = \tilde{\gamma}_l(k) + \\ +\nu(k) \left(\frac{m_l(k+1)(e^{\tilde{\boldsymbol{\gamma}}(k)})_l \mathbf{p}(k)^\top \mathrm{diag}(e^{-\tilde{\boldsymbol{\gamma}}(k)})\mathbf{q}(k)}{(\mathbf{p}(k)\circ\mathbf{q}(k))_l} - 1 \right).\end{aligned}$$

Now, the point $\tilde{\boldsymbol{\gamma}}'(k+1)$ after the above gradient update may be infeasible with respect to the constraint set of (58). We now project $\tilde{\boldsymbol{\gamma}}'(k+1)$ to the feasible set by adding a constant term $\log \mathsf{SIR}_l(\mathbf{p}(k))/\beta_l = 1/\rho(\mathrm{diag}(e^{\mathsf{SIR}(\mathbf{p}(k))/\boldsymbol{\beta}})\mathbf{B})$ (the value of the logarithmic weighted max-min SIR evaluated at $\mathbf{p}(k)$) to obtain the following update:

$$\tilde{\gamma}_l(k+1) = \tilde{\gamma}_l'(k+1) + \log(\mathsf{SIR}_l(\mathbf{p}(k))/\beta_l).$$

To show the feasibility of $\tilde{\gamma}_l(k+1)$, we have (let $\mathbf{x}' = \mathbf{x}(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B})$):

$$\begin{aligned}&\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}(k+1)})\mathbf{B}\mathbf{x}' = \mathrm{diag}(\mathsf{SIR}(\mathbf{p}(k))/\boldsymbol{\beta})\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B}\mathbf{x}' \\ &\leq \max_l \tfrac{\mathsf{SIR}_l(\mathbf{p}(k))}{\beta_l} \mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B}\mathbf{x}' \\ &\leq \tfrac{1}{\rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B})} \mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B}\mathbf{x}' \\ &= \tfrac{1}{\rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B})} \rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}'(k+1)})\mathbf{B})\mathbf{x}' = \mathbf{x}'.\end{aligned}$$

Hence, $\rho(\mathrm{diag}(e^{\tilde{\boldsymbol{\gamma}}(k+1)})\mathbf{B}) \leq 1$, i.e., $\tilde{\boldsymbol{\gamma}}(k+1)$ is feasible with respect to the constraint set of (58). Using Theorem 14, we conclude that $\tilde{\boldsymbol{\gamma}}(k)$ converges to the optimal solution as $k \to \infty$.

■

REFERENCES

[1] V. D. Blondel, L. Ninove, and P. Van Dooren. An affine eigenvalue problem on the nonnegative orthant. *Linear Algebra and its Applications*, 404:69–84, 2005.

[2] U. Krause. Concave Perron-Frobenius theory and applications. *Nonlinear analysis*, 47(2001):1457–1466, 2001.

[3] C. W. Tan, M. Chiang, and R. Srikant. Fast algorithms and performance bounds for sum rate maximization in wireless networks. *Proc. of IEEE Infocom*, 2009.

[4] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, 1998.

[5] S. Ulukus and R. Yates. Adaptive power control and MMSE interference suppression. *ACM Wireless Networks*, 4(6):489–496, 1998.

[6] P. Viswanath and D. N. C. Tse. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans. on Information Theory*, 49(8):1912–1921, 2003.

[7] W. Yu. Uplink-downlink duality via minimax duality. *IEEE Trans. on Information Theory*, 52(2):361–374, 2006.

[8] W. Yang and G. Xu. Optimal downlink power assignment for smart antenna systems. *Proc. of IEEE ICASSP*, 1998.

[9] M. Schubert and H. Boche. Solution of the multiuser downlink beamforming problem with individual SINR constraints. *IEEE Trans. on Vehicular Technology*, 53(1):18–28, 2004.

[10] 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.

[11] V. S. Annapureddy and V. V. Veeravalli. Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. *IEEE Trans. on Information Theory*, 55(7):3032–3050, 2009.

[12] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho. Joint design of Tx-Rx beamformers in MIMO downlink channel. *IEEE Trans. on Signal Processing*, 55(9):4639–4655, 2007.

[13] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1348, 1995.
[14] E. Visotsky and U. Madhow. Optimum beamforming using transmit antenna arrays. *Proc. of IEEE VTC*, 1999.
[15] P. Viswanath, V. Anantharam, and D. N. C. Tse. Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers. *IEEE Trans. on Information Theory*, 45(6):1968–1983, 1999.
[16] D. N. C. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 1st edition, 2005.
[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, USA, 1991.
[18] 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.
[19] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. on Information Theory*, 52(6):2508–2530, 2006.
[20] 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.
[21] S. Friedland. Convex spectral functions. *Linear and Multilinear Algebra*, 9(4):299–316, 1981.
[22] C. W. Tan. Nonconvex power control in multiuser communication systems. *Ph.D. Thesis, Princeton University*, November 2008.
[23] Y. K. Wong. Some mathematical concepts for linear economic models. *Chapter in Economic Activity Analysis*, Edited by O. Morgenstern, John Wiley & Sons, Inc.:283–339, 1954.
[24] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, USA, 1st edition, 1979.
[25] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, USA, 2nd edition, 2003.
[26] S. Friedland and C. W. Tan. Maximizing sum rates in Gaussian interference-limited channels. *arXiv*, 0806(2860v2), 2008.
[27] B. R. Marks and G. P. Wright. A general inner approximation algorithm for nonconvex mathematical programs. *Operations Research*, 26(4):681–683, 1978.
[28] 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.
[29] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM J. Optimization*, 10(3):627–642, 2000.