

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

Chee Wei Tan, *Member, IEEE*, Mung Chiang, *Senior Member, IEEE*, and R. Srikant, *Fellow, IEEE*

Abstract—Maximizing the minimum weighted signal-to-interference-and-noise ratio (SINR), minimizing the weighted sum mean-square error (MSE) and maximizing the weighted sum rate in a multiuser downlink system are three important performance objectives in nonconvex 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 SINR 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 weak interference regimes using fast algorithms. In the general case, we first establish 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 SINR problem. We then propose a distributed weighted proportional SINR algorithm that leverages our fast max-min weighted SINR algorithm to solve for local optimal solution of the two nonconvex problems, and give conditions under which global optimality is achieved. Numerical results are provided to complement the analysis.

Index Terms—Beamforming, multiple-inputmultiple-output (MIMO), optimization, uplink–downlink duality, wireless communication.

I. INTRODUCTION

WE consider 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. 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 the weighted sum rates and weighted sum mean-square error (MSE) as objectives are however challenging to solve, because these two problems are nonconvex. Further, the transmit beamformers are coupled across the 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 max-min weighted signal-to-interference-and-noise ratio (SINR) problem. While previous algorithms in the literature require centralized computation, e.g., iterative computation of an eigenvalue and eigenvector of a particular matrix at each step, we propose a fast distributed algorithm (without parameter configuration) that computes the optimal power and transmit beamformer in the max-min weighted SINR problem with geometric convergence rate. In addition, our algorithm reuses a power control module in [1], which has been used in practical CDMA cellular systems, e.g., IS-95 and CDMA2000, thus facilitating the implementation of our algorithm in practical systems. This is achieved by applying the nonlinear Perron–Frobenius theory in [2] and [3] and the uplink–downlink duality in [4]–[8] to optimize transmit beamformers in the downlink.

By considering simple and low-complexity transmitter design, e.g., linear beamformer, 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 SINR 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 (e.g., [9]) only solve these two nonconvex problems in a centralized manner and suboptimally. We solve these two nonconvex problems optimally under sufficiently weak interference conditions, and develop fast algorithms (independent of stepsize and no parameter configuration whatsoever) for their special cases.

We then turn to establishing the optimality conditions 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-square error (LMMSE) filter. This relates to earlier work on the optimality of the LMMSE filter in the max-min SINR problem and a related total power minimization problem [4], [7], [8]. Further theoretical and algorithmic connections between the two nonconvex problems and the max-min SINR problem are established using nonnegative matrix theory, and a fast algorithm (with minimal configuration) that leverages our fast max-min weighted SINR algorithm and the uplink–downlink duality is then proposed to solve for the local optimal solution of these two nonconvex problems in a distributed

Manuscript received April 24, 2011; revised July 11, 2011; accepted August 01, 2011. Date of publication August 18, 2011; date of current version November 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shahram Shahbazpanahi. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), South Korea, June 2009. This research has been supported in part by AFOSR FA9550-09-1-0643, ONR Grant N00014-07-1-0864, NSF CNS 0720570, NSF CNS-1011962, ARO MURI Award W911NF-08-1-0233 and City University Hong Kong project Grant 7008087.

C. W. Tan is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: cheewtan@cityu.edu.hk).

M. Chiang is with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544 USA (e-mail: mchiang@princeton.edu).

R. Srikant is with the Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801 USA (e-mail: rsrikant@illinois.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2011.2165065

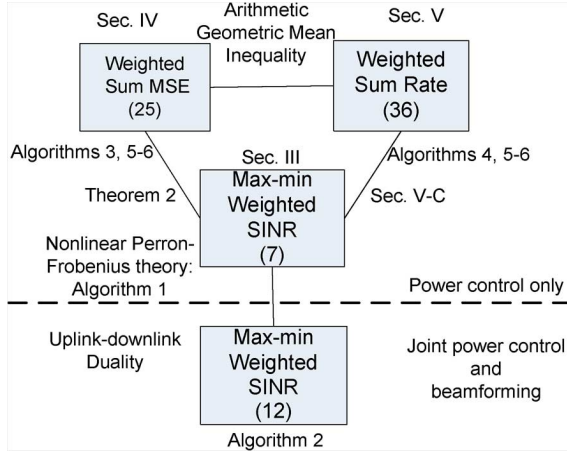


Fig. 1. Overview of the connection (solid lines) between the three optimization problems as revealed in the paper: i) Weighted sum MSE minimization in (25), ii) weighted sum rate maximization in (36), and iii) max-min weighted SINR in (7). The upper half of the dotted line considers power control only, while the lower half considers both power control and beamforming.

fashion. All the proposed algorithms in this paper are applicable to the MISO case.

This paper is organized as follows. We present the system model in Section II. In Section III, we look at the max-min weighted SINR 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 in the weak interference regime. In Section VI, we establish the optimality conditions for these two problems and propose a weighted proportional SINR 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 the Appendix.

We refer the readers to Fig. 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 component-wise inequality between nonnegative vectors \mathbf{u} and \mathbf{v} (nonnegative matrices \mathbf{B} and \mathbf{F}). 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 l th unit coordinate vector and \mathbf{I} denote the identity matrix. Let $\mathbf{x} \circ \mathbf{y}$ denote $\mathbf{x} \circ \mathbf{y} = [x_1 y_1, \dots, x_L y_L]^\top$ (Schur product). Let $\|\mathbf{w}\|_\infty^{\mathbf{x}}$ be the weighted maximum norm of a real vector with respect to the weight \mathbf{x} , i.e., $\|\mathbf{w}\|_\infty^{\mathbf{x}} = \max_l \frac{w_l}{x_l}$, $\mathbf{x} > \mathbf{0}$. Let $e^{\mathbf{x}}$ denote $e^{\mathbf{x}} = [e^{x_1}, \dots, e^{x_L}]^\top$. We let $(g(\mathbf{y}))_l$ denote the l th element of a function vector $g(\mathbf{y}) : \mathbb{R}^L \rightarrow \mathbb{R}^L$.

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:

$$\mathbf{y}_l = \mathbf{h}_l^\dagger \mathbf{x} + \mathbf{z}_l, \quad l = 1, \dots, L \quad (1)$$

where $\mathbf{y}_l \in \mathbb{C}^{1 \times 1}$ is the received signal of the l th user, $\mathbf{h}_l \in \mathbb{C}^{N \times 1}$ is the channel matrix between the base station and the l th user, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the transmitted signal vector, and \mathbf{z}_l 's are the i.i.d. additive complex Gaussian noise vectors with variance $\frac{\sigma_l^2}{2}$ 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 \mathbf{x} in the form of $\mathbf{x} = \sum_{l=1}^L d_l \hat{\mathbf{w}}_l$, where $\hat{\mathbf{w}}_l \in \mathbb{C}^{N \times 1}$ is the transmit beamformer that carries the information signal d_l of the l th user. We assume a total power constraint at the transmit antennas, i.e., $\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] = \bar{P}$ given in Watt (W). From (1), the received signal for the l th user can be expressed as $\mathbf{y}_l = (\mathbf{h}_l^\dagger \hat{\mathbf{w}}_l) d_l + \sum_{j \neq l} (\mathbf{h}_l^\dagger \hat{\mathbf{w}}_j) d_j + \mathbf{z}_l$. Next, we write $\hat{\mathbf{w}}_l = \sqrt{p_l} \mathbf{u}_l$, where p_l is the downlink transmit power (in Watt) and \mathbf{u}_l is the normalized transmit beamformer, i.e., $\mathbf{u}_l^\dagger \mathbf{u}_l = 1$, of the l th user. Now, the received SINR of the l th user in the downlink transmission can be given in terms of \mathbf{p} and $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L]$:

$$\text{SINR}_l(\mathbf{p}, \mathbf{U}) = \frac{p_l |\mathbf{h}_l^\dagger \mathbf{u}_l|^2}{\sum_{j \neq l} p_j |\mathbf{h}_l^\dagger \mathbf{u}_j|^2 + \sigma_l^2}.$$

This SINR function depends only on the L pairs of parameters $\{\mathbf{h}_l, \sigma_l\}$. Thus, an equivalent form can be obtained in terms of the normalized parameter pairs $\{\tilde{\mathbf{h}}_l, 1\}$, where $\tilde{\mathbf{h}}_l = \frac{\mathbf{h}_l}{\sigma_l}$ [10]. Let us define the matrix \mathbf{G} with entries $G_{lj} = |\tilde{\mathbf{h}}_l^\dagger \mathbf{u}_j|^2$. In terms of the beamforming matrix \mathbf{U} , we also define the (cross-channel interference) matrix $\mathbf{F}(\mathbf{U})$ with entries

$$F_{lj}(\mathbf{U}) = \begin{cases} 0, & \text{if } l = j \\ G_{lj}(\mathbf{U}), & \text{if } l \neq j \end{cases} \quad (2)$$

and

$$\mathbf{v}(\mathbf{U}) = \left(\frac{1}{G_{11}(\mathbf{U})}, \frac{1}{G_{22}(\mathbf{U})}, \dots, \frac{1}{G_{LL}(\mathbf{U})} \right)^\top. \quad (3)$$

Moreover, we assume that \mathbf{F} is irreducible, i.e., each link has at least an interferer. 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-B (beamforming optimization). In this paper, we shall consider the equivalent form of the $\text{SINR}_l(\mathbf{p}, \mathbf{U})$ as

$$\begin{aligned} \text{SINR}_l(\mathbf{p}, \mathbf{U}) &= \frac{p_l |\tilde{\mathbf{h}}_l^\dagger \mathbf{u}_l|^2}{\sum_{j \neq l} p_j |\tilde{\mathbf{h}}_l^\dagger \mathbf{u}_j|^2 + 1} \\ &= \frac{p_l}{(\text{diag}(\mathbf{v}(\mathbf{U}))(\mathbf{F}(\mathbf{U})\mathbf{p} + \mathbf{1}))_l} \end{aligned} \quad (4)$$

where we use \mathbf{F} and \mathbf{v} in (2) and (3), respectively.

In the following, we study optimization problems having two performance metrics that are functions of $\text{SINR}_l(\mathbf{p}, \mathbf{U})$, namely the MSE at the output of a LMMSE (scalar) filter of each user [11]:

$$\text{MSE}_l(\mathbf{p}) = \frac{1}{1 + \text{SINR}_l(\mathbf{p})} \quad (5)$$

and the throughput of each user (assuming the Shannon capacity formula) [12]:

$$r_l(\mathbf{p}) = \log(1 + \text{SINR}_l(\mathbf{p})). \quad (6)$$

III. MAX-MIN WEIGHTED SINR OPTIMIZATION

In this section, we first consider optimizing only power before we consider a joint optimization in power and transmit beamformers. Let $\boldsymbol{\beta}$ be a positive vector, where the l th entry β_l is assigned to the l th link to reflect some priority. Let us consider the following max-min weighted SINR problem:

$$\begin{aligned} & \text{maximize} \quad \min_l \frac{\text{SINR}_l(\mathbf{p})}{\beta_l} \\ & \text{subject to} \quad \mathbf{1}^\top \mathbf{p} \leq \bar{P}, \mathbf{p} \geq \mathbf{0} \\ & \text{variable :} \quad \mathbf{p}. \end{aligned} \quad (7)$$

Next, let us define the following nonnegative matrix:

$$\mathbf{B} = \mathbf{F} + (1/\bar{P})\mathbf{1}\mathbf{1}^\top. \quad (8)$$

We will extensively exploit the spectra of the product of a diagonal nonnegative matrix and \mathbf{B} (e.g., its spectral radius, its quasi-inverse and other properties) in our problem formulation, their solution and algorithm design in this paper. We now use the spectral property of $\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}$ to solve (7) in the following.

A. Optimal Solution and Algorithm

By exploiting a connection between the nonlinear Perron–Frobenius theory in [2] and [3] and the algebraic structure of (7), we can give a closed form solution to (7).¹

Lemma 1: The optimal value and solution of (7) is given by

$$1/\rho(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B})$$

and

$$(\bar{P}/\mathbf{1}^\top \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B})) \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B})$$

respectively.

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

$$\mathbf{p} = (\rho(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}) \mathbf{I} - \text{diag}(\mathbf{v})\mathbf{F})^{-1} \mathbf{v}. \quad (9)$$

The following algorithm computes the solution given in Lemma 1. We let k index discrete time slots.

Algorithm 1: Max-min Weighted SINR

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

$$p_l(k+1) = \left(\frac{\beta_l}{\text{SINR}_l(\mathbf{p}(k))} \right) p_l(k) \quad \forall l. \quad (10)$$

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

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P}/\mathbf{1}^\top \mathbf{p}(k+1). \quad (11)$$

¹A closed-form solution to (7) was first obtained in [7] using a nonnegative (increased dimension) matrix totally different from $\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}$. As such, the algorithmic solution to (7) in [7] is different and is mainly centralized. On the other hand, our solution exploits the DPC algorithm in [1] and is distributed.

Corollary 1: Starting from any initial point $\mathbf{p}(0)$, $\mathbf{p}(k)$ in Algorithm 1 converges geometrically fast to the optimal solution of (7), $(\bar{P}/\mathbf{1}^\top \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B})) \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B})$.

Remark 1: Interestingly, (10) in Algorithm 1 is simply the Distributed Power Control (DPC) algorithm in [1], where the l th user has a virtual SINR threshold of β_l in the downlink transmission. The normalization at Step 2 can be computed centrally at the base station. In case when it cannot be expected to know the total number of users in the system, the normalization can be made distributed using gossip algorithms to compute $\mathbf{1}^\top \mathbf{p}(k+1)$ at each user [13].

B. Uplink–Downlink Duality and Joint Optimization

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

$$\begin{aligned} & \text{maximize} \quad \min_l \frac{\text{SINR}_l(\mathbf{p}, \mathbf{U})}{\beta_l} \\ & \text{subject to} \quad \sum_{l=1}^L p_l \leq \bar{P}, p_l \geq 0, \mathbf{u}_l^\dagger \mathbf{u}_l = 1 \forall l, \\ & \text{variables:} \quad \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \quad \mathbf{p}. \end{aligned} \quad (12)$$

We first review the notion of uplink–downlink duality in [4]–[8]. Here, the uplink transmission uses the normalized channel vector $\tilde{\mathbf{h}}$ and unit noise power as in Section II. The duality theory states that, under a same total power constraint for all the users, the achievable SINR region for a downlink 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 the uplink problem 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.

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

$$\hat{\text{SINR}}_l(\mathbf{q}, \mathbf{U}) = \frac{q_l |\tilde{\mathbf{h}}_l^\dagger \mathbf{u}_l|^2}{\sum_{j \neq l} q_j |\tilde{\mathbf{h}}_j^\dagger \mathbf{u}_l|^2 + 1} \geq \beta_l \gamma \quad (13)$$

for all l . Since $\hat{\text{SINR}}_l$ in (13) only depends on the beamforming vector \mathbf{u}_l , the receive beamforming optimization, with the power fixed at \mathbf{q} , is solved by

$$\mathbf{u}_l^* = \arg \min_{\mathbf{u}_l} \sum_{j \neq l} \frac{G_{jl}(\mathbf{U})}{G_{ul}(\mathbf{U})} q_j + \frac{1}{G_{ul}(\mathbf{U})}, \quad (14)$$

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

$$\mathbf{u}_l^* = \left(\sum_{j \neq l} q_j \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^\dagger + \mathbf{I} \right)^{-1} \tilde{\mathbf{h}}_l. \quad (15)$$

Using this LMMSE receiver, the SINR constraint in (13) is always met with equality, i.e., $\text{SINR}_l(\mathbf{q}, \mathbf{U}) = \beta_l \gamma$ by choosing

$$\mathbf{q} = \arg \max_{\mathbf{1}^\top \mathbf{q} = \bar{P}} \min_l \hat{\text{SINR}}_l(\mathbf{q}, \mathbf{U}) / \beta_l.$$

By the uplink–downlink duality, the LMMSE receiver is also the optimal transmit beamformer in the downlink max-min weighted SINR problem given by (12).

Now, we are ready to use Algorithm 1 to solve the joint power control and beamforming problem in (12) in a fast and distributed fashion as given in the following:

Algorithm 2: Max-min Weighted SINR—Joint Optimization

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

$$q_l(k+1) = \left(\frac{\beta_l}{\hat{\text{SINR}}_l(\mathbf{q}(k), \mathbf{U}(k))} \right) q_l(k) \quad \forall l. \quad (16)$$

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

$$\mathbf{q}(k+1) \leftarrow \mathbf{q}(k+1) \cdot \bar{P} / \mathbf{1}^\top \mathbf{q}(k+1). \quad (17)$$

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

$$\begin{aligned} \mathbf{u}_l(k+1) &= \left(\sum_{j \neq l} q_j(k+1) \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^\dagger + \mathbf{I} \right)^{-1} \tilde{\mathbf{h}}_l \quad \forall l, \\ \mathbf{u}_l(k+1) &\leftarrow \mathbf{u}_l(k+1) / \|\mathbf{u}_l(k+1)\|_2 \quad \forall l. \end{aligned} \quad (18)$$

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

$$p_l(k+1) = \left(\frac{\beta_l}{\text{SINR}_l(\mathbf{p}(k), \mathbf{U}(k+1))} \right) p_l(k) \quad \forall l. \quad (19)$$

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

$$\mathbf{p}(k+1) \leftarrow \mathbf{p}(k+1) \cdot \bar{P} / \mathbf{1}^\top \mathbf{p}(k+1). \quad (20)$$

Theorem 1: Let the optimal power and beamforming matrix in (12) be \mathbf{p}^* and \mathbf{U}^* respectively. Then, starting from any initial point $\mathbf{q}(0)$ and $\mathbf{p}(0)$, $\mathbf{p}(k)$ in Algorithm 2 converges geometrically fast to

$$\mathbf{p}^* = \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}(\mathbf{U}^*)) \mathbf{B}(\mathbf{U}^*))$$

(unique up to a scaling constant).

Remark 2: In Algorithm 2, the uplink power $\mathbf{q}(k)$ converges geometrically fast to $(\bar{P} / \mathbf{1}^\top \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{B}^\top)) \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{B}^\top)$, which is also proportional to $\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{y}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{B})$.

Remark 3: Note that (16) and (19) of Algorithm 2 use the DPC algorithm in [1], where the l th user has a virtual SINR threshold of β_l in both the (virtual) uplink and downlink trans-

mission. When the uplink and downlink channels are reciprocal to each other, \mathbf{q} in (16) is the exact uplink transmit power, and only computing $\mathbf{1}^\top \mathbf{q}(k+1)$ in (17) requires a global coordination. Similar to Algorithm 1, $\mathbf{1}^\top \mathbf{q}(k+1)$ at Step 2 and $\mathbf{1}^\top \mathbf{p}(k+1)$ at Step 5 can be computed centrally at the base station or be made distributed using gossip algorithms [13]. Furthermore, in time-division duplex (TDD) system, where uplink and downlink channels are reciprocal, Step 1 can be implemented using local measurements in a distributed manner.

C. Nonlinear Perron–Frobenius Minimax Characterization

We first establish the following result based on the nonlinear Perron–Frobenius theory and the Friedland–Karlin inequality in [14] and [15] [see (62) in the Appendix], and then discuss how it provides further insight into the analytical solution of (7). The result is also useful when we consider a different reformulation of (7) in Section VI-A.

Lemma 2: Let $\mathbf{A} \in \mathbb{R}^{L \times L}$ be an irreducible nonnegative matrix, $\mathbf{b} \in \mathbb{R}^{L \times 1}$ a nonnegative vector and $\|\cdot\|$ a norm on \mathbb{R}^L with a corresponding dual norm $\|\cdot\|_D$. Then,

$$\begin{aligned} \log \rho(\mathbf{A} + \mathbf{b} \mathbf{c}_*^\top) &= \max_{\|\mathbf{c}\|_D=1} \log \rho(\mathbf{A} + \mathbf{b} \mathbf{c}^\top) \\ &= \max_{\boldsymbol{\lambda} \geq 0, \mathbf{1}^\top \boldsymbol{\lambda} = 1} \min_{\|\mathbf{p}\|=1} \sum_{l=1}^L \lambda_l \log \frac{(\mathbf{A} \mathbf{p} + \mathbf{b})_l}{p_l} \end{aligned} \quad (21)$$

$$= \min_{\|\mathbf{p}\|=1} \max_{\boldsymbol{\lambda} \geq 0, \mathbf{1}^\top \boldsymbol{\lambda} = 1} \sum_{l=1}^L \lambda_l \log \frac{(\mathbf{A} \mathbf{p} + \mathbf{b})_l}{p_l} \quad (22)$$

where the optimal \mathbf{p} in (21) and (22) are both given by $\mathbf{x}(\mathbf{A} + \mathbf{b} \mathbf{c}_*^\top)$, and the optimal $\boldsymbol{\lambda}$ in (21) and (22) 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 4: Lemma 2 is a general version of the Friedland–Karlin spectral radius minimax characterization in [14] and [15]. In particular, if $\mathbf{b} = \mathbf{0}$, we obtain [15, Theorem 3.2].

Using Lemma 2 (let $\mathbf{A} = \text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{F}$, $\mathbf{b} = (\frac{1}{\bar{P}}) \text{diag}(\boldsymbol{\beta}) \mathbf{v}$ and $\mathbf{c}_* = \mathbf{1}$ in Lemma 2), we deduce that the optimal SINR allocation in (7) is a weighted geometric mean of the optimal SINR, where the weights are the normalized Schur product of the uplink power and the downlink power (i.e., the Perron and left eigenvectors of $\text{diag}(\boldsymbol{\beta}) \mathbf{B}$, respectively)³:

$$\prod_{l=1}^L (\text{SINR}_l(\mathbf{p}) / \beta_l)^{p_l q_l / (\beta_l v_l)} = 1 / \rho(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{B}). \quad (23)$$

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

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

³The normalization of the Schur product is done such that $\sum_l \frac{p_l q_l}{\beta_l v_l} = 1$. The uplink and the downlink powers can be obtained as a sub-vector of the normalized Perron and left eigenvectors of a $(L+1) \times (L+1)$ extended coupling matrix, which is different from the $L \times L$ matrix $\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{B}$ [7], [8].

Uplink-downlink Duality Correspondence

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

Finally, Fig. 2 summarizes the connection between the uplink–downlink duality with the nonlinear Perron–Frobenius theorem and the Friedland–Karlin minimax characterization from Section III-C.

IV. WEIGHTED SUM MSE MINIMIZATION

In this section, we fix the beamformers and study the non-convex optimization problem of minimizing the weighted sum of the MSE's of individual data streams using two methods. In our first method (Section IV-B), we show that it can be solved optimally in 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 deduce optimality conditions and propose suboptimal algorithm that exploits the connection with max-min weighted SINR (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 the users. The weighted sum MSE at the output of the LMMSE receiver is given by [11]

$$\sum_{l=1}^L w_l \text{MSE}_l(\mathbf{p}) = \sum_{l=1}^L w_l \frac{1}{1 + \text{SINR}_l(\mathbf{p})} \quad (24)$$

where w_l is some positive weight assigned to the l th link to reflect some priority. Without loss of generality, we assume that \mathbf{w} is a probability vector. The weighted sum MSE minimization problem is given by

$$\begin{aligned} & \text{minimize} && \sum_{l=1}^L w_l \frac{1}{1 + \text{SINR}_l(\mathbf{p})} \\ & \text{subject to} && \sum_{l=1}^L p_l \leq \bar{P}, \quad p_l \geq 0 \quad \forall l, \\ & \text{variables:} && p_l \forall l. \end{aligned} \quad (25)$$

In this section, we denote the optimal power vector to (25) by \mathbf{p}^* .

B. Polynomial-Time Solution for Special Cases

We can rewrite (25) as

$$\text{minimize} \quad \sum_{l=1}^L w_l \left(\frac{\sum_{k \neq l} G_{lk} p_k + 1}{\sum_{k=1}^L G_{lk} p_k + 1} \right)$$

$$\text{subject to} \quad \sum_{l=1}^L p_l \leq \bar{P}, \quad p_l \geq 0 \quad \forall l. \quad (26)$$

It can be shown that the total power constraint in (25) and (26) are tight at optimality (see Appendix IX-E), which we exploit to transform (26) in the variables \mathbf{p} into another optimization problem that can be used to solve (26) optimally. To proceed further, we need to introduce the notion of quasi-invertibility of a nonnegative matrix in [16], which will be useful in solving (26) optimally.

Definition 1 (Quasi-Invertibility): A square nonnegative matrix \mathbf{A} is a quasi-inverse of a square nonnegative matrix $\tilde{\mathbf{B}}$ if $\mathbf{A} - \tilde{\mathbf{B}} = \mathbf{A}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}\mathbf{A}$. Furthermore, $(\mathbf{I} - \tilde{\mathbf{B}})^{-1} = \mathbf{I} + \mathbf{A}$ [16].

For a given square nonnegative matrix \mathbf{A} , we first compute either $\mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}$ or $(\mathbf{I} + \mathbf{A})^{-1}\mathbf{A}$. If this computed matrix is nonnegative, we say that $\tilde{\mathbf{B}}$ exists (as from Definition 1, this computed matrix is $\tilde{\mathbf{B}} \geq \mathbf{0}$). We now apply the definition of quasi-invertibility to the matrix $\text{diag}(\mathbf{v})\mathbf{B}$ with \mathbf{B} in (8), and study the existence of $\tilde{\mathbf{B}}$, which can interestingly be associated with the interference and SNR regime. In the case when the total power \bar{P} is large, (high SNR regime) or when interference (off-diagonals of \mathbf{F}) is large, it is deduced in the following that $\tilde{\mathbf{B}}$ does not exist.

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

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

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

Example 1: As a numerical example for a ten-user IEEE 802.11b network, we experiment with a total power constraint of 33 mW and 1 W (the largest possible value allowed in IEEE 802.11b) and equal noise power of 1 mW. Averaging over 10 000 random channel coefficient instances ($G_{ll} = 1.5$ and off-diagonal G_{lj} uniformly distributed between 0.01 and 0.3), the percentage of instances where $\tilde{\mathbf{B}}$ exists is 99% and 65% corresponding to the total power constraint of 33 mW and 1 W, respectively.

In the rest of Section IV, we focus on the case when $\tilde{\mathbf{B}}$ exists. We next solve (26) in the following. Let us define

$$\mathbf{z} = (\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B}) \mathbf{p}. \quad (27)$$

Then, we can rewrite (26) in terms of the new variable \mathbf{z} as

$$\begin{aligned} & \text{minimize} \quad \sum_{l=1}^L w_l \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} \\ & \text{subject to} \quad z_l \geq (\tilde{\mathbf{B}}\mathbf{z})_l, \quad l = 1, \dots, L \end{aligned} \quad (28)$$

where the constraints in (28) are due to the nonnegativity of \mathbf{p} , since, using Definition 1,

$$\mathbf{p} = (\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B})^{-1} \mathbf{z} = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z} \geq \mathbf{0}.$$

In the following, we leverage (28) to derive useful lower bounds to (25), investigate special cases, and finally solve (25) using polynomial-time algorithms.

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 (25).

Theorem 2: If $\tilde{\mathbf{B}}$ exists,

$$\begin{aligned} & \sum_{l=1}^L w_l \frac{1}{1 + \text{SINR}_l(\mathbf{p})} \\ & \geq \left(\frac{1}{1 + 1/\rho(\text{diag}(\mathbf{v})\mathbf{B})} \right) \|\mathbf{w}\|_{\infty}^{\mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B})\mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})} \end{aligned} \quad (29)$$

for all feasible \mathbf{p} in (25).

Equality is achieved if and only if

$$\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B}).$$

Thus, $\text{SINR}_l(\mathbf{p}^*) = \frac{1}{\rho(\text{diag}(\mathbf{v})\mathbf{B})}$ for all l . In this case, $\mathbf{p}^* = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B})$ solves (25).

Interestingly, Theorem 2 shows that solving (7) with $\beta = 1$ can be seen as an approximation method to solving (25) suboptimally, but with an approximation guarantee. In particular, by taking the logarithm of the objective function of (25), the positive quantity

$$\|\mathbf{w}\|_{\infty}^{\mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B})\mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})}$$

can be interpreted as the approximation ratio of Algorithm 1 with $\beta = 1$ in solving (25).

Remark 5: Based on Theorem 2, we obtain a connection between the min-max MSE problem, i.e., $\min_{\mathbf{p}} \max_l \frac{1}{(1 + \text{SINR}_l(\mathbf{p}))}$, and the weighted sum MSE optimization in (25). By the max-min characterization of $\rho(\tilde{\mathbf{B}})$ ⁴:

$$\max_{\mathbf{z} > \mathbf{0}} \min_l \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} = \min_{\mathbf{z} > \mathbf{0}} \max_l \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} = \rho(\tilde{\mathbf{B}}), \quad (30)$$

the optimal value of the min-max MSE problem is $\rho(\tilde{\mathbf{B}}) = 1/(1 + 1/\rho(\text{diag}(\mathbf{v})\mathbf{B}))$. It thus follows from Theorem 2 that (7) with $\beta = 1$ yields the equivalent solution as the min-max MSE problem and the optimal sum MSE with $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$.

We will establish further connections and equivalence results between the max-min weighted SINR problem and the two non-

convex problems (25) and (36) 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. Polynomial-Time Solution and Fast Algorithm

In fact, the existence of $\tilde{\mathbf{B}}$ allows us to delineate cases of (25) that can be solved efficiently [through (28)] from the general problem. Although (28) is nonconvex, it is equivalent to a convex problem by making a change of variables $\tilde{z}_l = \log z_l$ for all l . This is allowed since $\mathbf{z} > \mathbf{0}$. Hence, (28) is equivalent to the problem:

$$\begin{aligned} & \text{minimize} \quad \sum_{l=1}^L w_l (\tilde{\mathbf{B}}e^{\tilde{\mathbf{z}}})_l / e^{\tilde{z}_l} \\ & \text{subject to} \quad (\tilde{\mathbf{B}}e^{\tilde{\mathbf{z}}})_l / e^{\tilde{z}_l} \leq 1, \quad l = 1, \dots, L, \\ & \text{variables:} \quad \tilde{z}_l \quad \forall l, \end{aligned} \quad (31)$$

which is strictly convex in $\tilde{\mathbf{z}}$. Though the constraint set in (31) is unbounded, the optimal solution to (31) cannot have $\tilde{z}_l^* = -\infty$ for some l since, at optimality, $\mathbf{z}^* > \mathbf{0}$. Hence, (31) can be solved optimally by an interior-point method (e.g., see [18]), thus yielding the optimal solution to (28) after the change in variables. The following result gives the optimal solution to (25). A special case of (31) also yields a fast algorithm (Algorithm 3 below) that computes the optimal solution by leveraging the nonlinear Perron–Frobenius theory in [3] instead of using standard convex optimization algorithms to solve (31).

Theorem 3: If $\tilde{\mathbf{B}}$ exists, then the optimal solution to (25) is given by $\mathbf{p}^* = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^* \geq \mathbf{0}$, where \mathbf{z}^* is the optimal solution to (28). Furthermore, if $\mathbf{p}^* > \mathbf{0}$ (i.e., none of the constraints in (28) is tight), \mathbf{z}^* fulfills

$$z_l^* = \sqrt{\frac{w_l \sum_{j \neq l} \tilde{B}_{lj} z_j^*}{\sum_{j \neq l} w_j \tilde{B}_{jl} / z_j^*}} \quad (32)$$

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

Now, the righthand-side of (32) is a homogeneous and monotone concave function (cf. proof of Corollary 2). We can leverage the nonlinear Perron–Frobenius theory in [3] to propose the following (step size free) algorithm that computes \mathbf{z}^* in Theorem 3, and hence the optimal solution of (25).

Algorithm 3: Weighted Sum MSE Minimization

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

$$z_l(k+1) = \sqrt{\frac{w_l \sum_{j \neq l} \tilde{B}_{lj} z_j(k)}{\sum_{j \neq l} w_j \tilde{B}_{jl} / z_j(k)}} \quad \forall l. \quad (33)$$

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

$$\mathbf{z}(k+1) \leftarrow \mathbf{z}(k+1) / \mathbf{1}^\top \mathbf{z}(k+1). \quad (34)$$

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

$$\begin{aligned} p_l(k+1) &= \frac{\text{SINR}_l(\mathbf{p}(k))}{1 + \text{SINR}_l(\mathbf{p}(k))} z_l(k+1) \\ &\quad \times \bar{P} / \mathbf{1}^\top (\mathbf{I} - \tilde{\mathbf{B}}) \mathbf{z}(k+1) \quad \forall l. \end{aligned} \quad (35)$$

⁴The 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 [14], [17].

The following result establishes the convergence and optimality of Algorithm 3.

Corollary 2: If $\tilde{\mathbf{B}}$ exists and $\mathbf{p}^* > \mathbf{0}$ in (25), Algorithm 3 converges geometrically fast to the unique fixed point \mathbf{z}^* and \mathbf{p}^* in Theorem 3 from any initial point $\mathbf{z}(0)$.

V. WEIGHTED SUM RATE MAXIMIZATION

In this section, we fix the beamformers and consider the weighted sum rate of all the users as the next performance metric to be optimized. Similar to Section IV, we first show that it can be solved optimally when $\tilde{\mathbf{B}} \geq \mathbf{0}$, and then look at the general case in Section VI. Further, we quantify a connection of the weighted sum rate maximization and the weighted sum MSE minimization.

A. Problem Statement

The nonconvex weighted sum rate maximization problem is given by

$$\begin{aligned} & \text{maximize} && \sum_{l=1}^L w_l \log(1 + \text{SINR}_l(\mathbf{p})) \\ & \text{subject to} && \sum_{l=1}^L p_l \leq \bar{P}, p_l \geq 0 \quad \forall l, \\ & \text{variables:} && p_l \quad \forall l. \end{aligned} \quad (36)$$

In this section, we denote the optimal power vector to (36) by \mathbf{p}^* .

B. Exact Solution and Fast Algorithm

We can rewrite (36) to be equivalent to

$$\begin{aligned} & \text{minimize} && \prod_{l=1}^L \left(\frac{\sum_{k \neq l} G_{lk} p_k + 1}{\sum_{k=1}^L G_{lk} p_k + 1} \right)^{w_l} \\ & \text{subject to} && \sum_{l=1}^L p_l \leq \bar{P}, \quad p_l \geq 0 \forall l. \end{aligned} \quad (37)$$

Similar to Section IV, if $\tilde{\mathbf{B}}$ is the quasi-inverse of $\tilde{\mathbf{B}}$, we can rewrite (37) as

$$\begin{aligned} & \text{minimize} && \prod_{l=1}^L \left(\frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} \right)^{w_l} \\ & \text{subject to} && z_l \geq (\tilde{\mathbf{B}}\mathbf{z})_l, \quad l = 1, \dots, L \end{aligned} \quad (38)$$

where \mathbf{z} is given by (27). Similarly to (28), (38) can be solved optimally by first a logarithmic change in variables and then using convex optimization algorithms, e.g., the interior-point method. The following result gives the condition under which (36) is solved optimally when $\tilde{\mathbf{B}}$ exists (in the weak interference regime).

Theorem 4: If $\tilde{\mathbf{B}}$ exists, then the optimal solution to (36) is given by $\mathbf{p}^* = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^* \geq \mathbf{0}$, where \mathbf{z}^* is the optimal solution to (38). Furthermore, if $\mathbf{p}^* > \mathbf{0}$ (i.e., none of the constraints in (38) is tight), \mathbf{z}^* fulfills

$$z_l^* = \frac{w_l}{\sum_{j=1}^L w_j \tilde{B}_{jl} / (\tilde{\mathbf{B}}\mathbf{z}^*)_j} \quad (39)$$

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

As in the previous, using the nonlinear Perron–Frobenius theory, the following algorithm computes the optimal solution of (36) when $\mathbf{p}^* > \mathbf{0}$ in (36).

Algorithm 4: Weighted Sum Rate Maximization

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

$$z_l(k+1) = \frac{w_l}{\sum_{j=1}^L w_j \tilde{B}_{jl} / (\tilde{\mathbf{B}}\mathbf{z}(k))_j} \quad \forall l. \quad (40)$$

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

$$\mathbf{z}(k+1) \leftarrow \mathbf{z}(k+1) / \mathbf{1}^\top \mathbf{z}(k+1). \quad (41)$$

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

$$\begin{aligned} p_l(k+1) = & \frac{\text{SINR}_l(\mathbf{p}(k))}{1 + \text{SINR}_l(\mathbf{p}(k))} z_l(k+1) \\ & \times \bar{P} / \mathbf{1}^\top (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}(k+1) \quad \forall l. \end{aligned} \quad (42)$$

The following result establishes the convergence and optimality of Algorithm 4.

Corollary 3: If $\tilde{\mathbf{B}}$ exists and $\mathbf{p}^* > \mathbf{0}$ in (36), Algorithm 4 converges geometrically fast to the unique fixed point \mathbf{z}^* and \mathbf{p}^* in Theorem 4 from any initial point $\mathbf{z}(0)$.

Next, we connect the three optimization problems given in (25), (36) and (7) with $\beta = 1$.

C. Connection Between Weighted Sum Rate, Sum MSE and Max-Min SINR

Applying the arithmetic-geometric mean inequality (see (64) in the Appendix) to connect (28) and (38), we deduce that the weighted sum rate maximization has the same optimal power as the weighted sum MSE minimization when

$$\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B}).$$

Furthermore, from Remark 5, this optimal power is also the solution to the max-min weighted SINR problem (with $\beta = 1$).

Example 2: We give a simple illustrative example for the two user case. The channel gains are given by $G_{11} = 0.73, G_{12} = 0.04, G_{21} = 0.03, G_{22} = 0.89$ and the AWGN for the first and second user are 0.1 and 0.3 respectively. The total power is $2W$. It can be easily checked that $\tilde{\mathbf{B}}$ exists. We solve (7) with $\beta = 1$. We then set $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$ in both (25) and (36), and find their corresponding optimal solution by Algorithms 3 and 4. These are then plotted on the respective achievable MSE and rate region. Fig. 3 shows that the optimal weighted sum MSE point is the same as the weighted sum MSE point evaluated using the max-min SINR power control. Fig. 4 shows that the optimal weighted sum rate point is the same as the weighted sum rate point evaluated using the max-min SINR power control.

In Section VI-B, we continue to highlight this connection between (7) with $\beta = 1$ and the two problems (25) and (36) with $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$ in the general case, i.e., without the $\tilde{\mathbf{B}}$ existence condition.

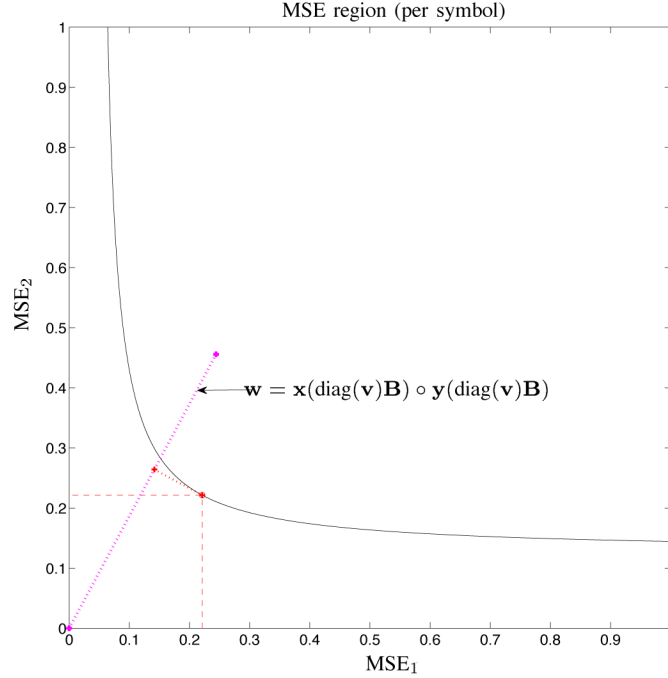


Fig. 3. A two-user example with the vector \mathbf{w} superimposed on the achievable MSE region. Its perpendicular coincides with the optimal point in the achievable region. From a geometrical perspective, the weighted sum MSE point evaluated at the solution of the max-min SINR power control finds the largest hypercube that supports the Pareto boundary of the achievable MSE region. When $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$, the weighted sum MSE problem and the max-min SINR problem have the same solution.

VI. DISTRIBUTED OPTIMIZATION IN SINR

In this section, we first study the optimality conditions of (25) and (36) 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.

A. Optimality Conditions in SINR

First, we reformulate both (36) and (25) as optimization problems having a new set of variables (in the SINR domain) and a spectral radius constraint involving \mathbf{B} in (8). The new formulation permits us to derive optimality conditions, propose fast algorithms and further connect the max-min SINR problem to (25) and (36). In particular, by leveraging the beamforming result in Section III-B, we also address the optimal beamformer to (25) and (36).

First, let us consider the following optimization problem:

$$\begin{aligned} & \text{maximize} && f(\boldsymbol{\gamma}) \\ & \text{subject to} && \rho(\text{diag}(\boldsymbol{\gamma} \circ \mathbf{v})\mathbf{B}) \leq 1 \\ & \text{variables:} && \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \quad \boldsymbol{\gamma}. \end{aligned} \quad (43)$$

Let us denote the optimal solution of (43) by $\boldsymbol{\gamma}^*$. Now, the optimal power to the weighted MSE minimization problem in (25) and the weighted sum rate maximization problem in (36) can be implicitly obtained by solving (43) using, respectively,

$$f(\boldsymbol{\gamma}) = -\sum_{l=1}^L w_l / (1 + \gamma_l) \quad (\text{weighted sum MSE}) \quad (44)$$

and

$$f(\boldsymbol{\gamma}) = \sum_{l=1}^L w_l \log(1 + \gamma_l) \quad (\text{weighted sum rate}) \quad (45)$$

in (43). We summarize this in the following result.

Lemma 5: For a feasible $\boldsymbol{\gamma}$ in (43), a feasible power in (25) and in (36) can be computed by

$$\mathbf{p} = (\mathbf{I} - \text{diag}(\boldsymbol{\gamma} \circ \mathbf{v})\mathbf{F})^{-1} \text{diag}(\boldsymbol{\gamma})\mathbf{v}. \quad (46)$$

This means that if we first solve (43) to obtain $\boldsymbol{\gamma}^*$, the optimal power in (25) and (36) can be obtained using (46) in Lemma 5.

We continue with a further change of variable technique in (43). For $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_L)^\top > 0$, let

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

i.e., $\gamma = e^{\tilde{\gamma}}$. Then, (43) is equivalent to

$$\begin{aligned} & \text{maximize} && f(e^{\tilde{\gamma}}) \\ & \text{subject to} && \log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \leq 0, \\ & \text{variables:} && \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L], \\ & && \tilde{\gamma} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_n)^\top \in \mathbb{R}^L. \end{aligned} \quad (47)$$

Note that the constraint set in (47) is an unbounded convex set [19] (see also [18, Sec. 4.5]). Let us denote the optimal solution of (47) by $\tilde{\boldsymbol{\gamma}}^*$. Note that $\tilde{\gamma}_l^* = \log \gamma_l^*$ for all l .

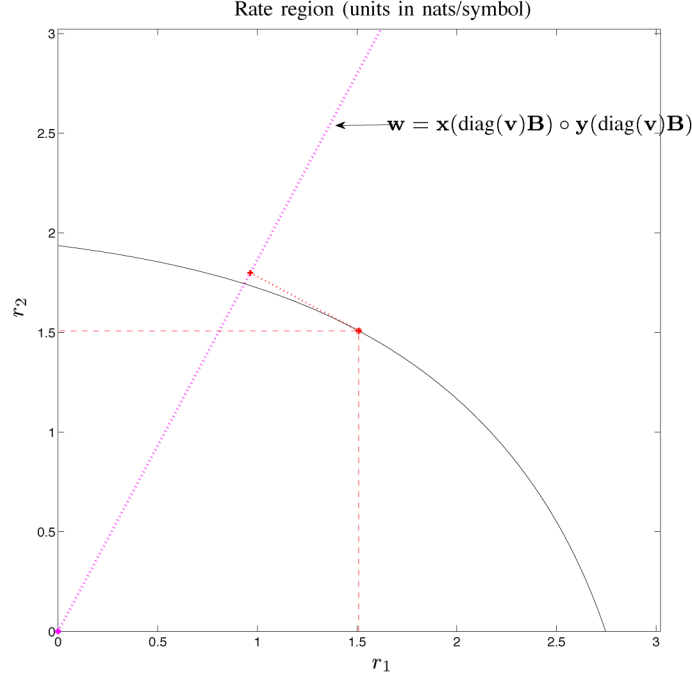


Fig. 4. A two-user example with the vector \mathbf{w} superimposed on the achievable rate region. Its perpendicular coincides with the optimal point in the achievable region. From a geometrical perspective, the weighted sum rate point evaluated at the solution of the max-min SINR power control finds the largest hypercube that is contained inside the achievable rate region. When $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$, the weighted sum rate problem and the max-min SINR problem have the same solution.

Suppose we know the optimal \mathbf{U} in (47). Define the first order derivative of the objective function in (47) with respect to $\tilde{\gamma}$ by $f'(e^{\tilde{\gamma}}) \in \mathbb{R}^L$. The optimality condition of (47) is given as follows.

Theorem 5: The optimal solution of (47), $\tilde{\gamma}^*$, satisfies

$$\mathbf{x}(\text{diag}(e^{\tilde{\gamma}^*} \circ \mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\gamma}^*} \circ \mathbf{v})\mathbf{B}) = \frac{f'(e^{\tilde{\gamma}^*})}{\mathbf{1}^\top f'(e^{\tilde{\gamma}^*})} \quad (48)$$

and

$$\rho(\text{diag}(e^{\tilde{\gamma}^*} \circ \mathbf{v})\mathbf{B}) = 1. \quad (49)$$

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

Our results in Sections IV and V established the connection between these three problems only in the weak interference or low signal-to-noise regime. Now, we establish the connection for the general case. From (48) in Theorem 5, we see that if the optimal SINR_l is equal for all l , then the optimal solution to (47) can be obtained by solving the max-min SINR problem in (12) since it means $\tilde{\gamma}^*$ is proportional to $\mathbf{1}$ and we have $\mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B}) = \mathbf{w}$ in (48) (cf. Remark 5).

Now, by extending this observation to the constraint set in (43), 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 (43) is given by the LMMSE filter in (15), where

$$\mathbf{q} = (\mathbf{I} - \text{diag}(\boldsymbol{\gamma}^* \circ \mathbf{v})\mathbf{F}^\top)^{-1} \text{diag}(\boldsymbol{\gamma}^*)\mathbf{v}$$

in (15).

Further, using (21) in Lemma 2, we see that a simple proof to the uplink–downlink duality is to observe that

$$\begin{aligned} \rho(\text{diag}(\boldsymbol{\gamma} \circ \mathbf{v})(\mathbf{F} + (1/\bar{P})\mathbf{1}\mathbf{1}^\top)) \\ = \rho(\text{diag}(\boldsymbol{\gamma} \circ \mathbf{v})(\mathbf{F}^\top + (1/\bar{P})\mathbf{1}\mathbf{1}^\top)) = 1 \end{aligned} \quad (50)$$

where we have used both the fact that $\rho(\mathbf{A}\mathbf{C}) = \rho(\mathbf{C}\mathbf{A})$ and $\rho(\mathbf{A}) = \rho(\mathbf{A}^\top)$ for any irreducible nonnegative matrices \mathbf{A} and \mathbf{C} [17]. The first and second spectral radius expression in (50) correspond to the uplink and the downlink systems, respectively. Note that (50) characterizes all *pareto efficient* achievable SINR in both the uplink and downlink systems (since the spectral radius increases monotonically in the entry of a matrix [17]).

Interestingly, using the weighted geometric mean equivalence result in (23), (12) is equivalent to an optimization problem given in the form of (43) by letting

$$f(\boldsymbol{\gamma}) = \sum_{l=1}^L x_l (\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}) y_l (\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}) \log \gamma_l, \quad (\text{max-min weighted SINR}) \quad (51)$$

and its corresponding transformed problem in (47), which is convex in $\tilde{\gamma}$, can be solved efficiently. We will now exploit this fact and our max-min weighted SINR algorithm in Section III-A to propose a fast algorithm that can be made distributed to solve for a local optimal solution to (25) and (36) in the following. Conditions under which they are global optimal are also given.

B. A Weighted Proportional SINR Algorithm

We first state our weighted proportional SINR algorithm, which uses Algorithm 1 as a sub-module and adapts the weight parameter $\boldsymbol{\beta}$ of Algorithm 1 iteratively to solve either (25) or (36).

Algorithm 5: Weighted Proportional SINR

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

$$\mathbf{m}(k+1) = \frac{f'(e^{\tilde{\gamma}(k)})}{\mathbf{1}^\top f'(e^{\tilde{\gamma}(k)})}. \quad (52)$$

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

$$\begin{aligned} & \text{maximize} \quad \sum_{l=1}^L m_l(k+1) \tilde{\gamma} \\ & \text{subject to} \quad \log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v}) \mathbf{B}) \leq 0. \end{aligned} \quad (53)$$

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

Depending on the $f'(\tilde{\gamma}(k))$ used in (52), Algorithm 5 can compute the optimal solution of (25) or (36) when the initial point is sufficiently close to the optimal solution. This is stated in the following result.

Theorem 6: For any $\tilde{\gamma}(0)$ in a sufficiently close neighborhood of $\tilde{\gamma}^*$, $\mathbf{p}(k+1)$ in Algorithm 5 converges to the optimal solution of (25) and (36) for $f'(e^{\tilde{\gamma}(k)})$ with, respectively,

$$\left(f'(e^{\tilde{\gamma}(k)})\right)_l = -w_l \frac{e^{\tilde{\gamma}_l(k)}}{(1 + e^{\tilde{\gamma}_l(k)})^2}$$

and

$$\left(f'(e^{\tilde{\gamma}(k)})\right)_l = w_l \frac{e^{\tilde{\gamma}_l(k)}}{1 + e^{\tilde{\gamma}_l(k)}}.$$

The convergence in Theorem 6 is proved only for initial points in the neighborhood of the optimal solution. However, we now state a result stronger than Theorem 6 (for a more relaxed initialization) that also highlights the connection between the max-min weighted SINR problem and the two nonconvex problems (25) and (36).

Corollary 4: If $\tilde{\gamma}_l^*$ are equal for all l and $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$, then $\mathbf{p}(k+1)$ in Algorithm 5 converges to the optimal solution of (25) and (36) using the respective $f'(e^{\tilde{\gamma}(k)})$ in Theorem 6, from any initial point $\tilde{\gamma}(0)$ such that $\tilde{\gamma}_l(0)$ are equal for all l .

Remark 6: This result is similar in flavor to the result in Section V-C, but is more general (without the existence condition of $\tilde{\mathbf{B}}$). Corollary 4 highlights a special case of (25) and (36) whose solution can be obtained by Algorithm 5 (a polynomial time algorithm) despite the nonconvexity.

C. Distributed Optimization

Although Algorithm 5 is centralized (due to computing (53)), it can be made distributed by leveraging the output of Algorithm 1. In particular, we use the approximate projected gradient method to obtain a distributed solution to solve (53). Recall that the gradient $\mathbf{g} \in \mathbb{R}^L$ of $\log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$ at $\tilde{\gamma}$ satisfies

$$\log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \geq \log \rho(\text{diag}(e^{\hat{\gamma}} \circ \mathbf{v})\mathbf{B}) + \mathbf{g}^\top (\tilde{\gamma} - \hat{\gamma}) \quad (54)$$

⁵The 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 $\tilde{\gamma}$.

for any feasible $\tilde{\gamma}$. In fact, \mathbf{g} is given by [15]

$$\mathbf{g} = \mathbf{x} \left(\text{diag}(e^{\hat{\gamma}} \circ \mathbf{v})\mathbf{B} \right) \circ \mathbf{y} \left(\text{diag}(e^{\hat{\gamma}} \circ \mathbf{v})\mathbf{B} \right), \quad (55)$$

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 $\boldsymbol{\beta} = e^{\tilde{\gamma}}$ in (7) (cf. Fig. 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 SINR

Set the step size $\nu(0) \in (0, 1)$.

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

$$\mathbf{m}(k+1) = \frac{f'(e^{\tilde{\gamma}(k)})}{\mathbf{1}^\top f'(e^{\tilde{\gamma}(k)})}. \quad (56)$$

2) Set the downlink power and uplink power output of Algorithm 2 using the input weight parameter ($\boldsymbol{\beta} = e^{\tilde{\gamma}(k)}$) as $\mathbf{p}(k)$ and $\mathbf{q}(k)$, respectively.

3) Each l th user computes:

$$\begin{aligned} \tilde{\gamma}_l(k+1) &= \tilde{\gamma}_l(k) + \log(\text{SINR}_l(\mathbf{p}(k)) / \beta_l) + \nu(k) \\ &\times \left(\frac{m_l(k+1) (e^{\tilde{\gamma}(k)} \circ \mathbf{v})_l \mathbf{p}(k)^\top \text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})^{-1} \mathbf{q}(k)}{(\mathbf{p}(k) \circ \mathbf{q}(k))_l} - 1 \right) \end{aligned} \quad (57)$$

for all l . Update $\nu(k+1)$ according to Theorem 7.

Theorem 7: Let $\mathbf{x}(k) = \mathbf{x}(\text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})\mathbf{B})$ and $\mathbf{y}(k) = \mathbf{y}(\text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})\mathbf{B})$ be the optimal solution of Algorithm 2 when $\boldsymbol{\beta} = e^{\tilde{\gamma}(k)}$. Let $\sqrt{\mathbf{m}(k+1)} = (\sqrt{m_1(k+1)}, \dots, \sqrt{m_L(k+1)})^\top$, where $m_l(k+1) = \frac{(f'(e^{\tilde{\gamma}(k)}))_l}{\mathbf{1}^\top f'(e^{\tilde{\gamma}(k)})}$ for all l . Suppose the output of Algorithm 2 (at Step 2 of Algorithm 6) satisfies

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

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. For any $\tilde{\gamma}(0)$ in a sufficiently close neighborhood of $\tilde{\gamma}^*$ and a sequence of step sizes $\nu(k) > 0$, $\{\nu(k)\}$, that satisfies

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

$\mathbf{p}(k+1)$ in Algorithm 6 converges to within a closed neighborhood of the optimal solution of (25) and (36) using the respective $f'(e^{\tilde{\gamma}(k)})$ in Theorem 6.

Remark 7: 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 SINR 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 7 states that these accumulated approximation

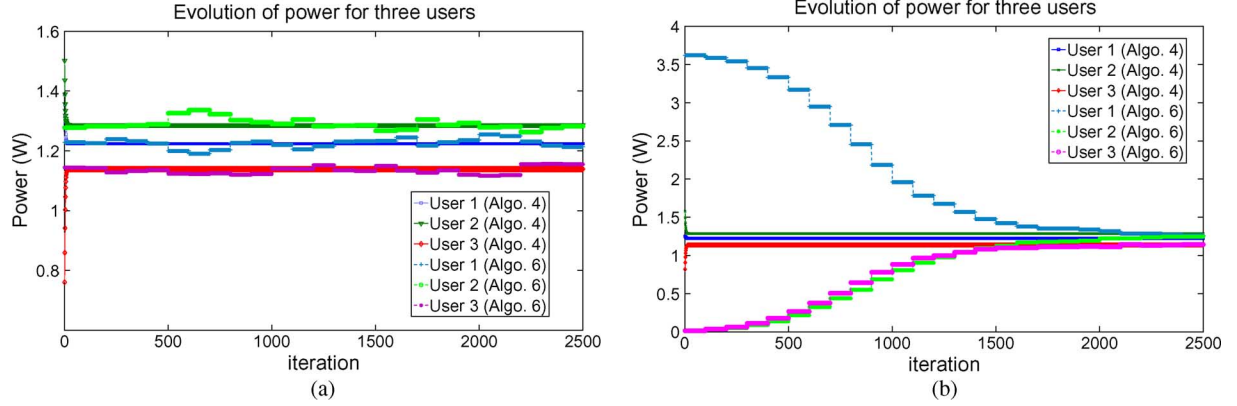


Fig. 5. Illustration of the convergence of Algorithm 3 and Algorithm 6 with an initial SINR vectors that are closer to, in (a), and further from, in (b), the optimal solution.

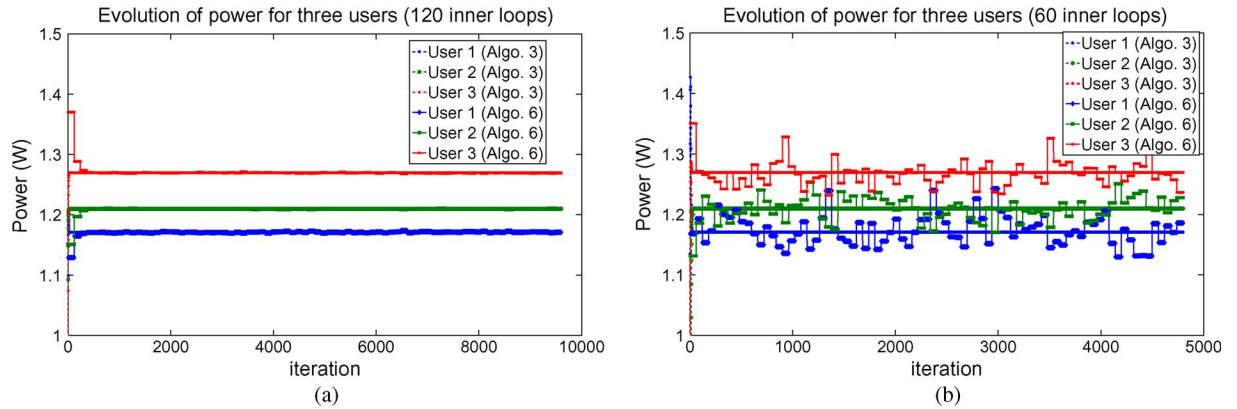


Fig. 6. Illustration of the convergence of Algorithm 4 and Algorithm 6 with different number of inner loops executed by Algorithm 2 of (a) 150 and (b) 70.

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. For example, in computing $\mathbf{p}(k)$, a practical stopping rule at Step 2 of Algorithm 6 is $\|\mathbf{p}(t) - \mathbf{p}(t-1)\| < \epsilon$, where t index discrete time slots of Algorithm 2 and $\epsilon > 0$ is sufficiently small, or a sufficiently long fixed number of inner loops. At Step 3, both the normalized $\mathbf{m}(k)$ and $\mathbf{p}(k)^\top \text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})^{-1} \mathbf{q}(k)$ can be computed centrally at the base station or be made distributed by a gossip algorithm.

VII. NUMERICAL EXAMPLES

In this section, we fix the beamformers and consider optimizing only powers. We evaluate the performance of Algorithm 3, Algorithm 4, and Algorithm 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:

$$\mathbf{G} = \begin{bmatrix} 0.73 & 0.14 & 0.13 \\ 0.15 & 0.69 & 0.12 \\ 0.15 & 0.12 & 0.79 \end{bmatrix}. \quad (58)$$

We set the total power constraint as $\bar{\mathbf{p}} = 3.65$ W and the noise power of each user be 1 W. The weight vector is given by $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$. It is easily verified that $\tilde{\mathbf{B}}$ exists. Moreover, both the optimal solution to (25) and (36) are attained at the equal SINR allocation of 0.673 for the three users (equal to the solution of (7) with $\beta = 1$), where

$\mathbf{p}^* = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) = [1.2238 \ 1.2870 \ 1.1392]^\top$ W. Thus, the optimal sum rate is 0.5144 nats/symbol.

Fig. 5 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 $\nu(k+1) = \frac{\nu(k)}{k+1}, k \geq 0$. In Fig. 5(a) and (b), we set the initial SINR 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 Corollary 3) 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 7). 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 SINR vector is closer to the optimal solution.

Next, we evaluate the performance of using a constant step size of 0.02 in Algorithm 6, and evaluate its performance to solve (25) for different number of inner loops executed at Step 2 of Algorithm 6. The weight vector is given by $\mathbf{w} = [\frac{1}{3} \ \frac{1}{3} \ \frac{1}{3}]^\top$. The optimal solution to (25) is attained at the point $[1.1739 \ 1.2106 \ 1.2656]^\top$. The initial SINR vectors in Algorithm 6 are set close to the optimal solution. Fig. 6 plots the evolution of the power for the three users that run Algorithm 3 and Algorithm 6. Fig. 6(a) and (b) illustrates the cases when we

run Algorithm 2 for 120 and 60 iterations before it terminates, respectively. We observe that if the number of inner loops to execute Algorithm 2 is smaller than a certain threshold, the iterates produced by Algorithm 6 tend towards points in the neighborhood of the optimal solution, but Algorithm 6 may not converge to the optimal solution due to the approximate gradient error. (In Fig. 6(b), less than 5% deviation from the optimal solution is observed.)

VIII. CONCLUSION

Maximizing the minimum weighted SINR, 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 precise connection between these three problems using nonnegative matrix theory, nonlinear Perron–Frobenius theory and the arithmetic-geometric mean inequality. Under sufficiently weak interference, we showed that the weighted sum MSE minimization and the sum rate maximization problems can be solved optimally using fast algorithms (and without any parameter 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 SINR algorithm that leveraged our fast max-min weighted SINR algorithm and the uplink–downlink duality to solve for the local optimal solution of these two nonconvex problems in a distributed manner. Numerical analysis highlighted the robust performance of our algorithms in finding the global optimal solution of these nonconvex problems. All the proposed algorithms in this paper are applicable to the MISO case. As future work, it will be interesting to extend our work to the MIMO case and go from empirical evidence to analytic understanding on why Algorithm 5 (centralized version) and Algorithm 6 (distributed version) never fail to find the global optimal solution from any initial point.

APPENDIX

A. Proof of Lemma 1

Proof: Our proof is based on the nonlinear Perron–Frobenius theory [2], [3]. We let the optimal weighted max-min SINR(\mathbf{p}^*) in (7) be τ^* . A key observation is that all the SINR constraints are tight at optimality. This implies, at optimality of (7),

$$\frac{(p_l^*/\sum_l p_l^*)}{\sum_{j \neq l} v_l F_{lj}(p_l^*/\sum_l p_l^*) + (v_l/\sum_l p_l^*)} = \beta_l \tau^* \quad (59)$$

for all l . Letting $\mathbf{s}^* = \left(\frac{1}{\sum_l p_l^*}\right) \mathbf{p}^*$, (59) can be rewritten as

$$(1/\tau^*) \mathbf{s}^* = \text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{F} \mathbf{s}^* + \left(1/\sum_l p_l^*\right) \text{diag}(\boldsymbol{\beta}) \mathbf{v}. \quad (60)$$

We first state the following conditional eigenvalue lemma:

Lemma 6 (Conditional Eigenvalue [2, 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 $\lambda \mathbf{s} = \mathbf{A} \mathbf{s} + \mathbf{b}$, $\lambda \in \mathbb{R}$, $\mathbf{s} \geq \mathbf{0}$, $\sum_l s_l = 1$, 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 $\lambda = \frac{1}{\tau^*}$, $\mathbf{A} = \text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{F}$, $\mathbf{b} = \left(\frac{1}{P}\right) \text{diag}(\boldsymbol{\beta}) \mathbf{v}$ in Lemma 6 and noting that $\sum_l s_l^* = 1$ shows that $\mathbf{p}^* = \left(\frac{\tau^*}{P}\right) \mathbf{x} \left(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) (\mathbf{F} + \left(\frac{1}{P}\right) \mathbf{1}\mathbf{1}^\top)\right)$ is a fixed point of (60).

B. Proof of Corollary 1

Proof: The fixed point in Lemma 6 is also a unique fixed point of [2], [3]

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

and can be obtained by

$$\mathbf{s}(k+1) = \frac{\mathbf{A} \mathbf{s}(k) + \mathbf{b}}{\|\mathbf{A} \mathbf{s}(k) + \mathbf{b}\|_1}.$$

Applying this to the system of equations in (60) shows that Algorithm 1 converges to the fixed point geometrically fast from any initial point. ■

C. Proof of Theorem 1

Proof: We will show that Algorithm 2 produces \mathbf{q}^* and \mathbf{U}^* (optimal uplink power and receive beamformer), and this \mathbf{U}^* in turn produces \mathbf{p}^* . The uplink max-min weighted SINR problem is given by

$$\max_{\mathbf{1}^\top \mathbf{q} = P} \min_l \text{SINR}_l(\mathbf{q}, \mathbf{U})/\beta_l$$

which is equivalent to

$$\min_{\mathbf{1}^\top \mathbf{q} = P} \max_l (\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}(\mathbf{U}))(\mathbf{B}(\mathbf{U}))^\top \mathbf{q})_l / q_l.$$

This uplink problem can be solved by applying the uplink–downlink duality and Lemma 6: \mathbf{U}^* is given by (15) as a function of \mathbf{q}^* , which can be solved by applying Lemma 6 to

$$\min_{\mathbf{1}^\top \mathbf{q} = P} \max_l (\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}(\mathbf{U}))(\mathbf{B}(\mathbf{U}))^\top \mathbf{q})_l / q_l.$$

By the uplink–downlink duality, the uplink max-min weighted SINR problem has the same optimal value (i.e., τ^*) as the downlink max-min weighted SINR problem. Further, \mathbf{U}^* is also the optimal transmit beamformer in the downlink max-min weighted SINR problem given by (12).

Let $\lambda = \frac{1}{\tau^*}$, $\mathbf{A} = \text{diag}(\boldsymbol{\beta} \circ \mathbf{v}) \mathbf{F}^\top$, $\mathbf{b} = \left(\frac{1}{P}\right) \text{diag}(\boldsymbol{\beta}) \mathbf{v}$ in Lemma 6. Hence, similarly to the proof of Lemma 1, we have

$$\mathbf{q}^* = \mathbf{x} \left(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v}(\mathbf{U}^*)) (\mathbf{B}(\mathbf{U}^*))^\top \right).$$

By applying the fixed point iteration in the proof of Corollary 1, we obtain Steps 1–2, where $\mathbf{q}(k)$ is computed and is used to compute the transmit beamforming matrix $\mathbf{U}(k)$ in Step 3. Lastly, Steps 4–5 keep the transmit beamformers fixed and compute the downlink power $\mathbf{p}(k)$. As $\lim_{k \rightarrow \infty} \mathbf{q}(k) \rightarrow \mathbf{q}^*$ and $\lim_{k \rightarrow \infty} \mathbf{U}(k) \rightarrow \mathbf{U}^*$, we have $\lim_{k \rightarrow \infty} \mathbf{p}(k) \rightarrow \mathbf{p}^*$. This completes the proof of Theorem 1. ■

D. 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 [2] 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:

$$\tau \mathbf{p} = \mathbf{A}\mathbf{p} + \mathbf{b}, \tau \in \mathbb{R}, \mathbf{p} > \mathbf{0}, \|\mathbf{p}\| = 1. \quad (61)$$

Next, we give further characterization of the above result by applying the Friedland-Karlin inequality and minimax theorem in [14] and [15] on $\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top$. The equality between (21) and (22) in Lemma 2 can be proved as follows. Observe that (22) 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. [20]), to solve the primal and dual problems, the optimal λ in (22) turns out to be the optimal dual variable.

The next step 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 λ 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 (61):

$$\tilde{\tau} = \log \tau.$$

Once (22) is solved completely, the equality between (21) and (22) follows from strong duality. The last step is to apply the Collatz-Wielandt characterization [cf. (30)] to relate (22) 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 [15] (see [15, Theorem 3.2]) to the matrix $\mathbf{A} + \mathbf{b}\mathbf{c}_*^\top$. ■

E. Total Power Constraint in (25) and (26) Tight at Optimality

Proof: Suppose $\mathbf{1}^\top \mathbf{p} < \mathbf{1}^\top \tilde{\mathbf{p}}$ at optimality. The objective function in (26) can be reduced by increasing the power of all the users proportionally such that $\mathbf{1}^\top \mathbf{p} = \mathbf{1}^\top \tilde{\mathbf{p}}$, since $\text{SINR}_l(\mathbf{p})$ for all l increases monotonically. But this contradicts the assumption, thus the constraint in (25) and (26) are tight at optimality. ■

F. Proof of Lemma 3

Proof: Let $\text{diag}(\mathbf{v})\mathbf{B} = \text{diag}(\mathbf{v})\mathbf{F}$, 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, $\text{diag}(\mathbf{v})\mathbf{F} \geq \tilde{\mathbf{B}}$, 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, $\text{diag}(\mathbf{v})\mathbf{F} - \tilde{\mathbf{B}} = \text{diag}(\mathbf{v})\mathbf{F}\tilde{\mathbf{B}}$. Thus, $\text{Tr}[\text{diag}(\mathbf{v})\mathbf{F}\tilde{\mathbf{B}}] = \text{Tr}[\tilde{\mathbf{B}}\text{diag}(\mathbf{v})\mathbf{F}] = 0$, where $\text{Tr}[\cdot]$ denotes the matrix trace. But, this cannot happen unless $\mathbf{F} = \mathbf{0}$ or $\tilde{\mathbf{B}} = \mathbf{0}$, which contradicts 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 $\text{diag}(\mathbf{v})\mathbf{B} = \mathbf{v}\mathbf{1}^\top$. 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 [17]. This leads to $\sum_{l=1}^{\infty} a^l (\mathbf{v}^\top \mathbf{1})^{l-1} = 1$, which yields $a = \frac{1}{(1+\mathbf{v}^\top \mathbf{1})}$. It can be checked, by definition 1, that $\text{diag}(\mathbf{v})\mathbf{B} - \tilde{\mathbf{B}} = \left(1 - \frac{1}{(1+\mathbf{v}^\top \mathbf{1})}\right) \mathbf{1}\mathbf{v}^\top = \text{diag}(\mathbf{v})\mathbf{B}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}\text{diag}(\mathbf{v})\mathbf{B} \geq \mathbf{0}$, thus proving Lemma 4. ■

H. Proof of Theorem 2

Proof: We first recall the following result in [14].

Theorem 8 [14, Theorem 3.1]: For any irreducible nonnegative matrix $\mathbf{A} \in \mathbb{R}^{L \times L}$,

$$\prod_{l=1}^L \left(\frac{(\mathbf{A}\mathbf{z})_l}{z_l} \right)^{(\mathbf{x} \circ \mathbf{y})_l} \geq \rho(\mathbf{A}) \quad (62)$$

for all strictly positive \mathbf{z} , where \mathbf{x} and \mathbf{y} are the Perron and left eigenvectors of \mathbf{A} , respectively.

Furthermore, if $\mathbf{z} \geq \mathbf{A}\mathbf{z}$ (which implies $\rho(\mathbf{A}) \leq 1$), then for any positive vector $\mathbf{w} \in \mathbb{R}^{L \times 1}$, (62) can be extended to

$$\prod_{l=1}^L \left(\frac{(\mathbf{A}\mathbf{z})_l}{z_l} \right)^{w_l} \geq (\rho(\mathbf{A}))^{||\mathbf{w}||_{\infty}^{\mathbf{x} \circ \mathbf{y}}}. \quad (63)$$

Clearly, (63) reduces to (62) when $\mathbf{w} = \mathbf{x} \circ \mathbf{y}$. Next, the arithmetic-geometric mean inequality states that

$$\sum_{l=1}^L \alpha_l v_l \geq \prod_{l=1}^L v_l^{\alpha_l} \quad (64)$$

where $\mathbf{v} > \mathbf{0}$ and $\alpha \geq \mathbf{0}, \mathbf{1}^\top \alpha = 1$. Equality is achieved in (64) if and only if $v_1 = v_2 = \dots = v_L$.

Suppose $\text{diag}(\mathbf{v})\mathbf{B}$ is the quasi-inverse of $\tilde{\mathbf{B}}$. Next, we compute $\rho(\tilde{\mathbf{B}})$ and its eigenvectors.

Lemma 7: If $\tilde{\mathbf{B}}$ exists, then $\tilde{\mathbf{B}}$ has the spectral radius

$$\rho(\tilde{\mathbf{B}}) = \frac{\rho(\text{diag}(\mathbf{v})\mathbf{B})}{1 + \rho(\text{diag}(\mathbf{v})\mathbf{B})},$$

with the corresponding left and Perron eigenvectors of $\text{diag}(\mathbf{v})\mathbf{B}$.

Proof: We first state the following lemma.

Lemma 8 (Splitting Lemma [17, Ch. 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, $\rho(\mathbf{H}) = \frac{\rho(\mathbf{A}^{-1}\mathbf{N})}{1 + \rho(\mathbf{A}^{-1}\mathbf{N})}$ if and only if $\mathbf{A}^{-1}\mathbf{N} \geq \mathbf{0}$. Letting $\mathbf{A} = (\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B})^{-1}$, $\mathbf{M} = \mathbf{I}$ and $\mathbf{N} = \tilde{\mathbf{B}}$ in Lemma 8, we have $\rho(\tilde{\mathbf{B}}) = \frac{\rho((\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B})\tilde{\mathbf{B}})}{1 + \rho((\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B})\tilde{\mathbf{B}})}$. But $(\mathbf{I} + \text{diag}(\mathbf{v})\mathbf{B})\tilde{\mathbf{B}} = \text{diag}(\mathbf{v})\mathbf{B}$, thus obtaining $\rho(\tilde{\mathbf{B}}) = \frac{\rho(\text{diag}(\mathbf{v})\mathbf{B})}{1 + \rho(\text{diag}(\mathbf{v})\mathbf{B})}$. Next, we multiply both sides of $\text{diag}(\mathbf{v})\mathbf{B} - \tilde{\mathbf{B}} =$

$\tilde{\mathbf{B}}\text{diag}(\mathbf{v})\mathbf{B}$ with the Perron eigenvector \mathbf{x} of $\text{diag}(\mathbf{v})\mathbf{B}$. After rearranging, we obtain $\tilde{\mathbf{B}}\mathbf{x} = \rho(\tilde{\mathbf{B}})\mathbf{x}$. Thus, $\tilde{\mathbf{B}}$ and $\text{diag}(\mathbf{v})\mathbf{B}$ have the same Perron eigenvector \mathbf{x} . A similar proof for the left eigenvector \mathbf{y} can also be shown. ■

Now, we are ready to prove (29). From the objective function of (28), we establish the following:

$$\begin{aligned} & \sum_{l=1}^L w_l(\tilde{\mathbf{B}}\mathbf{z})_l/z_l \\ & \stackrel{(a)}{\geq} \prod_{l=1}^L \left((\tilde{\mathbf{B}}\mathbf{z})_l/z_l \right)^{w_l} \\ & \stackrel{(b)}{\geq} \left(\rho(\tilde{\mathbf{B}}) \right)^{\|\mathbf{w}\|_{\infty}^{\mathbf{x}(\tilde{\mathbf{B}}) \circ \mathbf{y}(\tilde{\mathbf{B}})}} \\ & \stackrel{(c)}{=} \left(\frac{\rho(\text{diag}(\mathbf{v})\mathbf{B})}{1 + \rho(\text{diag}(\mathbf{v})\mathbf{B})} \right)^{\|\mathbf{w}\|_{\infty}^{\mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})}} \end{aligned} \quad (65)$$

where (a) is due to letting $v_l = \frac{(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l}$ for all l and $\alpha = \mathbf{w}$ in (64), (b) is due to letting $\mathbf{A} = \tilde{\mathbf{B}}$ in (63) since $\mathbf{z} \geq \tilde{\mathbf{B}}\mathbf{z}$, and (c) is due to Lemma 7. But, using (27), $\sum_{l=1}^L \frac{w_l(\tilde{\mathbf{B}}\mathbf{z})_l}{z_l} = \sum_{l=1}^L \frac{w_l}{1+\gamma_l}$. Thus, we establish (29).

To prove the second part, we note that both inequalities (a) and (b) in (65) become tight if and only if $\gamma_1 = \gamma_2 = \dots = \gamma_L$ and $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$ (required only for (b) to become tight). In particular, all the users receive a common SINR given by $\gamma_l^* = \frac{1}{\rho(\text{diag}(\mathbf{v})\mathbf{B})}$ for all l .

To establish the power vector corresponding to $\gamma_l^* = \frac{1}{\rho(\text{diag}(\mathbf{v})\mathbf{B})}$, we further note that equality is achieved in (29) of Theorem 2 by the max-min SINR power allocation. More precisely, $\gamma_l^* = \frac{1}{\rho(\text{diag}(\mathbf{v})\mathbf{B})}$ is in fact the total power constrained max-min SINR whose associated power vector is given in Lemma 1. This completes the proof of Theorem 2. ■

I. Proof of Theorem 3

Proof: Suppose $\text{diag}(\mathbf{v})\mathbf{B}$ is the quasi-inverse of $\tilde{\mathbf{B}}$. Lemma 7 shows that $\rho(\tilde{\mathbf{B}}) < 1$, thus $(\mathbf{I} - \tilde{\mathbf{B}})^{-1}$ is nonnegative as $\mathbf{I} - \tilde{\mathbf{B}}$ is an M-matrix [17]. Thus, there is a unique mapping between all feasible \mathbf{z} in (28) and feasible \mathbf{p} in (26), and the optimal solution to (25) is given by $\mathbf{p}^* = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^*$, where \mathbf{z}^* is the optimal solution to the convex problem (31). Note that $\mathbf{p}^* \geq \mathbf{0}$ since $\rho(\tilde{\mathbf{B}}) < 1$.

Next, suppose that $\mathbf{p}^* > \mathbf{0}$, i.e., the constraints in (31) are not tight, we take the first order derivative of $\sum_{l=1}^L \frac{w_l(\tilde{\mathbf{B}}\mathbf{z})_l}{e^{z_l}}$ with respect to z_l for each l and set it to zero. We obtain (32) after making the change of variables back to \mathbf{z} subject to the constraint that $\mathbf{1}^\top \mathbf{p}^* = \mathbf{1}^\top (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}^* = \bar{P}$. ■

Proof of Corollary 2:

Proof: Observe that the function on the right-hand side of (32) (let $(f(\mathbf{z}))_l = \sqrt{(w_l \sum_{j \neq l} \tilde{B}_{lj} z_j) / (\sum_{j \neq l} w_j \frac{\tilde{B}_{jl}}{z_j})}$) is positive, monotone, quasi-concave (see, e.g., [18]) in \mathbf{z} and homogeneous of degree one. Therefore, it is a concave self-mapping. We first state the following key theorem in [3].

Theorem 9 (Krause's Theorem [3]): Let $\|\cdot\|$ be a monotone norm on \mathbb{R}^L . For a concave mapping $f: \mathbb{R}_+^L \rightarrow \mathbb{R}_+^L$ with $f(\mathbf{z}) > \mathbf{0}$ 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, $\lim_{k \rightarrow \infty} \hat{f}(\mathbf{z}(k))$ converges geometrically fast to \mathbf{z}^* , where $\hat{f}(\mathbf{z}) = \frac{f(\mathbf{z})}{\|f(\mathbf{z})\|}$.

We constrain \mathbf{z} by a monotone norm $\|\mathbf{z}\|_1 = 1$. By Theorem 9, the convergence of the iteration

$$\mathbf{z}(k+1) = f(\mathbf{z}(k)) / \|f(\mathbf{z}(k))\|_1$$

to the unique fixed point $\mathbf{p} = \frac{f(\mathbf{p})}{\|f(\mathbf{p})\|_1}$ is geometrically fast, regardless of the initial point. Using Theorem 9 (let $\lambda^* = 1$), there is a unique solution to $f(\mathbf{z}) = \mathbf{z}$. Therefore, we can compute \mathbf{z}^* (cf. Steps 1–2 of Algorithm 3). Step 3 of Algorithm 3 is obtained from $\mathbf{p}(k) = (\mathbf{I} - \tilde{\mathbf{B}})\mathbf{z}(k)$ (with a normalized $\mathbf{z}(k)$ such that $\mathbf{1}^\top \mathbf{p}(k) = \bar{P}$ is enforced at each iteration). This completes the proof of Corollary 2. ■

J. Proof of Theorem 4

Proof: We make a change of variables $\tilde{z}_l = \log z_l$ for all l in (38), and thus (38) is equivalent to the convex problem:

$$\begin{aligned} & \text{minimize} \quad \prod_{l=1}^L \left((\tilde{\mathbf{B}}e^{\tilde{\mathbf{z}}})_l / e^{\tilde{z}_l} \right)^{w_l} \\ & \text{subject to} \quad (\tilde{\mathbf{B}}e^{\tilde{\mathbf{z}}})_l / e^{\tilde{z}_l} \leq 1 \quad l = 1, \dots, L, \\ & \text{variables:} \quad \tilde{z}_l, \forall l \end{aligned} \quad (66)$$

which is strictly convex in $\tilde{\mathbf{z}}$. Hence, $\mathbf{z}^* = e^{\tilde{\mathbf{z}}^*}$ can be obtained efficiently. If $\mathbf{p}^* > \mathbf{0}$, we take the first order derivative of $\log \prod_{l=1}^L \left(\frac{(\tilde{\mathbf{B}}e^{\tilde{\mathbf{z}}})_l}{e^{\tilde{z}_l}} \right)^{w_l}$ [the objective function of (66)] with respect to \tilde{z} and set it to zero. After making a change of variables back to the \mathbf{z} domain, we obtain (39). The optimal power vector \mathbf{p}^* in (36) is then recovered from \mathbf{z}^* in (39). ■

K. Proof of Corollary 3

Proof: Similarly to the proof of Corollary 2, Corollary 3 can be proved using the nonlinear Perron–Frobenius theory by observing that the function on the right-hand side of (39) is positive, monotone, homogeneous of degree one and concave in \mathbf{z} . ■

L. Proof of Lemma 5

Proof: It is easy to observe that for a given SINR value γ , the power to achieve γ is given by (46) [1]. Next, we will show that this power vector is feasible in (25) and (36), i.e., it satisfies the total power constraint. First, observe that $\text{diag}(\gamma \circ \mathbf{v})\mathbf{B}$ is an irreducible nonnegative matrix. Now, using (46), we have

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

Thus, $\text{diag}(\gamma \circ \mathbf{v})\mathbf{B}\mathbf{p} \leq \mathbf{p}$ for any \mathbf{p} in (46). Using the Perron–Frobenius theorem [17], this implies that $\rho(\text{diag}(\gamma \circ \mathbf{v})\mathbf{B}) \leq 1$. ■

M. Proof of Theorem 5

Proof: Theorem 5 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 (47) and write the Lagrangian

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

Now, the gradient of $\log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$ with respect to $\tilde{\gamma}$ is given by (unique up to a scaling constant) [15]: $\mathbf{x}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$, where $\mathbf{x}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$ and $\mathbf{y}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$ are scaled such that

$$\mathbf{x}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})^\top \mathbf{y}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) = 1.$$

Using this fact, we can compute $\frac{\partial L(\tilde{\gamma}, \lambda)}{\partial \tilde{\gamma}_l}$ for all l . In addition, we see that λ is unique and equals $\mathbf{1}^\top f'(e^{\tilde{\gamma}^*})$ at optimality. Hence, $\tilde{\gamma}^*$ satisfies

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

and

$$\rho(\text{diag}(e^{\tilde{\gamma}^*} \circ \mathbf{v})\mathbf{B}) = 1.$$

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

Lemma 9 [21]: Let $\mathbf{A} \in \mathbb{R}_+^{L \times L}$, $\mathbf{w} \in \mathbb{R}_+^L$ 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 $\mathbf{x}(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{A}) \circ \mathbf{y}(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{A}) = \mathbf{w}$. 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:

$$\begin{aligned} & \text{maximize} && \mathbf{w}^\top \boldsymbol{\eta} \\ & \text{subject to} && \log \rho(\text{diag}(e^{\boldsymbol{\eta}})\mathbf{A}) \leq 0, \\ & \text{variables:} && \boldsymbol{\eta}. \end{aligned} \quad (67)$$

■

N. Proof of Theorem 6

Proof: We use the fact that in a sufficiently close neighborhood of $\tilde{\gamma}^*$, the domain set is convex, and the objective function $f(e^{\tilde{\gamma}^*})$ is twice continuously differentiable. We then use a successive convex approximation technique to compute $\tilde{\gamma}^*$ assuming that the initial point is sufficiently close to $\tilde{\gamma}^*$. The convergence conditions for such a technique are given in [22] and [23]. Instead of solving (47) directly, we replace the objective function of (47) in a neighborhood of a feasible point $\tilde{\gamma}(0)$ by its Taylor series (up to the first-order terms)

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

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

$$\begin{aligned} & \text{maximize} && f'(e^{\tilde{\gamma}(k)})^\top (\tilde{\gamma} - \tilde{\gamma}(k)) \\ & \text{subject to} && \log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \leq 0 \end{aligned}$$

$$\begin{aligned} \text{variables: } & \mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_L] \\ & \tilde{\gamma} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_n)^\top \in \mathbb{R}^L \end{aligned} \quad (68)$$

where $\tilde{\gamma}(k)$ is the optimal solution of the k th approximation problem. This inner approximation technique converges to a local optimal solution [22], [23]. In addition, if $\tilde{\gamma}(0)$ is sufficiently close to $\tilde{\gamma}^*$, then $\lim_{k \rightarrow \infty} \tilde{\gamma}(k) = \tilde{\gamma}^*$.

Next, we leverage Lemma 9 in [21] and Algorithm 1 to solve (68). Note that, from (51), the solution to (7) satisfies the optimality condition

$$\begin{aligned} & \mathbf{x}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \\ & = \mathbf{x}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\boldsymbol{\beta} \circ \mathbf{v})\mathbf{B}) \end{aligned}$$

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

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

which converges to the optimal solution of (25) and (36) for, respectively, $f'(e^{\tilde{\gamma}(k)}) = \left(-w_1 \frac{e^{\tilde{\gamma}_1(k)}}{(1+e^{\tilde{\gamma}_1(k)})^2}, \dots, -w_L \frac{e^{\tilde{\gamma}_L(k)}}{(1+e^{\tilde{\gamma}_L(k)})^2} \right)^\top$ and $f'(e^{\tilde{\gamma}(k)}) = (w_1 \frac{e^{\tilde{\gamma}_1(k)}}{1+e^{\tilde{\gamma}_1(k)}}, \dots, w_L \frac{e^{\tilde{\gamma}_L(k)}}{1+e^{\tilde{\gamma}_L(k)}})^\top$. ■

O. Proof of Corollary 4

Proof: If $\tilde{\gamma}^* = \mathbf{1}$ (up to a scaling constant), and $\mathbf{w} = \mathbf{x}(\text{diag}(\mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(\mathbf{v})\mathbf{B})$, then the optimality conditions in Theorem 5 are clearly satisfied. Any initial condition $\tilde{\gamma}(0) = \mathbf{1}$ (up to a scaling constant), will satisfy (up to a scaling constant):

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

Thus, $\tilde{\gamma}(k) = \mathbf{1}$ (up to a scaling constant) for all k . This proves Corollary 4. ■

P. Proof of Theorem 7

Proof: Theorem 7 is proved by using the projected gradient method with error [20] to solve (53) and then applying a result on the convergence of approximate gradient method in [24]. We first look at computing an approximate projected gradient for (53), and then show that the approximate projected gradient method converges to a neighborhood 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 l th entry given by $\frac{1}{(\mathbf{x}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}) \circ \mathbf{y}(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}))_l}$. Now, based on the gradient of $\log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})$, $(\nabla \log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B}))_l = \frac{\partial \log \rho(\text{diag}(e^{\tilde{\gamma}} \circ \mathbf{v})\mathbf{B})}{\partial \tilde{\gamma}_l}$, in (55), an approximate gradient to $\tilde{\gamma}$ of (53) can be given by

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

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 pa-

parameter ($\beta = e^{\tilde{\gamma}(k)}$) are approximately optimal (in the sense of solving the downlink and uplink max-min weighted SINR 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 l th entry

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

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

We now show that the gradient method converges despite 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 [24] ([24, Proposition 1]; see also [20, Sec. 1.3, pp. 61]):

Theorem 10: 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 $\nabla \mathbf{f}$ satisfies the Lipschitz assumption, $\mathbf{d}(k)$ satisfies

$$c_1 \|\nabla \mathbf{f}(\mathbf{s}(k))\|^2 \leq \nabla \mathbf{f}(\mathbf{s}(k))^\top \mathbf{d}(k),$$

$$\|\mathbf{d}(k)\| \leq c_2 (1 + \|\nabla \mathbf{f}(\mathbf{s}(k))\|), \quad \forall k$$

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

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

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

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

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

The idea of Theorem 10 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 $\mathbf{x}(k) = \mathbf{x}(\text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})\mathbf{B})$ and $\mathbf{y}(k) = \mathbf{y}(\text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})\mathbf{B})$):

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

where

$$\alpha(k) = \frac{(\mathbf{x}(k) \circ \mathbf{y}(k))_l}{(\mathbf{p}(k) \circ \text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})^{-1} \mathbf{q}(k))_l}.$$

In Theorem 10, we let $\mathbf{s} = \tilde{\gamma}$, $\mathbf{f} = g(\tilde{\gamma})$, $\mathbf{d}_l(k) = g_l(\tilde{\gamma}(k))\alpha(k)$ and $\mathbf{e}_l(k) = g_l(\tilde{\gamma}(k)) - \left(\frac{m_l(k+1)}{(\mathbf{x}(k) \circ \mathbf{y}(k))_l} - \alpha(k) \right)$. It can be shown that

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

and

$$\|\mathbf{e}(k)\| \leq \|g_l(\tilde{\gamma}(k))\| + \left\| \sqrt{\mathbf{m}(k+1)} \right\|_{\infty}^{\mathbf{x}(k)} \left\| \sqrt{\mathbf{m}(k+1)} \right\|_{\infty}^{\mathbf{y}(k)} + \|\mathbf{x}(k)\|_{\mathbf{p}(k)}^{\mathbf{p}(k)} \|\mathbf{y}(k)\|_{\mathbf{q}(k)}^{\mathbf{q}(k)}$$

where $\sqrt{\mathbf{m}(k+1)} = (\sqrt{m_1(k+1)}, \dots, \sqrt{m_L(k+1)})^\top$, 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

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

and

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

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

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

in the following gradient update:

$$\tilde{\gamma}'_l(k+1) = \tilde{\gamma}_l(k) + \nu(k) \times \left(\frac{m_l(k+1) (e^{\tilde{\gamma}(k)} \circ \mathbf{v})_l \mathbf{p}(k)^\top \text{diag}(e^{\tilde{\gamma}(k)} \circ \mathbf{v})^{-1} \mathbf{q}(k)}{(\mathbf{p}(k) \circ \mathbf{q}(k))_l} - 1 \right).$$

Now, the point $\tilde{\gamma}'(k+1)$ after the above gradient update may be infeasible with respect to the constraint set of (53). We now project $\tilde{\gamma}'(k+1)$ to the feasible set by adding a constant term $\log \frac{\text{SINR}_l(\mathbf{p}(k))}{\beta_l}$ (the value of the logarithmic weighted SINR evaluated at $\mathbf{p}(k)$) to obtain the following update:

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

To show the feasibility of $\tilde{\gamma}_l(k+1)$, we first let $\mathbf{x}' = \mathbf{x}(\text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v})\mathbf{B})$. Then, we have

$$\begin{aligned} & \text{diag}(e^{\tilde{\gamma}(k+1)} \circ \mathbf{v}) \mathbf{B} \mathbf{x}' \\ &= \text{diag}(\text{SINR}(\mathbf{p}(k)) / \beta) \text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B} \mathbf{x}' \\ &\leq \max_l \frac{\text{SINR}_l(\mathbf{p}(k))}{\beta_l} \text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B} \mathbf{x}' \\ &\leq \frac{1}{\rho(\text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B})} \text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B} \mathbf{x}' \\ &= \frac{1}{\rho(\text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B})} \rho(\text{diag}(e^{\tilde{\gamma}'(k+1)} \circ \mathbf{v}) \mathbf{B}) \mathbf{x}' \\ &= \mathbf{x}'. \end{aligned}$$

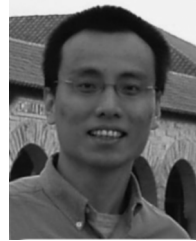
Hence, $\rho(\text{diag}(e^{\tilde{\gamma}(k+1)} \circ \mathbf{v})\mathbf{B}) \leq 1$, i.e., $\tilde{\gamma}(k+1)$ is feasible with respect to the constraint set of (53). Using Theorem 10, we conclude that $\tilde{\gamma}(k)$ converges to the optimal solution as $k \rightarrow \infty$. ■

ACKNOWLEDGMENT

The authors acknowledge helpful discussions with S. Low at Caltech, S. Friedland at UIC, and K. Tang at Cornell University.

REFERENCES

- [1] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, 1993.
- [2] V. D. Blondel, L. Ninove, and P. Van Dooren, "An affine eigenvalue problem on the nonnegative orthant," *Linear Algebra Its Appl.*, vol. 404, pp. 69–84, 2005.
- [3] U. Krause, "Concave Perron–Frobenius theory and applications," *Nonlinear Anal.*, vol. 47, no. 2001, pp. 1457–1466, 2001.
- [4] F. Rashid-Farrokh, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, 1998.
- [5] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.
- [6] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361–374, 2006.
- [7] W. Yang and G. Xu, "Optimal downlink power assignment for smart antenna systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 1998, vol. 6, pp. 3337–3340.
- [8] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, 2004.
- [9] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4639–4655, 2007.
- [10] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Jul. 1999, vol. 1, pp. 851–856.
- [11] 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. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, 1999.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
- [14] S. Friedland and S. Karlin, "Some inequalities for the spectral radius of non-negative matrices and applications," *Duke Math. J.*, vol. 42, no. 3, pp. 459–490, 1975.
- [15] S. Friedland, "Convex spectral functions," *Linear Multilinear Algebra*, vol. 9, no. 4, pp. 299–316, 1981.
- [16] Y. K. Wong, "Some mathematical concepts for linear economic models," in *Economic Activity Analysis*, O. Morgenstern, Ed. New York: Wiley, 1954, pp. 283–339.
- [17] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic, 1979.
- [18] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [19] J. F. C. Kingman, "A convexity property of positive matrices," in *Proc. Amer. Math. Soc.*, 1961, vol. 12, no. 2, pp. 283–284.
- [20] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 2003.
- [21] C. W. Tan, S. Friedland, and S. H. Low, "Nonnegative matrix inequalities and their application to nonconvex power control optimization," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 3, pp. 1030–1055, 2011.
- [22] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.
- [23] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, 2007.
- [24] D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," *SIAM J. Optim.*, vol. 10, no. 3, pp. 627–642, 2000.



Chee Wei Tan (M'08) received the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 2006 and 2008, respectively.

Previously, he was a Postdoctoral Scholar at the California Institute of Technology (Caltech), Pasadena. He is currently an Assistant Professor at City University Hong Kong. He was a Visiting Faculty at Qualcomm R&D, San Diego, CA, in 2011. His research interests are in wireless and broadband communications, signal processing and

nonlinear optimization.

Dr. Tan was the recipient of the 2008 Princeton University's Gordon Wu Prize for Excellence and 2011 IEEE ComSoc AP Outstanding Young Researcher Award.



Mung Chiang (M'03–SM'08) received the B.S. (Hons.), M.S., and Ph.D. degrees from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively.

He was an Assistant Professor from 2003 to 2008 and an Associate Professor from 2008 to 2011 at Princeton University, Princeton, NJ, where he is currently a Professor of electrical engineering and an affiliated faculty in applied and computational mathematics, and in computer science. His inventions resulted in five issued patents and several technology transfers to commercial adoption, and he founded the Princeton EDGE Lab in 2009. He is currently writing an undergraduate textbook *20 Questions About the Networked Life*.

Dr. Chiang received rewards for his research in networking, including the IEEE Tomiyasu Award, PECASE, TR35, ONR YIP, NSF CAREER, Princeton Wentz Faculty Award, and several best paper awards.



R. Srikant (S'90–M'91–SM'01–F'06) received the B.Tech. degree from the Indian Institute of Technology, Madras, in 1985 and the M.S. and Ph.D. degrees from the University of Illinois in 1988 and 1991, respectively, all in electrical engineering.

He was a Member of Technical Staff at AT&T Bell Laboratories, Holmdel, NJ, from 1991 to 1995. He is currently with the University of Illinois at Urbana-Champaign, where he is the Fredric G. and Elizabeth H. Nearing Professor in the Department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory. His research interests include communication networks, stochastic processes, queueing theory, information theory, and game theory.

He was an Associate Editor of *Automatica*, the *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, and the *IEEE/ACM TRANSACTIONS ON NETWORKING*. He has also served on the editorial boards of special issues of the *IEEE J. SEL. AREAS COMMUN.* and the *IEEE TRANSACTIONS ON INFORMATION THEORY*. He was the Chair of the 2002 IEEE Computer Communications Workshop in Santa Fe, NM, and a Program Co-Chair of IEEE INFOCOM 2007.