The Weighted Sum Rate Maximization in MIMO Interference Networks: Minimax Lagrangian Duality and Algorithm

Lijun Chen Seungil You

Abstract—**We take a new approach to the weighted sum-rate maximization in multiple-input multiple-output (MIMO) interference networks, by formulating an equivalent max-min problem. This reformulation has significant implications: the Lagrangian duality of the equivalent max-min problem provides an elegant way to establish the sum-rate duality between an interference network and its reciprocal, and more importantly, suggests a novel iterative minimax algorithm with monotonic convergence for the weighted sum-rate maximization. The design and the convergence proof of the algorithm use only general convex analysis. They apply and extend to other max-min problems with similar structure, and thus provide a general class of algorithms for such optimization problems. This paper presents a promising step and lends hope for establishing a general framework based on the minimax Lagrangian duality for developing efficient resource allocation and interference management algorithms for general MIMO interference networks.**

Index Terms—**Iterative minimax algorithm, Lagrangian duality, max-min optimization, weighted sum-rate maximization, interference networks, multiple-input multiple-output (MIMO).**

I. Introduction

The weighted sum-rate maximization, which aims to maximize the weighted sum-rate of all users or data links in a network, is an important problem that serves as a basis for many resource management and network design problems. It has a long history, with a rich literature from the classical water-filling structure for parallel Gaussian channels to more recent iterative weighted MMSE algorithm [2], [15] and polite water filling algorithm [11] for MIMO interference channels, to just name a few. The weighted sum-rate maximization is in general a highly nonconvex and NP hard problem, and despite its importance and long history, remains open for general channels/networks.

In this paper, we consider the weighted sum-rate maximization in a general MIMO interference network that consists of a set of interfering data links, each of them equipped with multiple antennas at the transmitter and receiver. The MIMO interference network, under many different names such as MIMO B-MAC and MIMO IBC, includes broadcast channels, multiple access channels, interference channels, small cell networks, and many other practical wireless networks as special cases. Specifically, we study the weighted sum-rate maximization with general linear power covariance matrix constraints, assuming Gaussian transmit signal, Gaussian noise, and the availability of channel state information at the transmitter (Section II). It typifies a class of problems that are key to the next generation wireless communication networks where the interference is a limiting factor.

Various solution approaches have been proposed for the weighted sum-rate maximization in the MIMO interference network or its special cases; see, e.g., [2]–[16], [19], [21]–[23], [25]–[27]. Broadly speaking, most of these approaches fall into the following three main categories, among others. The first category exploits the relation between the mutual information and the minimum mean square error (MMSE), and solves the weighted sum-rate maximization based on the weighted MMSE; see, e.g., [2], [15]. The second category exploits the water-filling structure or its variants at the optimum to solve iteratively for the KKT conditions; see, e.g., [11], [12]. The third category is based on the iterative convex approximation that at each iteration linearizes the nonconvex term around the point from the previous iteration; see, e.g., [3], [13], [21]. Many of the resulting algorithms are meta algorithms that require solving a large convex optimization problem at each iteration, which may incur a high computational complexity; and some of them do not even have guaranteed convergence.

In this paper, we take a different approach to the weighted sum-rate maximization in the MIMO interference network, which leads to a new and efficient algorithm with guaranteed monotonic convergence as well as an elegant way to establish the rate duality between an interference network and its reciprocal. Specifically, we reformulate the weighted sum-rate maximization as an equivalent max-min problem, by treating the interference-plus-noise covariance matrix definition as a constraint. We then construct an extended difference of logdet function and apply matrix analysis techniques from robust control to establish an explicit saddle point solution for the Lagrangian of the equivalent max-min optimization (Section III-A and the Appendix). When the explicit solution is applied to the optimal dual variable, the Lagrangian duality of the equivalent max-min problem provides an elegant way to establish the sum-rate duality between the interference network and its reciprocal (Section III). More importantly, the explicit saddle solution has significant algorithmic implication,

L. Chen is with Computer Science and Telecommunications, University of Colorado, Boulder, CO 80309, USA (email: lijun.chen@colorado.edu). S. You is with Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA (email: syou@caltech.edu).

and suggests a novel algorithm, termed *the iterative minimax algorithm*, for the weighted sum-rate maximization (Section IV). The design and the convergence proof of the algorithm use only general convex analysis. They apply and extend to other max-min problems where the objective function is concave in the maximizing variables and convex in the minimizing variables and the constraints are convex, and thus provide a general class of algorithms for such optimization problems.

The iterative minimax algorithm we design is based largely on an explicit saddle point solution for the Lagrangian of certain max-min optimization (Section III-A). This explicit solution has been identified for the case where the matrices involved are all square and invertible in [24]. In contrast, we establish the explicit solution for any general matrices, as long as the objective function is well-defined in a proper sense (the Appendix). Our proof uses only general matrix analysis, and the construction and techniques used are expected to find applications in handling singularity issues that arise from the matrix form capacity formula.

This paper benefits from the insight from the seminal work by Yu [24] that establishes uplink-downlink duality via minimax duality for the sum capacity of the Gaussian broadcast channel, and Section III can be seen as a substantial extension of [24]. Our model is more general and the results expect to find broad applications, and we establish the explicit saddle point solution for the Lagrangian of the max-min optimization with general matrices, and most importantly, we explore the algorithmic implication of the minimax duality and explicit saddle point solution to develop a novel algorithm for the weighted sum-rate maximization.

The paper is organized as follows. The next section presents details of the system model and problem formulation. Section III explores the minimax Lagrangian duality, and establishes the explicit saddle point solution for the Lagrangian of the equivalent max-min problem and the sum-rate duality. Section IV explores the algorithmic implication of the explicit saddle point solution, and presents the iterative minimax algorithm and its convergence analysis. Section V provides numerical examples to complement the theoretical analysis in the previous sections, and Section VI concludes the paper.

Notations. The capital letters such as L are used to denote sets, the capital letters in bold such as $\mathbf{H}$ are used to denote matrices, and the lower case letters in bold such as $\mathbf{x}$ are used to denote vectors. The identity matrix is denoted by $\mathbf{I}$, and the zero matrix is denoted by $\mathbf{0}$. The trace of matrix $\mathbf{A}$ is denoted by $\mathrm{Tr}\left(\mathbf{A}\right)$. For two $n \times n$ Hermitain matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{A} \succeq \mathbf{B}$ refer to the generalized inequalities under the positive semidefinite cone $\mathcal{S}_+^n$. The inequality $\mathbf{A} \succeq \mathbf{0}$ (or $\mathbf{A} \succ \mathbf{0}$) then means that matrix $\mathbf{A}$ is positive semidefinite (or positive definite), i.e., $\mathbf{A} \in \mathcal{S}_+^n$ (or $\mathbf{A} \in \mathcal{S}_{++}^n$).

II. System Model

Consider a general interference network $\mathcal{N}$ with a set L of MIMO data links (or users), with the transmitter t_l and receiver r_l of link $l \in L$ being equipped with n_l and m_l antennas respectively. Let $\mathbf{x}_l \in \mathbb{C}^{n_l \times 1}$ denote the transmit signal of link l, which is assumed to be circularly symmetric complex Gaussian.[1] The received signal $\mathbf{y}_l \in \mathbb{C}^{m_l \times 1}$ at the receiver r_l can be written as

$$\mathbf{y}_l = \sum_{k \in L} \mathbf{H}_{lk} \mathbf{x}_k + \mathbf{w}_l, \tag{1}$$

where $\mathbf{H}_{lk} \in \mathbb{C}^{m_l \times n_k}$ denotes the channel matrix from the transmitter t_k to the receiver r_l, and $\mathbf{w}_l \in \mathbb{C}^{m_l \times 1}$ denotes the additive circularly symmetric complex Gaussian noise with identity covariance matrix.

The interference network described above is very general and includes as special cases many practical channels and networks such as broadcast channels, multiple access channels, small cell networks, and heterogeneous networks, etc.

A. The power covariance matrix constraints

Denote by $\mathbf{\Sigma}_l \succeq \mathbf{0}$ the covariance matrix of the transmit signal $\mathbf{x}_l,\ l \in L$. We now specify the constraints on these power covariance matrices.

Assume that the links are grouped into a set S of non-empty subsets $L^s,\ s \in S$ that cover all of L. Each subset L^s may correspond to those links that are controlled or managed by a certain entity or for a certain purpose. These subsets may overlap with each other; and some of them may even be identical, corresponding to the situation where there may be multiple constraints on the same subset of links. For each link $l \in L$, denote by S^l the set of those subsets that include the link, i.e., $S^l = \{s \in S | l \in L^s\}$.

Each link $l \in L$ is associated with an $n_l \times n_l$ constraint matrix $\mathbf{Q}_l^s \succ \mathbf{0}$ for each $s \in S^l$; and two of these matrices may be identical. We assume that each group of links $L^s,\ s \in S$ is subject to a linear power covariance matrix constraint as follows:

$$\sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{\Sigma}_l \mathbf{Q}_l^s\right) \le 1,\ s \in S. \tag{2}$$

The constraint (2) is very general and captures all reasonable linear constraints on power. For instance, when there is only a budget P_T on the total power of all links as considered in many existing work such as [11], the cardinality $|S| = 1$ and $\mathbf{Q}_l^s = \frac{1}{P_T}\mathbf{I}$. When there is only a per-link power budget $p_l,\ l \in L$, each group L^s contains only one link and $\mathbf{Q}_l^s = \frac{1}{P_l}\mathbf{I}$. Each group $L^s, s \in S$ may also represent those links or users in a cell of a microcell network and each cell s is subject to a total power budget P_s. In this scenario, the subsets L^s are non-overlapping and $\mathbf{Q}_l^s = \frac{1}{P_s}\mathbf{I},\ \forall l \in L^s$.

Remark 1. *We have assumed linear constraints on the power covariance matrices. However, as will be seen later, our theory development and algorithm design are based on general convex analysis. So the results in this paper can be extended to the network with nonlinear convex power covariance matrix constraints, which we will investigate in future work.*

[1] The assumption of circularly symmetric complex signal can be dropped by applying the theory development and proposed algorithm to real Gaussian signal with twice the dimension.

B. The weighted sum-rate maximization

Assume that the channel state information is available at the transmitter. For given power covariance matrix $\mathbf{\Sigma}_l,\ l \in L$, an achievable rate c_l of the link l is given by

$$c_l = \log \left| \mathbf{I} + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+} \left(\mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+} \right)^{-1} \right|, \quad (3)$$

where $|\cdot|$ denotes the matrix determinant and the interferences from other links are treated as noise.[2] Assume that each link $l \in L$ is associated with a weight $w_l > 0$. We aim to allocate power for each link so as to maximize the weighted sum-rate subject to the power covariance matrix constraints:

$$\max_{\mathbf{\Sigma}_l \succeq \mathbf{0}} \quad \sum_{l \in L} w_l c_l \quad (4)$$

$$\text{s.t.} \quad \sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{\Sigma}_l \mathbf{Q}_l^s\right) \leq 1,\ s \in S. \quad (5)$$

The weighted sum-rate maximization is in general a hard non-convex problem. It is a fundamental problem that serves as a basis for many resource management and network design problems, while still remains open for general channels/networks.

III. The minimax Lagrangian duality

In this section, we will reformulate the weighted sum-rate maximization as an equivalent max-min problem, by treating the interference-plus-noise covariance matrix definition as a constraint. This seemingly trivial reformulation has significant implications: the Lagrangian duality of the equivalent max-min problem provides an elegant way to establish the sum-rate duality between an interference network and its reciprocal, and more importantly, suggests a new algorithm for the weighted sum-rate maximization.

A. The minimax Lagrangian duality

Denote by $\mathbf{\Omega}_l,\ l \in L$ the interference-plus-noise covariance matrix at the receiver r_l, i.e.,

$$\mathbf{\Omega}_l = \mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+}. \quad (6)$$

We can rewrite the weighted sum-rate maximization (4)-(5) equivalently as the following max-min problem:

$$\max_{\mathbf{\Sigma}_l \succeq \mathbf{0}} \min_{\mathbf{\Omega}_l \succeq \mathbf{0}} \quad \sum_{l \in L} w_l \left(\log\left|\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right| - \log|\mathbf{\Omega}_l| \right) \quad (7)$$

$$\text{s.t.} \quad \sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{\Sigma}_l \mathbf{Q}_l^s\right) \leq 1,\ s \in S, \quad (8)$$

$$\mathbf{\Omega}_l = \mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+},\ l \in L. \quad (9)$$

Note that, when $\mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}$ is not of full rank, the above problem is equivalent to a *truncated system* where $\mathbf{\Omega}_l$ is restricted to $\mathbf{\Omega}_l = \mathbf{H}_{ll}\mathbf{X}_l\mathbf{H}_{ll}^{+},\ \mathbf{X}_l \succeq \mathbf{0}$. Intuitively, this follows from the fact that when the signal at a channel is zero, it does not matter what the interference-plus-noise is, in terms of the achieved rate; mathematically, this causes technical difficulty regarding singular matrices; see the Appendix for more detail and insight.

The objective function of problem (7)-(9)

$$\mathcal{F}(\mathbf{\Sigma}, \mathbf{\Omega}) = \sum_{l \in L} w_l \left(\log\left|\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right| - \log|\mathbf{\Omega}_l| \right)$$

is concave in $\mathbf{\Sigma}$ and convex in $\mathbf{\Omega}$. Consider the Lagrangian

$$\begin{aligned} & \mathcal{L}(\mathbf{\Sigma}, \mathbf{\Omega}, \mathbf{\Lambda}, \boldsymbol{\mu}) \\ = \ & \mathcal{F}(\mathbf{\Sigma}, \mathbf{\Omega}) + \sum_{s \in S} \mu_s \left(1 - \sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{\Sigma}_l \mathbf{Q}_l^s\right) \right) \\ & + \sum_{l \in L} \mathrm{Tr}\left(\mathbf{\Lambda}_l (\mathbf{\Omega}_l - \mathbf{I} - \sum_{k \in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+}) \right), \end{aligned}$$

where $\boldsymbol{\mu} = \{\mu_s\}_{s \in S}$ with $\mu_s \geq 0$ the dual variable associated with the power covariance matrix constraint (8), and $\mathbf{\Lambda} = \{\mathbf{\Lambda}_l\}_{l \in L}$ with $\mathbf{\Lambda}_l \succeq \mathbf{0}$ the dual variable associated with the interference-plus-noise covariance matrix definition (9).[3] For any given $(\mathbf{\Lambda}, \boldsymbol{\mu})$, $\mathcal{L}$ is concave in $\mathbf{\Sigma}$ and convex in $\mathbf{\Omega}$ as $\mathcal{F}$ is. Thus, $\max_{\mathbf{\Sigma}} \min_{\mathbf{\Omega}} \mathcal{L} = \min_{\mathbf{\Omega}} \max_{\mathbf{\Sigma}} \mathcal{L}$, and the optimum in solving for the dual function is a saddle point.

Consider the first order condition (part of the KKT condition [1]) for the optimum:[4]

$$w_l \mathbf{H}_{ll}^{+} \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+} \right)^{-1} \mathbf{H}_{ll} = \mathbf{\Phi}_l, \quad (10)$$

$$w_l \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+} \right)^{-1} \right) = \mathbf{\Lambda}_l, \quad (11)$$

where

$$\mathbf{\Phi}_l = \sum_{s \in S_l} \mu_s \mathbf{Q}_l^s + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{kl}^{+}\mathbf{\Lambda}_k\mathbf{H}_{kl}.$$

For any given feasible dual variable $(\mathbf{\Lambda}, \boldsymbol{\mu})$, the above condition gives the saddle point condition of Lagrangian $\mathcal{L}$ as a function of $(\mathbf{\Sigma}, \mathbf{\Omega})$; and when $(\mathbf{\Lambda}, \boldsymbol{\mu})$ is a dual optimum, solving the equations (10)-(11) gives a primal optimum [1]. In the next section, we will exploit this fact to design a novel algorithm to solve the weighted sum-rate maximization.

Theorem 1. *Given feasible dual variables $(\mathbf{\Phi}, \boldsymbol{\mu})$, an explicit solution $(\mathbf{\Sigma}, \mathbf{\Omega})$ for the saddle point equations (10)-(11) is given by:*

$$\mathbf{\Omega}_l = w_l \mathbf{H}_{ll} \left(\mathbf{\Phi}_l + \mathbf{H}_{ll}^{+}\mathbf{\Lambda}_l\mathbf{H}_{ll} \right)^{-1} \mathbf{H}_{ll}^{+}, \quad (12)$$

$$\mathbf{\Sigma}_l = w_l \left(\mathbf{\Phi}_l^{-1} - \left(\mathbf{\Phi}_l + \mathbf{H}_{ll}^{+}\mathbf{\Lambda}_l\mathbf{H}_{ll} \right)^{-1} \right). \quad (13)$$

The solution (12)-(13) is motivated by [24] which focuses on a primal-dual optimum and where correspondingly the

[2] If the interference from link k to link l is completely cancelled using successive decoding and cancellation or dirty paper coding, we can simply set $\mathbf{H}_{lk} = 0$. This allows our model to cover a wide range of communication techniques.

[3] Even though the equation (9) is an equality constraint, the dual feasibility requires $\mathbf{\Lambda}_l \succeq \mathbf{0}$, as $\min_{\mathbf{\Omega}_l \succeq \mathbf{0}} \mathcal{L} = -\infty$ otherwise.

[4] Note that the first oder condition does not hold for all dual variables, but only for those that satisfy the dual feasibility condition. We only need to consider those feasible dual variables [1].

optimal power covariance matrix $\mathbf{\Sigma}_l$ and the interference-plus-noise matrix $\mathbf{\Omega}_l$ are assumed to be positive definite and the channel matrix $\mathbf{H}_{ll}$ is assumed to be square and invertible. Here, the explicit solution (12)-(13) is established for any given feasible dual variables, and the power covariance matrix and the interference-plus-noise matrix are positive semidefinite and the channel matrix can be any general matrix. However, the solution is for an equivalent, *truncated system* where we ignore the interference-plus-noise of a channel whose signal is zero, and "-1" denotes pseudo inverse if the matrix involved is singular. The proof of Theorem 1 is rather involved, and is presented in the Appendix.

The equations (10)-(11) and equations (12)-(13) have similar structures, which can be exploited to establish the sum-rate duality between an interference network and its reciprocal based on the Lagrangian dual of the (truncated) max-min problem (7)-(9).

Definition 1. *Consider an interference network $\mathcal{N}$ with a set L of MIMO links and channel matrix $\mathbf{H}_{kl}$ from the transmitter of link $l \in L$ to the receiver of link $k \in L$. Its reciprocal $\hat{\mathcal{N}}$ is defined as a network with the same set L of links but with reversed directions where the channel matrix $\hat{\mathbf{H}}_{kl}$ from the transmitter of link l to the receiver of link k is given by $\hat{\mathbf{H}}_{kl} = \mathbf{H}_{lk}^{+}$.*

The transmitter (receiver) of a link in the reciprocal network $\hat{\mathcal{N}}$ is the receiver (transmitter) of the corresponding link in the original network $\mathcal{N}$. For instance, for a broadcast channel with a channel matrix $\mathbf{H}$, its reciprocal is a multiple access channel with channel matrix $\mathbf{H}^{+}$, and vice versa [17], [18], [24].

Motivated by the structural parallel between the equations (10)-(11) and equations (12)-(13), define a weighted sum-rate maximization problem for the reciprocal network:

$$\max_{\hat{\mathbf{\Sigma}}_l \succeq \mathbf{0}} \min_{\hat{\mathbf{\Omega}}_l \succeq \mathbf{0}} \sum_{l\in L} w_l \left(\log\left|\hat{\mathbf{\Omega}}_l + \hat{\mathbf{H}}_{ll}\hat{\mathbf{\Sigma}}_l\hat{\mathbf{H}}_{ll}^{+}\right| - \log\left|\hat{\mathbf{\Omega}}_l\right|\right) \quad (14)$$

$$\text{s.t.} \quad \sum_{l\in L^s} \operatorname{Tr}\left(\frac{1}{|S_l|\mu_s}\hat{\mathbf{\Sigma}}_l\right) \le 1, \; s \in S, \quad (15)$$

$$\hat{\mathbf{\Omega}}_l = \sum_{s\in S_l} \mu_s \mathbf{Q}_l^s + \sum_{k\in L\setminus\{l\}} \hat{\mathbf{H}}_{lk}\hat{\mathbf{\Sigma}}_k\hat{\mathbf{H}}_{lk}^{+}, l \in L, \quad (16)$$

where $\mu_s, \; s \in S$ are the optimal duals associated with the constraints (8), the noise covariance matrix at link $l \in L$ is given by $\sum_{s\in S_l} \mu_s \mathbf{Q}_l^s$, and the power covariance matrices $\hat{\mathbf{\Sigma}}_l$ are constrained group-wise as in the original network. Denote by $\hat{\mu}_s, \; s \in S$ and $\hat{\mathbf{\Lambda}}_l, \; l \in L$ the dual variables associated with the constraints (15) and (16), respectively. By Theorem 1, the following primal-dual of the reciprocal network

$$\{\hat{\mathbf{\Sigma}}_l, \hat{\mathbf{\Omega}}_l; \hat{\mu}_s, \hat{\mathbf{\Lambda}}_l\} = \{\mathbf{\Lambda}_l, \mathbf{\Phi}_l; \mu_s, \mathbf{\Sigma}_l\} \quad (17)$$

satisfies the first-order condition for the optimum of the weighted sum-rate maximization (14)-(16) of the reciprocal network.

Even though the above problem achieves the same maximal sum-rate as problem (7)-(9) of the original network, its constraints depend on the optimum of problem (7)-(9) and also have a very different structure from those of problem (7)-(9), which makes the rate duality between the two networks less appealing. In the next subsection, we will study a few appealing cases with "strong" rate duality where the weighted sum-rate maximization problem of the reciprocal network has exactly the same structure as that of the original network.

However, this weighted sum-rate maximization problem may involve the optimal dual variables of the weighted sum-rate maximization of the original network, as follows:

B. Case studies

We now discuss two typical cases, and show how the minimax Lagrangian duality can be used to establish the strong rate duality between the interference network and its reciprocal.

1) The network with the per-link power constraints and without interlink interference: Here the set $S = L$, $\mathbf{\Omega}_l = \mathbf{I}$, and $\mathbf{Q}_l = \frac{\mathbf{I}}{P_l}$ with P_l the power budget at each link $l \in L$. As each link is independent, we can just focus on one link:

$$\max_{\mathbf{\Sigma}_l \succeq \mathbf{0}} \min_{\mathbf{\Omega}_l \succeq \mathbf{0}} \quad \log|\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}| - \log|\mathbf{\Omega}_l| \quad (18)$$

$$\text{s.t.} \quad \operatorname{Tr}\left(\frac{\mathbf{\Sigma}_l}{P_l}\right) \le 1, \quad \mathbf{\Omega}_l = \mathbf{I}. \quad (19)$$

The first order condition (10)-(11) reduces to

$$w_l \mathbf{H}_{ll}^{+}\left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\mathbf{H}_{ll} = \mu_l \frac{\mathbf{I}}{P_l},$$

$$w_l\left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right) = \mathbf{\Lambda}_l,$$

where $\mu_l \ge 0$ is the dual variable associated with the power covariance matrix constraint. Define

$$\hat{\mathbf{\Sigma}}_l = \frac{P_l}{\mu_l}\mathbf{\Lambda}_l,$$

$$\hat{\mathbf{\Omega}}_l = \mathbf{I}.$$

The first order condition becomes

$$w_l \mathbf{H}_{ll}^{+}\left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\mathbf{H}_{ll} = \frac{\mu_l}{P_l}\hat{\mathbf{\Omega}}_l, \quad (20)$$

$$w_l\left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right) = \frac{\mu_l}{P_l}\hat{\mathbf{\Sigma}}_l, \quad (21)$$

and the explicit solution (12)-(13) becomes

$$w_l \mathbf{H}_{ll}\left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^{+}\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}\right)^{-1}\mathbf{H}_{ll}^{+} = \frac{\mu_l}{P_l}\mathbf{\Omega}_l, \quad (22)$$

$$w_l\left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^{+}\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}\right)^{-1}\right) = \frac{\mu_l}{P_l}\mathbf{\Sigma}_l. \quad (23)$$

Comparing equations (20)-(21) and equations (22)-(23), we can conclude that the Lagrangian dual of the max-min problem (18)-(19) is also a max-min problem:

$$\max_{\hat{\mathbf{\Sigma}}_l \succeq \mathbf{0}} \min_{\hat{\mathbf{\Omega}}_l \succeq \mathbf{0}} \quad \log|\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^{+}\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}| - \log|\hat{\mathbf{\Omega}}_l| \quad (24)$$

$$\text{s.t.} \quad \operatorname{Tr}\left(\frac{\hat{\mathbf{\Sigma}}_l}{P_l}\right) \le 1, \quad \hat{\mathbf{\Omega}}_l = \mathbf{I}, \quad (25)$$

which is the sum-rate maximization problem defined on the reciprocal link with channel matrix $\mathbf{H}_l^{+}$. At the corresponding saddle points, the two problems achieve the same rate, since one is the dual of the other. Furthermore, introducing the dual variables $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Lambda}}_l$ for the problem (24)-(25), we have the following correspondence:

$$(\mathbf{\Sigma}_l; \mathbf{\Lambda}_l, \mu_l) = (\frac{P_l}{\hat{\mu}_l}\hat{\mathbf{\Lambda}}_l; \frac{\hat{\mu}_l}{P_l}\hat{\mathbf{\Sigma}}_l, \hat{\mu}_l), \tag{26}$$

$$(\hat{\mathbf{\Sigma}}_l; \hat{\mathbf{\Lambda}}_l, \hat{\mu}_l) = (\frac{P_l}{\mu_l}\mathbf{\Lambda}_l; \frac{\mu_l}{P_l}\mathbf{\Sigma}_l, \mu_l). \tag{27}$$

This recovers the well-known result in [17], [18], [24]. The difference from [24] is that we establish the explicit solution (22)-(23) and the correspondence (26)-(27) for general power covariance matrices and channel matrices and at any saddle points of the Lagrangian function (instead of only at an optimum).

2) The network with the total power constraint: Here $|S| = 1$ and $\mathbf{Q}_l = \frac{\mathbf{I}}{P_T}$, with P_T the total power budget. The max-min problem (7)-(9) reduces to

$$\max_{\mathbf{\Sigma}_l \succeq \mathbf{0}} \min_{\mathbf{\Omega}_l \succeq \mathbf{0}} \quad \sum_l w_l \left(\log|\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^+| - \log|\mathbf{\Omega}_l| \right) \tag{28}$$

$$\text{s.t.} \quad \sum_l \operatorname{Tr}\left(\frac{\mathbf{\Sigma}_l}{P_T}\right) \leq 1, \tag{29}$$

$$\mathbf{\Omega}_l = \mathbf{I} + \sum_{k \in L \setminus \{l\}} \mathbf{H}_{kl}\mathbf{\Sigma}_k\mathbf{H}_{lk}^+, \tag{30}$$

and the first order condition (10)-(11) reduces to

$$w_l \mathbf{H}_{ll}^+ \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^+\right)^{-1} \mathbf{H}_{ll} = \mathbf{\Phi}_l,$$
$$w_l \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^+\right)^{-1}\right) = \mathbf{\Lambda}_l,$$

with $\mathbf{\Phi}_l = \mu\frac{\mathbf{I}}{P_T} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{kl}^+\mathbf{\Lambda}_k\mathbf{H}_{kl}$, where $\mu \geq 0$ is the dual variable associated with the total power constraint. Define

$$\hat{\mathbf{\Sigma}}_l = \frac{P_T}{\mu}\mathbf{\Lambda}_l,$$
$$\hat{\mathbf{\Omega}}_l = \mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{kl}^+\hat{\mathbf{\Sigma}}_k\mathbf{H}_{kl}.$$

The first order condition becomes

$$w_l \mathbf{H}_{ll}^+ \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^+\right)^{-1} \mathbf{H}_{ll} = \frac{\mu}{P_T}\hat{\mathbf{\Omega}}_l, \tag{31}$$

$$w_l \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^+\right)^{-1}\right) = \frac{\mu}{P_T}\hat{\mathbf{\Sigma}}_l, \tag{32}$$

and the explicit solution (12)-(13) becomes

$$w_l \mathbf{H}_{ll} \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^+\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}\right)^{-1} \mathbf{H}_{ll}^+ = \frac{\mu}{P_T}\mathbf{\Omega}_l, \tag{33}$$

$$w_l \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^+\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}\right)^{-1}\right) = \frac{\mu}{P_T}\mathbf{\Sigma}_l. \tag{34}$$

Comparing equations (31)-(32) and equations (33)-(34), we can conclude that the Lagrangian dual of the max-min problem (28)-(30) is also a max-min problem:

$$\max_{\hat{\mathbf{\Sigma}}_l \succeq \mathbf{0}} \min_{\hat{\mathbf{\Omega}}_l \succeq \mathbf{0}} \quad \sum_l w_l \left(\log|\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^+\hat{\mathbf{\Sigma}}_l\mathbf{H}_{ll}| - \log|\hat{\mathbf{\Omega}}_l| \right) \tag{35}$$

$$\text{s.t.} \quad \sum_l \operatorname{Tr}\left(\frac{\hat{\mathbf{\Sigma}}_l}{P_T}\right) \leq 1, \tag{36}$$

$$\hat{\mathbf{\Omega}}_l = \mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{kl}^+\hat{\mathbf{\Sigma}}_k\mathbf{H}_{lk}, \tag{37}$$

which is the weighted sum-rate maximization problem defined on a network of reciprocal channels with channel matrix $\mathbf{H}^+$. At the corresponding saddle points, the two problems achieve the same weighted sum-rate, since one is the dual of the other. Furthermore, introducing the dual variables $\hat{\mu}$ and $\hat{\mathbf{\Lambda}}_l$ for the problem (35)-(37), we have the following correspondence:

$$(\mathbf{\Sigma}_l; \mathbf{\Lambda}_l, \mu) = (\frac{P_T}{\hat{\mu}}\hat{\mathbf{\Lambda}}_l; \frac{\hat{\mu}}{P_T}\hat{\mathbf{\Sigma}}_l, \hat{\mu}), \tag{38}$$

$$(\hat{\mathbf{\Sigma}}_l; \hat{\mathbf{\Lambda}}_l, \hat{\mu}) = (\frac{P_T}{\mu}\mathbf{\Lambda}_l; \frac{\mu}{P_T}\mathbf{\Sigma}_l, \mu). \tag{39}$$

This provides a simple proof of the weighted sum-rate duality for the MIMO interference network with the total power constraint identified in, e.g., [11].

IV. The iterative minimax algorithm

Motivated by the minimax Lagrangian duality, in this section we will design a novel algorithm for the weighted sum-rate maximization and establish its convergence properties.

A. The iterative minimax algorithm

Note that an optimum of the max-min problem (7)-(9) is a saddle point, and the first order condition (10)-(11) and its explicit solution (12)-(13) or part of them will give a saddle point, maximum, or minimum of Lagrangian $\mathcal{L}$ when certain subset of its variables is fixed and given. This motivates an iterative minimax algorithm to achieve an optimum, as follows.

1) Start with given $\mathbf{\Sigma}_l^n,\ l \in L$ that is feasible, i.e.,

$$\sum_{l \in L^s} \operatorname{Tr}\left(\mathbf{\Sigma}_l^n \mathbf{Q}_l^s\right) \leq 1,\ \ s \in S,$$

and $\mathbf{\Omega}_l^n = \mathbf{I} + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k^n\mathbf{H}_{lk}^+,\ \ l \in L$. By equation (11) that gives the condition for minimizing $\mathcal{L}$ over $\mathbf{\Omega}_l$, we choose $\mathbf{\Lambda}_l^n \succeq \mathbf{0}$ such that

$$\mathbf{\Lambda}_l^n = w_l \left((\mathbf{\Omega}_l^n)^{-1} - \left(\mathbf{\Omega}_l^n + \mathbf{H}_{ll}\mathbf{\Sigma}_l^n\mathbf{H}_{ll}^+\right)^{-1} \right). \tag{40}$$

Therefore, for any $\mathbf{\Omega} \succeq \mathbf{0}$, we have

$$\begin{aligned} \mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) &\leq \mathcal{L}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\ &\leq \mathcal{L}(\mathbf{\Sigma}^n, \mathbf{\Omega}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n), \end{aligned} \tag{41}$$

where $\boldsymbol{\mu}^n \geq 0$ will be determined later. Define

$$\mathbf{\Phi}_l^n = \sum_{s \in S^l} \mu_s^n \mathbf{Q}_l^s + \sum_{k \in L\setminus\{l\}} \mathbf{H}_{kl}^+\mathbf{\Lambda}_k^n\mathbf{H}_{kl}. \tag{42}$$

Note that $\mathbf{\Phi}_l^n$ does not necessary satisfy equation (10).

2) Given the above $(\mathbf{\Lambda}_l^n, \mathbf{\Phi}_l^n)$ and μ^n, by equations (12)-(13), we choose $(\tilde{\mathbf{\Sigma}}^{n+1}, \tilde{\mathbf{\Omega}}^{n+1})$ such that

$$\tilde{\mathbf{\Sigma}}_l^{n+1} = w_l \left((\mathbf{\Phi}_l^n)^{-1} - \left(\mathbf{\Phi}_l^n + \mathbf{H}_{ll}\mathbf{\Lambda}_l^n\mathbf{H}_{ll}^+\right)^{-1} \right). \tag{43}$$

$$\tilde{\mathbf{\Omega}}_l^{n+1} = w_l \mathbf{H}_{ll} \left(\mathbf{\Phi}_l^n + \mathbf{H}_{ll}\mathbf{\Lambda}_l^n\mathbf{H}_{ll}^+\right)^{-1} \mathbf{H}_{ll}^+. \tag{44}$$

Plug $\mathbf{\Omega} = \tilde{\mathbf{\Omega}}^{n+1}$ into inequality (41), we have

$$\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) \leq \mathcal{L}(\mathbf{\Sigma}^n, \tilde{\mathbf{\Omega}}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n). \tag{45}$$

By the first order condition (10)-(11), $(\tilde{\mathbf{\Sigma}}^{n+1}, \tilde{\mathbf{\Omega}}^{n+1})$ is the saddle point of $\mathcal{L}(\mathbf{\Sigma}, \mathbf{\Omega}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n)$. Thus,

$$\begin{aligned}\mathcal{L}(\mathbf{\Sigma}^n, \tilde{\mathbf{\Omega}}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) &\leq \mathcal{L}(\tilde{\mathbf{\Sigma}}^{n+1}, \tilde{\mathbf{\Omega}}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\ &\leq L(\tilde{\mathbf{\Sigma}}^{n+1}, \mathbf{\Omega}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \quad (46)\end{aligned}$$

for any $\mathbf{\Omega} \succeq \mathbf{0}$.

3) The matrix $\tilde{\mathbf{\Sigma}}_l^{n+1}$ is a function of $\mu_s^n,\ s \in S^l$, denoted explicitly by $\tilde{\mathbf{\Sigma}}_l^{n+1}(\mu_s^n; s \in S^l)$. Define the set T such that

$$T = \{s \in S | \sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{Q}_l^s \tilde{\mathbf{\Sigma}}_l^{n+1}(\mu_{\bar{s}}^n = 0^+; \bar{s} \in S^l)\right) \geq 1\}.$$

For each $s \in S \backslash T$, we set $\mu_s^n = 0$. For those $s \in T$, we choose μ_s^n such that

$$\sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{Q}_l^s \tilde{\mathbf{\Sigma}}_l^{n+1}(\mu_{\bar{s}}^n; \bar{s} \in S^l)\right) = 1,\ s \in T. \quad (47)$$

Note that $\mathrm{Tr}\left(\tilde{\mathbf{\Sigma}}_l^{n+1}(\mu_s^n; s \in S^l)\right)$ is decreasing in μ_s^n, and there are $|T|$ equations for $|T|$ variables. So, there exists a solution to equation (47). With the afore choice of $\mu_s^n,\ s \in S$, we can see that

$$\mu_s^n \left(1 - \sum_{l \in L^s} \mathrm{Tr}\left(\tilde{\mathbf{\Sigma}}_l^{n+1} \mathbf{Q}_l^s\right)\right) = 0. \quad (48)$$

The above is a complementary slackness condition (part of the KKT condition) that is required at an optimum [1], but in our algorithm we enforce this condition at each iteration.

4) Let

$$\lambda = \max_{s \in S} \sum_{l \in L^s} \mathrm{Tr}\left(\tilde{\mathbf{\Sigma}}_l^{n+1} \mathbf{Q}_l^s\right).$$

We see that $0 < \lambda \leq 1$. We then choose $(\mathbf{\Sigma}_l^{n+1}, \mathbf{\Omega}_l^{n+1})$ such that

$$\mathbf{\Sigma}_l^{n+1} = \frac{\tilde{\mathbf{\Sigma}}_l^{n+1}}{\lambda}, \quad (49)$$

$$\mathbf{\Omega}_l^{n+1} = \mathbf{I} + \sum_{k \in L \backslash \{l\}} \mathbf{H}_{lk} \mathbf{\Sigma}_k^{n+1} \mathbf{H}_{lk}^+. \quad (50)$$

Plug $\tilde{\mathbf{\Sigma}}^{n+1} = \lambda \mathbf{\Sigma}^{n+1}$ and $\mathbf{\Omega} = \lambda \mathbf{\Omega}_l^{n+1}$ into the inequality (46) and combine with the inequality (45), we have

$$\begin{aligned}\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) &\leq \mathcal{L}(\lambda \mathbf{\Sigma}^{n+1}, \lambda \mathbf{\Omega}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\ &= \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}) + \sum_{l \in L} (\lambda - 1) \mathrm{Tr}\left(\mathbf{\Lambda}_l^n\right) \\ &\quad + \sum_{s \in S} \mu_s^n \left(1 - \sum_{l \in L^s} \mathrm{Tr}\left(\tilde{\mathbf{\Sigma}}_l^{n+1} \mathbf{Q}_l^s\right)\right) \\ &= \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}) + \sum_{l \in L} (\lambda - 1) \mathrm{Tr}\left(\mathbf{\Lambda}_l^n\right) \\ &\leq \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}), \quad (51)\end{aligned}$$

where the second equality follows from equation (48) and the last inequality follows from the fact that $\lambda \leq 1$.

5) Repeat 1-4, we will obtain a monotone increasing sequence $\{\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n)\}$, based on which we can conclude that $(\mathbf{\Sigma}^n, \mathbf{\Omega}^n)$ converges to a saddle point of the max-min problem (7)-(9) and thus an (local) optimum of the weighted sum-rate maximization (4)-(5).

We call the above algorithm the *iterative minimax algorithm*; see Table I for a formal description. Different from many existing algorithms mentioned in Section I, the iterative minimax algorithm is not a meta algorithm that requires solving a large convex optimization problem at each iteration. The only step of the algorithm that may potentially be complicated is Step 8) that may require solving sets of coupled equations when the subsets $L^s,\ s \in S$ overlap. But Step 8) can be solved efficiently using, e.g., the bisection search method. Moreover, in practice, we expect that the subsets $L^s,\ s \in S$ seldom overlap.

TABLE I
THE ITERATIVE MINIMAX ALGORITHM

1) Initialize $\mathbf{\Sigma}_l,\ l \in L$ such that $\sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{\Sigma}_l \mathbf{Q}_l^s\right) \leq 1,\ s \in S$
2) $\mathbf{\Omega}_l \leftarrow \mathbf{I} + \sum_{k \in L \backslash \{l\}} \mathbf{H}_{lk} \mathbf{\Sigma}_k \mathbf{H}_{lk}^+,\ l \in L$
3) $\mathbf{\Lambda}_l \leftarrow w_l \left((\mathbf{\Omega}_l)^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll} \mathbf{\Sigma}_l \mathbf{H}_{ll}^+\right)^{-1}\right),\ l \in L$
4) $\mathbf{\Phi}_l \leftarrow \sum_{s \in S^l} \mu_s \mathbf{Q}_l^s + \sum_{k \in L \backslash \{l\}} \mathbf{H}_{kl}^+ \mathbf{\Lambda}_k \mathbf{H}_{kl},\ l \in L$
5) $\tilde{\mathbf{\Sigma}}_l \leftarrow w_l \left((\mathbf{\Phi}_l)^{-1} - \left(\mathbf{\Phi}_l + \mathbf{H}_{ll} \mathbf{\Lambda}_l \mathbf{H}_{ll}^+\right)^{-1}\right),\ l \in L$
6) $T \leftarrow \{s \in S \mid \sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{Q}_l^s \tilde{\mathbf{\Sigma}}_l(\mu_{\bar{s}} = 0^+; \bar{s} \in S^l)\right) \geq 1\}$
7) $\mu_s \leftarrow 0$ if $s \in S \backslash T$
8) For $s \in T$, choose μ_s such that $\sum_{l \in L^s} \mathrm{Tr}\left(\mathbf{Q}_l^s \tilde{\mathbf{\Sigma}}_l(\mu_{\bar{s}}; \bar{s} \in S^l)\right) = 1,\ s \in T$
9) $\lambda \leftarrow \max_{s \in S} \sum_{l \in L^s} \mathrm{Tr}\left(\tilde{\mathbf{\Sigma}}_l \mathbf{Q}_l^s\right)$
10) $\mathbf{\Sigma}_l \leftarrow \frac{\tilde{\mathbf{\Sigma}}_l}{\lambda},\ l \in L$
11) Go to 2)

B. The convergence analysis

We now study the convergence properties of the iterative minimax algorithm. The following result is immediate.

Theorem 2. *The iterative minimax algorithm converges to a saddle point* $(\mathbf{\Sigma}^*, \mathbf{\Omega}^*)$ *of the max-min problem (7)-(9); and* $\mathbf{\Sigma}^*$ *is a (local) optimum of the weighted sum-rate maximization (4)-(5).*

Proof. By inequality (51), we have

With Lemma **??**, to show the convergence of the iterative minimax algorithm, it is enough to show that if $\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) = \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1})$, then $(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) = (\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1})$.

Seen form the derivation of the inequality (51), if

$\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) = \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1})$, then

$$\begin{aligned}
\mathcal{F}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n) &= \mathcal{L}(\mathbf{\Sigma}^n, \mathbf{\Omega}^n, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\
&= \mathcal{L}(\mathbf{\Sigma}^n, \tilde{\mathbf{\Omega}}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\
&= \mathcal{L}(\tilde{\mathbf{\Sigma}}^{n+1}, \tilde{\mathbf{\Omega}}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\
&= \mathcal{L}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n) \\
&= \mathcal{F}(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}).
\end{aligned}$$

It follows that both $(\mathbf{\Sigma}^n, \mathbf{\Omega}^n, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n)$ and $(\mathbf{\Sigma}^{n+1}, \mathbf{\Omega}^{n+1}, \mathbf{\Lambda}^n, \boldsymbol{\mu}^n)$ satisfy the KKT condition (the first order condition, the primal feasibility, the dual feasibility, and the complementary slackness [1]) of the max-min problem (7)-(9), and thus both are saddle points of the max-min problem. Furthermore, for any given dual variables, the Lagrangian $\mathcal{L}$ is strictly concave in $\mathbf{\Sigma}$. So, $\mathbf{\Sigma}^n = \mathbf{\Sigma}^{n+1}$, and $\mathbf{\Omega}^n = \mathbf{\Omega}^{n+1}$ follows. Therfore, the iterative minimax algorithm converges *monotonically* to a saddle point of the max-min problem (7)-(9). The second part of the theorem follows from the equivalence between the max-min problem and the weighted sum-rate maximization problem. □

Remark 2. *The iterative minimax algorithm converges fairly fast and can be implemented realtime. As each link knows its own power covariance matrix and can measure/estimate its interference-plus-noise covariance matrix, the algorithm admits a distributed implementation if used as a realtime algorithm.*

Remark 3. *The design and the convergence proof of the iterative minimax algorithm use only general convex analysis. They apply and extend to any max-min problems where the objective function is concave in the maximizing variables and convex in the minimizing variables and the constraints are convex, and thus provide a general class of algorithms for such optimization problems, which we will investigate in detail in future work.*

C. Case studies

We now discuss two typical cases and the corresponding iterative minimax algorithms.

1) The network with the total power constraint: As mentioned in Section III-B2, here $|S| = 1$, and $\mathbf{Q}_l = \frac{\mathbf{I}}{P_T}$ with P_T the total power budget. The matrix $\tilde{\mathbf{\Sigma}}_l$ defined in Section IV-A is a function of μ, the dual variable associated with the total power constraint. The iterative minimax algorithm reduces to that described in Table II.

We have proposed another algorithm for the network with the total power constraint in a previous work [10], which uses the fact that the total power constraint is tight at an optimum, and normalizes μ such that $\sum_{l\in L} \mathrm{Tr}\left(\frac{\mu}{P_T}\tilde{\mathbf{\Sigma}}_l\right) = 1$, i.e., the algorithm enforces the tightness of the total power constraint at the initial point and each iteration. In contrast, the algorithm in Table II enforces the complementary slackness condition at each iteration and can start with any feasible $\mathbf{\Sigma}$. Moreover, the algorithm in [10] hardly offers any insight on the algorithm design for the network with general linear power covariance matrix constraints.

TABLE II
THE ITERATIVE MINIMAX ALGORITHM FOR THE NETWORK WITH THE TOTAL POWER CONSTRAINT

1) Initialize $\mathbf{\Sigma}_l,\ l \in L$ such that $\sum_{l\in L} \mathrm{Tr}\left(\frac{\mathbf{\Sigma}_l}{P_T}\right) \le 1$
2) $\mathbf{\Omega}_l \leftarrow \mathbf{I} + \sum_{k\in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+},\ l \in L$ --
3) $\mathbf{\Lambda}_l \leftarrow w_l\left((\mathbf{\Omega}_l)^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right),\ l \in L$
4) $\mathbf{\Phi}_l \leftarrow \mu\mathbf{I} + \sum_{k\in L\setminus\{l\}} \mathbf{H}_{kl}^{+}\mathbf{\Lambda}_k\mathbf{H}_{kl},\ l \in L$
5) $\tilde{\mathbf{\Sigma}}_l \leftarrow w_l\left((\mathbf{\Phi}_l)^{-1} - \left(\mathbf{\Phi}_l + \mathbf{H}_{ll}\mathbf{\Lambda}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right),\ l \in L$
6) $\mu \leftarrow 0$ if $\sum_{l\in L} \mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l(0^+)}{P_T}\right) < 1$;
otherwise, choose μ such that $\sum_{l\in L} \mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l(\mu)}{P_T}\right) = 1$
7) $\lambda \leftarrow \sum_{l\in L} \mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l}{P_T}\right)$
8) $\mathbf{\Sigma}_l \leftarrow \frac{\tilde{\mathbf{\Sigma}}_l}{\lambda},\ l \in L$
9) Go to 2)

2) The network with the per-link power constraints: Here the set $S = L$, and $\mathbf{Q}_l = \frac{\mathbf{I}}{P_l}$ with P_l the power budget at each link $l \in L$. The matrix $\tilde{\mathbf{\Sigma}}_l$ defined in Section IV-A is a function of μ_l, the dual variable associated with the power constraint at link l. The iterative minimax algorithm reduces to that described in Table III.

TABLE III
THE ITERATIVE MINIMAX ALGORITHM FOR THE NETWORK WITH THE PER-LINK POWER CONSTRAINTS

1) Initialize $\mathbf{\Sigma}_l,\ l \in L$ such that $\mathrm{Tr}\left(\frac{\mathbf{\Sigma}_l}{P_l}\right) \le 1$
2) $\mathbf{\Omega}_l \leftarrow \mathbf{I} + \sum_{k\in L\setminus\{l\}} \mathbf{H}_{lk}\mathbf{\Sigma}_k\mathbf{H}_{lk}^{+},\ l \in L$
3) $\mathbf{\Lambda}_l \leftarrow w_l\left((\mathbf{\Omega}_l)^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll}\mathbf{\Sigma}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right),\ l \in L$
4) $\mathbf{\Phi}_l \leftarrow \mu\mathbf{I} + \sum_{k\in L\setminus\{l\}} \mathbf{H}_{kl}^{+}\mathbf{\Lambda}_k\mathbf{H}_{kl},\ l \in L$
5) $\tilde{\mathbf{\Sigma}}_l \leftarrow w_l\left((\mathbf{\Phi}_l)^{-1} - \left(\mathbf{\Phi}_l + \mathbf{H}_{ll}\mathbf{\Lambda}_l\mathbf{H}_{ll}^{+}\right)^{-1}\right),\ l \in L$
6) $\mu_l \leftarrow 0$ if $\mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l(0^+)}{P_l}\right) < 1$;
otherwise, choose μ_l such that $\mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l(\mu_l)}{P_l}\right) = 1$
7) $\lambda \leftarrow \max_{l\in L} \mathrm{Tr}\left(\frac{\tilde{\mathbf{\Sigma}}_l}{P_T}\right)$
8) $\mathbf{\Sigma}_l \leftarrow \frac{\tilde{\mathbf{\Sigma}}_l}{\lambda},\ l \in L$
9) Go to 2)

V. NUMERICAL EXAMPLES

In this section, we provide numerical examples to complement the analysis in the previous sections. Consider a network where each link is equipped with 3 (4) antennas at its transmitter (receiver) and the channel matrices have zero-mean, unit-variance, i.i.d. complex Gaussian entries. We will consider and compare the networks with low, moderate, and high interference, which are characterized by scaling the interference channel matrices $\mathbf{H}_{ij},\ i \neq j$ with a factor of 0.1, 1, and 5 respectively. The weights w_l's are uniformly drawn from $[0.5, 1]$.

The implementation of our iterative minimax algorithm is straightforward. It only uses basic matrix operations, except for finding $\boldsymbol{\mu}$ for which we use a bisection search method. This is different from many other algorithms that need additional problem parser or use the interior point method which are often hard to implement in practical applications.

A. The network with the total power constraint

Figures 1, 2 and 3 show the monotonic convergence of our algorithm (Table II) in a network with $|L| = 10$ interfering links and a total power constraint $P_T = 10$. Overall, the algorithm shows very fast convergence. We see that the convergence speed depends on the strength of interference. As the interference becomes stronger, the weighted sum-rate becomes highly non-convex. This intrinsic difficulty of the problem makes the convergence slower.

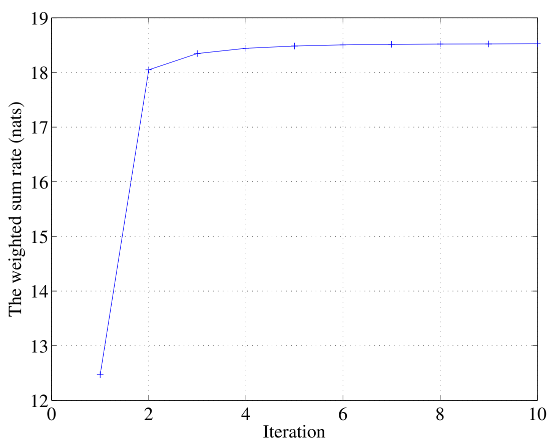

Fig. 1. The network with low interference and total power constraint.

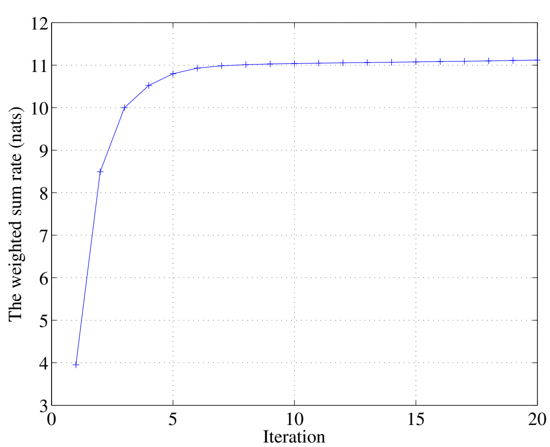

Fig. 2. The network with moderate interference and total power constraint.

B. The network with the per-link power constraints

Figures 4, 5 and 6 show the monotonic convergence of our algorithm (Table III) in a network with $|L| = 10$ interfering links and per-link power constraints where P_l's are uniformly drawn from $\{1, 2, \cdots, 10\}$. Again, we see that the stronger the interference, the slower the algorithm converges; but overall, the algorithm shows fast convergence.

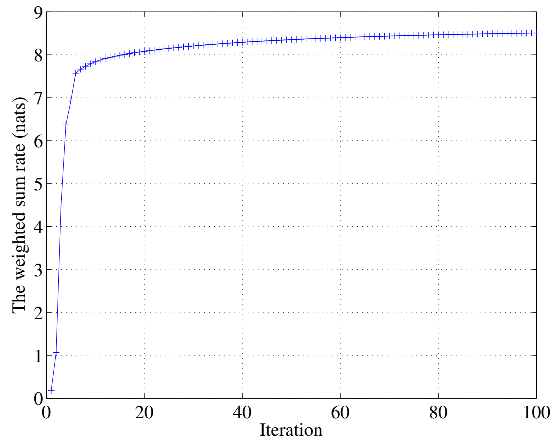

Fig. 3. The network with high interference and total power constraint.

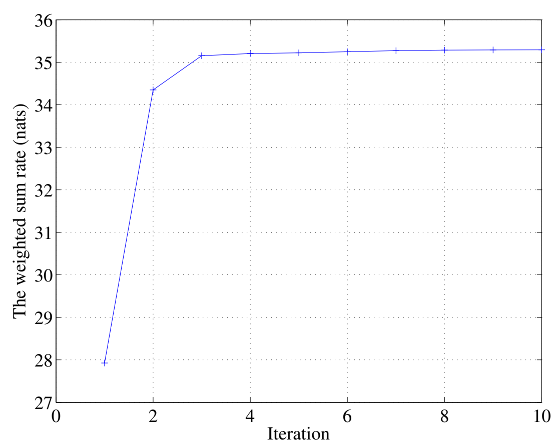

Fig. 4. The network with low interference and per-link power constraints.

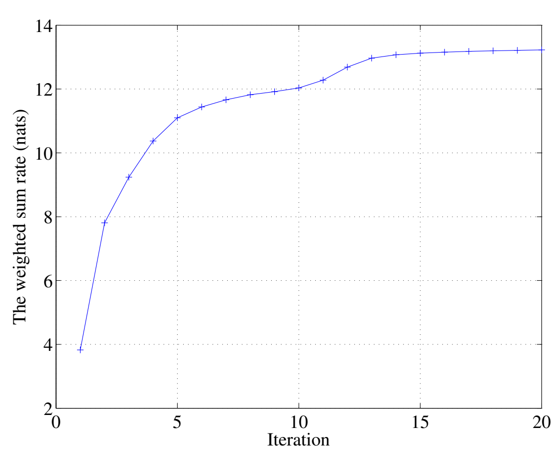

Fig. 5. The network with moderate interference and per-link power constraints.

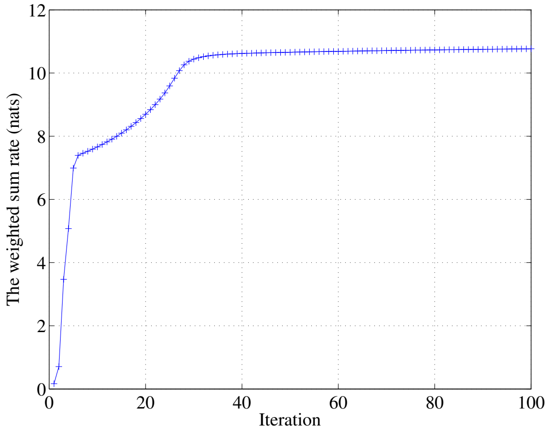

Fig. 6. The network with high interference and per-link power constraints.

C. The network with the general linear power constraints

We have tested and evaluated the general iterative minimax algorithm (Table I) for large networks with general linear power covariance constraints. Here we present results on a network with $|L| = 30$ interfering links, i.e., the network serves the 30 links or users simultaneously, without using any time, frequency, or code multiplexing. These links are partitioned into five groups: $\mathcal{S} = \{\{1\}, \{2\}, \{3, \cdots, 7\}, \{8, \cdots, 20\}, \{21, \cdots, 30\}\}$. We can view these different groups of links as being served by different "cells". We consider cases with different combinations of intracell interference (i.e., interference between the links within a cell) and intercell interference (i.e., interference between the links of different cells). The power constraint matrix $\mathbf{Q}_l^s$ is set to the (scaled) identity matrix without loss of generality, and the power budget of each cell is randomly drawn from $\{1, \cdots, 10\}$ and then post-multiplied by the cardinality of the cell.

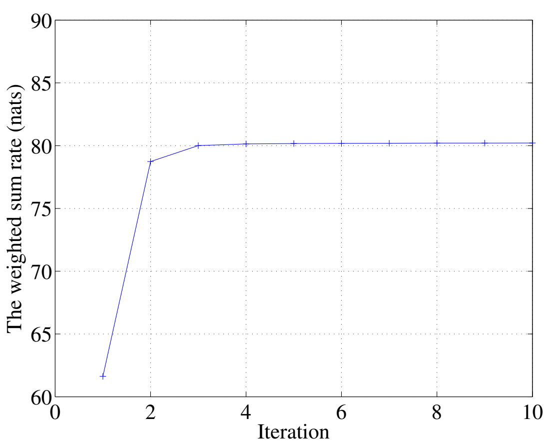

Fig. 7. The network with low intracell interference and low intercell interference.

Figures 7–9 show the monotonic convergence of our algorithm for the cases where the intracell interference – intercell interference are low – low, moderate – low, and high – moderate, respectively. As expected, the algorithm shows fast convergence.

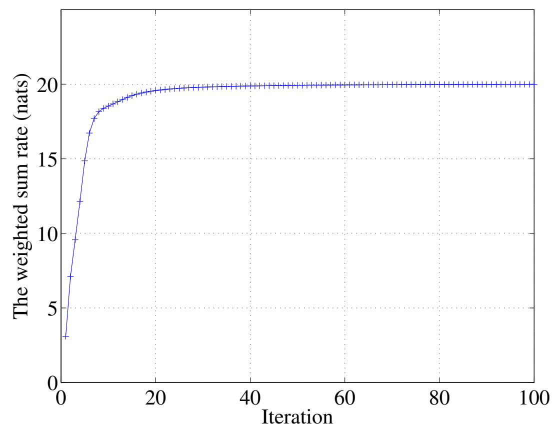

Fig. 8. The network with moderate intracell interference and low intercell interference.

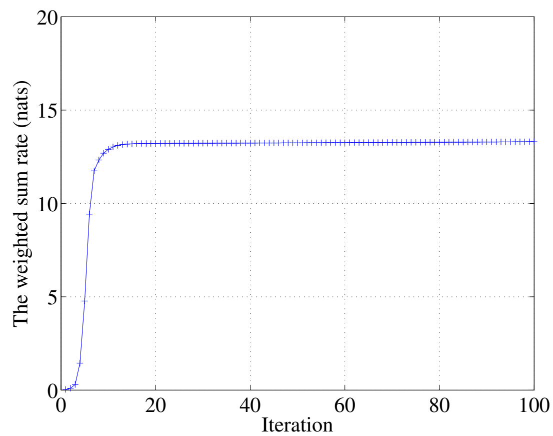

Fig. 9. The network with high intracell interference and moderate intercell interference.

Note that in our numerical experiments, we draw parameters such as the power budgets for the network randomly for each run of the simulation. So, the sum-rate achieved for specific examples may not appear proportional to the network size. For example, compared to Figures 5–6 of the 10-link network, Figures 8–9 of the 30-link network show a much lower rate per-link, which is because examples shown in Figures 8–9 happen to have much lower aggregate power.

D. Complexity analysis

We have evaluated in the above the monotonic convergence of the iterative minimax algorithm in terms of the number of iterations. We now analyze the complexity of *each iteration*. Recall that L is the number of data links, and for simplicity, assume that each link has N transmit (and receive) antennas, so the resulting $\mathbf{\Sigma}_l$ is an $N \times N$ matrix. Suppose that we use the straightforward matrix multiplication and inversion, then the complexity of these operations are $O(N^3)$. In each iteration, $\mathbf{\Omega}_l$ incurs a complexity of $O(LN^3)$, and so does $\mathbf{\Omega}_l + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n+1)}\mathbf{H}_{l,l}^+$. Furthermore, $\mathbf{\Phi}_l$ incurs a complexity of $O(LN^3)$, and so do $\tilde{\mathbf{\Sigma}}_l$ and $\mathbf{\Sigma}_l$. Since we need L of these operations, the total complexity is $O(L^2N^3)$. If we use faster matrix multiplication such as the one in [20] that has a complexity of $O(N^{2.3727})$, we can reduce computational complexity at each iteration to $O(L^2N^{2.3727})$.

VI. Conclusion

We have taken a new approach to the weighted sum-rate maximization in the MIMO interference networks, by formulating an equivalent max-min problem. The Lagrangian duality of the equivalent max-min problem provides an elegant way to establish the sum-rate duality between an interference network and its reciprocal, and more importantly, suggests a novel iterative minimax algorithm with monotonic convergence for the weighted sum-rate maximization. The design and the convergence proof of the iterative minimax algorithm use only general convex analysis and matrix analysis. They apply and extend to any max-min problems where the objective function is concave in the maximizing variables and convex in the minimizing variables and the constraints are convex, and thus provide a general class of algorithms for such optimization problems. This paper presents a promising step and lends hope for establishing a general framework based on the minimax Lagrangian duality for characterizing the weighted sum-rate and developing efficient resource allocation and interference management algorithms for general MIMO interference networks.

As further work, we will study the practical and distributed implementation of the iterative minimax algorithm and evaluate its performance under realistic characteristics of wireless channels. We are also studying to exploit the special structure (i.e., the objective function is convex in minimizing variables and concave in maximizing variables) of the problem to design algorithm that is guaranteed to converge to a global optimum. We will also investigate the application of the minimax duality and algorithm to other resource management and design problems in wireless networks. Lastly, we are exploring the algorithmic implication of the minimax duality for general nonconvex optimization problems that can be (re-)formulated equivalently as a max-min problem.

Appendix: Proof of Theorem 1

Before we present the proof, we first define an extended difference of logdet function. Let $\mathbf{A}, \mathbf{B} \in \mathcal{S}_+^n$, the difference of logdet function

$$F(\mathbf{A}, \mathbf{B}) = \log|\mathbf{A}+\mathbf{B}| - \log|\mathbf{B}|$$

is not well-defined if $\mathbf{B}$ is not positive definite. If there exists a nonsingular square matrix $\mathbf{T}$ such that

$$\mathbf{T}^+\mathbf{A}\mathbf{T} = \begin{bmatrix} \mathbf{A}_1 & \\ & \mathbf{0} \end{bmatrix}, \quad \mathbf{T}^+\mathbf{B}\mathbf{T} = \begin{bmatrix} \mathbf{B}_1 & \\ & \mathbf{0} \end{bmatrix}$$

where $\mathbf{A}_1 \in \mathcal{S}_+^m$, $\mathbf{B}_1 \in \mathcal{S}_{++}^m$ for some $m \le n$, then we can define an extended difference of logdet function:

$$F(\mathbf{A}, \mathbf{B}) := \log|\mathbf{A}_1+\mathbf{B}_1| - \log|\mathbf{B}_1|.$$

With the definition of the above extended function, matrix inverse resulting from the derivative of logdet function is pseudo inverse when the matrix involved is singular. In the following, *a difference of logdet function is meant to be the extended difference of logdet function, and matrix inverse is pseudo inverse when the matrix involved is singular*.

We now come to the proof of Theorem 1. For simplicity of presentation and without loss of generality, we reload notations and consider the following problem:

$$\max_{\mathbf{\Sigma}\succeq 0}\min_{\mathbf{\Omega}\succeq 0} \quad \log|\mathbf{\Omega}+\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+| - \log|\mathbf{\Omega}| + \mathrm{Tr}(\mathbf{\Lambda}\mathbf{\Omega}) - \mathrm{Tr}(\mathbf{\Phi}\mathbf{\Sigma}) \tag{52}$$

where $\mathbf{\Lambda} \succeq \mathbf{0}$ and $\mathbf{\Phi} \succeq \mathbf{0}$. The key idea of the proof is to show that problem (52) is equivalent to a problem with $\mathbf{\Omega}$ restricted to $\mathbf{\Omega} = \mathbf{H}\mathbf{X}\mathbf{H}^+,\ \mathbf{X} \succeq \mathbf{0}$.

Lemma 1. *The problem* (52) *is equivalent to the following problem:*

$$\max_{\mathbf{\Sigma}\succeq 0}\min_{\mathbf{\Omega}\succeq 0} \quad \log|\mathbf{\Omega}+\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+| - \log|\mathbf{\Omega}| + Tr(\mathbf{\Lambda}\mathbf{\Omega}) - Tr(\mathbf{\Phi}\mathbf{\Sigma}) \tag{53}$$

$$s.t. \quad \mathbf{\Omega} = \mathbf{H}\mathbf{X}\mathbf{H}^+, \quad \mathbf{X} \succeq \mathbf{0}. \tag{54}$$

Proof. Since $\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+ \succeq \mathbf{0}$ and $\mathbf{\Lambda} \succeq \mathbf{0}$, there exists a nonsingular square matrix $\mathbf{T}$ such that

$$\mathbf{T}\mathbf{\Lambda}\mathbf{T}^+ = \begin{bmatrix} \mathbf{S}_1 & & & \\ & \mathbf{0} & & \\ & & \mathbf{S}_3 & \\ & & & \mathbf{0} \end{bmatrix},$$

$$(\mathbf{T}^+)^{-1}\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{S}_1 & & & \\ & \mathbf{S}_2 & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix},$$

where $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ are diagonal and positive definite; see, e.g., Theorem 3.22 in [28]. Let $\mathbf{\Omega} = \mathbf{T}^+\tilde{\mathbf{\Omega}}\mathbf{T}$, problem (52) becomes

$$\max_{\mathbf{\Sigma}\succeq 0}\min_{\tilde{\mathbf{\Omega}}\succeq 0} \quad \log|\tilde{\mathbf{\Omega}} + (\mathbf{T}^+)^{-1}\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+\mathbf{T}^{-1}| - \log|\tilde{\mathbf{\Omega}}| + \mathrm{Tr}\left(\mathbf{T}\mathbf{\Lambda}\mathbf{T}^+\tilde{\mathbf{\Omega}}\right) - \mathrm{Tr}(\mathbf{\Phi}\mathbf{\Sigma}).$$

Now, consider those terms in the objective function that depend on $\tilde{\mathbf{\Omega}}$:

$$\begin{aligned}
&\tilde{\mathcal{L}}(\tilde{\mathbf{\Omega}}) \\
=\ & \log|\tilde{\mathbf{\Omega}} + (\mathbf{T}^+)^{-1}\mathbf{H}\mathbf{\Sigma}\mathbf{H}^+\mathbf{T}^{-1}| - \log|\tilde{\mathbf{\Omega}}| + \mathrm{Tr}\left(\mathbf{T}\mathbf{\Lambda}\mathbf{T}^+\tilde{\mathbf{\Omega}}\right) \\
=\ & \log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11}+\mathbf{S}_1 & \tilde{\mathbf{\Omega}}_{12} & \tilde{\mathbf{\Omega}}_{13} & \tilde{\mathbf{\Omega}}_{14} \\ \tilde{\mathbf{\Omega}}_{12}^+ & \tilde{\mathbf{\Omega}}_{22}+\mathbf{S}_2 & \tilde{\mathbf{\Omega}}_{23} & \tilde{\mathbf{\Omega}}_{24} \\ \tilde{\mathbf{\Omega}}_{13}^+ & \tilde{\mathbf{\Omega}}_{23}^+ & \tilde{\mathbf{\Omega}}_{33} & \tilde{\mathbf{\Omega}}_{34} \\ \tilde{\mathbf{\Omega}}_{14}^+ & \tilde{\mathbf{\Omega}}_{24}^+ & \tilde{\mathbf{\Omega}}_{34}^+ & \tilde{\mathbf{\Omega}}_{44} \end{bmatrix}\right| \\
& - \log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11} & \tilde{\mathbf{\Omega}}_{12} & \tilde{\mathbf{\Omega}}_{13} & \tilde{\mathbf{\Omega}}_{14} \\ \tilde{\mathbf{\Omega}}_{12}^+ & \tilde{\mathbf{\Omega}}_{22} & \tilde{\mathbf{\Omega}}_{23} & \tilde{\mathbf{\Omega}}_{24} \\ \tilde{\mathbf{\Omega}}_{13}^+ & \tilde{\mathbf{\Omega}}_{23}^+ & \tilde{\mathbf{\Omega}}_{33} & \tilde{\mathbf{\Omega}}_{34} \\ \tilde{\mathbf{\Omega}}_{14}^+ & \tilde{\mathbf{\Omega}}_{24}^+ & \tilde{\mathbf{\Omega}}_{34}^+ & \tilde{\mathbf{\Omega}}_{44} \end{bmatrix}\right| \\
& + \mathrm{Tr}\left(\mathbf{S}_1\tilde{\mathbf{\Omega}}_{11}\right) + \mathrm{Tr}\left(\mathbf{S}_3\tilde{\mathbf{\Omega}}_{33}\right)
\end{aligned}$$

and its minimization over $\tilde{\mathbf{\Omega}} \succeq \mathbf{0}$. By the determinant formula for block matrix, when A is invertible $\left|\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right| = |A|\left|D - CA^{-1}B\right|$, and the fact that the determinant is a continuous function, we have

$$\begin{aligned}\tilde{\mathcal{L}}(\tilde{\mathbf{\Omega}}) \;\geq\; & \log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11}+\mathbf{S}_1 & \tilde{\mathbf{\Omega}}_{12} & \tilde{\mathbf{\Omega}}_{13} \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22}+\mathbf{S}_2 & \tilde{\mathbf{\Omega}}_{23} \\ \tilde{\mathbf{\Omega}}_{13}^{+} & \tilde{\mathbf{\Omega}}_{23}^{+} & \tilde{\mathbf{\Omega}}_{33} \end{bmatrix}\right| \\ & -\log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11} & \tilde{\mathbf{\Omega}}_{12} & \tilde{\mathbf{\Omega}}_{13} \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22} & \tilde{\mathbf{\Omega}}_{23} \\ \tilde{\mathbf{\Omega}}_{13}^{+} & \tilde{\mathbf{\Omega}}_{23}^{+} & \tilde{\mathbf{\Omega}}_{33} \end{bmatrix}\right| \\ & +\mathrm{Tr}\left(\mathbf{S}_1\tilde{\mathbf{\Omega}}_{11}\right) + \mathrm{Tr}\left(\mathbf{S}_3\tilde{\mathbf{\Omega}}_{33}\right),\end{aligned}$$

where the equality can be achieved when $\tilde{\mathbf{\Omega}}_{i4} = \mathbf{0}$ for all $i = 1,2,3,4$.[5] We will restrict $\tilde{\mathbf{\Omega}}$ to those with $\tilde{\mathbf{\Omega}}_{i4} = \mathbf{0}$ for all $i = 1,2,3,4$, as the equality is achieved at one of those matrices.

Since $\mathbf{S}_3 \succeq \mathbf{0}$ and $\tilde{\mathbf{\Omega}}_{33} \succeq \mathbf{0}$, $\mathrm{Tr}\left(\mathbf{S}_3\tilde{\mathbf{\Omega}}_{33}\right) \geq \mathbf{0}$. We further have

$$\begin{aligned}\tilde{\mathcal{L}}(\tilde{\mathbf{\Omega}}) \;\geq\; & \log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11}+\mathbf{S}_1 & \tilde{\mathbf{\Omega}}_{12} \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22}+\mathbf{S}_2 \end{bmatrix}\right| \\ & -\log\left|\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11} & \tilde{\mathbf{\Omega}}_{12} \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22} \end{bmatrix}\right| + \mathrm{Tr}\left(\mathbf{S}_1\tilde{\mathbf{\Omega}}_{11}\right),\end{aligned}$$

where the equality is achieved when additionally $\tilde{\mathbf{\Omega}}_{i3} = \mathbf{0}$ for all $i = 1,2,3$. Therefore, we conclude that there exists a minimizer $\tilde{\mathbf{\Omega}}^{\star}$ with the form:

$$\tilde{\mathbf{\Omega}}^{\star} = \begin{bmatrix} \tilde{\mathbf{\Omega}}_{11} & \tilde{\mathbf{\Omega}}_{12} & & \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22} & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix}.$$

The above manipulation is to restrict the problem to an equivalent, *truncated system* where we ignore the interference-plus-noise of a channel whose signal is zero. As mentioned in Section III-A, intuitively, the equivalence of this truncated system to the original max-min problem follows from the fact that when the signal is zero it does not matter what the interference-plus-noise is.

Now, consider a vector v such that $\mathbf{H}^{+}v = \mathbf{0}$, we have

$$\begin{aligned} v^{+}\mathbf{H}\mathbf{\Sigma}\mathbf{H}^{T}v &= v^{+}\mathbf{T}^{+}\begin{bmatrix} \mathbf{S}_1 & & & \\ & \mathbf{S}_2 & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix}\mathbf{T}v \\ &= 0,\end{aligned}$$

which implies

$$\mathbf{T}v = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \bar{\mathbf{v}}_3 & \bar{\mathbf{v}}_4 \end{bmatrix}^{T}.$$

[5]By the determinant formula and the Sylvester's criterion for positive semidefinite matrix, if $\tilde{\mathbf{\Omega}}_{44} = \mathbf{0}$, then $\tilde{\mathbf{\Omega}}_{i4} = \mathbf{0}$ for all $i = 1,2,3$.

Therefore,

$$\begin{aligned} v^{+}\mathbf{\Omega}^{\star}v &= v^{+}\mathbf{T}^{+}\tilde{\mathbf{\Omega}}^{\star}\mathbf{T}v \\ &= v^{+}\mathbf{T}^{+}\begin{bmatrix} \tilde{\mathbf{\Omega}}_{11} & \tilde{\mathbf{\Omega}}_{12} & & \\ \tilde{\mathbf{\Omega}}_{12}^{+} & \tilde{\mathbf{\Omega}}_{22} & & \\ & & \mathbf{0} & \\ & & & \mathbf{0} \end{bmatrix}\mathbf{T}v \\ &= 0.\end{aligned}$$

This implies the null space $\mathcal{N}(H^{+}) \subset \mathcal{N}(\mathbf{\Omega}^{\star})$, and further, the range $\mathcal{R}(H) \supset \mathcal{R}(\mathbf{\Omega}^{\star +}) = \mathcal{R}(\mathbf{\Omega}^{\star})$. Therefore, there exists a matrix $\mathbf{X} \succeq \mathbf{0}$ such that

$$\mathbf{\Omega}^{\star} = \mathbf{H}\mathbf{X}\mathbf{H}^{+}.$$

We conclude that there exists an optimal solution with $\mathbf{\Omega}^{\star} = \mathbf{H}\mathbf{X}\mathbf{H}^{+}$, and thus problem (52) and problem (53)-(54) are equivalent. $\square$

With Lemma 1, we are ready to present the explicit saddle point solution. Consider the logdet terms in the objective function:

$$\begin{aligned} & \log|\mathbf{\Omega}+\mathbf{H}\mathbf{\Sigma}\mathbf{H}^{+}| - \log|\mathbf{\Omega}| \\ =\; & \log|\mathbf{H}\mathbf{X}\mathbf{H}^{+}+\mathbf{H}\mathbf{\Sigma}\mathbf{H}^{+}| - \log|\mathbf{H}\mathbf{X}\mathbf{H}^{+}| \\ =\; & \log|\mathbf{H}^{+}\mathbf{H}(\mathbf{X}+\mathbf{\Sigma})| - \log|\mathbf{H}^{+}\mathbf{H}\mathbf{X}| \\ =\; & \log|\mathbf{X}+\mathbf{\Sigma}| - \log|\mathbf{X}|.\end{aligned}$$

The singularity issue comes out when $\mathbf{H}^{+}\mathbf{H}$ is not invertible, but this can be handled by adding a small term $\kappa\mathbf{I}$, $\kappa > 0$ to $\mathbf{H}^{+}\mathbf{H}$ and then taking the limit $\kappa \to 0$. Thus, we can transform problem (52) into the following simple one:

$$\max_{\mathbf{\Sigma}\succeq 0}\min_{\mathbf{X}\succeq 0} \; \log|\mathbf{X}+\mathbf{\Sigma}| - \log|\mathbf{X}| + \mathrm{Tr}\left(\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H}\mathbf{X}\right) - \mathrm{Tr}\left(\mathbf{\Phi}\mathbf{\Sigma}\right). \tag{55}$$

By the first order optimality condition for the saddle point, we have

$$\begin{aligned} (\mathbf{X}+\mathbf{\Sigma})^{-1} - \mathbf{\Phi} &= \mathbf{0}, \\ (\mathbf{X}+\mathbf{\Sigma})^{-1} - \mathbf{X}^{-1} + \mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H} &= \mathbf{0},\end{aligned}$$

from which we obtain the following explicit saddle point solution:

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{\Phi}^{-1} - (\mathbf{\Phi}+\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H})^{-1}, \qquad (56) \\ \mathbf{X} &= (\mathbf{\Phi}+\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H})^{-1},\end{aligned}$$

and in terms of Ω we have

$$\mathbf{\Omega} = \mathbf{H}\mathbf{X}\mathbf{H}^{+} = \mathbf{H}(\mathbf{\Phi}+\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H})^{-1}\mathbf{H}^{+}. \tag{57}$$

Note that problem (55) is well-defined only when $\mathbf{\Sigma}$, $\mathbf{X}$ satisfy the property specified for matrices $\mathbf{A}$, $\mathbf{B}$ in the beginning of this Appendix. This is verified as follows.

Proposition 1. *The objective function in problem* (55) *is well-defined for matrices* $\mathbf{\Sigma}$, $\mathbf{X}$ *that are given by* (56)-(57)*.*

Proof. Let $\mathbf{\Psi} = \mathbf{\Phi} + \mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H}$, we have that the null space $\mathcal{N}(\mathbf{\Psi}) \subset \mathcal{N}(\mathbf{\Phi})$. To see this, note that $\mathbf{\Phi} \succeq \mathbf{0}$ and $\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H} \succeq \mathbf{0}$. Suppose $v \in \mathcal{N}(\mathbf{\Psi})$, then $v^{T}\mathbf{\Psi}v = v^{T}\mathbf{\Phi}v + v^{T}\mathbf{H}^{+}\mathbf{\Lambda}\mathbf{H}v = 0$. As each term is nonnegative, $v^{T}\mathbf{\Phi}v = 0$, i.e., $v \in \mathcal{N}(\mathbf{\Phi})$.

Since $\mathcal{N}(\mathbf{\Psi}) \subset \mathcal{N}(\mathbf{\Phi})$, there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{U}^{+}\mathbf{\Psi}\mathbf{U} = \begin{bmatrix} \mathbf{\Psi}_1 & \\ & \mathbf{0} \end{bmatrix}, \quad \mathbf{U}^{+}\mathbf{\Phi}\mathbf{U} = \begin{bmatrix} \mathbf{\Phi}_1 & \\ & \mathbf{0} \end{bmatrix}$$

where $\mathbf{\Psi}_1 \succ \mathbf{0}$ and $\mathbf{\Phi}_1 \succeq \mathbf{0}$. By equations (56)-(57),

$$\mathbf{X} = \mathbf{U} \begin{bmatrix} \mathbf{\Psi}_1^{-1} & \\ & \mathbf{0} \end{bmatrix} \mathbf{U}^{+}, \ \mathbf{\Sigma} = \mathbf{U} \begin{bmatrix} \mathbf{\Phi}_1^{-1} - \mathbf{\Psi}_1^{-1} & \\ & \mathbf{0} \end{bmatrix} \mathbf{U}^{+}.$$

Note that $\mathbf{\Psi}_1^{-1} \succ \mathbf{0}$, so by the definition of the extended difference of logdet function, the objective function in problem (55) is well-defined. □

With (56)-(57), we can easily recover the explicit solution (12)-(13). This concludes the proof of Theorem 1.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[2] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi. Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design. *IEEE Transactions on Wireless Commmunications*, 7(12):4792 – 4799, December 2008.

[3] M. Hong, Y.-F. Liu Q. Li, and Z.-Q. Luo. Decomposition by Successive Convex Approximation: A Unifying Approach for Linear Transceiver Design in Interfering Heterogenous Networks. *arXiv preprint arXiv:1210.1507*, 2012.

[4] H. Huh, H. Papadopoulos, and G. Caire. MIMO Broadcast Channel Optimization under General Linear Constraints. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2009.

[5] H. Huh, H. C. Papadopoulos, and G. Caire. Multiuser MISO Transmitter Optimization for Intercell Interference Mitigation. *IEEE Transactions on Signal Processing*, 58(8):4272 –4285, August 2010.

[6] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith. Sum Power Iterative Water-filling for Multi-antenna Gaussian Broadcast Channels. *IEEE Transactions on Information Theory*, 51(4):1570–1580, April 2005.

[7] S. Kaviani, O. Simeone, W. A. Krzymien, and S. Shamai (Shitz). Linear Precoding and Equalization for Network MIMO with Partial Cooperation. *arXiv preprint arXiv:1302.2187*, 2013.

[8] M. G. Kibria and H. Murata. An Efficient Algorithm for Weighted Sum-rate Maximization in Multicell Downlink Beamforming. *arXiv preprint arXiv:1309.4203*, 2013.

[9] S.-J. Kim and G.B. Giannakis. Optimal Resource Allocation for MIMO Ad Hoc Cognitive Radio Networks. *IEEE Transactions on Information Theory*, 57(5):3117 – 3131, May 2011.

[10] X. Li, S. You, L. Chen, A. Liu, and Y. Liu. A New Algorithm for the Weighted Sum Rate Maximization in MIMO Interference Networks. *IEEE International Symposium on Information Theory (ISIT)*, submitted, 2014.

[11] A. Liu, Y. Liu, H. Xiang, and W. Luo. Duality, Polite Water-filling, and Optimization for MIMO B-MAC Interference Networks and iTree Networks. *arXiv preprint arXiv:1004.2484*, 2010.

[12] A. Liu, Y. Liu, H. Xiang, and W. Luo. Polite Water-filling for Weighted Sum-rate Maximization in MIMO B-MAC Networks under Multiple Linear Constraints. *IEEE Transactions on Signal Processing*, 60(2):834–847, 2012.

[13] C. TK Ng and H. Huang. Linear precoding in Cooperative MIMO Cellular Networks with Limited Coordination Clusters. *IEEE Journal on Selected Areas in Communications*, 28(9):1446–1454, December 2010.

[14] C. Shi, R. A. Berry, and M. L. Honig. Monotonic Convergence of Distributed Interference Pricing in Wireless Networks. *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, June 2009.

[15] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He. An Iteratively Weighted MMSE Approach to Distributed Sum-utility Maximization for a MIMO Interfering Broadcast Channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, 2011.

[16] C. W. Tan, M. Chiang, and R. Srikant. Maximizing Sum Rate and Minimizing MSE on Multiuser Downlink: Optimality, Fast Algorithms and Equivalence via Max-min SINR. *IEEE Transactions on Signal Processing*, 59(12):6127 – 6143, August 2011.

[17] P. Viswanath, N. Jindal, and A. Goldsmith. Duality, Achievable Rates, and Sum-Rate Capacity of of Gaussian MIMO Broadcast Channels. *IEEE Transactions on Information Theory*, 49(10):2658–2668, 2003.

[18] P. Viswanath and D. N. C. Tse. Sum Capacity of the Multiple Antenna Gaussian Broadcast Channel and Uplink-Downlink Duality. *IEEE Transactions on Information Theory*, 49(8):1912–1921, 2003.

[19] P. C. Weeraddana, M. Codreanu, and M. Latva-aho. *Weighted Sum-Rate Maximization in Wireless Networks: A Review*. Now Publishers Inc., Hanover, MA, USA, 2012.

[20] V. V. Williams. Multiplying Matrices Faster than Coppersmith-Winograd. In *Proceedings of the 44th Symposium on Theory of Computing*, pages 887–898. ACM, 2012.

[21] S. You, L. Chen, and Y. Liu. The Weighted Sum Rate Maximization in MIMO Interference Networks: A Convex-Concave Procedure Approach. *IEEE International Symposium on Information Theory (ISIT)*, submitted, 2014.

[22] W. Yu. Sum-Capacity Computation for the Gaussian Vector Broadcast Channel via Dual Decomposition. *IEEE Transactions on Information Theory*, 52(2):754–759, February 2006.

[23] W. Yu. Multiuser Water-filling in the Presence of Crosstalk. In *Proceedings of Information Theory and Applications Workshop*, pages 414 –420, February 2007.

[24] W. Yu. Uplink-Downlink Duality Via Minimax Duality. *IEEE Transactions on Information Theory*, 52(2):361–374, February 2006.

[25] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi. Iterative Water-Filling for Gaussian Vector Multiple-Access Channels. *IEEE Transactions on Information Theory*, 50(1):145–152, January 2004.

[26] L. Zhang, Y.-C. Liang, Y. Xin, R. Zhang, and H.V. Poor. On Gaussian MIMO BC-MAC Duality with Multiple Transmit Covariance Constraints. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, June 2009.

[27] L. Zhang, Y. Xin, and Y.-C. Liang. Weighted Sum Rate Optimization for Cognitive Radio MIMO Broadcast Channels. *IEEE Transactions on Wireless Communications*, 8(6):2950 –2959, June 2009.

[28] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, New Jersey, 1995.

PLACE PHOTO HERE

Lijun Chen (M'05) is an Assistant Professor of Computer Science and Telecommunications at University of Colorado at Boulder. He received the Ph.D. from California Institute of Technology. He was a co-recipient of the Best Paper Award at the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) in 2007. His current research interests include optimization and control of networked systems, distributed optimization and control, convex relaxation and parsimonious solutions, and game theory and its engineering application. He is currently an editor of IEEE Transactions on Communications.

PLACE PHOTO HERE

Seungil You (SM'12) received the B.S. degree in electrical engineering from the Seoul National University, Seoul, Korea, in 2011. He is currently a Ph.D. student at the Department of Computing and Mathematical Science, California Institute of Technology. His research interests include information theory, control theory, statistical signal processing, smart grids and optimization.