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

Lijun Chen  Seungil You

Abstract—We take a new perspective on the weighted sum-rate maximization in multiple-input multiple-output (MIMO) interference networks, by formulating an equivalent max-min problem. 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 when such a duality exists, and more importantly, suggests a novel iterative minimax algorithm for the weighted sum-rate maximization. Moreover, the design and convergence proof of the algorithm use only general convex analysis. They apply and extend to any 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 characterizing the weighted sum-rate and developing efficient 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 is a fundamental problem in information theory and communications, and 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 polite water filling algorithm [7] and iterative weighted MMSE algorithm [12] 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 channel, 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 constraints, assuming Gaussian transmit signal, Gaussian noise, and the availability of channel state information at the transmitter (Section III). It typifies a class of problems that are key to the next generation wireless communication networks where the interference is a limiting factor; and various algorithms have been proposed for this problem or its special cases, see, e.g., [2]–[8], [10]–[12], [17], [19]–[22].

In a recent work [6], we and our coauthors propose a new algorithm for the weighted sum-rate maximization in the MIMO interference network with the total power constraint and establish its convergence. The convergence proof suggests certain general structure behind the problem and its possible connection to the Lagrangian duality. In this paper, we identify such a structure and establish its connection to the minimax Lagrangian duality for the weighted sum-rate maximization with general linear constraints, and explore its implications for the sum-rate characterization and algorithm design. 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. 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 when such a duality exists (Section III), and more importantly, suggests a novel algorithm, termed the iterative minimax algorithm, for the weighted sum-rate maximization (Section IV). Moreover, the design and convergence proof of the 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 provides a general class of algorithms for such optimization problems.

The iterative minimax algorithm is based partially on an explicit saddle point solution of certain max-min optimization (Section III-A). This explicit solution has been identified for the case where the matrices involved are all invertible in [18]. In contrast, we prove 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 in and to some extent can be seen as a substantial extension of the seminal work by Yu [18] that establishes uplink-downlink duality via minimax duality for the sum capacity of the Gaussian broadcast channel.
Our model is much more general and the results expect to find broad applications, and we establish the explicit saddle point solution for the max-min optimization with general matrices, and more importantly, we develop a novel algorithm for the weighted sum-rate maximization.

II. SYSTEM MODEL

Consider a general interference network with a set \( L \) of MIMO 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 \( x_l \in \mathbb{C}^{n_l \times 1} \) denote the transmit signal of link \( l \), which is assumed to be circularly symmetric complex Gaussian. The received signal \( y_l \in \mathbb{C}^{m_l \times 1} \) at the receiver \( r_l \) can be written as

\[
y_l = \sum_{k \in L} H_{lk} x_k + w_{tl},
\]

where \( H_{lk} \in \mathbb{C}^{m_l \times n_l} \) denotes the channel matrix from the transmitter \( t_k \) to the receiver \( r_l \), and \( w_{tl} \in \mathbb{C}^{m_l \times 1} \) denotes the additive circularly symmetric complex Gaussian noise with identity covariance matrix.

The interference network defined 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 constraints

Denote by \( \Sigma_l \geq 0 \) the covariance matrix of the transmit signal \( x_l \), \( l \in L \). 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. 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 \( Q^l > 0 \) for each \( s \in S^l \); and any 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 constraint as follows:

\[
\sum_{l \in L^s} \text{Tr}(\Sigma_l Q^l_l) \leq 1, \quad s \in S.
\]

The constraint (2) is very general and captures all reasonable linear power constraints. For example, when there is only a budget \( P_T \) on the total power of all links as considered in many existing work such as [6], the cardinality \(|S| = 1\) and \( Q^l \equiv \frac{1}{P_T} I \). When there is only a per-link power budget \( p_l \), \( l \in L \), each group \( L^s \) contains only one link and \( Q^l \equiv \frac{1}{P_l} I \). Each group \( L^s \), \( s \in S \) may also represent those links or users in a cell of a microcell network and each cell is subject to a total power budget \( P_s \). In this scenario, the subsets \( L^s \) are non-overlapping and \( Q^l \equiv \frac{1}{P_s} I \), \( \forall l \in L^s \).

Remark 1: We have assumed linear power covariance constraints. 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 constraints.

B. The weighted sum-rate maximization

Assume that the channel state information is known. For given power covariance matrix \( \Sigma_l \), \( l \in L \), an achievable rate \( R_l \) of the link \( l \) is given by

\[
R_l = \log \left| I + H_{ll} \Sigma_l H_{ll}^+ \right|, \quad l \in L.
\]

where \( | \cdot | \) denotes the matrix determinant and the interferences from other links are treated as noise. 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 constraints:

\[
\max_{\Sigma_l \succeq 0} \sum_{l \in L} w_l R_l
\]

s.t.

\[
\sum_{l \in L^s} \text{Tr}(\Sigma_l Q^l_l) \leq 1, \quad s \in S.
\]

The weighted sum-rate maximization is in general a hard non-convex problem. It is a fundamental problem in information theory and communications and 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 when such a duality exists, and more importantly, suggests a new algorithm for the weighted sum-rate maximization.

A. The minimax Lagrangian duality

Denote by \( \Omega_l \), \( l \in L \) the interference-plus-noise covariance matrix at the receiver \( r_l \), i.e.,

\[
\Omega_l = I + \sum_{k \in L \setminus \{l\}} H_{lk} \Sigma_k H_{lk}^+.
\]

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

\[
\max_{\Sigma_l \succeq 0} \min_{H_{ll}} \sum_{l \in L} w_l \left( \log |\Omega_l + H_{ll} \Sigma_l H_{ll}^+| - \log |\Omega_l| \right)
\]

s.t.

\[
\sum_{l \in L^s} \text{Tr}(\Sigma_l Q^l_l) \leq 1, \quad s \in S,
\]

\[
\Omega_l = I + \sum_{k \in L \setminus \{l\}} H_{lk} \Sigma_k H_{lk}^+.
\]

Note that, when \( H_{ll} \Sigma_l H_{ll}^+ \) is not of full rank, the above problem is equivalent to a truncated system where \( \Omega_l \) is restricted to \( \Omega_l = H_{ll} X_l H_{ll}^+ \), \( X_l \succeq 0 \). 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 \(\text{(1)}-\text{(3)}\)
\[
F(\Sigma, \Omega) = \sum_{l \in L} w_l \left( \log |\Omega_l + H_{il} \Sigma_l H_{il}^H| - \log |\Omega_l| \right)
\]
is concave in \(\Sigma\) and convex in \(\Omega\). So, the max-min is equal to min-max, and the optimum is a saddle point. Consider the Lagrangian
\[
L(\Sigma, \Omega, \Lambda, \mu) = F(\Sigma, \Omega) + \sum_{s \in S} \mu_s \left( 1 - \sum_{l \in L^s} \text{Tr}(\Sigma_s Q_l^s) \right) + \sum_{l \in L} \text{Tr} \left( \Lambda_l (\Omega_l - I - \sum_{k \in L \setminus \{l\}} H_{lk} \Sigma_k H_{lk}^H) \right),
\]
where \(\mu_s \geq 0\) is the dual variable associated with the power constraint (8) and \(\Lambda_l \geq 0\) is the dual variable associated with the interference-plus-noise covariance matrix definition (9). For any given \((\Lambda, \mu)\), \(L\) is concave in \(\Sigma\) and convex in \(\Omega\) as \(F\) is.

Consider the first order condition (part of the KKT condition [1]) for the optimum \(\tilde{\Sigma}_l\)
\[
\begin{align*}
w_l H_{il}^+ \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} H_{il} & = \Phi_l, \\
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} \right) & = \Lambda_l,
\end{align*}
\]
where
\[
\Phi_l = \sum_{s \in S_l} \mu_s Q_l^s + \sum_{k \in L \setminus \{l\}} H_{lk}^+ A_k H_{lk}.
\]
For any given feasible dual variable \((\Lambda, \mu)\), the above condition gives the saddle point condition of Lagrangian \(L\) as a function of \((\Sigma, \Omega)\); and when \((\Lambda, \mu)\) is a dual optimum, solving \((10)-(11)\) 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.

Lemma 1: Given feasible dual variables \((\Phi, \mu)\), an explicit solution \((\tilde{\Sigma}, \tilde{\Omega})\) for the saddle point equations \((10)-(11)\) is given by
\[
\begin{align*}
w_l \left( \tilde{H}_{il}^+ \left( \tilde{\Phi}_l + H_{il}^+ A_l H_{il} \right)^{-1} H_{il}^+ \right) & = \tilde{\Omega}_l, \\
w_l \left( \tilde{\Phi}_l^{-1} - \left( \tilde{\Phi}_l + H_{il}^+ A_l H_{il} \right)^{-1} \right) & = \tilde{\Sigma}_l,
\end{align*}
\]
The solution \((12)-(13)\) is motivated by [18] that focuses on an (primal-dual) optimum and where correspondingly the optimal power covariance matrix \(\Sigma_l\) and the interference-plus-noise matrix \(\Omega_l\) are assumed to be positive definite and the channel matrix \(H_{il}\) is assumed to be square and invertible. Here, the solution is 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 Lemma 1 is rather involved, and is presented in the Appendix.

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 \((1)-(3)\).

B. Case studies

We now discuss two typical cases, and show how the minmax Lagrangian duality can be used to establish the 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 \(S = L\), and \(\Omega_l = I\) and \(Q_l = \frac{1}{P_l}\); with \(P_l\) the power budget at each link \(l \in L\). As each link is independent, we can just focus on one link:
\[
\begin{align*}
\max_{\Sigma \geq 0} & \sum_{l \in L} \log |\Omega_l + H_{il} \Sigma_l H_{il}^H| - \log |\Omega_l| \\
\text{s.t.} & \quad \text{Tr}(\Sigma_l) \leq 1, \quad \Omega_l = I,
\end{align*}
\]
The first order condition \((10)-(11)\) reduces to
\[
\begin{align*}
w_l H_{il}^+ \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} H_{il} & = \mu_l \frac{I}{P_l}, \\
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} \right) & = A_l,
\end{align*}
\]
where \(\mu_l \geq 0\) is the dual variable associated with the power constraint. Define
\[
\begin{align*}
\Sigma_l & = \frac{P_l}{\mu_l} A_l, \\
\Omega_l & = I.
\end{align*}
\]
The first order condition becomes
\[
\begin{align*}
w_l H_{il}^+ \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} H_{il} & = \mu_l \frac{\Omega_l}{P_l}, \\
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} \right) & = \mu_l \frac{\Sigma_l}{P_l},
\end{align*}
\]
and the explicit solution \((12)-(13)\) becomes
\[
\begin{align*}
w_l H_{il}^+ \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} H_{il} & = \frac{\mu_l}{P_l} \Omega_l, \\
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il} \Sigma_l H_{il}^H \right)^{-1} \right) & = \frac{\mu_l}{P_l} \Sigma_l.
\end{align*}
\]

Compare equations \((16)-(17)\) and equations \((18)-(19)\), we can conclude that the Lagrangian dual of the max-min problem
is also a max-min problem:

\[
\max_{\Sigma_l \succeq 0} \min_{\Omega_l \succeq 0} \log |\Omega_l + H_{il}^T \Sigma_l H_{il}| - \log |\Omega_l| \tag{20}
\]

\[
\text{s.t. } \text{Tr} \left( \frac{\Sigma_l}{P_l} \right) \leq 1, \quad \Omega_l = I, \tag{21}
\]

which is the sum-rate maximization problem defined on the reciprocal link with channel matrix \(H_i^T\). 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 \(\hat{\mu}_l\) and \(\hat{\Lambda}_l\) for the problem (20)-(21), we have the following correspondence:

\[
(\Sigma_l; \Lambda_l, \mu_l) = \left( \frac{P_l}{\hat{\mu}_l} \hat{\Lambda}_l; \frac{\hat{\mu}_l}{P_l} \Sigma_l, \mu_l \right), \tag{22}
\]

\[
(\hat{\Sigma}_l; \hat{\Lambda}_l, \hat{\mu}_l) = \left( \frac{P_l}{\hat{\mu}_l} \hat{\Lambda}_l; \frac{\hat{\mu}_l}{P_l} \Sigma_l, \mu_l \right). \tag{23}
\]

This recovers the well-known result in \([14], [15], [18]\). The difference from \([18]\) is that we establish the explicit solution \([13]-[19]\) and the correspondence \((22)-(23)\) 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 \(Q_l = \frac{1}{P_l}\), with \(P_T\) the total power budget. The max-min problem (7)-(9) reduces to

\[
\max_{\Sigma_l \succeq 0} \min_{\Omega_l \succeq 0} \sum_l w_l \left( \log |\Omega_l + H_{il}^T \Sigma_l H_{il}| - \log |\Omega_l| \right) \tag{24}
\]

\[
\text{s.t. } \sum_l \text{Tr} \left( \frac{\Sigma_l}{P_l} \right) \leq 1, \tag{25}
\]

\[
\Omega_l = I + \sum_{k \in L \setminus \{l\}} H_{kl}^+ \Sigma_k H_{kl}, \tag{26}
\]

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

\[
w_l H_{il}^+ \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} H_{il} = \Phi_l, \tag{27}
\]

\[
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} \right) = \Lambda_l, \tag{28}
\]

with \(\Phi_l = \frac{\mu_l}{P_l} \sum_{k \in L \setminus \{l\}} H_{kl}^+ \Sigma_k H_{kl}\), where \(\mu_l \geq 0\) is the dual variable associated with the total power constraint. Define

\[
\hat{\Sigma}_l = \frac{P_T}{\mu_l} \Lambda_l, \tag{29}
\]

\[
\hat{\Omega}_l = I + \sum_{k \in L \setminus \{l\}} H_{kl}^+ \hat{\Sigma}_k H_{kl}, \tag{30}
\]

The first order condition becomes

\[
w_l H_{il}^+ \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} H_{il} = \frac{\mu_l}{P_T} \hat{\Omega}_l, \tag{31}
\]

\[
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} \right) = \frac{\mu_l}{P_T} \hat{\Sigma}_l, \tag{32}
\]

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

\[
w_l H_{il}^+ \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} H_{il}^+ = \frac{\mu_l}{P_T} \hat{\Omega}_l, \tag{33}
\]

\[
w_l \left( \Omega_l^{-1} - \left( \Omega_l + H_{il}^T \Sigma_l H_{il}^T \right)^{-1} \right) = \frac{\mu_l}{P_T} \hat{\Sigma}_l. \tag{34}
\]

We can conclude that the Lagrangian dual of the max-min problem (24)-(26) is also a max-min problem:

\[
\max_{\Sigma_l \succeq 0} \min_{\Omega_l \succeq 0} \sum_l w_l \left( \log |\Omega_l + H_{il}^T \Sigma_l H_{il}| - \log |\Omega_l| \right) \tag{35}
\]

\[
\text{s.t. } \sum_l \text{Tr} \left( \frac{\Sigma_l}{P_l} \right) \leq 1, \tag{36}
\]

\[
\Omega_l = I + \sum_{k \in L \setminus \{l\}} H_{kl}^+ \Sigma_k H_{kl}, \tag{37}
\]

which is the weighted sum-rate maximization problem defined on a network of reciprocal channels with channel matrix \(H^+\). 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 \(\hat{\hat{\mu}}_l\) and \(\hat{\hat{\Lambda}}_l\) for the problem (31)-(33), we have the following correspondence:

\[
(\Sigma_l; \Lambda_l, \mu_l) = \left( \frac{P_T}{\hat{\hat{\mu}}_l} \hat{\hat{\Lambda}}_l; \frac{\hat{\hat{\mu}}_l}{P_T} \Sigma_l, \mu_l \right), \tag{38}
\]

\[
(\hat{\Sigma}_l; \hat{\hat{\Lambda}}_l, \hat{\hat{\mu}}_l) = \left( \frac{P_T}{\hat{\hat{\mu}}_l} \hat{\hat{\Lambda}}_l; \frac{\hat{\hat{\mu}}_l}{P_T} \Sigma_l, \mu_l \right). \tag{39}
\]

This provides a simple proof of the weighted sum-rate duality identified in, e.g., \([7]\).

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. Our algorithm applies/extends 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 provides a general class of algorithms for such optimization problems.

A. The iterative minimax algorithm

Note that the optimum of the max-min problem (7)-(9) is a saddle point, and the first order condition \((10)-(11)\) or part of it will give a saddle point, maximum or minimum of Lagrangian \(L\) 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 \(\Sigma_l^0, l \in L\) that is feasible, i.e.,

\[
\sum_{l \in L_s} \text{Tr} \left( \Sigma_l^0 Q_s^0 \right) \leq 1, \quad s \in S,
\]

and \(\Omega_l^0 = I + \sum_{k \in L \setminus \{l\}} H_{kl}^+ \Sigma_k^0 H_{kl}, l \in L\). By equation \((11)\) that gives the condition for minimizing \(L\) over \(\Omega_l\), we choose \(\Lambda_l^0 \geq 0\) such that

\[
\Lambda_l^0 = w_l \left( \left( \Omega_l^0 \right)^{-1} - \left( \Omega_l^0 + H_{il}^T \Sigma_l^0 H_{il}^T \right)^{-1} \right). \tag{40}
\]

Therefore, for any \(\Omega \geq 0\), we have

\[
F(\Sigma^0, \Omega^0) \leq L(\Sigma^0, \Omega^0, \Lambda^0, \mu^0) \leq L(\Sigma, \Omega, \Lambda, \mu), \tag{41}
\]
where $\mu^n \geq 0$ will be determined later. Define
\begin{equation}
\Phi_l^n = \sum_{s \in S} \mu^n_s Q_l^s + \sum_{k \in L \setminus \{l\}} H_{kl}^{+} \Lambda^n_{k} H_{kl}. \tag{38}
\end{equation}
Note that $\Phi_l^n$ does not necessary satisfy equation (10).

2) Given the above $(\Lambda_l^n, \Phi_l^n)$ and $\mu^n_s$, by equations (12)-(13), we choose $(\Sigma_l^{n+1}, \Omega_l^{n+1})$ such that
\begin{align}
\Sigma_l^{n+1} &= w_l \left( \Phi_l^n - (\Phi_l^n + H_l \Lambda_l^n H_l^+)^{-1} \right), \tag{39}
\Omega_l^{n+1} &= w_l H_l \left( \Phi_l^n + H_l \Lambda_l^n H_l^+ \right)^{-1} H_l^T. \tag{40}
\end{align}
Plug $\Omega = \Omega_l^{n+1}$ into inequality (37), we have
\begin{equation}
\mathcal{F}(\Sigma^n, \Omega^n) \leq L(\Sigma^n, \Omega_l^{n+1}, \Lambda^n, \mu^n). \tag{41}
\end{equation}
By the first order condition (10)-(11), $(\Sigma_l^{n+1}, \Omega_l^{n+1})$ is the saddle point of $\mathcal{L}(\Sigma, \Omega, \Lambda^n, \mu^n)$. Thus,
\begin{align*}
\mathcal{L}(\Sigma^n, \Omega_l^{n+1}, \Lambda^n, \mu^n) &\leq \mathcal{L}(\Sigma_l^{n+1}, \Omega_l^{n+1}, \Lambda^n, \mu^n) \\
&\leq L(\Sigma_l^{n+1}, \Omega_l^{n+1}, \Lambda^n, \mu^n) \tag{42}
\end{align*}
for any $\Omega \succeq 0$.

3) The matrix $\Sigma_l^{n+1}$ is a function of $\mu^n_s$, $s \in S$, denoted explicitly by $\Sigma_l^{n+1}(\mu^n_s; s \in S)$. Define the set $T$ such that
\begin{equation}
T = \{ s \in S | \sum_{l \in L^s} \text{Tr} \left( Q_l^n \Sigma_l^{n+1}(\mu^n_s = 0^+; s \in S) \right) \geq 1 \}. \tag{43}
\end{equation}
For each $s \in S \setminus T$, we set $\mu^n_s = 0$. For those $s \in T$, we choose $\mu^n_s$ such that
\begin{equation}
\sum_{l \in L^s} \text{Tr} \left( Q_l^n \Sigma_l^{n+1}(\mu^n_s; s \in S) \right) = 1, \ s \in T. \tag{44}
\end{equation}
Note that $\text{Tr} \left( \Sigma_l^{n+1}(\mu^n_s; s \in S) \right)$ is decreasing in $\mu^n_s$, and there are $|T|$ equations for $|T|$ variables. So, there exists a solution to equation (43). With the afore choice of $\mu^n_s$, $s \in S$, we can see that
\begin{equation}
\mu^n_s \left( 1 - \sum_{l \in L^s} \text{Tr} \left( \Sigma_l^{n+1} Q_l^n \right) \right) = 0. \tag{44}
\end{equation}
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
\begin{equation}
\lambda = \max_{s \in S} \sum_{l \in L^s} \text{Tr} \left( \Sigma_l^{n+1} Q_l^n \right). 
\end{equation}
We see that $0 < \lambda \leq 1$. We then choose $(\Sigma_l^{n+1}, \Omega_l^{n+1})$ such that
\begin{align}
\Sigma_l^{n+1} &= \frac{\Sigma_l^{n+1}}{\lambda}, \tag{45}
\Omega_l^{n+1} &= I + \sum_{k \in L \setminus \{l\}} H_{lk} \Sigma_k^{n+1} H_{lk}^T. \tag{46}
\end{align}
Plug $\Sigma_l^{n+1} = \lambda \Sigma_l^{n+1}$ and $\Omega = \lambda \Omega_l^{n+1}$ into the inequality (42) and combine with the inequality (41), we have
\begin{align*}
\mathcal{F}(\Sigma^n, \Omega^n) &\leq L(\Sigma^n, \lambda \Omega_l^{n+1}, \lambda \Lambda^n, \mu^n) \\
&= \mathcal{F}(\Sigma^n, \Omega_l^{n+1}) + \sum_{l \in L} (\lambda - 1) \text{Tr} (\Lambda_l^n) \\
&+ \sum_{s \in S} \mu^n_s \left( 1 - \sum_{l \in L^s} \text{Tr} \left( \Sigma_l^{n+1} Q_l^n \right) \right) \\
&= \mathcal{F}(\Sigma^n, \Omega_l^{n+1}) + \sum_{l \in L} (\lambda - 1) \text{Tr} (\Lambda_l^n) \\
&\leq \mathcal{F}(\Sigma^n, \Omega_l^{n+1}),
\end{align*}
where the second equality follows from equation (44) and the last inequality follows from the fact that $\lambda \leq 1$.

5) Repeat 1)-4), we obtain a monotone increasing sequence $\{ \mathcal{F}(\Sigma^n, \Omega^n) \}$, based on which we can conclude that $(\Sigma^n, \Omega^n)$ 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.

| TABLE I | THE ITERATIVE MINIMAX ALGORITHM |
|---|---|
| 1) | Initialize $\Sigma_0$, $l \in L$, such that $\sum_{l \in L^s} \text{Tr} \left( \Sigma_0 Q_l^s \right) \leq 1$, $s \in S$ |
| 2) | $\Omega_l \leftarrow 1 + \sum_{k \in L \setminus \{l\}} H_{lk} \Sigma_l H_{lk}^+$, $l \in L$ |
| 3) | $A_l \leftarrow w_l \left( (\Omega_l)^{-1} - (\Omega_l + H_l \Sigma_l H_l^+)^{-1} \right)$, $l \in L$ |
| 4) | $\Phi_l \leftarrow w_l \left( (\Phi_l)^{-1} + (\Phi_l + H_l \Lambda_l H_l^+)^{-1} \right)$, $l \in L$ |
| 5) | $\Sigma_l \leftarrow w_l \left( (\Phi_l)^{-1} - (\Phi_l + H_l \Lambda_l H_l^+)^{-1} \right)$, $l \in L$ |
| 6) | $T \leftarrow \{ s \in S | \sum_{l \in L^s} \text{Tr} \left( Q_l^n \Sigma_l(\mu^n_s = 0^+; s \in S) \right) \geq 1 \}$ |
| 7) | $\mu^n_s \leftarrow 0$ if $s \in S \setminus T$ |
| 8) | For $s \in T$, choose $\mu^n_s$ such that $\sum_{l \in L^s} \text{Tr} \left( Q_l^n \Sigma_l(\mu^n_s; s \in S) \right) = 1, s \in T$ |
| 9) | $\lambda \leftarrow \max_{s \in S} \sum_{l \in L^s} \text{Tr} \left( Q_l^n \Sigma_l(\mu^n_s; s \in S) \right)$ |
| 10) | $\Sigma_l \leftarrow \frac{\Sigma_l}{\lambda}$, $l \in L$ |
| 11) | Go to 2 |

### B. The convergence analysis

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

**Lemma 2:** Under the iterative minimax algorithm, the sequence $\{ C^n = \mathcal{F}(\Sigma^n, \Omega^n) \}$ converges to a limit point $C^*$. 

**Proof:** Since $\mathcal{F}(\Sigma, \Omega)$ is a continuous function and its domain (specified by the constraints (8)-(9)) is a compact set, $C^n$ is bounded above. By inequality (17), the sequence $\{ R^n \}$ is a monotone increasing sequence. Therefore, there exists a limit point $C^*$ such that $\lim_{n \to \infty} C^n = C^*$. \[\square\]

**Theorem 1:** The iterative minimax algorithm converges to a saddle point $(\Sigma^*, \Omega^*)$ of the max-min problem (7)-(9); and $\Sigma^*$ is an optimum of the weighted sum-rate maximization (4)-(5). 

**Proof:** With Lemma 2 to show the convergence of the iterative minimax algorithm, it is enough to show
that if \( \mathcal{F}(\Sigma^n, \Omega^n) = \mathcal{F}(\Sigma^{n+1}, \Omega^{n+1}) \), then \( (\Sigma^n, \Omega^n) = (\Sigma^{n+1}, \Omega^{n+1}) \).

From the derivation of inequality (47), if \( \mathcal{F}(\Sigma^n, \Omega^n) = \mathcal{F}(\Sigma^{n+1}, \Omega^{n+1}) \), then

\[
\mathcal{F}(\Sigma^n, \Omega^n) = \mathcal{L}(\Sigma^n, \Omega^n, \Lambda^n, \mu^n) = \mathcal{L}(\Sigma^{n+1}, \Omega^{n+1}, \Lambda^n, \mu^n) = \mathcal{L}(\Sigma^{n+1}, \Omega^{n+1}, \Lambda^n, \mu^n) = \mathcal{F}(\Sigma^{n+1}, \Omega^{n+1}).
\]

It follows that both \( (\Sigma^n, \Omega^n, \Lambda^n, \mu^n) \) and \( (\Sigma^{n+1}, \Omega^{n+1}, \Lambda^n, \mu^n) \) satisfy the KKT (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 maximin problem. Furthermore, for any given dual variables, the Lagrangian \( \mathcal{L} \) is strictly concave in \( \Sigma \). So, \( \Sigma^n = \Sigma^{n+1} \), and \( \Omega^n = \Omega^{n+1} \) 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 design and 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.

Remark 3: 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.

C. Case studies

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

1) The network with the total power constraint: As mentioned in Section III-B,2, here \(|S| = 1 \) and \( Q_l = \frac{1}{P_T} \), with \( P_T \) the total power budget. The matrix \( \Sigma_l \) defined in Section IV-A is a function of \( \mu \), the dual variable associated with the total power constraint. The iterative minimax algorithm reduces to that described in Table III.

The above algorithm is different from the algorithm proposed in the previous work [6]. The algorithm in [6] uses the fact that the total power constraint is tight at an optimum, and normalizes \( \mu \) such that \( \sum_{l \in L} \text{Tr} \left( \frac{\Sigma_l}{\mu} \right) = 1 \), i.e., the algorithm enforces the tightness of the total power constraint at the initial point and each iteration. In contrast, our algorithm enforces the complementary slackness condition at each iteration and can start with any feasible \( \Sigma \).

2) The network with the per-link power constraints: Here the set \( S = L \) and \( Q_l = \frac{1}{P_l} \), with \( P_l \) the power budget at each link \( l \in L \). The matrix \( \Sigma_l \) defined in Section IV-A is a function of \( \mu_l \), the dual variable associated with the power constraint at link \( l \). The iterative minimax algorithm reduces to that described in Table III.

V. Numerical Examples

In this section, we provide numerical examples to complement the analysis in the previous sections. Consider a network with \( L = 10 \) links, corresponding to 10 transmitter-receiver pairs that interfere with each other. Each link is equipped with 3 (4) antennas at its transmitter (receiver). 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 \( H_{ij}, i \neq j \) with a factor of 0.1, 1, and 5 respectively. The weights \( w_l \)'s are uniformly drawn from \([0, 1]\), the case with total power constraint \( P_T = 10 \), and for the case with the per-link power constraints \( P_l \)'s are uniformly drawn from \([1, 2, \cdots, 10]\).

For the computation, we use SDPT3 [13] combined with the problem parser YALMIP [9]. The algorithm implementation is straightforward except for finding \( \mu \), for which we use a bisection search method.
a) The network with the total power constraint: Figures 1, 2, and 3 show the monotonic convergence of our algorithm in a network with the total power constraint. 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 slow. However, in the network with low and moderate interference, the algorithm shows very fast convergence. Also note that the stronger the interference, the smaller the weighted sum-rate is.

b) The network with the per-link power constraints: Figures 4, 5, and 6 show the monotonic convergence of our algorithm in a network with the per-link power constraints. Again, we see that the stronger the interference, the slower the algorithm converges; but in the network with low and moderate interference, the algorithm shows fast convergence.

c) Complexity Analysis: We have evaluated in the above the monotonic convergence of our 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 $\Sigma_i$ 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, $\Omega_i$ incurs a complexity of $O(LN^3)$, and so does $\Omega_i + H_i\Sigma_i^{(n+1)}H_i^+$. Furthermore, $\Phi_i$ incurs a complexity of $O(LN^3)$, and so do $\Sigma_i$ and $\Sigma_i$. 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 [16] that has a complexity of $O(N^{2.3727})$, we can reduce computational
lish the sum-rate duality between an interference network and an equivalent max-min problem. The Lagrangian duality of the minimization in the MIMO interference network, by formulating variables and the constraints are convex, and thus provides an analysis and matrix analysis. They apply and extend to any equivalent max-min problem provides an elegant way to establish developing efficient algorithms for general MIMO interference networks.

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

VI. CONCLUSION

We take a new perspective on the weighted sum-rate maximization in the MIMO interference network, 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 when such a duality exists, and more importantly, suggests a novel iterative minimax algorithm for the weighted sum-rate maximization. The design and 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 provides 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 algorithms for general MIMO interference networks.

APPENDIX: PROOF OF LEMMA 1

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

\[
F(A, B) = \log |A + B| - \log |B|
\]

is not well-defined if \( B \) is not positive definite. If there exists a nonsingular square matrix \( T \) such that

\[
T^+ AT = \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad T^+ BT = \begin{bmatrix} B_1 & 0 \\ 0 & 0 \end{bmatrix}
\]

where \( A_1 \in \mathcal{S}_+^{m}, B_1 \in \mathcal{S}_+^{n-m} \) for some \( m \leq n \), then we can define an extended difference of logdet function:

\[
F(A, B) := \log |A_1 + B_1| - \log |B_1|.
\]

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 Lemma 1. For simplicity of presentation and without loss of generality, we reload notations and consider the following problem:

\[
\max_{\Sigma \succeq 0} \min_{\Omega \succeq 0} \log |\Omega + H\Sigma H^+| - \log |\Omega| + \text{Tr}(\Lambda \Omega) - \text{Tr}(\Phi \Sigma)
\]

(48)

where \( \Lambda \succeq 0 \) and \( \Phi \succeq 0 \). The key idea of the proof is to show that problem (48) is equivalent to a problem with \( \Omega \) restricted to \( \Omega = HXH^+, \; X \succeq 0 \).

**Lemma 3:** The problem (48) is equivalent to the following problem:

\[
\max_{\Sigma \succeq 0} \min_{\Omega \succeq 0} \log |\Omega + H\Sigma H^+| - \log |\Omega| + \text{Tr}(\Lambda \Omega) - \text{Tr}(\Phi \Sigma)
\]

\[
\text{s.t.} \quad \Omega = HXH^+, \quad X \succeq 0.
\]

(50)

**Proof:** Since \( H\Sigma H^+ \succeq 0 \) and \( \Lambda \succeq 0 \), there exists a nonsingular square matrix \( T \) such that

\[
T^+ HT = \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix},
\]

\[
(T^+)^{-1} H\Sigma H^+ T^{-1} = \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix},
\]

where \( S_1, S_2, S_3 \) are diagonal and positive definite; see, e.g., Theorem 3.22 in [23]. Let \( \Omega = T^+\Omega T \), problem (48) becomes

\[
\max_{\Sigma \succeq 0} \min_{\Omega \succeq 0} \log |\Omega + (T^+)^{-1} H\Sigma H^+ T^{-1}| - \log |\Omega| + \text{Tr}(T^+ \Omega T) - \text{Tr}(\Phi \Sigma).
\]

Now, consider those terms in the objective function that depend on \( \Omega \):

\[
\hat{L}(\Omega)
\]

\[
= \log |\Omega + (T^+)^{-1} H\Sigma H^+ T^{-1}| - \log |\Omega| + \text{Tr}(T^+ \Omega T)
\]

\[
= \log \begin{bmatrix} \Omega_{11} + S_{11} & \Omega_{12} + S_{12} & \Omega_{13} + S_{13} & \Omega_{14} + S_{14} \\ \Omega_{12} + S_{12} & \Omega_{22} + S_{22} & \Omega_{23} + S_{23} & \Omega_{24} + S_{24} \\ \Omega_{13} + S_{13} & \Omega_{23} + S_{23} & \Omega_{33} + S_{33} & \Omega_{34} + S_{34} \\ \Omega_{14} + S_{14} & \Omega_{24} + S_{24} & \Omega_{34} + S_{34} & \Omega_{44} + S_{44} \end{bmatrix}
\]

\[
- \log \begin{bmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} & \hat{\Omega}_{13} & \hat{\Omega}_{14} \\ \hat{\Omega}_{12} & \hat{\Omega}_{22} & \hat{\Omega}_{23} & \hat{\Omega}_{24} \\ \hat{\Omega}_{13} & \hat{\Omega}_{23} & \hat{\Omega}_{33} & \hat{\Omega}_{34} \\ \hat{\Omega}_{14} & \hat{\Omega}_{24} & \hat{\Omega}_{34} & \hat{\Omega}_{44} \end{bmatrix}
\]

\[
+ \text{Tr} \left( S_{11} \hat{\Omega}_{11} + \text{Tr} \left( S_{22} \hat{\Omega}_{22} \right) \right)
\]

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

\[
\tilde{L}(\tilde{\Omega}) \geq \log \left[ \begin{bmatrix} \tilde{\Omega}_{11} + S_1 & \tilde{\Omega}_{12} & \tilde{\Omega}_{13} \\ \tilde{\Omega}_{12}^+ & \tilde{\Omega}_{22} + S_2 & \tilde{\Omega}_{23} \\ \tilde{\Omega}_{13}^+ & \tilde{\Omega}_{23} & \tilde{\Omega}_{33} \end{bmatrix} \right] - \log \left[ \begin{bmatrix} \tilde{\Omega}_{11}^+ & \tilde{\Omega}_{12}^+ & \tilde{\Omega}_{13}^+ \\ \tilde{\Omega}_{12} & \tilde{\Omega}_{22} & \tilde{\Omega}_{23} \\ \tilde{\Omega}_{13} & \tilde{\Omega}_{23} & \tilde{\Omega}_{33} \end{bmatrix} \right] + \text{Tr} \left( S_1 \tilde{\Omega}_{11} \right) + \text{Tr} \left( S_3 \tilde{\Omega}_{33} \right),
\]

where the equality holds when \(\tilde{\Omega}_{44} \to \infty\) but can be achieved when \(\tilde{\Omega}_{44} = 0\) for all \(i = 1, 2, 3, 4\). We will restrict \(\tilde{\Omega}\) to those with \(\tilde{\Omega}_{44} = 0\) for all \(i = 1, 2, 3, 4\), as the equality is achieved at one of those matrices.

Since \(S_3 \geq 0\) and \(\tilde{\Omega}_{33} \geq 0\), \(\text{Tr} \left( S_3 \tilde{\Omega}_{33} \right) \geq 0\). We further have

\[
\tilde{L}(\tilde{\Omega}) \geq \log \left[ \begin{bmatrix} \tilde{\Omega}_{11} + S_1 & \tilde{\Omega}_{12} \\ \tilde{\Omega}_{22} + S_2 & \tilde{\Omega}_{23} \end{bmatrix} \right] - \log \left[ \begin{bmatrix} \tilde{\Omega}_{11}^+ & \tilde{\Omega}_{12}^+ \\ \tilde{\Omega}_{22}^+ & \tilde{\Omega}_{23}^+ \end{bmatrix} \right] + \text{Tr} \left( S_1 \tilde{\Omega}_{11} \right),
\]

where the equality is achieved when additionally \(\tilde{\Omega}_{33} = 0\) for all \(i = 1, 2, 3, 4\). Therefore, we conclude that there exists a minimizer \(\tilde{\Omega}^*\) with the form:

\[
\tilde{\Omega}^* = \begin{bmatrix} \tilde{\Omega}_{11} & \tilde{\Omega}_{12} \\ \tilde{\Omega}_{22} & 0 \end{bmatrix},
\]

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 \(H^+ v = 0\), we have

\[
v^+ H \Sigma H^T v = v^+ T^+ \begin{bmatrix} S_1 & S_2 \\ 0 & 0 \end{bmatrix} T v = 0,
\]

which implies

\[
T v = \begin{bmatrix} 0 & 0 & v_3 & v_4 \end{bmatrix}^T.
\]

Therefore,

\[
v^+ \Omega^* v = v^+ T^+ \begin{bmatrix} \tilde{\Omega}_{11} & \tilde{\Omega}_{12} \\ \tilde{\Omega}_{22} & 0 \end{bmatrix} T v = v^+ T^+ \begin{bmatrix} \tilde{\Omega}_{11} & \tilde{\Omega}_{12} \\ \tilde{\Omega}_{22} & 0 \end{bmatrix} T v = 0.
\]

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

\[
\Omega^* = HXH^+.
\]

We conclude that there exists an optimal solution with \(\Omega^* = HXH^+\), and thus problem (48) and problem (49)-(50) are equivalent.

With Lemma 3 we are ready to present the explicit saddle point solution. Consider the logdet terms in the objective function:

\[
\log |\Omega + H \Sigma H^+| - \log |\Omega| = \log |HXH^+ + H \Sigma H^+| - \log |HXH^+| = \log |H^+ H(X + \Sigma)| - \log |H^+ HX| = \log |X + \Sigma| - \log |X|.
\]

The singularity issue comes out when \(H^+ H\) is not invertible, but this can be handled by adding a small term \(\kappa I\), \(\kappa > 0\) to \(H^+ H\) and then taking the limit \(\kappa \to 0\). Thus, we can transform problem (48) into the following simple one:

\[
\max_{\Sigma \succeq 0} \min_{X \succeq 0} \log |X + \Sigma| - \log |X| + \text{Tr} \left( H^+ \Lambda HX \right) - \text{Tr} \left( \Phi \Sigma \right).
\]

(51)

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

\[
(X + \Sigma)^{-1} - \Phi = 0, \quad (X + \Sigma)^{-1} - X^{-1} + H^+ \Lambda H = 0,
\]

from which we obtain the following explicit saddle point solution:

\[
\Sigma = (\Phi + H^+ \Lambda H)^{-1}, \quad X = (\Phi + H^+ \Lambda H)^{-1},
\]

(52)

and in terms of \(\Omega\) we have

\[
\Omega = HXH^+ = H(\Phi + H^+ \Lambda H)^{-1}H^+.
\]

(53)

Note that problem (51) is well-defined only when \(\Sigma, X\) satisfy the property specified for matrices \(A, B\) in the beginning of this Appendix. This is verified as follows.

**Proposition 1:** The objective function in problem (51) is well-defined for matrices \(\Sigma, X\) that are given by (52)-(53).

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

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

\[
U^+ \Psi U = \begin{bmatrix} \Psi_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad U^+ \Phi U = \begin{bmatrix} \Phi_1 & 0 \\ 0 & 0 \end{bmatrix}
\]

where \(\Psi_1 > 0\) and \(\Phi_1 \succeq 0\). By equations (52)-(53),

\[
X = U \begin{bmatrix} \Psi_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^+, \quad \Sigma = U \begin{bmatrix} \Phi_1^{-1} - \Psi_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^+.
\]
Note that $\Psi^{-1}_{\text{cov}} \succ 0$, so by the definition of the extended difference of logdet function, the objective function in problem (51) is well-defined. \hfill □

With (52)-(53), we can easily recover the explicit solution (12)-(13). This concludes the proof of Lemma 1.

REFERENCES

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

[2] H. Huh, H. Papadopoulo, and G. Caire. MIMO Broadcast Channel Optimization under General Linear Constraints. In Proc. IEEE Int. Symp. on Info. Theory (ISIT), 2009.

[3] 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, Aug. 2010.

[4] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith. Sum power iterative water-filling for multi-antenna Gaussian broadcast channels. IEEE Trans. Info. Theory, 51(4):1570–1580, Apr. 2005.

[5] S.-J. Kim and G.B. Giannakis. Optimal resource allocation for MIMO ad hoc cognitive radio networks. IEEE Trans. Info. Theory, 57(5):3117 – 3131, may 2011.

[6] X. Li, S. You, A. Liu, and C. Zhao. A new algorithm for the weighted sum rate maximization in mimo interference networks, submitted. IEEE InfoCom, 2014.

[7] 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.

[8] 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. Signal Processing, IEEE Transactions on, 60(2):834–847, 2012.

[9] J. Lofberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In Computer Aided Control Systems Design, 2004 IEEE International Symposium on, pages 284–289. IEEE, 2004.

[10] C. TK Ng and H. Huang. Linear precoding in cooperative MIMO cellular networks with limited coordination clusters. Selected Areas in Communications, IEEE Journal on, 28(9):1446–1454, 2010.

[11] C. Shi, R. A. Berry, and M. L. Honig. Monotonic convergence of distributed interference pricing in wireless networks. in Proc. IEEE ISIT, Seoul, Korea, June 2009.

[12] 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. Signal Processing, IEEE Transactions on, 59(9):4331–4340, 2011.

[13] K.-C. Toh, M. J Todd, and R. H. Tütüncü. SDPT3a MATLAB software package for semidefinite programming, version 1.3. Optimization Methods and Software, 11(1-4):545–581, 1999.

[14] P. Viswanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of guassian mimo broadcast channels. Transactions on Information Theory, 49(10):2658–2668, 2003.

[15] P. Viswanath and D. N. C. Tse. Sum capacity of the multiple antenna gaussian broadcast channel and uplink-downlink duality. Transactions on Information Theory, 49(8):1912–1921, 2003.

[16] V. V. Williams. Multiplying matrices faster than coppersmith-winograd. In Proceedings of the 44th symposium on Theory of Computing, pages 887–898. ACM, 2012.

[17] W. Yu. Sum-capacity computation for the gaussian vector broadcast channel via dual decomposition. IEEE Trans. Inform. Theory, 52(2):754–759, Feb. 2006.

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

[19] W. Yu. Multiuser water-filling in the presence of crosstalk. Information Theory and Applications Workshop, San Diego, CA, U.S.A, pages 414 –420, 29 2007-feb. 2 2007.

[20] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi. Iterative water-filling for Gaussian vector multiple-access channels. IEEE Trans. Info. Theory, 50(1):145–152, 2004.

[21] L. Zhang, Y.-C. Liang, Y. Xin, R. Zhang, and H.V. Poor. On gaussian MIMO BC-MAC duality with multiple transmit covariance constraints. pages 2502 –2506, June 2009.

[22] L. Zhang, Y. Xin, and Y.-C. Liang. Weighted sum rate optimization for cognitive radio MIMO broadcast channels. IEEE Trans. Wireless Commun., 8(6):2950–2959, 2009.

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