A New Algorithm for the Weighted Sum Rate Maximization in MIMO Interference Networks

Xing Li[1], Seungil You[2], Lijun Chen[1], An Liu[3], Youjian (Eugene) Liu[1]

***Abstract*—We propose a new algorithm to solve the non-convex weighted sum-rate maximization problem in general MIMO interference networks. With the Gaussian input assumption, the previous state-of-the-art algorithms are the WMMSE algorithm and the polite water-filling (PWF) algorithm. The WMMSE algorithm is provably convergent, while the PWF algorithm converges faster in most situations but sometimes oscillates. Thus, it is highly desirable to design an algorithm that takes advantage of the optimal transmit signal structure to ensure fast convergence while being provably convergent. We present such an algorithm and prove its monotonic convergence. Moreover, our convergence proof uses very general convex analysis as well as a scaling invariance property of the weighted sum-rate maximization problem. We expect that the scaling invariance holds for and our proof technique applies to many non-convex problems in communication networks.**

***Index Terms*—MIMO, Interference Network, Weighted Sum-rate Maximization, Duality, Scaling Invariance, Optimization**

I. Introduction

One of the most effective approaches to accommodating the explosive growth in mobile data is to reduce the cell size and increase the base station/access point density. However, the path loss versus distance curve is flatter at shorter distance, as opposed to being steep at relatively long distance. As a result, the inter-cell interference becomes significant as the cell size/coverage area shrinks. Therefore, joint transmit signal design for interference networks is a key technology for the next generation wireless communication systems.

In this paper, we consider the joint transmit signal design for a general interference network called MIMO B-MAC network [5]. The MIMO B-MAC network consists of multiple mutually interfering data links between multiple transmitters and multiple receivers that are equipped with multiple antennas (MIMO). It includes broadcast channel (BC), multiaccess channel (MAC), interference channels, X networks, and many practical wireless networks as special cases.

Specifically, we study the problem of jointly optimizing the transmit signals of all transmitters in order to maximize the weighted sum-rate of the data links for the MIMO B-MAC network, assuming Gaussian transmit signal and availability of all channel state information. It typifies a class of problems that are key to the next generation wireless communication networks where the interference is the limiting factor. This problem is non-convex, and various algorithms have been proposed for this problem or its special cases over the years; see, e.g., [1], [2], [3], [4], [5], [8], [9], [12], [13], [14], [15], [16].

In particular, we have recently proposed the polite water-filling (PWF) algorithm for the MIMO B-MAC network [5]. It is based on the identification of a polite water-filling structure that the optimal transmit signal possesses. It is an iterative algorithm with the forward link polite water-filling followed by the virtual reverse link polite water-filling. Because it takes advantage of the optimal signal structure, the PWF algorithm has nearly the lowest complexity and the fastest convergence when it converges. For instance, it converges to the optimal solution in half iteration for parallel channels. However, the convergence of the PWF algorithm is only guaranteed for the special case of interference tree networks [5] and it may oscillate under certain strong interference conditions. Another state-of-the-art algorithm is the WMMSE algorithm in [9]. It is proposed for the MIMO interfering broadcast channels but could be readily applied to the more general B-MAC networks. It transforms the weighted sum-rate maximization into an equivalent weighted sum-MSE cost minimization problem, which has three sets of variables and is convex when any two variable sets are fixed. With the block coordinate descent technique, the WMMSE algorithm is guaranteed to converge to a stationary point, though the convergence is observed in simulations to be slower than the PWF algorithm.

Thus, it is highly desirable to have an algorithm with the advantages of both PWF and WMMSE algorithms, i.e., fast convergence by taking advantage of the optimal transmit signal structure and provable convergence for the general MIMO B-MAC network. One major contribution of this paper is to propose such an algorithm. A key finding in proving the polite water-filling structure is to identify that a term involving interference in the KKT condition of the weighted sum-rate maximization problem is equal to the reverse link signal covariance of the dual network at a stationary point [5, Theorem 22]. Then, solving the KKT condition reveals that the optimal transmit signal has the polite water-filling structure, which reduces to the well known water-filling structure for the MIMO parallel channels. Instead of using this key finding to solve the KKT condition, our new algorithm uses it directly to alternatively update the forward link and reverse link signal covariance matrices. Numerical experiments demonstrate that the new algorithm could be a few iterations or more than ten

[1]College of Engineering and Applied Science, University of Colorado at Boulder
[2]Department of Control and Dynamical Systems, California Institute of Technology
[3]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology

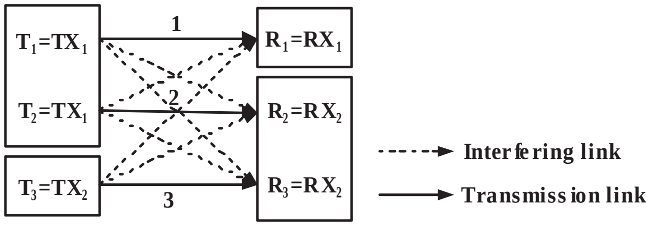

Fig. 1. An example of B-MAC network

iterations faster than the WMMSE algorithm, depending on the desired numerical accuracy. Note that being faster even by a few iterations will be significant in distributed implementation where the overhead of each iteration costs significant signaling resources between the transmitters and the receivers. Indeed, the new algorithm is highly scalable and suitable for distributed implementation with distributed channel estimation because for each data link, only its own channel state and the aggregated interference plus noise covariance needs to be estimated no matter how many interferers are there. The distributed implementation of the new algorithm will be developed in the future work.

Besides proposing the new algorithm with fast convergence, another major contribution of this paper is to present an elegant proof of the monotonic convergence of the algorithm. The proof uses only very general convex analysis, as well as a particular scaling invariance property that we identify for the weighted sum-rate maximization problem. We expect that the scaling invariance holds for and our proof technique applies to many non-convex problems in communication networks that involve the rate or throughput maximization.

The rest of this paper is organized as follows. Section II presents the system model and formulates the problem. Section III briefly reviews the related results on the polite water-filling and presents the proposed new algorithm. Its monotonic convergence is established in Section IV. Section V presents numerical examples and complexity analysis. Section VI concludes.

II. System Model and Problem Formulation

We consider a general interference network named MIMO B-MAC network [5], [6]. A transmitter in the MIMO B-MAC network may send independent data to different receivers, like in BC, and a receiver may receive independent data from different transmitters, like in MAC. Assume there are totally L mutually interfering data links. The set of physical transmitter labels is $\mathcal{T} = \{\mathrm{TX}_1, \mathrm{TX}_2, \mathrm{TX}_3, ...\}$ and the set of physical receiver labels is $\mathcal{R} = \{\mathrm{RX}_1, \mathrm{RX}_2, \mathrm{RX}_3, ...\}$. Define transmitter T_l of link l as a mapping from l to link l's physical transmitter label in $\mathcal{T}$. Define receiver R_l as a mapping from l to link l's physical receiver label in $\mathcal{R}$. The numbers of antennas at T_l and R_l are L_{T_l} and L_{R_l} respectively. Fig. 1 shows a B-MAC network with three data links. When multiple data links have the same receiver or the same transmitter, techniques such as successive decoding and cancellation or dirty paper coding can be applied [5]. The received signal at R_l is

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

where $\mathbf{x}_k \in \mathbb{C}^{L_{T_k}\times 1}$ is the transmit signal of link k and is assumed to be circularly symmetric complex Gaussian; $\mathbf{H}_{l,k} \in \mathbb{C}^{L_{R_l}\times L_{T_k}}$ is the channel matrix between T_k and R_l; and $\mathbf{w}_l \in \mathbb{C}^{L_{R_l}\times 1}$ is a circularly symmetric complex Gaussian noise vector with identity covariance matrix. The circularly symmetric assumption of the transmit signal can be dropped easily by applying the proposed algorithm to real Gaussian signals with twice the dimension.

Assuming the availability of channel state information at the transmitters, an achievable rate of link l is

$$\mathcal{I}_l\left(\mathbf{\Sigma}_{1:L}\right) = \log\left|\mathbf{I} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l\mathbf{H}_{l,l}^{\dagger}\mathbf{\Omega}_l^{-1}\right| \tag{2}$$

where $\mathbf{\Sigma}_l$ is the covariance matrix of $\mathbf{x}_l$; the interferences from other links are treated as noise; and $\mathbf{\Omega}_l$ is the interference-plus-noise covariance matrix of the l^{th} link, which is

$$\mathbf{\Omega}_l = \mathbf{I} + \sum_{\substack{k=1\\k\neq l}}^{L} \mathbf{H}_{l,k}\mathbf{\Sigma}_k\mathbf{H}_{l,k}^{\dagger}. \tag{3}$$

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 $\mathbf{H}_{l,k} = \mathbf{0}$ in (3). It allows this paper to cover a wide range of communication techniques.

The optimization problem that we want to solve is the weighted sum-rate maximization under total power constraint:

$$\begin{aligned} \textbf{WSRM\_TP}: \quad & \max_{\mathbf{\Sigma}_{1:L}} \quad \sum_{l=1}^{L} w_l \mathcal{I}_l\left(\mathbf{\Sigma}_{1:L}\right) \\ & \text{s.t.} \quad \mathbf{\Sigma}_l \succeq 0, \ \forall l, \\ & \qquad \sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) \leq P_T \end{aligned} \tag{4}$$

where $w_l > 0$ is the weight for link l. The generalization to multiple linear constraints can be done similarly to [7] and will be considered in future works.

III. The Optimization Algorithm

In this section, we will review briefly some related results on the polite water-filling and propose a new algorithm for the weighted sum-rate problem (4) that has fast, monotonic convergence.

A. A Review on the Polite Water-filling Structure and Algorithm

Although the problem (4) is non-convex and cannot be solved directly, the optimal transmit signal has a polite water-filling structure, based on which an efficient algorithm can be designed [5]. The results in [5] are briefly reviewed here.

We first give the duality results. Let

$$\left(\left[\mathbf{H}_{l,k}\right],\ \sum_{l=1}^{L}\mathrm{Tr}\left(\mathbf{\Sigma}_l\right)\leq P_T\right) \tag{5}$$

denote a network with total power constraint and channel matrices $[\mathbf{H}_{l,k}]$ as in the problem (4). An achievable rate region of (5) is defined as

$$\begin{aligned}\mathcal{R}\left(P_T\right) \triangleq & \bigcup_{\mathbf{\Sigma}_{1:L}:\sum_{l=1}^{L}\mathrm{Tr}(\mathbf{\Sigma}_l)\leq P_T}\left\{\mathbf{r}\in\mathbb{R}_+^L:\right. \\ & \left. r_l\leq\mathcal{I}_l\left(\mathbf{\Sigma}_{1:L}\right),1\leq l\leq L\right\}.\end{aligned} \tag{6}$$

Its dual network or reverse links is defined as

$$\left(\left[\mathbf{H}_{k,l}^{\dagger}\right],\ \sum_{l=1}^{L}\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_l\right)\leq P_T\right) \tag{7}$$

where the roles of all transmitters and receivers are reversed, and the channel matrices are replaced with their conjugate transpose. $\hat{}$ denote the corresponding terms in the reverse links. Similarly, the interference-plus-noise covariance matrix of reverse link l is

$$\hat{\mathbf{\Omega}}_l = \mathbf{I}+\sum_{\substack{k=1\\k\neq l}}^{L}\mathbf{H}_{k,l}^{\dagger}\hat{\mathbf{\Sigma}}_k\mathbf{H}_{k,l}; \tag{8}$$

the achievable rate of reverse link l is

$$\hat{\mathcal{I}}_l\left(\hat{\mathbf{\Sigma}}_{1:L}\right)=\log\left|\mathbf{I}+\mathbf{H}_{l,l}^{\dagger}\hat{\mathbf{\Sigma}}_l\mathbf{H}_{l,l}\hat{\mathbf{\Omega}}_l^{-1}\right|; \tag{9}$$

the reverse link achievable rate region is defined as

$$\begin{aligned}\hat{\mathcal{R}}\left(P_T\right) \triangleq & \bigcup_{\hat{\mathbf{\Sigma}}_{1:L}:\sum_{l=1}^{L}\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_l\right)\leq P_T}\left\{\hat{\mathbf{r}}\in\mathbb{R}_+^L:\right. \\ & \left. \hat{r}_l\leq\hat{\mathcal{I}}_l\left(\hat{\mathbf{\Sigma}}_{1:L}\right),1\leq l\leq L\right\}.\end{aligned} \tag{10}$$

The rate duality states that the achievable rate regions of the forward and reverse links are the same [5, Theorem 9] . A *covariance transformation* [5, (18)] calculates the reverse link input covariance matrices from the forward ones. The rate duality is proved by showing that the reverse link input covariance matrices calculated from the covariance transformation achieves equal or higher rates than the forward link rates under the same value of linear constraint P_T [5, Lemma 11] .

The Lagrange function of problem (4) is

$$\begin{aligned}& L\left(\mu,\mathbf{\Theta}_{1:L},\mathbf{\Sigma}_{1:L}\right) \\ = & \sum_{l=1}^{L}w_l\log\left|\mathbf{I}+\mathbf{H}_{l,l}\mathbf{\Sigma}_l\mathbf{H}_{l,l}^{\dagger}\mathbf{\Omega}_l^{-1}\right|+\sum_{l=1}^{L}\mathrm{Tr}\left(\mathbf{\Sigma}_l\mathbf{\Theta}_l\right) \\ & +\mu\left(P_T-\sum_{l=1}^{L}\mathrm{Tr}\left(\mathbf{\Sigma}_l\right)\right),\end{aligned}$$

where $\mathbf{\Theta}_{1:L}$ and μ are Lagrange multipliers. The KKT conditions are

$$\begin{aligned}& \nabla_{\mathbf{\Sigma}_l}L \\ = & w_l\mathbf{H}_{l,l}^{\dagger}\left(\mathbf{\Omega}_l+\mathbf{H}_{l,l}\mathbf{\Sigma}_l\mathbf{H}_{l,l}^{\dagger}\right)^{-1}\mathbf{H}_{l,l}+\mathbf{\Theta}_l-\mu\mathbf{I} \\ & -\sum_{k\neq l}w_k\mathbf{H}_{k,l}^{\dagger}\left(\mathbf{\Omega}_k^{-1}-\left(\mathbf{\Omega}_k+\mathbf{H}_{k,k}\mathbf{\Sigma}_k\mathbf{H}_{k,k}^{\dagger}\right)^{-1}\right)\mathbf{H}_{k,l} \\ = & \mathbf{0},\end{aligned} \tag{11}$$

$$\mu\left(P_T-\sum_{l=1}^{L}\mathrm{Tr}\left(\mathbf{\Sigma}_l\right)\right)=0,$$

$$\mathrm{tr}\left(\mathbf{\Sigma}_l\mathbf{\Theta}_l\right)=0,$$

$$\mathbf{\Sigma}_l,\ \mathbf{\Theta}_l\succcurlyeq 0,\ \mu\geq 0.$$

The polite water-filling structure is given as follows [5, Theorem 22]. A key finding leading to the polite water-filling structure is that at a stationary point, the dual input covariance matrices $\hat{\mathbf{\Sigma}}_{1:L}$ calculated from the covariance transformation satisfies

$$\hat{\mathbf{\Sigma}}_l=\frac{w_l}{\mu}\left(\mathbf{\Omega}_l^{-1}-\left(\mathbf{\Omega}_l+\mathbf{H}_{l,l}\mathbf{\Sigma}_l\mathbf{H}_{l,l}^{\dagger}\right)^{-1}\right),\ l=1,\ldots,L. \tag{12}$$

We substitute it into (11) to obtain the *polite water-filling structure,*

$$\mathbf{Q}_l=\mathbf{G}_l\mathbf{D}_l\mathbf{G}_l^{\dagger}, \tag{13}$$

$$\mathbf{D}_l=\left(\nu_l\mathbf{I}-\mathbf{\Delta}_l^{-2}\right)^{+}, \tag{14}$$

where $\mathbf{Q}_l\triangleq\hat{\mathbf{\Omega}}_l^{\frac{1}{2}}\mathbf{\Sigma}_l\hat{\mathbf{\Omega}}_l^{\frac{1}{2}}$ is the equivalent input covariance matrix of the link l; $\mathbf{G}_l$ and $\mathbf{\Delta}_l$ are from SVD decomposition of the equivalent single-user channel $\bar{\mathbf{H}}_l=\mathbf{F}_l\mathbf{\Delta}_l\mathbf{G}_l$; $\bar{\mathbf{H}}_l$ is given by $\bar{\mathbf{H}}_l=\mathbf{\Omega}_l^{-\frac{1}{2}}\mathbf{H}_{l,l}\hat{\mathbf{\Omega}}_l^{-\frac{1}{2}}$; $\hat{\mathbf{\Sigma}}_{1:L}$ is obtained from $\mathbf{\Sigma}_{1:L}$ by the covariance transformation in [5, Definition 4] and is used to calculate the corresponding $\hat{\mathbf{\Omega}}_{1:L}$; $\nu_l=\frac{w_l}{\mu}\geq 0$ is the water-filling level. That is the link l's equivalent input covariance matrix $\mathbf{Q}_l$ is a water-filling over the equivalent channel $\bar{\mathbf{H}}_l$.

In addition, at a stationary point, the $\hat{\mathbf{\Sigma}}_{1:L}$ obtained from the covariance transformation also have the polite water-filling structure [5, Theorem 21] and satisfy the following KKT conditions (16) of the following dual problem (15):

$$\begin{aligned}\textbf{WSRM\_TP\_D}:\quad \max_{\hat{\mathbf{\Sigma}}_{1:L}} \quad & \sum_{l=1}^{L}w_l\hat{\mathcal{I}}_l\left(\hat{\mathbf{\Sigma}}_{1:L}\right) \\ \text{s.t.}\quad & \hat{\mathbf{\Sigma}}_l\succeq 0,\ \forall l, \\ & \sum_{l=1}^{L}\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_l\right)\leq P_T,\end{aligned} \tag{15}$$

whose KKT conditions are

$$\begin{aligned} & w_l \mathbf{H}_{l,l} \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{l,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{l,l} \right)^{-1} \mathbf{H}_{l,l}^{\dagger} + \hat{\mathbf{\Theta}}_l - \hat{\mu} \mathbf{I} \\ & - \sum_{k \neq l} w_k \mathbf{H}_{k,l} \left(\hat{\mathbf{\Omega}}_k^{-1} - \left(\hat{\mathbf{\Omega}}_k + \mathbf{H}_{k,k}^{\dagger} \hat{\mathbf{\Sigma}}_k \mathbf{H}_{k,k} \right)^{-1} \right) \mathbf{H}_{k,l}^{\dagger} \\ = & \ \mathbf{0}, \end{aligned} \tag{16}$$

$$\hat{\mu} \left(P_T - \sum_{l=1}^{L} \mathrm{Tr} \left(\hat{\mathbf{\Sigma}}_l \right) \right) = 0,$$

$$\mathrm{tr} \left(\hat{\mathbf{\Sigma}}_l \hat{\mathbf{\Theta}}_l \right) = 0,$$

$$\hat{\mathbf{\Sigma}}_l, \, \hat{\mathbf{\Theta}}_l \succcurlyeq 0, \, \hat{\mu} \geq 0.$$

Similarly to (12), from the polite water-filling structure on reverse links, we have

$$\mathbf{\Sigma}_l = \frac{w_l}{\hat{\mu}} \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{l,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{l,l} \right)^{-1} \right), \, l = 1, \ldots, L. \tag{17}$$

It is well known that $\sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) = P_T$ when $\mathbf{\Sigma}_{1:L}$ is a stationary point of problem (4), e.g., [6, Theorem 8 (item 3)]. This indicates that the full power should always be used. Since the covariance transformation preserves total power, we also have $\sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) = \sum_{l=1}^{L} \mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_l\right) = P_{\mathrm{T}}$ [5, Lemma 8]. It can also be proved that $\mathrm{Tr}\left(\mathbf{Q}_l\right) = \mathrm{Tr}\left(\hat{\mathbf{Q}}_l\right)$, where $\hat{\mathbf{Q}}_l \triangleq \mathbf{\Omega}_l^{\frac{1}{2}} \hat{\mathbf{\Sigma}}_l \mathbf{\Omega}_l^{\frac{1}{2}}$ are the reverse link equivalent covariance matrices [5, Lemma 20].

The polite water-filling algorithm, Algorithm PP in [5], works as follows. After initializing the reverse link interference plus noise covariance matrices $\hat{\mathbf{\Omega}}_{1:L}$, we perform a forward link polite water-filling using (13, 14) followed by a reverse link polite water-filling, which is defined to be one iteration. The iterations stops when the change of the objective function is less than a threshold or when a predetermined number of iterations is reached. Because the algorithm enforces the optimal signal structure at each iteration, it converges very fast. In particular, for parallel channels, it reduces to the traditional water-filling and gives the optimal solution in half an iteration with initial values $\hat{\mathbf{\Omega}}_l = \mathbf{I}, \ \forall l$. Unfortunately, this remarkable algorithm sometimes does not converge and the objective function has some oscillation, especially in very strong interference cases.

B. The New Algorithm

Instead of using (12) and (17) to solve the KKT condition for the polite water-filling structure, we can directly use these two equations to update $\hat{\mathbf{\Sigma}}_{1:L}$ and $\mathbf{\Sigma}_{1:L}$. Note that, since the full power is used, Lagrange multiplier μ can be chosen to satisfy the power constraint $\sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) = P_{\mathrm{T}}$ as

$$\mu = \frac{1}{P_{\mathrm{T}}} \sum_{l=1}^{L} w_l \mathrm{tr} \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{l,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{l,l} \right)^{-1} \right). \tag{18}$$

Algorithm 1 Dual Polite Water-filling Algorithm

1. Initialize $\mathbf{\Sigma}_l$'s, s.t. $\sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) = P_{\mathrm{T}}$
2. $R \Leftarrow \sum_{l=1}^{L} w_l \mathcal{I}_l \left(\mathbf{\Sigma}_{1:L}\right)$
3. Repeat
4. $R' \Leftarrow R$
5. $\mathbf{\Omega}_l \Leftarrow \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{lk} \mathbf{\Sigma}_k \mathbf{H}_{lk}^{\dagger}$
6. $\hat{\mathbf{\Sigma}}_l \Leftarrow \frac{P_{\mathrm{T}} w_l \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll} \mathbf{\Sigma}_l \mathbf{H}_{ll}^{\dagger} \right)^{-1} \right)}{\sum_{l=1}^{L} w_l \mathrm{tr} \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{ll} \mathbf{\Sigma}_l \mathbf{H}_{ll}^{\dagger} \right)^{-1} \right)}$
7. $\hat{\mathbf{\Omega}}_l \Leftarrow \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{kl}^{\dagger} \hat{\mathbf{\Sigma}}_k \mathbf{H}_{kl}$
8. $\mathbf{\Sigma}_l = \frac{P_{\mathrm{T}} w_l \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{ll} \right)^{-1} \right)}{\sum_{l=1}^{L} w_l \mathrm{tr} \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{ll}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{ll} \right)^{-1} \right)}$
9. $R \Leftarrow \sum_{l=1}^{L} w_l \mathcal{I}_l \left(\mathbf{\Sigma}_{1:L}\right)$
10. until $\left| R - R' \right| \leq \epsilon$.

This gives a new algorithm, Algorithm 1, that takes advantage of the structure of the weighted sum-rate maximization problem and, as confirmed by the analytical analysis and numerical experiments, has fast monotonic convergence. It converges to a stationary point of both problem (4) and problem (15) simultaneously, and at the stationary point, both (12) and (17) achieve the same sum-rate. We will analyze the convergence properties of Algorithm 1 in the next section.

IV. Convergence of Algorithm

In this section, we will prove the monotonic convergence of Algorithm 1. As will be seen later, the proof uses only very general convex analysis, as well as a particular scaling invariance property that we identify for the weighted sum-rate maximization problem. We expect that the scaling invariance holds for and our proof technique applies to many non-convex problems in communication networks that involve the rate or throughput maximization.

A. Preliminaries

1) Lagrangian of the weighted sum-rate function: The weighted sum-rate maximization problem (4) is equivalent to the following problem:

$$\begin{aligned} \max_{\mathbf{\Sigma}_{1:L}, \mathbf{\Omega}_{1:L}} \quad & \sum_{l=1}^{L} w_l \left(\log \left| \mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right| - \log \left| \mathbf{\Omega}_l \right| \right) \\ s.t. \quad & \mathbf{\Sigma}_l \succeq 0, \, \forall l, \\ & \sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l\right) \leq P_T, \\ & \mathbf{\Omega}_l = \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k \mathbf{H}_{l,k}^{\dagger}, \forall l, \end{aligned}$$

which is still non-convex. Consider the Lagrangian of the above problem

$$
\begin{aligned}
& F(\mathbf{\Sigma}, \mathbf{\Omega}, \mathbf{\Lambda}, \mu) \\
= & \sum_{l=1}^{L} w_l \left(\log \left| \mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right| - \log |\mathbf{\Omega}_l| \right) \\
& + \mu \{ P_T - \sum_{l=1}^{L} \mathrm{Tr}(\mathbf{\Sigma}_l) \} \\
& + \sum_{l=1}^{L} \mathrm{Tr} \left(\mathbf{\Lambda}_l \left(\mathbf{\Omega}_l - \mathbf{I} - \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k \mathbf{H}_{l,k}^{\dagger} \right) \right),
\end{aligned}
$$

where the domain of F is $\{\mathbf{\Sigma},\mathbf{\Omega},\mathbf{\Lambda},\mu | \mathbf{\Sigma}_l \in \mathbb{H}_{+}^{L_{R_l} \times L_{R_l}}, \mathbf{\Omega}_l \in \mathbb{H}_{++}^{L_{R_l} \times L_{R_l}}, \mathbf{\Lambda}_l \in \mathbb{H}^{L_{R_l} \times L_{R_l}}, \mu \in \mathbb{R}^{+}\}$. Here $\mathbb{H}^{n \times n}$, $\mathbb{H}_{+}^{n \times n}$, and $\mathbb{H}_{++}^{n \times n}$ mean n by n Hermitian matrix, positive semidefinite matrix, and positive definite matrix respectively.

One can easily verify that the function F is concave in $\mathbf{\Sigma}$ and convex in $\mathbf{\Omega}$. Furthermore, the partial derivatives are given by

$$
\begin{aligned}
\frac{\partial F}{\partial \mathbf{\Sigma}_l} &= w_l \mathbf{H}_{l,l}^{\dagger} \left(\mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right)^{-1} \mathbf{H}_{l,l} \\
& \quad - \mu \mathbf{I} - \sum_{k \neq l} \mathbf{H}_{k,l}^{\dagger} \mathbf{\Lambda}_l \mathbf{H}_{k,l}, \\
\frac{\partial F}{\partial \mathbf{\Omega}_l} &= w_l \left(\left(\mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right)^{-1} - \mathbf{\Omega}_l^{-1} \right) + \mathbf{\Lambda}_l.
\end{aligned}
$$

Now suppose that we have the pair $(\mathbf{\Sigma}, \mathbf{\Omega})$ such that

$$
\begin{aligned}
\sum_{l=1}^{L} \mathrm{Tr}(\mathbf{\Sigma}_l) &= P_T, \\
\mathbf{\Omega}_l &= \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k \mathbf{H}_{l,k}^{\dagger},
\end{aligned}
$$

then,

$$
\begin{aligned}
& F(\mathbf{\Sigma}, \mathbf{\Omega}, \mathbf{\Lambda}, \mu) \\
= & \sum_{l=1}^{L} w_l \left(\log \left| \Omega_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right| - \log |\mathbf{\Omega}_l| \right),
\end{aligned}
$$

which is the original weighted sum-rate function. For notational simplicity, denote the weighted sum-rate function by $V(\Sigma)$, i.e.,

$$
\begin{aligned}
& V(\mathbf{\Sigma}) \\
= & \sum_{l=1}^{L} w_l \left(\log \left| \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k \mathbf{H}_{l,k}^{\dagger} + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right| \right. \\
& \left. - \log \left| \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k \mathbf{H}_{l,k}^{\dagger} \right| \right).
\end{aligned}
$$

2) Solution of the first-order condition: Suppose that we want to solve the following system of equations in terms of $(\mathbf{\Sigma}, \mathbf{\Omega})$ for given $(\mathbf{\Lambda}, \mu)$:

$$
\begin{aligned}
\frac{\partial F}{\partial \mathbf{\Sigma}_l} &= 0, \\
\frac{\partial F}{\partial \mathbf{\Omega}_l} &= 0.
\end{aligned}
$$

Define

$$
\begin{aligned}
\hat{\mathbf{\Sigma}}_l &= \frac{1}{\mu} \mathbf{\Lambda}_l, \\
\hat{\mathbf{\Omega}}_l &= \mathbf{I} + \sum_{k \neq n} \mathbf{H}_{k,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{k,l},
\end{aligned}
$$

the above system of equations becomes

$$
\begin{aligned}
w_l \mathbf{H}_{l,l}^{\dagger} \left(\mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right)^{-1} \mathbf{H}_{l,l} &= \mu \hat{\mathbf{\Omega}}_l, \\
w_l \left(\mathbf{\Omega}_l^{-1} - \left(\mathbf{\Omega}_l + \mathbf{H}_{l,l} \mathbf{\Sigma}_l \mathbf{H}_{l,l}^{\dagger} \right)^{-1} \right) &= \mu \hat{\mathbf{\Sigma}}_l.
\end{aligned}
$$

A solution to this system of equations is given by

$$
\mathbf{\Sigma}_l = \frac{w_l}{\mu} \left(\hat{\mathbf{\Omega}}_l^{-1} - \left(\hat{\mathbf{\Omega}}_l + \mathbf{H}_{l,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{l,l} \right)^{-1} \right) \tag{19}
$$

$$
\mathbf{\Omega}_l = \frac{w_l}{\mu} \mathbf{H}_{l,l} \left(\mathbf{H}_{l,l}^{\dagger} \hat{\mathbf{\Sigma}}_l \mathbf{H}_{l,l} + \hat{\mathbf{\Omega}}_l \right)^{-1} \mathbf{H}_{l,l}^{\dagger}. \tag{20}
$$

The proof can be found in [11].

B. Convergence Results

We are ready to present the following two main convergence results regarding Algorithm 1. Denote by $\mathbf{\Sigma}^{(n)}$ the $\mathbf{\Sigma}$ value at the nth iteration of Algorithm 1.

Theorem 1. *The objective value, i.e., the weighted sum-rate, is monotonically increasing in Algorithm 1, $i.e.$,*

$$
V(\mathbf{\Sigma}^{(n)}) \leq V(\mathbf{\Sigma}^{(n+1)}).
$$

From the above theorem, the following conclusion is immediate.

Corollary 2. *The sequence* $V_n = V(\mathbf{\Sigma}^{(n)})$ converges to some limit point V_∞.

Proof: Since $V(\mathbf{\Sigma})$ is a continuous function and its domain $\{\mathbf{\Sigma} | \mathbf{\Sigma} \succeq \mathbf{0}, \mathrm{Tr}(\mathbf{\Sigma}) \leqq P_T\}$ is a compact set, V_n is bounded above. From Theorem 1, $\{V_n\}$ is a monotone increasing sequence, therefore there exists a limit point V_∞ such that $\lim_{n \to \infty} V_n = V_\infty$. ■

If we define a stationary point $(\mathbf{\Sigma}^\star)$ of Algorithm 1 as, $\mathbf{\Sigma}^{(n)} = \mathbf{\Sigma}^\star$ implies $\mathbf{\Sigma}^{(n+k)} = \mathbf{\Sigma}^\star$ for all $k = 0, 1, \cdots$, then we have the following result.

Theorem 3. *Algorithm 1 converges to a stationary point $\mathbf{\Sigma}^\star$.*

The proof of Theorems 1 and 3 will be presented later in this section. Before that, we first establish a few inequalities and identify a particular scaling property of the Lagrangian F.

1) The first inequality: Suppose that we have a feasible point $\mathbf{\Sigma}^{(\mathrm{n})} \succeq 0$, and $\sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Sigma}_l^{(n)}\right) = P_T$. In Algorithm 1, we generate $\mathbf{\Omega}_l^{(n)}$ such that

$$\mathbf{\Omega}_l^{(n)} = \mathbf{I} + \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k^{(n)} \mathbf{H}_{l,k.}^{\dagger}$$

Now we have a pair $(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)})$. Using this pair, we can compute $(\mathbf{\Lambda}^{(n)}, \mu^{(n)})$ such that

$$\begin{aligned} \mathbf{\Lambda}_l^{(n)} &= w_l \left(\mathbf{\Omega}_l^{(n)^{-1}} - \left(\mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l} \mathbf{\Sigma}_l^{(n)} \mathbf{H}_{l,l}^{\dagger} \right)^{-1} \right), \\ \mu^{(n)} &= \frac{1}{P_T} \sum_{l=1}^{L} \mathrm{Tr}\left(\mathbf{\Lambda}_l^{(n)} \right). \end{aligned}$$

Note that $\hat{\mathbf{\Sigma}}^{(n)}$ in Algorithm 1 is equal to

$$\hat{\mathbf{\Sigma}_l}^{(n)} = \frac{\mathbf{\Lambda}_l^{(n)}}{\mu^{(n)}}.$$

From this, the gradient of F with respect to $\mathbf{\Omega}$ at the point $(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)})$ vanishes, i.e.,

$$\frac{\partial F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}, \mathbf{\Lambda}^{(n)}, \mu^{(n)})}{\partial \mathbf{\Omega}}|_{\mathbf{\Omega}^{(n)}} = 0.$$

Since F is convex in $\mathbf{\Omega}$, if we fix $\mathbf{\Sigma} = \mathbf{\Sigma}^{(n)}$, then $\mathbf{\Omega}^{(n)}$ is a global minimizer of F. In other words,

$$F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Lambda}^{(n)}, \mu^{(n)}) \leq F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}, \mathbf{\Lambda}^{(n)}, \mu^{(n)}) \quad (21)$$

for all $\mathbf{\Omega} \succ 0$.

2) Scaling invariance of F: We will identify a remarkable scaling invariance property of F, which plays a key role in the convergence proof of Algorithm 1. For given $(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Lambda}^{(n)}, \mu^{(n)})$, we have

$$\begin{aligned} & F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \frac{1}{\alpha}\mathbf{\Omega}^{(n)}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ = \ & F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Lambda}^{(n)}, \mu^{(n)}) \end{aligned}$$

for all α >0. To show this scaling invariance property, note that

$$\begin{aligned} \mathbf{\Omega}_l^{(n)} - \sum_{k \neq l} \mathbf{H}_{l,k} \mathbf{\Sigma}_k^{(n)} \mathbf{H}_{l,k}^{\dagger} &= \mathbf{I}, \\ \sum_{l=1}^{L} \mathrm{Tr}(\mathbf{\Sigma}_l^{(n)}) &= P_T, \\ P_T \mu^{(n)} &= \sum_{l=1}^{L} \mathrm{Tr}(\mathbf{\Lambda}_l^{(n)}). \end{aligned}$$

Applying the above equalities and some mathematical manipulations, we have

$$\begin{aligned} & F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \frac{1}{\alpha}\mathbf{\Omega}^{(n)}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ = & \sum_{l=1}^{L} w_l \left(\log\left| \mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right| - \log\left| \mathbf{\Omega}_l^{(n)} \right| \right) \\ & + \alpha\mu^{(n)}\{P_T - \frac{1}{\alpha}P_T\} + \sum_{l=1}^{L} \mathrm{Tr}\left(\alpha\mathbf{\Lambda}_l^{(n)} \left(\frac{1}{\alpha}\mathbf{I} - \mathbf{I} \right) \right) \\ = & \sum_{l=1}^{L} w_l \left(\log\left| \mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right| - \log\left| \mathbf{\Omega}_l^{(n)} \right| \right) \\ & + (\alpha - 1)\mu^{(n)}P_T + (1-\alpha)\sum_{l=1}^{L} \mathrm{Tr}(\mathbf{\Lambda}_l^{(n)}) \\ = & \sum_{l=1}^{L} w_l \left(\log\left| \mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right| - \log\left| \mathbf{\Omega}_l^{(n)} \right| \right) \\ = & F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Lambda}^{(n)}, \mu^{(n)}), \end{aligned}$$

where the first equality uses the fact that

$$\begin{aligned} & \log\left| \frac{1}{\alpha}\left(\mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right) \right| - \log\left| \frac{1}{\alpha}\mathbf{\Omega}_l^{(n)} \right| \\ = \ & \log\left| \mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right| - \log\left| \mathbf{\Omega}_l^{(n)} \right|. \end{aligned}$$

Furthermore,

$$\begin{aligned} & \frac{\partial F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)})}{\partial \mathbf{\Omega}}|_{\frac{1}{\alpha}\mathbf{\Omega}^{(n)}} \\ = & w_l \left(\left(\frac{1}{\alpha}\mathbf{\Omega}_l^{(n)} + \mathbf{H}_{l,l}\frac{1}{\alpha}\mathbf{\Sigma}_l^{(n)}\mathbf{H}_{l,l}^{\dagger} \right)^{-1} - \left(\frac{1}{\alpha}\mathbf{\Omega}_l^{(n)} \right)^{-1} \right) \\ & + \alpha\mathbf{\Lambda}_l^{(n)} \\ = & \alpha\frac{\partial F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}, \mathbf{\Lambda}^{(n)}, \mu^{(n)})}{\partial \mathbf{\Omega}}|_{\mathbf{\Omega}^{(n)}} \\ = & 0. \end{aligned}$$

Therefore, $\frac{1}{\alpha}\mathbf{\Omega}^{(n)}$ is a global minimizer of $F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)})$, as F is convex in Ω.

3) The second and third inequalities: Given $(\alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)})$, we generate $\tilde{\mathbf{\Sigma}}, \tilde{\mathbf{\Omega}}$ using equation (19) and (20). If we choose α so that $\sum_{l=1}^{L} \tilde{\mathbf{\Sigma}}_l = P_T$, then $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma}^{(n+1)}$ in Algorithm 1. Since $(\mathbf{\Sigma}^{(n+1)}, \tilde{\mathbf{\Omega}})$ is chosen to make partial derivatives zero:

$$\begin{aligned} \frac{\partial F(\mathbf{\Sigma}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)})}{\partial \mathbf{\Sigma}_l}|_{\mathbf{\Sigma}^{(n+1)}} &= 0, \\ \frac{\partial F(\mathbf{\Sigma}^{(n+1)}, \mathbf{\Omega}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)})}{\partial \mathbf{\Omega}_l}|_{\tilde{\mathbf{\Omega}}} &= 0, \end{aligned}$$

we conclude that $\mathbf{\Sigma}^{(n+1)}$ is a global maximizer, i.e.,

$$F(\mathbf{\Sigma}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n+1)}, \alpha\mu^{(n+1)}) \leq F(\mathbf{\Sigma}^{(n+1)}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \quad (22)$$

for all $\mathbf{\Sigma} \succeq 0$; and $\tilde{\mathbf{\Omega}}$ is a global minimizer, i.e.,

$$F(\mathbf{\Sigma}^{(n+1)}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \leq F(\mathbf{\Sigma}^{(n+1)}, \mathbf{\Omega}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \quad (23)$$

for all $\mathbf{\Omega} \succ 0$.

4) Proof of Theorem 1: With the three inequalities (21)-(23) obtained above, we are ready to prove Theorem 1. As in Algorithm 1 $\mathbf{\Omega}_l^{(n+1)} = \mathbf{I} + \sum_{k\neq l} \mathbf{H}_{l,k}\mathbf{\Sigma}_k^{(n+1)}\mathbf{H}_{l,k}^\dagger$, we have

$$\begin{aligned} & V(\mathbf{\Sigma}^{(n+1)}) \\ = \;& F(\mathbf{\Sigma}^{(n+1)}, \mathbf{\Omega}^{(n+1)}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ \geq \;& F(\mathbf{\Sigma}^{(n+1)}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ \geq \;& F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \tilde{\mathbf{\Omega}}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ \geq \;& F(\frac{1}{\alpha}\mathbf{\Sigma}^{(n)}, \frac{1}{\alpha}\mathbf{\Omega}^{(n)}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n)}) \\ = \;& F(\mathbf{\Sigma}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Lambda}^{(n)}, \mu^{(n)}) \\ = \;& V(\mathbf{\Sigma}^{(n)}), \end{aligned}$$

where the first inequality follows from (23), and the second and third inequalities follows from (22) and (21).

5) Proof of Theorem 3: We have shown in Corollary 2 that V_n converges to a limit point under Algorithm 1. To show the convergence of the algorithm, it is enough to show that if $V(\mathbf{\Sigma}^{(n)}) = V(\mathbf{\Sigma}^{(n+1)})$, then $\mathbf{\Sigma}^{(n+1)} = \mathbf{\Sigma}^{(n+k)}$ for all $k = 1, 2, \cdots$. Suppose $V(\mathbf{\Sigma}^{(n)}) = V(\mathbf{\Sigma}^{(n+1)})$, then from the proof in the above, we have

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

Since $\tilde{\mathbf{\Omega}}$ is a global minimizer, the above equality implies $\mathbf{\Omega}^{(n+1)}$ is a global minimizer too. From the first order condition for optimality, we have

$$\begin{aligned} & \frac{\partial F(\mathbf{\Sigma}^{(n+1)}, \mathbf{\Omega}, \alpha\mathbf{\Lambda}^{(n)}, \alpha\mu^{(n+1)})}{\partial \mathbf{\Omega}_l}|_{\mathbf{\Omega}^{(n+1)}} \\ = \;& w_l \left(\left(\mathbf{\Omega}_l^{(n+1)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n+1)}\mathbf{H}_{l,l}^\dagger \right)^{-1} - \{\mathbf{\Omega}_l^{(n+1)}\}^{-1} \right) \\ & + \alpha\mathbf{\Lambda}_l^{(n)} \\ = \;& 0. \end{aligned}$$

On the other hand, we generate $\mathbf{\Lambda}^{(n+1)}$ such that

$$\begin{aligned} & \mathbf{\Lambda}_l^{(n+1)} \\ = \;& w_l \left(\mathbf{\Omega}_l^{(n+1)^{-1}} - \left(\mathbf{\Omega}_l^{(n+1)} + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n+1)}\mathbf{H}_{l,l}^\dagger \right)^{-1} \right) \\ = \;& \alpha\mathbf{\Lambda}_l^{(n)}. \end{aligned}$$

This shows $\hat{\mathbf{\Sigma}}^{(n+1)} \propto \hat{\mathbf{\Sigma}}^{(n)}$. However, since the trace of each matrix is same, we conclude that

$$\hat{\mathbf{\Sigma}}^{(n+1)} = \hat{\mathbf{\Sigma}}^{(n)}.$$

From this it is obvious that $\hat{\mathbf{\Sigma}}^{(n)} = \hat{\mathbf{\Sigma}}^{(n+1)} = \cdots$ and $\mathbf{\Sigma}^{(n+1)} = \mathbf{\Sigma}^{(n+2)} = \cdots$.

V. Numerical Examples and Complexity Analysis

A. Numerical Examples

In this section, we provide numerical examples to verify the analysis in the previous sections and compare the proposed new algorithm against the PWF algorithm [5] and the WMMSE algorithm [9]. Consider a B-MAC network with $L = 10$ links among 10 transmitter-receiver pairs that fully interfere with each other. Each link has 3 transmit antennas and 4 receive antennas. For each simulation, the channel matrices are independently generated by one realization of $\mathbf{H}_{l,k} = \sqrt{g_{l,k}}\mathbf{H}_{l,k}^{(\text{W})}, \forall k, l$, where $\mathbf{H}_{l,k}^{(\text{W})}$ has zero-mean i.i.d. complex Gaussian entries with unit variance and $g_{l,k}$ is the average channel gain. The weights w_l's are uniformly chosen from 0.5 to 1. The total transmit power P_{T} in the network is 100.

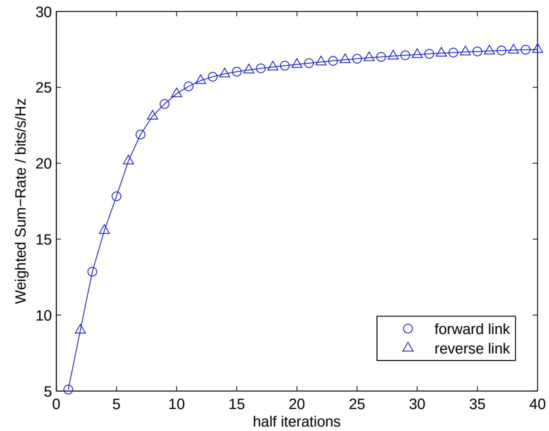

Fig. 2. The monotonic convergence of the forward and reverse link weighted sum-rates of the new algorithm with $P_{\text{T}} = 100$ and $g_{l,k} = 0\text{dB}, \forall l, k$.

Fig. 2 shows the convergence of the new algorithm for a network with $g_{l,k} = 0\text{dB}, \forall l, k$. From the proof of Theorem 1, the weighted sum-rate of the forward links and that of the reverse links not only increase monotonically over iterations, but also increase over each other over half iterations. In Algorithm 1, the reverse link transmit signal covariance matrices are updated in the first half of each iteration (line 6), and the forward link transmit signal covariance matrices are updated in the second half (line 8). From Fig. 2, we clearly see that the weighted sum-rates of the forward links and reverse links increase in turns until they converge to the same value, which also confirms that problem (4) and its dual problem (15) reach their stationary points at the same time.

We compare the performance of the new algorithm with the PWF and WMMSE algorithms under different channel conditions: weak, moderate, and strong interference. Fig. 3 shows a comparison under the weak interference condition. We see that the PWF algorithm converges slightly faster than the new algorithm, and it is close to the stationary point in three iterations. The reason for this remarkable convergence is that under the weak interference condition, the channels in the network are close to parallel channels and the PWF algorithm can converges to the optimal solution in half an iteration. Since the new algorithm is also based on the polite water-filling structure, it is not surprising that it has a fast convergence. In contrast, the WMMSE algorithm's convergence under such a

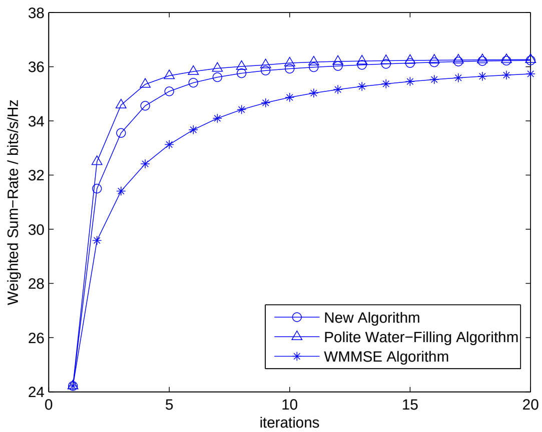

Fig. 3. PWF algorithm vs. WMMSE algorithm vs. new algorithm under weak interference with $P_{\rm T} = 100$, $g_{l,l} = 0$dB and $g_{l,k} = -10$dB for $l \neq k$.

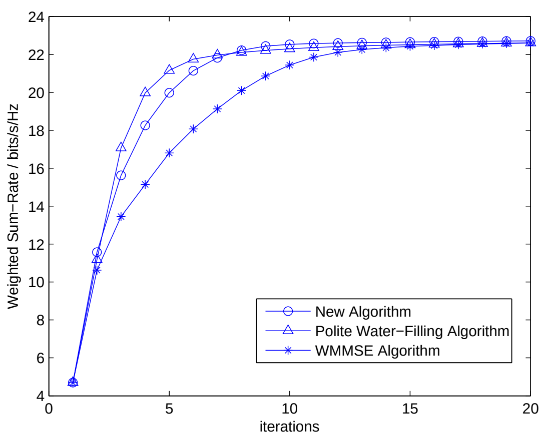

Fig. 4. PWF algorithm vs. WMMSE algorithm vs. new algorithm under moderate interference with $P_{\rm T} = 100$ and $g_{l,k} = 0$dB, $\forall l, k$.

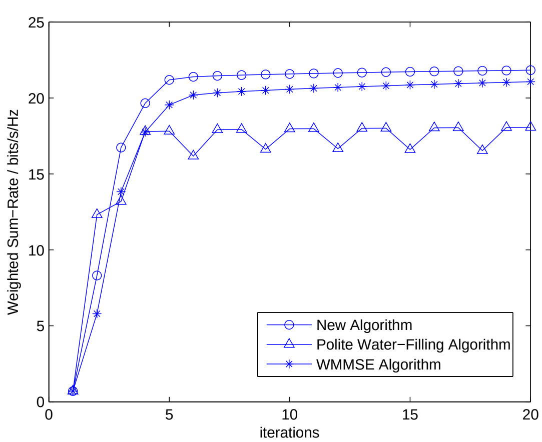

Fig. 5. PWF algorithm vs. WMMSE algorithm vs. new algorithm under strong interference with $P_{\rm T} = 100$, $g_{l,l} = 0$dB and $g_{l,k} = 10$dB for $l \neq k$.

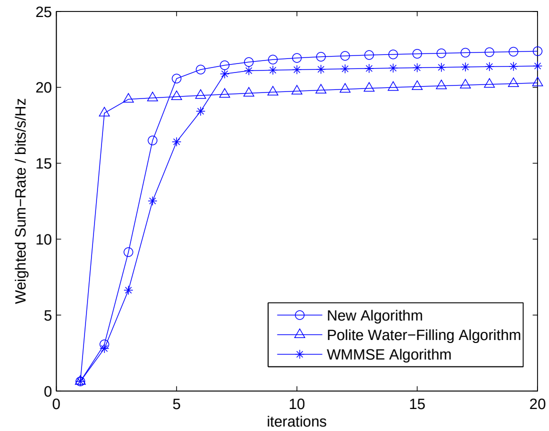

Fig. 6. An example that PWF, WMMSE, and new algorithms converge to different stationary points. $P_{\rm T} = 100$, $g_{l,l} = 0$dB and $g_{l,k} = 10$dB for $l \neq k$.

channel condition is significantly slower; e.g., more than ten iterations slower to reach some high value of the weighted sum-rate (i.e., if a high numerical accuracy is desired). When the gain of the interfering channels are comparable to that of the desired channel, as shown in Fig. 4, the difference in the convergence speed between the PWF/new algorithm and the WMMSE algorithm is less than that of the weak interference case. But around five iteration difference for some high value of the weighted sum-rate is still significant. Under some strong interference conditions, as shown in Fig. 5, the PWF algorithm oscillates around certain points and no longer converges, while the other two algorithms still have the guaranteed convergence. The new algorithm still converges faster than the WMMSE algorithm and the difference is significant if high value of the weighted sum-rate is desired. Even in the case that one is satisfied with smaller weighted sum-rate, being faster by a few iterations will be significant in distributed implementation where the overhead of each iteration costs significant signaling resources between the transmitters and the receivers.

Combining the three cases together, we see that while both algorithms are provably convergent, the new algorithm outperforms the WMMSE algorithm in all situations, especially in the weak interference case. Although the PWF algorithm has better convergence over the new algorithm, it does not converge under certain strong interference channels. We can conclude that the new algorithm preserves the fast convergence of the PWF algorithm and has the desired convergence property as the WMMSE algorithm as well.

Note that given the same initial point, these three algorithms may converge to different stationary points, as shown in Fig. 6. Since the original weighted sum-rate maximization problem is non-convex, a stationary point is not necessarily a global maximum. In practical applications, we may run such algorithms multiple times starting from different initial points and pick the stationary point that gives the largest weighted sum-rate.

We may also create a hybrid algorithm by combining the new algorithm and the PWF algorithm, by applying the PWF algorithm until the weighted sum-rate starts to drop, then switching to the new algorithm which will converge to a stationary point. In this way, the hybrid algorithm will retain the PWF algorithm's fast convergence and also will monotonically converge under those strong interference channel conditions where the PWF algorithm may oscillate.

B. Complexity Analysis

We have evaluated in the above convergence properties of the proposed new algorithm, the PWF algorithm and the WMMSE algorithm in terms of the number of iterations. We now analyze the complexity of each iteration for these algorithms.

Recall that L is the number of users or links, and for simplicity, assume that each user has N transmit (and receive) antennas, so the resulting $\mathbf{\Sigma}_l$ (and $\hat{\mathbf{\Sigma}}_l$) is a $N \times N$ matrix. Suppose that we use the straightforward matrix multiplication and inversion. Then the complexity of these operations are $O(N^3)$. For the new algorithm, at each iteration, $\mathbf{\Omega}_l$ incurs a complexity of $O(LN^3)$ and $\mathbf{\Omega}_l + \mathbf{H}_{l,l}\mathbf{\Sigma}_l^{(n+1)}\mathbf{H}_{l,l}^\dagger$ incurs a complexity of $O(LN^3)$. To obtain $\hat{\mathbf{\Sigma}}_l$, we have to invert a $N \times N$ matrix, which incurs a complexity of $O(N^3)$. Therefore, the total complexity for calculating a $\hat{\mathbf{\Sigma}}_l$ is given by $O(LN^3)$, and the complexity of generating $\hat{\mathbf{\Sigma}}$ is given by $O(L^2N^3)$. As calculating $\mathbf{\Sigma}$ incurs the same complexity as calculating $\hat{\mathbf{\Sigma}}$, the complexity of the new algorithm is $O(L^2N^3)$ for each iteration.

The PWF algorithm uses the same calculation to generate $\mathbf{\Omega}_l$ and incurs a complexity of $O(LN^3)$ for each $\mathbf{\Omega}_l$. Then, it uses the singular value decomposition of $\mathbf{\Omega}_l^{-\frac{1}{2}}\mathbf{H}_{l,l}\hat{\mathbf{\Omega}}_l^{-\frac{1}{2}}$, which incurs a complexity of $O(N^3)$. Since we need L of these operations, the total complexity of the PWF algorithm is $O(L^2N^3)$. For the WMMSE algorithm, it is shown in [9] that its complexity is $O(L^2N^3)$.

So, all three algorithms have the same computational complexity per iteration if we use $O(N^3)$ matrix multiplication. Recently, Williams [10] presents an $O(N^{2.3727})$ matrix multiplication and inversion method. If we use this algorithm, then the new algorithm and the WMMSE algorithm have $O(L^2N^{2.3727})$ complexity since the N^3 factor comes from the matrix multiplication and inversion. However, in addition to L^2 number of matrix multiplications and inversions, the PWF algorithm has L number of N by N matrix singular value decompositions. Therefore the complexity of the PWF algorithm is $O(L^2N^{2.3727} + LN^3)$.

VI. Conclusion

We have proposed a new algorithm to solve the weighted sum-rate maximization problem in general interference networks. Based on the polite water-filling results and the rate duality [5], the new algorithm updates the transmit signal covariance matrices in the forward and reverse links in a symmetric manner and has fast and guaranteed convergence. We present an elegant convergence proof for this otherwise hard problem. Simulations demonstrate that the new algorithm has convergence speed close to the state-of-the-art polite water-filling algorithm, which is however not guaranteed to converge. Compared with another state-of-the-art WMMSE algorithm, which is guaranteed to converge, the convergence speed is a few iterations or more than ten iterations faster, depending on the desired numerical accuracy. Being faster even by a few iterations will be significant in distributed implementation where the overhead of each iteration costs significant signaling resources between the transmitters and the receivers. Indeed, the new algorithm is highly scalable and suitable for distributed implementation with distributed channel estimation because for each data link, only its own channel state and the aggregated interference plus noise covariance needs to be estimated no matter how many interferers are there (See lines 6 and 8 of Algorithm 1). The distributed implementation of the new algorithm will be developed in the future work.

References

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

[2] 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.

[3] 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.

[4] Seung-Jun 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.

[5] An Liu, Youjian Liu, Haige Xiang, and Wu Luo. Duality, polite water-filling, and optimization for MIMO B-MAC interference networks and iTree networks. *submitted to IEEE Trans. Info. Theory*, Apr. 2010; revised Oct. 2010.

[6] An Liu, Youjian Liu, Haige Xiang, and Wu Luo. MIMO B-MAC interference network optimization under rate constraints by polite water-filling and duality. *IEEE Trans. Signal Processing*, 59(1):263 –276, Jan. 2011.

[7] An Liu, Youjian Liu, Haige Xiang, and Wu Luo. Polite water-filling for weighted sum-rate maximization in B-MAC networks under multiple linear constraints. *IEEE Trans. Signal Processing*, 60(2):834 –847, Feb. 2012.

[8] 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.

[9] Qingjiang Shi, M. Razaviyayn, Zhi-Quan Luo, and Chen He. An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel. *IEEE Trans. Signal Processing*, 59(9):4331 –4340, sept. 2011.

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

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

[12] 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.

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

[14] Wei 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.

[15] Lan Zhang, Ying-Chang Liang, Yan Xin, Rui Zhang, and H.V. Poor. On gaussian MIMO BC-MAC duality with multiple transmit covariance constraints. pages 2502 –2506, June 2009.

[16] Lan Zhang, Yan Xin, and Ying chang Liang. Weighted sum rate optimization for cognitive radio MIMO broadcast channels. *IEEE Trans. Wireless Commun.*, 8(6):2950 –2959, 2009.