Optimality Conditions of Performance-Guaranteed Power Minimization in MIMO Networks: A Distributed Algorithm and Its Feasibility

Guojun Xiong, Taejoon Kim, David J. Love, and Erik Perrins

***Abstract*—A distributed approach is proposed to the problem of signal-to-interference-plus-noise-ratio (SINR)-guaranteed power minimization (SGPM) for multicell multiuser (MCMU) multiple-input multiple-output (MIMO) systems. Unlike prior SGPM approaches, the proposed technique is based on solving necessary and sufficient optimality conditions, which are derived by decomposing the original problem into forward and backward (FB) subproblems, while ensuring the strong duality of each subproblem. The proposed distributed SGPM algorithm makes use of FB adaptation and Jacobi recursion, respectively, for iterative filter design and power allocation. A sufficient condition for the feasibility of the proposed distributed algorithm is analyzed, based on the matrix inverse-positive theory. Unlike the existing fully distributed FB filter update algorithms, the proposed approach guarantees target SINR performance as well as its convergence to a stationary point. Simulation results illustrate the enhanced power efficiency with the performance guarantees of the proposed method compared to the existing distributed techniques.**

***Index Terms*—Multiple-input multiple-output (MIMO) network, performance-guaranteed power minimization, forward-backward (FB) iteration, Jacobi recursion.**

I. Introduction

The deluge of wireless data traffic catalyzed by the growing number of data-intensive devices has motivated the deployment of small-cells in fifth-generation (5G) networks [2]–[4]. Simultaneously, as spectrum dedicated to wireless communications continues to be in high demand, dense small-cell deployment in sub-28 GHz millimeter wave (mmWave) bands[1] becomes increasingly important [5]–[7]. Unlike the above-28 GHz spectrum, in sub-28 GHz bands, fully digital processing is feasible and dense urban channels follow Rayleigh/Rician fading [8]. Thus, in both sub-6 GHz and sub-28 GHz mmWave bands, interference is still a limiting factor [7], [9] and poses challenges as accumulated intra-cell and inter-cell interference can severely deteriorate the link quality.

An approach to addressing the interference issue is employing advanced multiple-input multiple-output (MIMO) precoding techniques in order to maximize the network sum rate [7], [10]–[12]. However, these approaches do not provide user-wise quality of service (QoS) guarantee because sum-rate maximization often results in unexpected interference amplification. Reducing or turning off the transmit power of strong users can effectively mitigate the interference [13], [14]. A drawback of such approaches is insufficient utilization of the available degrees of freedom. A rather prudent strategy is minimizing the transmit power while guaranteeing a certain level of quality of service (QoS) for each user [15]. In this category, the designs of MIMO precoders and/or combiners that minimize the total transmit power subject to signal-to-interference-plus-noise-ratio (SINR) constraints have been popularly studied [15]–[27]. We refer to the latter strategy as SINR-guaranteed power minimization (SGPM). The SGPM problem in MIMO is non-convex, so it is not directly solvable.

Optimal approaches to SGPM have been studied for single-user MIMO systems, based on the application of the majorization theorem [16], second-order-cone programming (SOCP) [17], and uplink-downlink (UD) duality[2] [17]. It is possible to extend UD duality and SOCP to multicell multiuser (MCMU) MIMO beamforming scenarios [18], which can be solved by primal decomposition [24], [25], dual decomposition [26], or more advanced alternating direction method of multipliers [27]. However, UD duality and SOCP only hold when the receive combiner is held fixed. Once joint precoder and combiner design is involved, UD duality and SOCP does not hold [19], and the above algorithms [17], [18], [24]–[27] cannot be directly extended. Hence, many have turned to iterative heuristics that combine SOCP (or UD duality) with separate combiner design algorithms [19], [21]–[23]. Although the convergence to the first-order necessary Karush-Kuhn-Tucker (KKT) conditions has been presented [23], the sufficient and necessary optimality conditions of MCMU MIMO SGPM have not been fully addressed. Therefore, in this paper, we establish the sufficient and necessary optimality conditions of SGPM when joint precoding and combining are considered.

Global channel state information (CSI) was a prerequisite of prior SGPM techniques for power allocation [18]–[23]. In practice, a large fraction of MCMU MIMO networks will not have access to global CSI. This has motivated the investigation of fully distributed MCMU MIMO precoding and combining techniques [7], [13], [14], [28]–[30] that rely only on local CSI available at each communication end. They

G. Xiong, T. Kim, and E. Perrins are with the Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA (e-mail: gjxiong@ku.edu; taejoonkim@ku.edu; esp@ku.edu). D. J. Love is with the School of Electrical and computer Engineering, Purdue University, West Lafayette, IN, USA (e-mail: djlove@ecn.purdue.edu). T. Kim was supported in part by the National Science Foundation (NSF) under grants CNS 1955561 and AST 2037864. D. J. Love was supported in part by the NSF under grant CNS1642982. Parts of this work were previously presented at the IEEE Vehicular Technology Conference, Toronto, Canada 2017 [1].

[1]In 5G, sub-28 GHz bands, e.g., X-band (8–12 GHz), Ku-band (12–18 GHz), and Ka-band (24–28 GHz) have been popularly discussed.

[2]UD duality refers to the situation when the primal downlink SGPM problem shares the same set of SINR values as the dual uplink SGPM problem.

employ forward and backward (FB) iteration to update the combiners in the downlink and the precoders in the uplink, based on the channel reciprocity in time division duplexing (TDD). They can resort to tractable subproblem formulations, such as maximum (max)-SINR [13], [28], interference leakage minimization (ILM) [14], [28], and weighted minimum mean square error (WMMSE) [29], [30] subject to user-wise power constraints. Except for max-SINR [13], [28], the convergence of FB ILM [14], [28] and WMMSE [29], [30] has been established. However, the lack of user-wise QoS guarantees is a major drawback.

Some recent research on distributed power minimization has addressed QoS guarantee such as per-user rate constraints [24]. By relaxing the original problem to tractable convex programs, the work in [24] established the objective convergence of the proposed distributed algorithm. In [24], both precoding and combning were also jointly considered. However, optimality and feasibility conditions have not been identified. The feasibility of distributed power minimization is determined by the target QoS values that must be set prior to iterate the algorithm. If these values are improperly set, the algorithm can return non-positive power values.

In this paper, we propose a fully distributed SGPM framework for MCMU MIMO precoding and combining systems. We establish the optimality conditions of the MCMU MIMO SGPM problem and based on these, derive a fully distributed SGPM algorithm and its feasibility conditions, which relies on local CSI while providing QoS guarantees. We begin by decomposing the original SGPM problem into forward and backward non-convex subproblems and show that strong duality holds for each subproblem. Obtaining the necessary and sufficient optimality conditions of each subproblems, it can be readily shown that the joint optimality conditions of the forward and backward subproblems are indeed equivalent to those of the original SGPM problem. Our proposed distributed FB SGPM algorithm is centered on solving these optimality conditions.

In particular, the proposed algorithm makes use of Jacobi recursion [31] for the power allocation in conjunction with the FB precoder and combiner adaptation to iteratively solve the optimality conditions. The Jacobi method refines the transmit power values by leveraging one scalar feedback introduced between the users and base stations (BSs). We provide a distributed approach to predetermine the target SINR values at each communication end to satisfy feasibility based on the application of the matrix inverse-positive theory [32], [33]. The convergence to a stationary point of the proposed distributed algorithm is established according to the first-order stationary conditions. Any target SINR value chosen to meet the feasibility condition ensures the convergence of the proposed algorithm to a stationary point.

We numerically demonstrate substantially enhanced power efficiency and QoS guarantees of the proposed algorithm, especially, for dense small-cell networks. It is revealed that the proposed technique combines the best features of both prior SGPM strategies [15]–[23] and fully distributed methods [7], [13], [14], [28]–[30] while addressing their individual drawbacks. Possible variations of the proposed distributed algorithm are also presented by extending the proposed technique to the distributed rank adaptation (RA) and QoS adaptation (QA).

In this paper, we have approached the problem of MCMU MIMO SGPM transceiver design from a linear-transceiver-optimization point-of-view, with a primary focus on establishing the optimality conditions and the feasibility guarantee. Recently, there has been active development of general non-linear transceiver design such as widely linear transceiver [34], time-sharing [35], [36], and rate-splitting [37], [38]. These non-linear approaches can lead to improved QoS with moderately increased complexity especially in the interference-limited regime. Moreover, they are flexible to be combined with linear precoding to inspire further improved methods in different scenarios [39]. Investigation of ways of introducing non-linear structure to the proposed SGPM to refine performance and feasibility is an interesting topic for future work.

The rest of the paper is organized as follows. In Section II, the system model and preliminaries are present. In Section III, the SGPM subproblem formulations and their optimality analysis are enunciated. Section IV devises the distributed SGPM algorithm. Section V presents the convergence analysis. In Section VI, we discuss several practical issues of the proposed algorithm. The numerical simulation results and the concluding remarks are presented in Section VII and Section VIII, respectively.

Notation: A bold lower case $\mathbf{a}$ is a vector, a bold upper case $\mathbf{A}$ is a matrix. $\mathbf{A}(u,v)$ is the uth row and vth column entry of $\mathbf{A}$. $\mathbf{1}_M \in \mathbb{R}^{M\times 1}$, $\mathbf{0}_{M\times N} \in \mathbb{R}^{M\times N}$, and $\mathbf{I}_M \in \mathbb{R}^{M\times M}$ are the all-one column vector, all-zero matrix, and M-dimensional identity matrix, respectively. We denote, respectively, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_\infty$, $\|\mathbf{A}\|$, $\mathbf{A}^*$, $\mathrm{tr}(\mathbf{A})$, and $\mathrm{rank}(\mathbf{A})$ as the Frobenius norm, infinity norm that is defined by $\max_u \sum_v |A(u,v)|$, arbitrary induced matrix norm, conjugate transpose, trace of $\mathbf{A}$, and rank of $\mathbf{A}$. $\nu_{1:d}(\mathbf{A})$ extracts the first d dominant eigenvectors of $\mathbf{A}$. $\rho(\mathbf{A}) = \max_i |\lambda_i(\mathbf{A})|$ is the spectral radius of $\mathbf{A}$ with $\lambda_i(\mathbf{A})$ being the ith eigenvalue of $\mathbf{A}$. $\mathcal{R}(\mathbf{A})$ is the range space of $\mathbf{A}$. $\mathbb{R}_+$ is the set of non-negative real numbers and $\succeq$ denotes an elements-wise vector or matrix inequality.

II. System Model and Preliminaries

In this section, we discuss the system model and problem formulation under consideration. We also introduce the inverse-positive matrix theory [32], [33] and derive some preliminary results, which will be used in the rest of the paper.

A. System Model

Consider a MCMU MIMO small cell network consisting of L cells with each cell having one base station (BS) that simultaneously serves K users. We assume channel reciprocity in the TDD setting. Users and BSs are equipped with M and N antennas, respectively. We use the notation ℓ_k to denote the kth user in the ℓth cell where $k \in \{1,...,K\}$ and $\ell \in \{1,...,L\}$. Both downlink (forward direction) and uplink (backward direction) transmission are considered. The downlink signal model is first presented.

1) Forward Direction: The received signal in the downlink at user ℓ_k is modeled as

$$\mathbf{y}_{\ell_k}^{(DL)} = \mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}\mathbf{s}_{\ell_k}^{(DL)} + \sum_{i_j \neq \ell_k} \mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}\mathbf{s}_{i_j}^{(DL)} + \mathbf{n}_{\ell_k}^{(DL)}, \quad (1)$$

where $\mathbf{H}_{\ell_k,i} \in \mathbb{C}^{M\times N}$ is the channel from BS i to user ℓ_k and $\mathbf{n}_{\ell_k}^{(DL)} \in \mathbb{C}^{M\times 1}$ is the noise vector at user ℓ_k with each entry being an independent and identically distributed (i.i.d.) zero mean and σ^2 variance Gaussian random variable, i.e., $\mathbf{n}_{\ell_k}^{(DL)} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$. $\widetilde{\mathbf{V}}_{\ell_k} \in \mathbb{C}^{N\times d}$ is the linear precoder for user ℓ_k, where d is the number of data streams. It can be decomposed to $\widetilde{\mathbf{V}}_{\ell_k} = \sqrt{\alpha_{\ell_k}}\mathbf{V}_{\ell_k}$, where $\alpha_{\ell_k} \in \mathbb{R}_+$ is the transmit power value and $\|\mathbf{V}_{\ell_k}\|_F = 1$. $\mathbf{s}_{\ell_k}^{(DL)} \in \mathbb{C}^{d\times 1}$ is the signal vector and follows $\mathbf{s}_{\ell_k}^{(DL)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_d)$. The second term on the right hand side (r.h.s.) of (1) corresponds to the intra- and inter-cell interference at user ℓ_k. A linear combiner $\mathbf{U}_{\ell_k} \in \mathbb{C}^{M\times d}$ is employed at user ℓ_k to restore $\mathbf{s}_{\ell_k}^{(DL)}$ from $\mathbf{y}_{\ell_k}^{(DL)}$ such that $\hat{\mathbf{s}}_{\ell_k}^{(DL)} = \mathbf{U}_{\ell_k}^*\mathbf{y}_{\ell_k}^{(DL)}$. The rank of $\mathbf{V}_{\ell_k}$ and $\mathbf{U}_{\ell_k}$ is assumed to be d, i.e., $\mathrm{rank}(\mathbf{V}_{\ell_k}) = \mathrm{rank}(\mathbf{U}_{\ell_k}) = d$. Under the assumption that joint encoding is used across streams for each user, joint decoding of each user's streams is used at the users, and interference is treated as noise, the Shannon rate of user ℓ_k is

$$R_{\ell_k} = \log\left|\mathbf{I}_d + \mathbf{U}_{\ell_k}^*\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^*\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\right|, \quad (2)$$

where $\mathbf{R}_{\ell_k} \triangleq \mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}\left(\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}\right)^*$ and $\mathbf{Q}_{\ell_k} \triangleq \sum_{i_j\neq\ell_k}\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}\left(\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}\right)^* + \sigma^2\mathbf{I}_M$ denote the desired signal and the interference-plus-noise covariance matrices of user ℓ_k, respectively.

We define the SINR as the sum signal power across user ℓ_k divided by the sum interference power. Then, the downlink SINR of user ℓ_k is expressed as

$$\gamma_{\ell_k}^{(DL)} = \frac{\alpha_{\ell_k}\left\|\mathbf{U}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\right\|_F^2}{\sum_{i_j\neq\ell_k}\alpha_{i_j}\left\|\mathbf{U}_{\ell_k}^*\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\right\|_F^2 + \sigma^2\left\|\mathbf{U}_{\ell_k}\right\|_F^2}, \quad (3)$$

which is invariant to power-scaling $\mathbf{U}_{\ell_k}$. Without loss of generality, we assume $\|\mathbf{U}_{\ell_k}\|_F = 1$. The downlink SINR in (3) has a direct relation to the Shannon rate defined in (2) as

$$R_{\ell_k} \geq \log(d(1+\gamma_{\ell_k}^{(DL)})), \quad (4)$$

whose proof can be found in Appendix A.

2) Backward Direction: In the uplink, the users become the transmitters and the BSs become the receivers. Then, the conjugate uplink signal received at BS ℓ is modeled by

$$\mathbf{y}_{\ell_k}^{(UL)} = \mathbf{H}_{\ell_k,\ell}^*\widetilde{\mathbf{U}}_{\ell_k}\mathbf{s}_{\ell_k}^{(UL)} + \sum_{i_j\neq\ell_k}\mathbf{H}_{i_j,\ell}^*\widetilde{\mathbf{U}}_{i_j}\mathbf{s}_{i_j}^{(UL)} + \mathbf{n}_{\ell_k}^{(UL)},$$

where $\widetilde{\mathbf{U}}_{\ell_k} = \sqrt{\omega_{\ell_k}}\,\mathbf{U}_{\ell_k}$ and $\omega_{\ell_k} \in \mathbb{R}_+$ is the uplink transmit power. Other notations are defined in the same manner as the downlink. Similarly, we can express the uplink SINR of user ℓ_k at BS ℓ as

$$\gamma_{\ell_k}^{(UL)} = \frac{\omega_{\ell_k}\left\|\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*\mathbf{U}_{\ell_k}\right\|_F^2}{\sum_{i_j\neq\ell_k}\omega_{i_j}\left\|\mathbf{V}_{\ell_k}^*\mathbf{H}_{i_j,\ell}^*\mathbf{U}_{i_j}\right\|_F^2 + \sigma^2\left\|\mathbf{V}_{\ell_k}\right\|_F^2}. \quad (5)$$

We assume local CSI is available at each end. This means that user ℓ_k has knowledge of the effective channel matrix $\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}$ and the receive covariance matrix $\sum_{i_j}\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}\widetilde{\mathbf{V}}_{i_j}^*\mathbf{H}_{\ell_k,i}^* + \sigma^2\mathbf{I}_M$. Similarly, the BS ℓ has knowledge of its effective channels $\left\{\mathbf{H}_{\ell_k,\ell}^*\widetilde{\mathbf{U}}_{\ell_k}\right\}_{k=1}^K$ and the receive covariance matrix $\sum_{i_j}\mathbf{H}_{i_j,\ell}^*\widetilde{\mathbf{U}}_{i_j}\widetilde{\mathbf{U}}_{i_j}^*\mathbf{H}_{i_j,\ell} + \sigma^2\mathbf{I}_N$.

We employ distributed FB iteration [7], [13], [14], [28]–[30] to iteratively update the combiners in the downlink and the precoders in the uplink. There are instantaneous and error-free feedback links between a BS and its serving users. Each feedback link transfers a scalar parameter per channel use and will be utilized for the distributed power allocation devised in Section IV.

B. Inverse-Positive Matrix Theory

The inverse-positive matrix theory [32], [33] is a key tool that we will use to characterize the feasibility of the proposed problem. Suppose a real linear system

$$\mathbf{F}\mathbf{x} = \mathbf{b}, \quad (6)$$

where $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$, and $\mathbf{b} \in \mathbb{R}_+^{n\times 1} \cap \mathcal{R}(\mathbf{F})$. We first present some definitions.

Definition 1 (Inverse-positive Matrix): A square matrix $\mathbf{F}$ is said to be inverse-positive if $\mathbf{F}^{-1}$ exists and $\mathbf{F}^{-1} \succeq \mathbf{0}_{n\times n}$. If $\mathbf{F}$ is inverse-positive, the system in (6) has a non-negative solution $\mathbf{x} = \mathbf{F}^{-1}\mathbf{b} \succeq \mathbf{0}_{n\times 1}$.

Definition 2 (Z-matrix): A square matrix $\mathbf{F}$ whose off-diagonal elements are non-positive, i.e., $\mathbf{F}(u,v) \leq 0$ for $u \neq v$, is called a Z-matrix.

Definition 3 (Positive Regular Splitting): If a square matrix $\mathbf{F}$ can be divided into $\mathbf{F} = \mathbf{D} - \mathbf{C}$, where $\mathbf{D} \succeq \mathbf{0}_{n\times n}$, $\mathbf{C} \succeq \mathbf{0}_{n\times n}$, and $\mathbf{D}$ is inverse-positive such that $\mathbf{D}^{-1} \succeq \mathbf{0}_{n\times n}$, the decomposition $\mathbf{F} = \mathbf{D} - \mathbf{C}$ is called positive regular splitting.

It is shown in [33] that if $\mathbf{F}$ is a Z-matrix, there exists a positive regular splitting $\mathbf{F} = \mathbf{D} - \mathbf{C} = (\mathbf{I} - \mathbf{C}\mathbf{D}^{-1})\mathbf{D}$. Because $\mathbf{D}$ is inverse-positive, so is $\mathbf{F}$ if and only if $\mathbf{I} - \mathbf{C}\mathbf{D}^{-1}$ is inverse-positive, which is closely related to the spectral radius of the positive matrix $\mathbf{C}\mathbf{D}^{-1} \succeq \mathbf{0}_{n\times n}$.

Theorem 1 ([33]): Let $\mathbf{F} = \mathbf{D} - \mathbf{C}$ be a positive regular splitting. Then, the following three statements are equivalent: (i) $\mathbf{F}$ is inverse-positive; (ii) $\rho(\mathbf{C}\mathbf{D}^{-1}) < 1$; (iii) there exists a non-negative solution $\mathbf{x} \succeq \mathbf{0}_{n\times 1}$ of (6).

Theorem 1 reveals that if $\mathbf{F}$ is a Z-matrix, showing that $\mathbf{F} = \mathbf{D} - \mathbf{C}$ is an inverse-positive matrix is equivalent to verifying either (ii) or (iii). While a sufficient condition for Theorem 1 has been identified in [40] where its proof relies on stochastic matrix analysis and PF theorem [41], this sufficient condition can be tractably indentified using a simple spectral radius bound as shown below.

Lemma 1 ([40]): Let a square matrix $\mathbf{F}$ be a Z-matrix with positive diagonal elements. Then, if $\mathbf{F}$ is strictly diagonal dominant (SDD), i.e., $|\mathbf{F}(u,u)| > \sum_{v\neq u}|\mathbf{F}(u,v)|, \forall u$, $\mathbf{F}$ is inverse-positive.

Proof: Because $\mathbf{F}$ is a Z-matrix with positive diagonal elements, there exists a positive regular splitting $\mathbf{F} = \mathbf{D} - \mathbf{O}$, where $\mathbf{D} \succeq \mathbf{0}_{n\times n}$ is the diagonal matrix sharing the same diagonal elements as $\mathbf{F}$, and $\mathbf{O} \succeq \mathbf{0}_{n\times n}$ has the zero diagonal

elements and is formed by taking the absolute value of the off-diagonal elements of $\mathbf{F}$. We need to show that if $\mathbf{F}$ is SDD, then $\rho(\mathbf{O}\mathbf{D}^{-1}) < 1$ by Theorem 1. If $\mathbf{F}$ is SDD, i.e., $\mathbf{D}(i,i) > \sum_j \mathbf{O}(i,j)$, it follows $\|\mathbf{D}^{-1}\mathbf{O}\|_\infty < 1$. Then, invoking a bound $\|\mathbf{D}^{-1}\mathbf{O}\|_\infty \geq \rho(\mathbf{D}^{-1}\mathbf{O})$ and the equality $\rho(\mathbf{D}^{-1}\mathbf{O}) = \rho(\mathbf{O}\mathbf{D}^{-1})$ leads to $\rho(\mathbf{O}\mathbf{D}^{-1}) < 1$. This completes the proof. ∎

Lemma 1 is useful because it characterizes an inverse-positive matrix without computing its spectral radius (the condition (ii) in Theorem 1). Leveraging Lemma 1 is the key to obtaining the feasibility condition of the proposed distributed algorithm in Section IV.

C. General Statement of Technique

We let $\xi_{\ell_k} > 0$ be the target downlink SINR value of user ℓ_k. The MCMU MIMO SGPM problem is then formulated as

$$\begin{aligned} \min_{\{\mathbf{V}_{\ell_k}\},\{\mathbf{U}_{\ell_k}\},\{\alpha_{\ell_k}\}} \ & \sum_{\ell_k} \alpha_{\ell_k} \\ \text{subject to} \ & \gamma_{\ell_k}^{(DL)}(\{\alpha_{\ell_k}\},\{\mathbf{V}_{\ell_k}\},\mathbf{U}_{\ell_k}) \geq \xi_{\ell_k}, \ \forall \ell_k, \end{aligned} \tag{7}$$

where $\gamma_{\ell_k}^{(DL)}\left(\{\alpha_{\ell_k}\},\{\mathbf{V}_{\ell_j}\},\mathbf{U}_{\ell_k}\right)$ is the downlink SINR in (3) expressed in terms of $\{\alpha_{\ell_k}\}$, $\{\mathbf{V}_{\ell_k}\}$ and $\mathbf{U}_{\ell_k}$. The problem in (7) jointly optimizes the precoders, combiners, and transmit power values by minimizing the total transmit power budget subject to the downlink SINR constraints.

As mentioned in Section I, UD duality does not hold for the problem in (7). Our goal is to devise a fully distributed and feasible framework that is based on solving the optimality conditions of (7). This will require stringent analysis to characterize the optimality criteria of (7). Moreover, the aimed distributed SGPM must choose the target SINR values $\{\xi_{\ell_k}\}$ relying on local CSI to ensure its feasibility. The challenge is the coupled nature of (7) does not allow for amenable analysis. As we will see in the next sections, the key to address the challenges lies in how to alter the original problem in (7) to get tractable subproblems that admit analysis.

III. Optimality Conditions

In this section, the original problem in (7) is decomposed into forward and backward SGPM subproblems, and the optimality conditions of (7) are obtained, based on the combined analysis of the two subproblems. The forward subproblem is discussed first.

A. Forward SGPM Subproblem

We fix the precoders $\{\mathbf{V}_{\ell_k}\}$ and reformulate the problem in (7) in terms of the combiners $\{\mathbf{U}_{\ell_k}\}$ and the transmit power values $\{\alpha_{\ell_k}\}$, yielding the forward SGPM subproblem:

$$\begin{aligned} \min_{\{\mathbf{U}_{\ell_k}\},\{\alpha_{\ell_k}\}} \ & \sum_{\ell_k} \alpha_{\ell_k} \\ \text{subject to} \ & \gamma_{\ell_k}^{(DL)}\left(\{\alpha_{\ell_k}\},\mathbf{U}_{\ell_k}\right) \geq \xi_{\ell_k}, \ \forall \ell_k. \end{aligned} \tag{8}$$

Even though the precoders $\{\mathbf{V}_{\ell_k}\}$ are fixed, (8) is still non-convex. This is in contrast with the prior work [17]–[23], in which the optimization was with respect to the precoders $\{\mathbf{V}_{\ell_k}\}$ and transmit power values $\{\alpha_{\ell_k}\}$, in which the problem is separable and thus can be optimally solved, for example, by using a SOCP framework.

Though non-convex, strong duality holds for the forward SGPM in (8) and hence the optimality conditions are available, which is formally stated in the following theorem.

Theorem 2: If the primal problem (8) is feasible, the strong duality of (8) holds and the optimal solution obeys the following fixed-point equations:

$$\left(\sum_{i_j \neq \ell_k} \alpha_{i_j} \mathbf{H}_{\ell_k,i} \mathbf{V}_{i_j} \mathbf{V}_{i_j}^* \mathbf{H}_{\ell_k,i}^* + \sigma^2 \mathbf{I}_M - \frac{\alpha_{\ell_k}}{\xi_{\ell_k}} \mathbf{H}_{\ell_k,\ell} \mathbf{V}_{\ell_k} \mathbf{V}_{\ell_k}^* \mathbf{H}_{\ell_k,\ell}^* \right) \mathbf{U}_{\ell_k} = \mathbf{0}_{M\times d}, \forall \ell_k, \tag{9}$$

and

$$\operatorname{tr}\left(\mathbf{U}_{\ell_k}^* \Big(\sum_{i_j \neq \ell_k} \alpha_{i_j} \mathbf{H}_{\ell_k,i} \mathbf{V}_{i_j} \mathbf{V}_{i_j}^* \mathbf{H}_{\ell_k,i}^* + \sigma^2 \mathbf{I}_M - \frac{\alpha_{\ell_k}}{\xi_{\ell_k}} \mathbf{H}_{\ell_k,\ell} \mathbf{V}_{\ell_k} \mathbf{V}_{\ell_k}^* \mathbf{H}_{\ell_k,\ell}^* \Big) \mathbf{U}_{\ell_k} \right) = 0, \forall \ell_k. \tag{10}$$

Proof: See Appendix B.

B. Backward SGPM Subproblem

In the uplink, the backward SGPM subproblem is formulated by fixing $\{\mathbf{U}_{\ell_k}\}$ in the uplink SINR in (5) as

$$\begin{aligned} \min_{\{\mathbf{V}_{\ell_k}\},\{\omega_{\ell_k}\}} \ & \sum_{\ell_k} \omega_{\ell_k} \\ \text{subject to} \ & \gamma_{\ell_k}^{(UL)}\left(\{\omega_{\ell_k}\},\mathbf{V}_{\ell_k}\right) \geq \xi_{\ell_k}, \ \forall \ell_k. \end{aligned} \tag{11}$$

The exact same strong duality as Theorem 2 holds for the backward subproblem in (11). Its optimality conditions are summarized below.

Corollary 1: If the backward SGPM subproblem in (11) is feasible, the strong duality of (11) holds and its optimal solution satisfies the following fixed-point equations:

$$\left(\sum_{i_j \neq \ell_k} \omega_{i_j} \mathbf{H}_{i_j,\ell}^* \mathbf{U}_{i_j} \mathbf{U}_{i_j}^* \mathbf{H}_{i_j,\ell}^* + \sigma^2 \mathbf{I}_N - \frac{\omega_{\ell_k}}{\xi_{\ell_k}} \mathbf{H}_{\ell_k,\ell}^* \mathbf{U}_{\ell_k} \mathbf{U}_{\ell_k}^* \mathbf{H}_{\ell_k,\ell} \right) \mathbf{V}_{\ell_k} = \mathbf{0}_{N\times d}, \ \forall \ell_k \tag{12}$$

and

$$\operatorname{tr}\left(\mathbf{V}_{\ell_k}^* \Big(\sum_{i_j \neq \ell_k} \omega_{i_j} \mathbf{H}_{i_j,\ell}^* \mathbf{U}_{i_j} \mathbf{U}_{i_j}^* \mathbf{H}_{i_j,\ell}^* + \sigma^2 \mathbf{I}_N - \frac{\omega_{\ell_k}}{\xi_{\ell_k}} \mathbf{H}_{\ell_k,\ell}^* \mathbf{U}_{\ell_k} \mathbf{U}_{\ell_k}^* \mathbf{H}_{\ell_k,\ell} \Big) \mathbf{V}_{\ell_k} \right) = 0, \ \forall \ell_k \tag{13}$$

The proof of Corollary 1 follows the exact same procedure as the proof of Theorem 2, so we omit it. The forward and backward subproblems are coupled and related to the original SGPM problem in (7). Regarding their joint optimality to the original problem, we have the following theorem.

Theorem 3: If the SGPM problem in (7) is feasible, the optimal solution of (7) coincides with the joint optimal solution

of the forward and backward SGPM subproblems in (8) and (11), respectively. Solving for the joint optimality conditions (9)–(10) and (12)–(13) leads to an optimal solution to the original SGPM problem in (7).

Proof: See Appendix C.

Theorem 3 reveals that solving the original SGPM problem in (7) is equivalent to solving for the fixed-point conditions in (9)–(10) and (12)–(13). However, closed-form solutions satisfying (9)–(10) and (12)–(13) are difficult to obtain due to unresolvable coupling. Moreover, directly solving (9)–(10) and (12)–(13) demands centralized approaches with global CSI, which is impractical. As previously mentioned, the goal of this work is to develop a fully distributed framework, based on solving the optimality conditions in (9)–(10) and (12)–(13). The coupling among $\{\mathbf{V}_{\ell_k}\}$, $\{\mathbf{U}_{\ell_k}\}$, $\{\alpha_{\ell_k}\}$, $\{w_{\ell_k}\}$ and the local CSI requirement make distributed FB iteration, also known as block coordinate descent (BCD), an ideal approach to iteratively solve for the variables in (9)–(10) and (12)–(13).

IV. Distributed SGPM

Based on the analysis in Section III, we devise in this section a distributed FB SGPM algorithm with a feasibility guarantee. The overall FB iteration structure is introduced first and detailed algorithmic descriptions follow afterward.

A. FB Iteration Structure

We let n be the FB iteration index. At the first FB iteration $n = 1$, the algorithm initializes the parameters $\{\mathbf{V}_{\ell_k}^{(0)}\}$, $\{\mathbf{U}_{\ell_k}^{(0)}\}$, $\{\alpha_{\ell_k}^{(0)}\}$, and $\{\omega_{\ell_k}^{(0)}\}$, and determines the feasible target SINR values $\{\xi_{\ell_k}\}$. At each FB iteration, the algorithm solves for the forward optimality conditions (9) and (10) in the downlink and the backward optimality conditions (12) and (13) in the uplink. Assuming that $\{\mathbf{V}_{\ell_k}^{(n-1)}\}$, $\{\mathbf{U}_{\ell_k}^{(n-1)}\}$, $\{\alpha_{\ell_k}^{(n-1)}\}$, and $\{w_{\ell_k}^{(n-1)}\}$ have been designed at the $(n-1)$th FB iteration, we focus on the nth FB iteration, in which $\{\mathbf{V}_{\ell_k}^{(n)}\}$, $\{\mathbf{U}_{\ell_k}^{(n)}\}$, $\{\alpha_{\ell_k}^{(n)}\}$, and $\{w_{\ell_k}^{(n)}\}$ are optimized by solving the following subproblems:

- (F1) Fix the precoders $\{\mathbf{V}_{\ell_k}^{(n-1)}\}$ and downlink power values $\{\alpha_{\ell_k}^{(n-1)}\}$, users optimize the combiners $\{\mathbf{U}_{\ell_k}^{(n)}\}$ to satisfy (9) in the downlink;
- (P1) Fix the combiners $\{\mathbf{U}_{\ell_k}^{(n)}\}$ and precoders $\{\mathbf{V}_{\ell_k}^{(n-1)}\}$, users optimize the downlink power values $\{\alpha_{\ell_k}^{(n)}\}$ to satisfy (10) in the downlink and feed them back to the BSs;
- (F2) Fix the combiners $\{\mathbf{U}_{\ell_k}^{(n)}\}$ and uplink power values $\{\omega_{\ell_k}^{(n-1)}\}$, BSs optimize the precoders $\{\mathbf{V}_{\ell_k}^{(n)}\}$ to satisfy (12) in the uplink;
- (P2) Fix the precoders $\{\mathbf{V}_{\ell_k}^{(n)}\}$ and combiners $\{\mathbf{U}_{\ell_k}^{(n)}\}$, BSs optimize the uplink power values $\{\omega_{\ell_k}^{(n)}\}$ to satisfy (13) in the uplink and feed then back to the users.

A conceptual diagram of the proposed FB iteration is present in Fig. 1. In what follows, we give detailed algorithmic descriptions for (F1), (P1), (F2), and (P2). We first discuss the combiner design (F1) and downlink power allocation (P1) in the forward iteration.

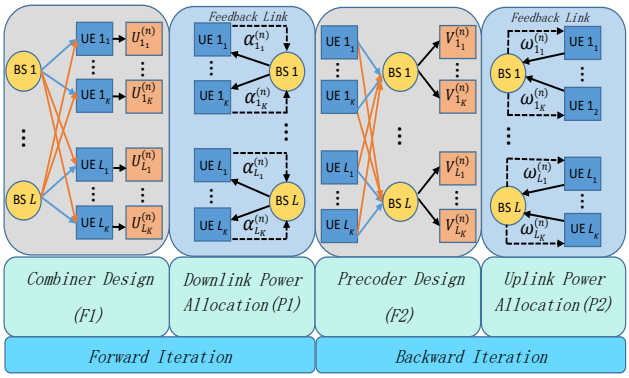

Fig. 1. Block Diagram of the proposed algorithm for the FB iteration.

B. Forward Iteration

1) Combiner Design (F1): The combiner obeying (9) does not admit a unique solution. For instance, if $\mathbf{U}_{\ell_k}^{(n)} \in \mathbb{C}^{N\times d}$ with $\mathrm{rank}(\mathbf{U}_{\ell_k}^{(n)}) = d$ satisfies (9), right multiplication of $\mathbf{U}_{\ell_k}^{(n)}$ with $\mathbf{\Psi} \in \mathbb{C}^{d\times d}$, i.e., $\mathbf{U}_{\ell_k}^{(n)}\mathbf{\Psi}$, where $\mathrm{rank}(\mathbf{\Psi}) \leq d$, also satisfies (9). However, $\mathbf{U}_{\ell_k}^{(n)}\mathbf{\Psi}$ fails to resolve the d distinct streams because $\mathrm{rank}(\mathbf{U}_{\ell_k}^{(n)}\mathbf{\Psi}) \leq d$. To decouple the d distinct streams, we assume full-rank filters. Then, the following lemma provides the solution to (F1).

Lemma 2: Suppose the problem (F1) in the nth FB iteration:

$$\begin{aligned} \text{find} \quad & \mathbf{U}_{\ell_k}^{(n)} \\ \text{subject to} \quad & \left(\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)} - \tfrac{1}{\xi_{\ell_k}}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\right)\mathbf{U}_{\ell_k}^{(n)} = \mathbf{0}_{M\times d}, \\ & \|\mathbf{U}_{\ell_k}^{(n)}\|_F = 1, \text{ and } \mathrm{rank}\left(\mathbf{U}_{\ell_k}^{(n)}\right) = d, \ \forall \ell_k, \end{aligned} \tag{14}$$

where $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)} \triangleq \sum_{i_j\neq\ell_k}\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(n-1)}\left(\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(n-1)}\right)^* + \sigma^2\mathbf{I}_M$ and $\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)} \triangleq \mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\left(\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\right)^*$ are the interference-plus-noise and the desired signal covariance matrices of user ℓ_k, respectively, where $\widetilde{\mathbf{V}}_{i_j}^{(n-1)} = \sqrt{\alpha_{i_j}^{(n-1)}}\mathbf{V}_{i_j}^{(n-1)}$. We let the Cholesky decomposition of $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)}$ be $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)} \triangleq \mathbf{L}^*\mathbf{L}$, where $\mathbf{L} \in \mathbb{C}^{M\times M}$ is an upper triangular matrix. The optimal solution of (14) is then given by

$$\mathbf{U}_{\ell_k}^{(n)\star} = \frac{\mathbf{L}^{-1}\nu_{1:d}\left((\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}\right)}{\left\|\mathbf{L}^{-1}\nu_{1:d}\left((\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}\right)\right\|_F}, \ \forall \ell_k, \tag{15}$$

where $\nu_{1:d}(\mathbf{A})$ extracts the first d dominant eigenvectors of a symmetric matrix $\mathbf{A}$.

Proof: See Appendix D.

Note that the constraint $\|\mathbf{U}_{\ell_k}^{(n)}\|_F = 1$ in (14) makes (15) invariant to ξ_{ℓ_k}. Moreover, the first constraint of (14) is a generalized eigenvalue problem, which can be equivalently formulated as the Rayleigh quotient problem

$$\begin{aligned} \max_{\mathbf{U}_{\ell_k}^{(n)}} \quad & \gamma_{\ell_k}^{(DL)}\left(\{\alpha_{\ell_k}^{(n-1)}\}, \mathbf{U}_{\ell_k}^{(n)}\right) \\ \text{subject to} \quad & \mathbf{U}_{\ell_k}^{(n)*}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{U}_{\ell_k}^{(n)} \text{ diagonal}, \end{aligned} \tag{16}$$

whose solution is also (15) [42, Algorithm 8.7.1]. Due to this equivalence and the SINR maximization criterion in (16), we

conclude that $\mathbf{U}_{\ell_k}^{(n)^\star}$ does not decrease the downlink SINR and

$$\gamma_{\ell_k}^{(DL)}\big(\{\alpha_{\ell_k}^{(n-1)}\},\mathbf{U}_{\ell_k}^{(n-1)^\star}\big)\leq\gamma_{\ell_k}^{(DL)}\big(\{\alpha_{\ell_k}^{(n-1)}\},\mathbf{U}_{\ell_k}^{(n)^\star}\big). \quad (17)$$

The monotonicity in (17) will be useful when characterizing the feasibility of the power allocation.

2) Downlink Power Allocation (P1): To facilitate a distributed implementation of (P1), we propose the following power update structure

$$\alpha_{\ell_k}^{(n)}=\delta_{\ell_k}^{(n)}\alpha_{\ell_k}^{(n-1)}, \quad (18)$$

where $\delta_{\ell_k}^{(n)}\in\mathbb{R}_+$ is a power scaling factor to be optimized to satisfy (10). For notational simplicity, we define the downlink precoding and combining factor as

$$\beta_{i_j,(n-1)}^{\ell_k,(n)}\triangleq\big\|\big(\mathbf{U}_{\ell_k}^{(n)}\big)^*\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(n-1)}\big\|_F^2. \quad (19)$$

Then (10) can be rewritten as

$$\frac{\delta_{\ell_k}^{(n)}\beta_{\ell_k,(n-1)}^{\ell_k,(n)}}{\sum_{i_j\neq\ell_k}\delta_{i_j}^{(n)}\beta_{i_j,(n-1)}^{\ell_k,(n)}+\sigma^2}=\xi_{\ell_k},\ \forall\ell_k. \quad (20)$$

Collecting the LK equations in (20) yields

$$\mathbf{A}_{(n-1)}^{(n)}\mathbf{a}^{(n)}=\mathbf{q}, \quad (21)$$

where $\mathbf{q}=\sigma^2[\xi_{1_1},\xi_{1_2},\ldots,\xi_{L_K}]^T\in\mathbb{R}^{LK\times1}$, $\mathbf{a}^{(n)}=[\delta_{1_1}^{(n)},\delta_{1_2}^{(n)},\ldots\delta_{L_K}^{(n)}]^T\in\mathbb{R}^{LK\times1}$, and

$$\mathbf{A}_{(n-1)}^{(n)}=\begin{pmatrix}\beta_{1_1,(n-1)}^{1_1,(n)} & -\beta_{1_2,(n-1)}^{1_1,(n)}\xi_{1_1} & \cdots & -\beta_{L_K,(n-1)}^{1_1,(n)}\xi_{1_1}\\ -\beta_{1_1,(n-1)}^{1_2,(n)}\xi_{1_2} & \beta_{1_2,(n-1)}^{1_2,(n)} & \cdots & -\beta_{L_K,(n-1)}^{1_2,(n)}\xi_{1_2}\\ \vdots & \ddots & \vdots & \cdots\\ -\beta_{1_1,(n-1)}^{L_K,(n)}\xi_{L_K} & -\beta_{1_2,(n-1)}^{L_K,(n)}\xi_{L_K} & \cdots & \beta_{L_K,(n-1)}^{L_K,(n)}\end{pmatrix}. \quad (22)$$

(P1) is now formulated as

$$\begin{aligned}&\text{find} && \mathbf{a}^{(n)}\\ &\text{subject to} && \mathbf{A}_{(n-1)}^{(n)}\mathbf{a}^{(n)}=\mathbf{q}\ \text{and}\ \mathbf{a}^{(n)}\succeq\mathbf{0}_{LK\times1}.\end{aligned} \quad (23)$$

3) Feasibility Condition of (23)*:* The prerequisite of solving (23) is to guarantee the feasibility, i.e., showing that the set

$$\mathcal{R}_{\mathbf{a}^{(n)}}\triangleq\{\mathbf{a}^{(n)}\succeq\mathbf{0}_{LK\times1}|\mathbf{A}_{(n-1)}^{(n)}\mathbf{a}^{(n)}=\mathbf{q}\} \quad (24)$$

is non-empty. The condition ensuring the feasibility is provided in the following theorem.

Theorem 4: For fixed $\{\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\}$ and $\{\mathbf{U}_{\ell_k}^{(n)}\}$, if the target SINR ξ_{ℓ_k} satisfies

$$0<\xi_{\ell_k}<\beta_{\ell_k,(n-1)}^{\ell_k,(n)}\Big/\sum_{i_j\neq\ell_k}\beta_{i_j,(n-1)}^{\ell_k,(n)},\ \forall\ell_k, \quad (25)$$

the (P1) in (23) is feasible.

Proof: By the definition of $\mathbf{A}_{(n-1)}^{(n)}$ in (22), $\mathbf{A}_{(n-1)}^{(n)}$ is a Z-matrix with the positive diagonal entries. Note that the inequalities in (25) correspond to $\mathbf{A}_{(n-1)}^{(n)}$ being SDD. Then, $\mathbf{A}_{(n-1)}^{(n)}$ is inverse-positive by Lemma 1. It is concluded from Condition (iii) of Theorem 1 that $\mathcal{R}_{\mathbf{a}^{(n)}}$ in (24) is non-empty. This completes the proof. ■

To solve (23) in a distributed manner, we propose a Jacobi-based power allocation algorithm.

Jacobi-Based Power Allocation: Jacobi recursion [31] refers to an iterative procedure to solve a linear system equation without matrix inversion. We let $\mathbf{D}_{(n-1)}^{(n)}=\mathrm{diag}\left(\beta_{1_1,(n-1)}^{1_1,(n)},\cdots,\beta_{L_K,(n-1)}^{L_K,(n)}\right)\in\mathbb{R}^{LK\times LK}$ be the diagonal matrix whose entries are taken from $\mathbf{A}_{(n-1)}^{(n)}$. Then, (21) can be rewritten as $\mathbf{D}_{(n-1)}^{(n)}\mathbf{a}^{(n)}=(\mathbf{D}_{(n-1)}^{(n)}-\mathbf{A}_{(n-1)}^{(n)})\mathbf{a}^{(n)}+\mathbf{q}$. Multiplying both sides with $\left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}$ leads to the Jacobi recursion formula,

$$\mathbf{a}_{(m)}^{(n)}=\Big(\mathbf{I}_{LK}-\Big(\mathbf{D}_{(n-1)}^{(n)}\Big)^{-1}\mathbf{A}_{(n-1)}^{(n)}\Big)\mathbf{a}_{(m-1)}^{(n)}+\Big(\mathbf{D}_{(n-1)}^{(n)}\Big)^{-1}\mathbf{q}, \quad (26)$$

where $m\geq1$ is the Jacobi-iteration index. If $\big\|\mathbf{I}_{LK}-\left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\big\|_\infty<1$, for an initial $\mathbf{a}_{(0)}^{(n)}$, any sequence $\mathbf{a}_{(0)}^{(n)},\mathbf{a}_{(1)}^{(n)},\ldots$ produced by (26) converges to the solution of $\mathbf{A}_{(n-1)}^{(n)}\mathbf{a}^{(n)}=\mathbf{q}$ [31].

Regarding the convergence of the Jacobi recursion in (26) to the solution of (23), we have the following corollary.

Corollary 2: If the condition in (25) holds, the Jacobi recursion in (26) converges to the solution of (23).

Proof: $\mathbf{A}_{(n-1)}^{(n)}$ is SDD due to (25). Therefore, the Jacobi recursion converges to the solution of (23) because the inequality $\big\|\mathbf{I}_{LK}-\left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\big\|_\infty<1$ holds. ■

By Corollary 2, the achievable SINR region of the downlink Jaboci power allocation is given by

$$\mathcal{S}_{\ell_k}^{(n)}=\left\{\xi_{\ell_k}\middle|\forall\xi_{\ell_k}\ \text{satisfying (25)}\right\},\ \forall\ell_k. \quad (27)$$

Remark 1: Due to the monotonicity of the downlink SINR, i.e., (17), the nested property of the achievable SINR region implies that $\mathcal{S}_{\ell_k}^{(n-1)}\subseteq\mathcal{S}_{\ell_k}^{(n)}$ for $n>1$, holds. This means that if $\xi_{\ell_k}\in\mathcal{S}_{\ell_k}^{(1)}$, the forward direction becomes feasible. Hence, characterizing the achievable SINR region $\mathcal{S}_{\ell_k}^{(1)}$ at the first FB iteration is sufficient to guarantee the feasibility of the overall algorithm.

In the proposed approach, the Jacobi recursion in (26) is run in a distributed manner using feedback as detailed below. When $m=1$, BSs initialize $\delta_{\ell_k,(0)}^{(n)}=1$ in (18), $\forall\ell_k$. Then user ℓ_k measures its local CSI, i.e., $\beta_{\ell_k,(n-1)}^{\ell_k,(n)}$ and $\sum_{i_j\neq\ell_k}\beta_{i_j,(n-1)}^{\ell_k,(n)}$. If $n=1$, user ℓ_k chooses $\xi_{\ell_k}\in\mathcal{S}_{\ell_k}^{(1)}$. The power update formula at user ℓ_k is then given by the $((\ell-1)K+k)$th row of (26),

$$\delta_{\ell_k,(m)}^{(n)}=\xi_{\ell_k}\Big(\sum_{i_j\neq\ell_k}\delta_{i_j,(m-1)}^{(n)}\beta_{i_j,(n-1)}^{\ell_k,(n)}+\sigma^2\Big)/\beta_{\ell_k,(n-1)}^{\ell_k,(n)}. \quad (28)$$

After the update, user ℓ_k feeds back $\delta_{\ell_k,(m)}^{(n)}$ to BS ℓ. At the $(m+1)$th iteration, BSs form the precoders $\{\sqrt{\delta_{\ell_k,(m)}^{(n)}}\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\}$ using the fed back $\{\delta_{\ell_k,(m)}^{(n)}\}$. On the downlink reception, user ℓ_k updates its power value by (28) and feeds back $\delta_{\ell_k,(m+1)}^{(n)}$ to BS ℓ. This closed-loop recursion continues until (28) converges. Denoting the converged power scaling factor as $\delta_{\ell_k}^{(n)}$, the downlink power $\alpha_{\ell_k}^{(n)}$ update follows (18). A formal description of the closed-loop Jacobi algorithm is outlined in Algorithm 1.

Algorithm 1 Closed-loop Jacobi iteration for downlink power allocation (P1)

Require: Initialize $\{\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\}$, $\{\mathbf{U}_{\ell_k}^{(n)}\}$, $\{\delta_{\ell_k,(0)}^{(n)}=1\}$, $m=1$.

1: Begin iteration
2: BSs transmit in the downlink with the precoders $\{\sqrt{\delta_{\ell_k,(m-1)}^{(n)}}\widetilde{\mathbf{V}}_{\ell_k}^{(n-1)}\}$;
3: **if** $n=1$ and $m=1$ **then** User ℓ_k sets $\xi_{\ell_k}\in\mathcal{S}_{\ell_k}^{(1)}$ in (27);
4: **end if**
5: User ℓ_k updates $\delta_{\ell_k,(m)}^{(n)}$ by (28) and feeds back it to BS ℓ;
6: $m=m+1$ and go to Step 2;
7: Repeat until converge, set $\delta_{\ell_k}^{(n)}=\delta_{\ell_k,(m)}^{(n)}$.
8: Set the power $\alpha_{\ell_k}^{(n)}$ at BS ℓ by (18).

Algorithm 2 Closed-loop Jacobi iteration for uplink power allocation (P2)

Require: Initialize $\{\mathbf{U}_{\ell_k}^{(n)}\}$, $\{\widetilde{\mathbf{V}}_{\ell_k}^{(n)}\}$, $\{\varrho_{\ell_k,(0)}^{(n)}=1\}$, $m=1$.

1: Begin iteration
2: **if** $n=1$ **then** User ℓ_k feeds back ξ_{ℓ_k} to BS ℓ;
3: **end if**
4: Users transmit in the uplink with $\{\sqrt{\varrho_{\ell_k,(m-1)}^{(n)}}\widetilde{\mathbf{U}}_{\ell_k}^{(n)}\}$;
5: BS ℓ updates $\varrho_{\ell_k,(m)}^{(n)}$ by (35) and feeds back it to users ℓ_k;
6: $m=m+1$ and go to Step 4;
7: Repeat until converge and set $\varrho_{\ell_k}^{(n)}=\varrho_{\ell_k,(m)}^{(n)}$;
8: Set the power $\omega_{\ell_k}^{(n)}$ at user ℓ_k by (32).

C. Backward Iteration

The exact same procedures for solving (F1) and (P1) apply to (F2) and (P2). Hence, we briefly describe their algorithmic procedures.

1) Precoder Design (F2): We define the desired signal and interference-plus-noise covariance matrices as $\overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}\triangleq\mathbf{H}_{\ell_k,\ell}^{*}\widetilde{\mathbf{U}}_{\ell_k}^{(n)}\widetilde{\mathbf{U}}_{\ell_k}^{(n)^*}\mathbf{H}_{\ell_k,\ell}$ and $\overleftarrow{\mathbf{Q}}_{\ell_k}^{(n)}\triangleq\sum_{i_j\neq\ell_k}\mathbf{H}_{i_j,\ell}^{*}\widetilde{\mathbf{U}}_{i_j}^{(n)}\left(\mathbf{H}_{i_j,\ell}^{*}\widetilde{\mathbf{U}}_{i_j}^{(n)}\right)^{*}+\sigma^2\mathbf{I}_N$, respectively, at the nth backward iteration, where $\widetilde{\mathbf{U}}_{\ell_k}^{(n)}=\sqrt{\omega_{\ell_k}^{(n-1)}}\mathbf{U}_{\ell_k}^{(n)}$. Then, the problem (F2) is reformulated as

$$\begin{aligned}\text{find}\quad & \mathbf{V}_{\ell_k}^{(n)}\\ \text{subject to}\quad & \left(\overleftarrow{\mathbf{Q}}_{\ell_k}^{(n)}-\tfrac{1}{\xi_{\ell_k}}\overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}\right)\mathbf{V}_{\ell_k}^{(n)}=\mathbf{0}_{N\times d},\\ & \|\mathbf{V}_{\ell_k}^{(n)}\|_F=1,\ \text{and}\ \operatorname{rank}\left(\mathbf{V}_{\ell_k}^{(n)}\right)=d,\ \forall\ell_k.\end{aligned}\tag{29}$$

Given the Cholesky decomposition $\overleftarrow{\mathbf{Q}}_{\ell_k}^{(n)}=\mathbf{M}^{*}\mathbf{M}$, where $\mathbf{M}\in\mathbb{C}^{N\times N}$ is an upper triangular matrix, the solution to (29) is given by

$$\mathbf{V}_{\ell_k}^{(n)^\star}=\frac{\mathbf{M}^{-1}\nu_{1:d}\left((\mathbf{M}^{*})^{-1}\overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{M}^{-1}\right)}{\left\|\mathbf{M}^{-1}\nu_{1:d}\left((\mathbf{M}^{*})^{-1}\overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{M}^{-1}\right)\right\|_F},\ \forall\ell_k.\tag{30}$$

Following the same reasoning as (17), the precoder in (30) does not decrease the uplink SINR values,

$$\gamma_{\ell_k}^{(UL)}\left(\{\omega_{\ell_k}^{(n-1)}\},\mathbf{V}_{\ell_k}^{(n-1)^\star}\right)\leq\gamma_{\ell_k}^{(UL)}\left(\{\omega_{\ell_k}^{(n-1)}\},\mathbf{V}_{\ell_k}^{(n)^\star}\right).\tag{31}$$

2) Uplink Power Allocation (P2): Similar to (P1), we employ the following power update structure

$$\omega_{\ell_k}^{(n)}=\varrho_{\ell_k}^{(n)}\omega_{\ell_k}^{(n-1)},\tag{32}$$

where $\varrho_{\ell_k}^{(n)}\in\mathbb{R}_{+}$ is a power scaling factor to be designed. The uplink precoding and combining factor is defined as $\kappa_{i_j,(n)}^{\ell_k,(n)}\triangleq\left\|\left(\mathbf{V}_{\ell_k}^{(n)}\right)^{*}\mathbf{H}_{i_j,\ell}^{*}\widetilde{\mathbf{U}}_{i_j}^{(n)}\right\|_F^2$, which is related to (19) as $\kappa_{i_j,(n)}^{\ell_k,(n)}=\beta_{\ell_k,(n)}^{i_j,(n)}\omega_{i_j}^{(n-1)}/\alpha_{\ell_k}^{(n)}$. Then, the LK conditions in (13) are compactly put into a linear system:

$$\mathbf{B}_{(n)}^{(n)}\mathbf{b}^{(n)}=\mathbf{q},\tag{33}$$

where $\mathbf{q}=\sigma^2[\xi_{1_1},\xi_{1_2},\ldots,\xi_{L_K}]^T\in\mathbb{R}^{LK\times1}$, $\mathbf{b}^{(n)}=[\varrho_{1_1}^{(n)},\varrho_{1_2}^{(n)},\ldots,\varrho_{L_K}^{(n)}]^T\in\mathbb{R}^{LK\times1}$, and

$$\mathbf{B}_{(n)}^{(n)}=\begin{pmatrix}\kappa_{1_1,(n)}^{1_1,(n)} & -\kappa_{1_2,(n)}^{1_1,(n)}\xi_{1_1} & \cdots & -\kappa_{L_K,(n)}^{1_1,(n)}\xi_{1_1}\\ -\kappa_{1_1,(n)}^{1_2,(n)}\xi_{1_2} & \kappa_{1_2,(n)}^{1_2,(n)} & \cdots & -\kappa_{L_K,(n)}^{1_2,(n)}\xi_{1_2}\\ \vdots & \ddots & \vdots & \cdots\\ -\kappa_{1_1,(n)}^{L_K,(n)}\xi_{L_K} & -\kappa_{1_2,(n)}^{L_K,(n)}\xi_{L_K} & \cdots & \kappa_{L_K,(n)}^{L_K,(n)}\end{pmatrix},$$

and (P2) is formulated as

$$\begin{aligned}\text{find}\quad & \mathbf{b}^{(n)}\\ \text{subject to}\quad & \mathbf{B}_{(n)}^{(n)}\mathbf{b}^{(n)}=\mathbf{q}\ \text{and}\ \mathbf{b}^{(n)}\succeq\mathbf{0}_{LK\times1}.\end{aligned}\tag{34}$$

3) Feasibility Condition of (34)*:* The feasibility of (P2) in (34) is described by the following corollary.

Corollary 3: If the target SINR values $\{\xi_{\ell_k}\}$ satisfy (25), (P2) in (34) is also feasible, i.e., $\mathcal{R}_{\mathbf{b}^{(n)}}\triangleq\{\mathbf{b}^{(n)}\succeq\mathbf{0}_{LK\times1}|\mathbf{B}_{(n)}^{(n)}\mathbf{b}^{(n)}=\mathbf{q}\}$ is non-empty.

Proof: Prior to the precoder update (F2) in the backward iteration, the following equality holds $\kappa_{i_j,(n)}^{\ell_k,(n-1)}=\beta_{\ell_k,(n-1)}^{i_j,(n)}\omega_{i_j}^{(n-1)}/\alpha_{\ell_k}^{(n-1)}$, which can be put into a matrix equality $\mathbf{B}_{(n)}^{(n-1)}=(\mathbf{D}_1\mathbf{D}_2\mathbf{A}_{(n-1)}^{(n)}\mathbf{D}_3\mathbf{D}_4)^T$, where $\mathbf{D}_1,\mathbf{D}_2,\mathbf{D}_3$, and $\mathbf{D}_4\in\mathbb{R}^{LK\times LK}$ are diagonal matrices with the $((\ell-1)K+k)$th entry being $\omega_{\ell_k}^{(n-1)}$, $1/\xi_{\ell_k}$, ξ_{ℓ_k}, and $1/\alpha_{\ell_k}^{(n-1)}$, respectively. Due to Theorem 4, $\mathbf{A}_{(n-1)}^{(n)}$ is SDD. Thus, it is inverse positive. Moreover, because $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$, and $\mathbf{D}_4$ are positive diagonal matrices, there exists a non-negative solution $\mathbf{b}^{(n)}$ such that $\mathbf{B}_{(n)}^{(n-1)}\mathbf{b}^{(n)}=\mathbf{q}$ is feasible for the target SINR values $\{\xi_{\ell_k}\}$. Optimizing the filters in (F2) updates $\mathbf{B}_{(n)}^{(n-1)}$ to $\mathbf{B}_{(n)}^{(n)}$. Then, by the nested property of the achievable SINR region in (31), updating the filters $\{\mathbf{V}_{\ell_k}^{(n)}\}$ does not alter the feasibility and, thereby, $\mathbf{B}_{(n)}^{(n)}\mathbf{b}^{(n)}=\mathbf{q}$ has a non-negative solution. This completes the proof. ■

The Jacobi recursion to solve (34) for user ℓ_k at BS ℓ is given by

$$\varrho_{\ell_k,(m)}^{(n)}=\xi_{\ell_k}\left(\sum_{i_j\neq\ell_k}\varrho_{i_j,(m-1)}^{(n)}\kappa_{i_j,(n)}^{\ell_k,(n)}+\sigma^2\right)/\kappa_{\ell_k,(n)}^{\ell_k,(n)}.\tag{35}$$

TABLE I
SUMMARY OF INFORMATION AND VARIABLES NEEDED AT EACH SUBPROBLEM AND EACH FB ITERATION.

Subproblems	(F1)	(P1)	(F2)	(P2)
Nodes	User ℓ_k	User ℓ_k	BS ℓ	BS ℓ
Required information	$\vec{\mathbf{R}}_{\ell_k}^{(n)}, \vec{\mathbf{Q}}_{\ell_k}^{(n)}$	$\mathbf{U}_{\ell_k}^{(n)}, \vec{\mathbf{R}}_{\ell_k}^{(n)}, \vec{\mathbf{Q}}_{\ell_k}^{(n)}$	$\overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}, \overleftarrow{\mathbf{Q}}_{\ell_k}^{(n)}$	$\mathbf{V}_{\ell_k}^{(n)}, \overleftarrow{\mathbf{R}}_{\ell_k}^{(n)}, \overleftarrow{\mathbf{Q}}_{\ell_k}^{(n)}$
Update	$\mathbf{U}_{\ell_k}^{(n-1)} \longmapsto \mathbf{U}_{\ell_k}^{(n)}$	$\alpha_{\ell_k}^{(n-1)} \longmapsto \alpha_{\ell_k}^{(n)}$	$\mathbf{V}_{\ell_k}^{(n-1)} \longmapsto \mathbf{V}_{\ell_k}^{(n)}$	$\omega_{\ell_k}^{(n-1)} \longmapsto \omega_{\ell_k}^{(n)}$
Feedback	None	User $\ell_k \overset{\delta_{\ell_k,(m)}^{(n)}}{\longrightarrow}$ BS ℓ	None	If $n = m = 1$, User $\ell_k \overset{\xi_{\ell_k}}{\longrightarrow}$ BS ℓ BS $\ell \overset{\varrho_{\ell_k,(m)}^{(n)}}{\longrightarrow}$ user ℓ_k

After the update in (35), BS ℓ feeds back $\varrho_{\ell_k,(m)}^{(n)}$ to user ℓ_k. By Corollary 3, (35) converges to a feasible solution after a suitable number of the closed-loop iterations. The uplink Jacobi iteration is outlined in Algorithm 2.

Remark 2: If the target SINR values $\{\xi_{\ell_k}\}$ are chosen to ensure the feasibility of the forward SGPM, it is now a direct consequence of Corollary 3 that $\{\xi_{\ell_k}\}$ also ensures the feasibility of the backward SGPM. For this reason, the users and BSs share the same $\{\xi_{\ell_k}\}$ via feedback in Step 2 of Algorithm 2.

The entire operation of the proposed distributed FB SGPM is now outlined in Algorithm 3. What information is needed and what variables are updated at each iteration are summarized in Table I.

Algorithm 3 Proposed Distributed FB SGPM

Require: Initialize $\{\mathbf{V}_{\ell_k}^{(0)}\}$, $\{\mathbf{U}_{\ell_k}^{(0)}\}$, $\{\alpha_{\ell_k}^{(0)}\}$, and $\{\omega_{\ell_k}^{(0)}\}$.

1: Begin iteration, $n = 1$
2: Users update $\{\mathbf{U}_{\ell_k}^{(n)}\}$ by (15) and set $\{\widetilde{\mathbf{U}}_{\ell_k}^{n)} = \sqrt{\omega_{\ell_k}^{(n-1)}}\mathbf{U}_{\ell_k}^{(n)}\}$ in the forward iteration;
3: Users update $\{\alpha_{\ell_k}^{(n)}\}$ according to Algorithm 1 and feed back them to BSs;
4: BSs update $\{\mathbf{V}_{\ell_k}^{(n)}\}$ by (30) and set $\{\widetilde{\mathbf{V}}_{\ell_k}^{(n)} = \sqrt{\alpha_{\ell_k}^{(n)}}\mathbf{V}_{\ell_k}^{(n)}\}$ in the backward iteration;
5: BSs update $\{\omega_{\ell_k}^{(n)}\}$ according to Algorithm 2 and feed back them to users;
6: $n = n + 1$ and go to Step 2;
7: Repeat until converge.

V. CONVERGENCE ANALYSIS

In general, distributed FB iteration with power allocation does not admit tractable convergence analysis [13]. Nevertheless, this is not the case of the proposed distributed algorithm.

Theorem 5: The FB iteration in Algorithm 3 does not increase the total transmit power

$$\sum_{\ell_k} \alpha_{\ell_k}^{(n-1)} \geq \sum_{\ell_k} \alpha_{\ell_k}^{(n)}, \ \forall n > 1,$$

and it converges to a stationary point.

Proof: See Appendix E.

Theorem 5 guarantees the convergence of $\sum_{\ell_k} \alpha_{\ell_k}^{(n)}$ to a stationary point. The stationary point is indeed a local optimum because we impose the full-rank precoders and combines in (30) and (15), respectively.

In this section, we further analyze what factors affect the convergence speed of the closed-loop Jacobi power allocation in Algorithms 1 and 2. We focus on the forward direction, keeping in mind that the exact same observation can be made for the backward direction. It is shown in [31] that the error $\|\mathbf{a}^{(n)} - \mathbf{a}_{(m)}^{(n)}\|_\infty$ of the Jacobi recursion in (26) is bounded by

$$\left\|\mathbf{a}^{(n)} - \mathbf{a}_{(m)}^{(n)}\right\|_\infty \leq \left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty^m. \quad (36)$$

The smaller the value $\left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty < 1$, the faster the convergence will be. Due to Remark 1, the feasibility in Corollary 2 implies $\left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty < 1$. Recalling the structure of $\mathbf{A}_{(n-1)}^{(n)}$ in (21), the following two observations can be made:

(a) The quantity $\left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty$ in (36) is proportional to the target SINR values $\{\xi_{\ell_k}\}$, i.e., for fixed $\{\beta_{ij,(n-1)}^{\ell_k,(n)}\}$ in $\mathbf{A}_{(n-1)}^{(n)}$, if $\xi_{\ell_k} \geq \xi'_{\ell_k}$, we have $g\big(\{\xi_{\ell_k}\}, \{\beta_{ij,(n-1)}^{\ell_k,(n)}\}\big) \geq g\big(\{\xi'_{\ell_k}\}, \{\beta_{ij,(n-1)}^{\ell_k,(n)}\}\big)$, $\forall \ell_k$, where $g\big(\{\xi_{\ell_k}\}, \{\beta_{ij,(n-1)}^{\ell_k,(n)}\}\big) \triangleq \left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty$.

(b) The matrix $\left(\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right)$ has diagonal entries of zero and its off-diagonal entries are $\xi_{\ell_k}\frac{\beta_{ij,(n-1)}^{\ell_k,(n)}}{\beta_{\ell_k,(n-1)}^{\ell_k,(n)}}$, $\forall \ell_k, i_j$, where the numerator $\beta_{ij,(n-1)}^{\ell_k,(n)} = \|(\mathbf{U}_{\ell_k}^{(n)})^* \mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(n-1)}\|_F^2$, $i_j \neq \ell_k$, corresponds to the interference power at user ℓ_k leaked from BS i. Hence, the quantity $\left\|\mathbf{I}_{LK} - \left(\mathbf{D}_{(n-1)}^{(n)}\right)^{-1}\mathbf{A}_{(n-1)}^{(n)}\right\|_\infty$ is proportional to the interference power.

The observations (a) and (b) reveal that the lower the target SINR values $\{\xi_{\ell_k}\}$ are and the greater the number of FB iterations to monotonically increase the SINR values in (17), a faster convergence of the Jacobi power allocation results.

VI. PRACTICAL ASPECTS

In this section, we underline some practical aspects of the proposed algorithm.

A. Distributed CSI Acquisition

To acquire the local CSI at each FB iteration, a channel estimation phase prior to the execution of the

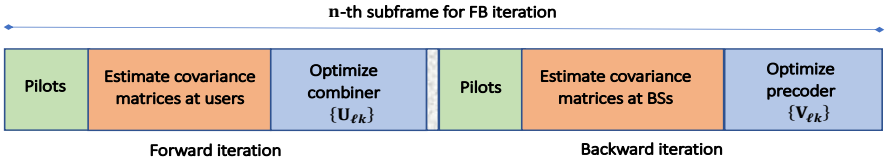

Fig. 2. The structure of FB CSI acquisition.

algorithm is needed. Specifically, in the forward iteration, user ℓ_k first estimates the signal covariance matrix $\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}\widetilde{\mathbf{V}}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*$ and the interference-plus-noise covariance matrix $\sum_{i_j\neq\ell_k}\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}\widetilde{\mathbf{V}}_{i_j}^*\mathbf{H}_{\ell_k,i}^*+\sigma^2\mathbf{I}_M$; in the backward iteration, the BS ℓ estimates its signal covariance matrix $\left\{\mathbf{H}_{\ell_k,\ell}^*\widetilde{\mathbf{U}}_{\ell_k}\widetilde{\mathbf{U}}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}\right\}_{k=1}^K$ and the interference-plus-noise covariance matrix $\sum_{i_j\neq\ell_k}\mathbf{H}_{i_j,\ell}^*\widetilde{\mathbf{U}}_{i_j}\widetilde{\mathbf{U}}_{i_j}^*\mathbf{H}_{i_j,\ell}+\sigma^2\mathbf{I}_N$. A practical CSI acquisition method based on orthogonal pilot-assisted channel estimation technique has been independently studied in [30]. The user-wise orthogonal pilot signals in [30] is also called "User-specific reference signal" in 4G Long-Term Evolution (LTE) standards [43], [44]. The structure of the FB CSI acquisition and filter updates is illustrated in Fig. 2. According to Algorithm 3, the FB iterations are carried out until the algorithm converges. When implemented in practice, however, the total number of iterations is limited by the channel coherence time. While low-overhead and robust channel estimation issues have to be considered as well, such matters are outside the scope of our current work.

B. Rank Adaptation

In severely interference-limited scenarios (i.e., improper network), a rank adaptation (RA) mechanism [14], [45] can be introduced to enhance the performance of the proposed algorithm by turning down some spatial data streams. A simple RA scheme in the downlink can be designed by taking the optimality condition in (9) and the filter design criterion in (14) such that

$$\begin{aligned}\max_{r_{\ell_k}}\quad & r_{\ell_k}\\ \text{subject to}\quad & \left(\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)}-\tfrac{1}{\xi_{\ell_k}}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\right)\mathbf{U}_{\ell_k}^{(n)}=\mathbf{0}_{M\times d},\\ & \|\mathbf{U}_{\ell_k}^{(n)}\|_F=1,\ \ \mathrm{rank}\left(\mathbf{U}_{\ell_k}\right)=r_{\ell_k},\\ & \text{and } r_{\ell_k}\leq d,\ \forall\ell_k.\end{aligned}\tag{37}$$

Because $\mathbf{U}_{\ell_k}^{(n)}$ is extracted from the null space of $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)}-\frac{1}{\xi_{\ell_k}}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}$ according to (62), i.e., $\mathbf{L}^*\Big(\mathbf{I}_M-\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}\Big)\mathbf{L}\mathbf{U}_{\ell_k}^{(n)}=\mathbf{0}_{M\times d}$ in (62), the problem in (37) is equivalent to finding the first r_{ℓ_k} largest eigenvalues of $((\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}\mathbf{L}^{-1})$. By turning off some eigenmodes of $((\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}\mathbf{L}^{-1})$, whose eigenvalues are below a tolerance level $c>0$, the r_{ℓ_k} at user ℓ_k can be determined by

$$r_{\ell_k}^\star=\mathrm{Card}\left(\{i|\lambda_i((\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}\mathbf{L}^{-1})>c\}\right),\tag{38}$$

where $\mathrm{Card}(\mathcal{S})$ denotes the cardinality of set $\mathcal{S}$ and c is a threshold value. Similarly, in the backward direction, each BS adjusts the rank of the precoders according to

$$z_{\ell_k}^\star=\mathrm{Card}\left(\{i|\lambda_i((\mathbf{M}^*)^{-1}\overleftarrow{\mathbf{R}}_{\ell_k}\mathbf{M}^{-1})>c\}\right).\tag{39}$$

The proposed Algorithm 3 can easily adopt the above RA approaches by adding: i) one procedure before Step 2, where each user finds the optimal $r_{\ell_k}^\star$ by (38) and feeds back it to its serving BS, and ii) another procedure before Step 4, where each BS finds the optimal $z_{\ell_k}^\star$ according to (39) and chooses the smallest value between $r_{\ell_k}^\star$ and $z_{\ell_k}^\star$. BSs then feed back the adapted rank values to corresponding users. These procedures ensure the rank synchronization between the users and BSs.

However, it should be mentioned that the latter RA scheme demands additional coordination overhead for the rank synchronization. In addition, the threshold value c in (38) and (39) must be pre-determined so that the devised RA ensures a performance improvement, especially, when the network is improper. We will provide performance insight into the above RA scheme in Section VII.

C. QoS Adaptation

As aforementioned in Remark 1 and Remark 2, the feasible QoS chosen at the first iteration guarantees the feasibility of the entire algorithm. Another important property revealed by Remark 1 is that the target QoS value can be also increased along with the FB iterations. This makes sense because if we let the target SINR at the $(n-1)$th FB iteration be $\xi_{\ell_k}^{(n-1)}$ and $\xi_{\ell_k}^{(n-1)}\in\mathcal{S}_{\ell_k}^{(n-1)}$, the nested property $\mathcal{S}_{\ell_k}^{(n-1)}\subseteq\mathcal{S}_{\ell_k}^{(n)}$ for $n>1$, implies that there exists $\xi_{\ell_k}^{(n)}$ such that $\xi_{\ell_k}^{(n)}\in\mathcal{S}_{\ell_k}^{(n)}$ and $\xi_{\ell_k}^{(n-1)}\leq\xi_{\ell_k}^{(n)}$ for $n>1$. Thus, it is desired that $\xi_{\ell_k}^{(n)}$ chosen close to the upper boundary of $\mathcal{S}_{\ell_k}^{(n)}$ achieves a high feasible QoS level. However, $\mathcal{S}_{\ell_k}^{(n)}$ in (27) is an open set and one cannot explicitly define the exact boundary. One simply approach is to find a strict lower bound of the upper bound of (27). Because σ^2 is strictly positive, one can set

$$\xi_{\ell_k}^{(n)}=\beta_{\ell_k,(n-1)}^{\ell_k,(n)}\Big/\left(\sum_{i_j\neq\ell_k}\beta_{i_j,(n-1)}^{\ell_k,(n)}+\sigma^2\right),\tag{40}$$

which satisfies $\xi_{\ell_k}^{(n)}\in\mathcal{S}_{\ell_k}^{(n)}$.

One requirement of the QoS adaptation (QA) is that each BS and user need to share the updated $\xi_{\ell_k}^{(n)}$ per iteration, resulting in large feedback overhead. From the distributed algorithm perspective, this can be relaxed by updating the target SINR

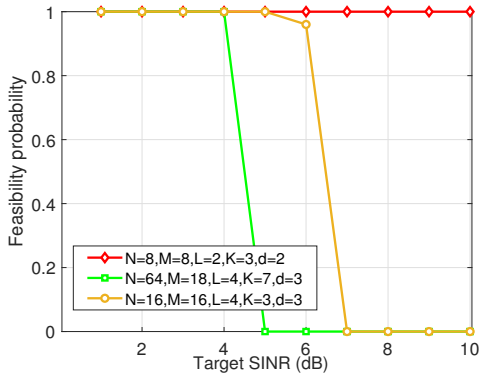

Fig. 3. Feasibility probability vs. target SINR for different networks configurations for $X = 1$ and $Y = 1$.

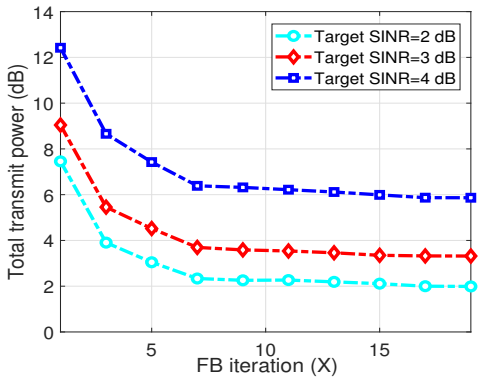

Fig. 4. Total transmit power vs. required iteration numbers for $N = M = 8, L = 2, K = 3, d = 2$, and $Y = 3$.

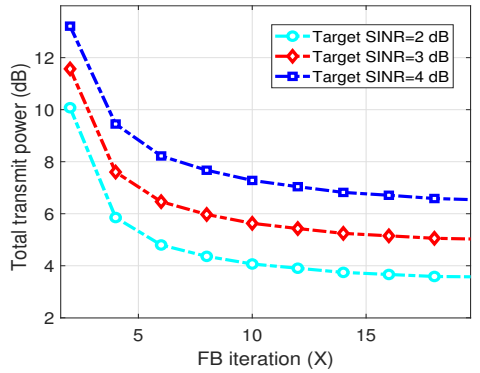

Fig. 5. Total transmit power vs. required iteration numbers for $N = 64, M = 18, L = 4, K = 7, d = 3$, and $Y = 3$.

in every $\Delta > 1$ interval such that $\xi_{\ell_k}^{(1+(l-1)\Delta)}$ is updated according to (40) for $l = 1, 2, \ldots$. We will evaluate the QA scheme, based on (40) and describe its potential performance gainin Section VII.

VII. Simulation Results

In this section, we evaluate the performance of the proposed distributed FB SGPM in Algorithm 3. We also explore the extensions of Algorithm 3 by incorporating the distributed RA and QA schemes proposed in Section VI. Throughout the simulation, we assume MCMU MIMO downlink scenarios where each channel matrix is realized to have i.i.d. entries following $\mathcal{CN}(0, 1)$ and the noise power is set to $\sigma^2 = 1$. In the proposed Algorithm 3, each entry of the precoders $\{\mathbf{V}_{\ell_k}^{(0)}\}$ and combiners $\{\mathbf{U}_{\ell_k}^{(0)}\}$ follows i.i.d. $\mathcal{CN}(0, 1)$. The initial value of power budget $\sum_{\ell_k} \alpha_{\ell_k}^{(0)}$ is set to 10 dB with equal power allocation. We consider both proper and improper network configurations for evaluation, in which a MIMO network becomes proper, if $M + N > (LK + 1)d$, and improper, otherwise [46][3]. Throughout the simulations, we use the integer numbers X and Y to denote the numbers of FB iterations and Jacobi iterations, respectively.

A. Feasibility vs. Target SINR

In Fig. 3, we evaluate the feasibility probability across the target SINR value ξ_{ℓ_k} for different network configurations. Because the feasibility of the proposed algorithm is determined if $\xi_{\ell_k} \in \mathcal{S}_{\ell_k}^{(1)}$ by Remark 1, the feasibility probability is defined by

$$P_f(\{\xi_{\ell_k}\}) = \Pr\left(\bigcap_{\ell_k=1_1}^{L_K} \mathcal{S}_{\ell_k}^{(1)}\right).$$

If any of $\mathcal{S}_{\ell_k}^{(1)}$ is empty, we declare that Algorithm 3 is infeasible. In Fig. 3, one proper case, i.e., $N = M = 8, L = 2, K = 3, d = 2$, and two improper cases, i.e., $N = 64, M = 18, L = 4, K = 7, d = 3$ and $N = M = 16, L = 4, K = 3, d = 3$, are evaluated for $X = 1$ and $Y = 1$. For the proper network, the proposed distributed FB SGMP becomes feasible with high probability. However, the feasibility rate decreases as the network becomes improper and dense. These results provide guidance on how to choose the target SINR values for a specific network configuration.

B. Convergence

In this set of simulations, we evaluate the convergence of Algorithm 3 asserted in Theorem 5. Figs. 4 and 5 show the total transmit power budget $\sum_{\ell_k} \alpha_{\ell_k}^{(n)}$ across different numbers of FB iteration X, while keeping the number of Jacobi iteration constant $Y = 3$. We choose the two network configurations in Fig. 3 for evaluation, one for a proper network with $N = M = 8, L = 2, K = 3, d = 2$, and the other for an improper network with $N = 64, M = 18, L = 4, K = 7, d = 3$. The curves are evaluated for three different target SINR values, 2, 3, and 4 dB that are chosen to be feasible with high probability in Fig. 3. Specifically, for both configurations, the total transmit power values monotonically decrease. The figures clearly show that to guarantee the same target QoS level, an improper network requires a higher total transmit power. The curves quickly converge to stationary points when the network is proper, while the convergence become slow for the improper network.

In the next set of simulations, the network configurations are labeled as $(N, M, L, K, d, X\text{-}Y)$ for brevity. In Fig. 6,

[3]If a network is proper, the interference could be made arbitrarily small [46].

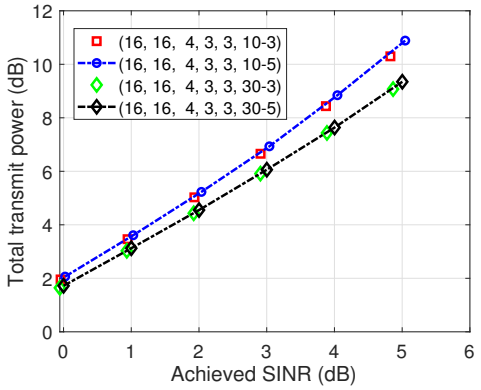

Fig. 6. Total transmit power vs. target SINR for different numbers of FB iteration X and Jacobi iteration Y.

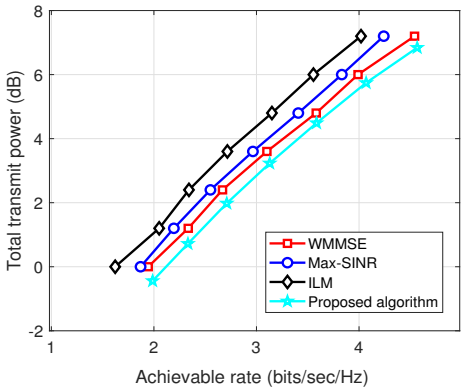

Fig. 8. Total transmit power vs. achievable rate for $N = M = 8, L = 2, K = 3, d = 2, X = 10, Y = 3$.

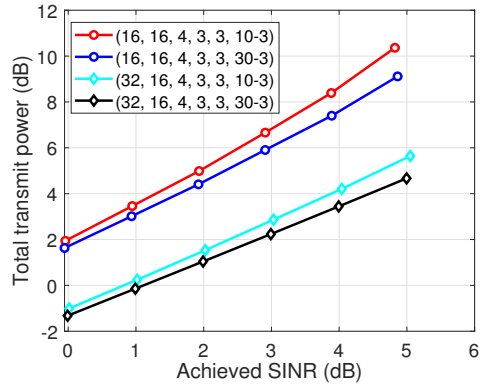

Fig. 7. Total transmit power vs. target SINR for different numbers of BS antennas N and user antennas M.

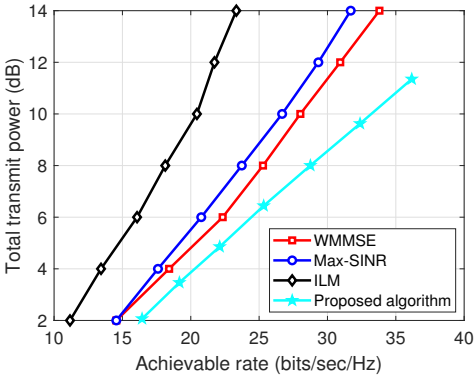

Fig. 9. Total transmit power vs. achievable rate for $N = 64, M = 18, L = 4, K = 7, d = 3, X = 10, Y = 3$.

we evaluate the effects of allocating different numbers of FB iteration X and Jacobi iteration Y to the power efficiency and SINR guarantees. The target SINR values are set to $0, 1, 2, 3, 4$, and 5 dB based on the improper configuration $N = M = 16, L = 4, K = 3, d = 3$ in Fig. 3. The curves are evaluated for the 5% SINR outage probability,

$$P_{out}(\{\xi_{\ell_k}\}) = \Pr\left(\bigcap_{\ell_k=1_1}^{L_K} \left\{\gamma_{\ell_k}^{(DL)} < \xi_{\ell_k}\right\}\right) = 5\%, \quad (41)$$

where $\gamma_{\ell_k}^{(DL)}$ denotes the achieved downlink SINR value. From Fig. 6, it can be seen that increasing the number of Jacobi iteration, while holding the number of FB iteration, allows Algorithm 3 to achieve a higher target SINR value. For instance, when $X = 10$, the proposed algorithm can achieve up to 5 dB SINR with $Y = 5$, while it only achieves 4.8 dB SINR when $Y = 3$. Moreover, allocating more number of FB iteration to update precoders and combiners improves the power efficiency in Fig. 6. The same setting applies to Fig. 7 except for that the numbers of antennas increase from $N = M = 16$ to $N = 32, M = 16$. From Fig. 7, it can be seen that having larger numbers of BS and user antennas can greatly improve the power efficiency. Both Figs. 6 and 7 reveal that the power efficiency and the SINR guarantee of the proposed algorithm can be effectively controlled by properly adjusting the numbers of antennas (N, M) and the numbers of FB iteration X and Jaboci iteration Y.

C. Power Efficiency vs. Achievable Rate

We demonstrate the improved power efficiency and rate guarantees of the proposed algorithm by comparing it with prior fully distributed MCMU MIMO precoding and combining techniques, such as max-SINR [13], ILM [28], and WMMSE [29]. We assume the exact same proper and improper network configurations as in Figs. 4 and 5. In Figs. 8 and 9, the curves are displayed by evaluating the required total transmit power vs. the achieved rate per user, which satisfies 5% SINR outage probability in (41). Because those benchmark algorithms have different objectives, i.e., maximizing weighted sum-rate [29], maximizing SINR [13], and minimizing interference [28], under the same per-user power constraint, we cannot predefine the target SINR values for them. Instead we fix the total transmit power values and evaluate the average achievable rates per user of the benchmarks.

For the proposed Algorithm 3, we run $X = 10$ FB iterations. For each FB iteration, the downlink (respectively, uplink) Jacobi iterations occupies $Y = 3$ (resp., 3) iterations, resulting in a total of 60 (i.e., $10 \times (3 + 3)$) iterations (i.e., channel uses). The same amount of channel use overhead is assumed for all benchmarks. In Fig. 8, the total transmit power values of the benchmarks are predefined by 0, 1.2, 2.4, 3.6, 4.8, 6.0, and 7.2 dB. If the system is proper, the benchmarks and

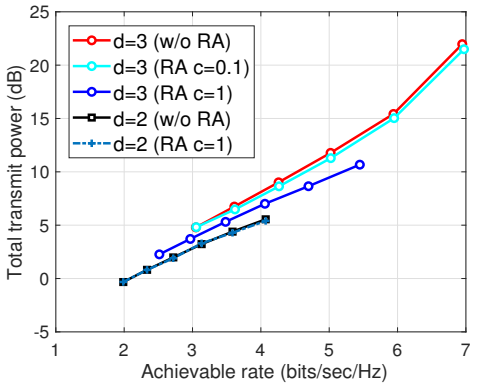

Fig. 10. Exploration of RA: total transmit power vs. achievable rate for $N = 8, M = 8, L = 2, K = 3, X = 20$ and $Y = 10$.

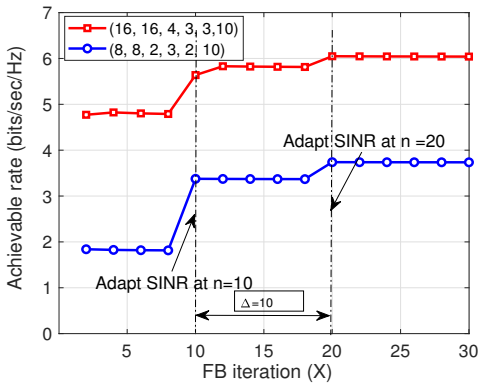

Fig. 11. Exploration of QA with $\Delta = 10$: achievable rate vs. FB iteration.

the proposed algorithm achieve similar performance, in which the proposed technique shows a slightly improved power efficiency compared to other benchmarks. However, when the network becomes dense and improper, the proposed algorithm significantly outperforms the benchmarks as shown in Fig. 9. For instance, in Fig. 9, to achieve a rate per user of around 7.2 bit/sec/Hz, the proposed algorithm only consumes about 10 dB of total transmit power. In contrast, WMMSE requires around 14 dB of total transmit power to achieve the same rate. The total transmit power values for the benchmarks in Fig. 9 are predefined by 2, 4, 6, 8, 10, 12, and 14 dB. This reveals significantly improved power efficiency of Algorithm 3 when the network is improper. It should be mentioned that WMMSE [29] in this setting requires a total of $LK = 40$ weight matrices with each matrix having dimensions 18×64, shared in both downlink and uplink per FB iteration. This clearly results in impractical feedback overhead.

D. RA and QA Exploration

In this set of simulations, we evaluate the RA and QA algorithms in Sections VI-B and VI-C, respectively. In Fig. 10, a network with $N = M = 8$, $L = 2$, and $K = 3$ is assumed. We fix the numbers of FB and Jacobi iteration as $X = 20$ and $Y = 10$, respectively. Fig. 10 demonstrates total transmit power vs. achievable rate per user satisfying 5% SINR outage probability in (41) when RA is performed based on (38) and (39). Each curve is evaluated as the target SINR values increase from 0 to 5 dB. As seen from Fig. 10, when $d = 3$ (i.e., an improper network), RA can significantly improve the power efficiency, compared to the case without RA. The improvement of power efficiency largely relies on the choice of the threshold value c. When $d = 2$ (i.e., a proper network), RA, however, rarely improve the power efficiency. This is because sufficient degrees of freedom are provided for the proposed FB algorithm to coordinate interference when the network is proper.

In Fig. 11, we denote the network configuration as (N, M, L, K, d, Y) for brevity and demonstrate the distributed QA algorithm in Section VI-C for two different network configurations, where $(8, 8, 2, 3, 2, 10)$ models a proper network while $(16, 16, 4, 3, 3, 10)$ models an improper network. The curves in Fig. 11 depict the achieved rate per user vs. FB iteration, which obeys 5% SINR outage probability in (41). In every $\Delta = 10$ iterations, the algorithm adaptively adjusts the target SINR value according to (40). Seen from Fig. 11, the proposed QA can achieve gradually increasing achievable rate per user along with the number of FB iterations.

VIII. Conclusions

This paper studied the general MCMU MIMO precoding and combining SGPM problem. Necessary and sufficient conditions for the optimality of the MCMU MIMO SGPM problem have been characterized by dividing the original problem into forward and backward SGPM subproblems. The derived optimality conditions offer useful clues for designing a fully distributed FB SGPM algorithm. The proposed distributed algorithm is based on solving the four fixed-point equations stemming from the optimality conditions. The sufficient feasibility conditions for the distributed algorithm have been established by leveraging the matrix inverse-positive theory. We proposed closed-loop Jacobi power allocation and FB iterative precoder and combiner design algorithms. The convergence of the proposed algorithm was analytically verified. Numerical studies demonstrated the greatly improved power efficiency of the proposed algorithm for densely populated improper networks. It was shown that the proposed distributed technique combines power efficiency and SGPM feasibility to address the convergence and SINR guarantees of the MCMU MIMO spatial multiplexing systems. Practical issues such as distributed RA and QA were also discussed. The proposed distributed framework holds great potential to be adopted to next generation small-cell networks.

Appendix A
Justification for the Considered QoS

The rate formula in (2) can be lower-bounded by

$$R_{\ell_k} = \sum_{i=1}^{d} \log\Big(\lambda_i\big(\mathbf{I}_d + \mathbf{U}_{\ell_k}^{*}\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^{*}\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\big)\Big) \quad (42)$$

$$\geq \log\left(d + \sum_{i=1}^{d}\lambda_i\left(\mathbf{U}_{\ell_k}^{*}\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^{*}\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\right)\right), \quad (43)$$

where the inequality in (43) is due to the positive semi-definite property of $\mathbf{R}_{\ell_k}$ and $\mathbf{Q}_{\ell_k}$, and it holds for an arbitrary $\mathbf{U}_{\ell_k}$. Thus, maximizing the r.h.s of (43) with respect to $\mathbf{U}_{\ell_k}$ does not alter the inequality in (43), yielding

$$R_{\ell_k} \geq \max_{\mathbf{U}_{\ell_k}} \log\left(d+\sum_{i=1}^{d}\lambda_i\left(\mathbf{U}_{\ell_k}^*\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^*\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\right)\right) \quad (44)$$

$$= \max_{\mathbf{U}_{\ell_k}} \log\Big(d+\mathrm{tr}\left(\mathbf{U}_{\ell_k}^*\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^*\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\right)\Big), \quad (45)$$

which is equivalent to the following optimization problem

$$\begin{aligned}\max_{\mathbf{U}_{\ell_k}}\ & \mathrm{tr}\Big(\mathbf{U}_{\ell_k}^*\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}(\mathbf{U}_{\ell_k}^*\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k})^{-1}\Big)\\ \text{subject to}\ & \mathrm{rank}\left(\mathbf{U}_{\ell_k}\right)=d.\end{aligned} \quad (46)$$

A solution to the problem in (46) is given by [42][Algorithm 8.7.1]

$$\mathbf{U}_{\ell_k}^\star = \mathbf{L}^{-1}\nu_{1:d}\left((\mathbf{L}^*)^{-1}\mathbf{R}_{\ell_k}\mathbf{L}^{-1}\right), \quad (47)$$

where $\mathbf{Q}_{\ell_k}=\mathbf{L}^*\mathbf{L}$ is the Cholesky decomposition of $\mathbf{Q}_{\ell_k}$ and $\nu_{1:d}(\mathbf{A})$ extracts the first d dominant eigenvectors of a matrix $\mathbf{A}$. Given the solution in (47), the following holds

$$\mathrm{tr}\left(\mathbf{U}_{\ell_k}^{\star^*}\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}^\star(\mathbf{U}_{\ell_k}^{\star^*}\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k}^\star)^{-1}\right) = d\left(\frac{\mathrm{tr}\left(\mathbf{U}_{\ell_k}^{\star^*}\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}^\star\right)}{\mathrm{tr}\left(\mathbf{U}_{\ell_k}^{\star^*}\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k}^\star\right)}\right) \quad (48)$$

$$\geq d\left(\frac{\mathrm{tr}\left(\mathbf{U}_{\ell_k}^*\mathbf{R}_{\ell_k}\mathbf{U}_{\ell_k}\right)}{\mathrm{tr}\left(\mathbf{U}_{\ell_k}^*\mathbf{Q}_{\ell_k}\mathbf{U}_{\ell_k}\right)}\right), \quad (49)$$

The equality in (48) is due to $\mathbf{U}_{\ell_k}^\star$ in (47). Substituting $\mathbf{U}_{\ell_k}^\star$ for arbitrary $\mathbf{U}_{\ell_k}$ leads to (49). Noticing that the trace ratio in (49) is the SINR measure in (3), the Shannon rate is now lower bounded by

$$R_{\ell_k} \geq \log(d(1+\gamma_{\ell_k})). \quad (50)$$

This completes the proof. ∎

Appendix B
Proof of Theorem 2

The Lagrangian of (8) is

$$\mathcal{L}(\{\lambda_{\ell_k}\},\{\alpha_{\ell_k}\},\{\mathbf{U}_{\ell_k}\}) \triangleq \sum_{\ell_k}\alpha_{\ell_k}+\sum_{\ell_k}\lambda_{\ell_k}\Big(\sigma^2 + \sum_{i_j\neq\ell_k}\alpha_{i_j}\|\mathbf{U}_{\ell_k}^*\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\|_F^2-\frac{\alpha_{\ell_k}}{\xi_{\ell_k}}\|\mathbf{U}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\|_F^2\Big), \quad (51)$$

where $\{\lambda_{\ell_k}\}$ are the Lagrangian multipliers. Rearranging the order of the summations in (51) gives

$$\mathcal{L}(\{\lambda_{\ell_k}\},\{\alpha_{\ell_k}\},\{\mathbf{U}_{\ell_k}\}) = \sum_{\ell_k}\lambda_{\ell_k}\sigma^2+\sum_{\ell_k}\lambda_{\ell_k}\,\mathrm{tr}\Bigg(\mathbf{U}_{\ell_k}^*\Big(\frac{\alpha_{\ell_k}}{\lambda_{\ell_k}}\mathbf{I}_M + \sum_{i_j\neq\ell_k}\alpha_{i_j}\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\mathbf{V}_{i_j}^*\mathbf{H}_{\ell_k,i}^*-\frac{\alpha_{\ell_k}}{\xi_{\ell_k}}\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*\Big)\mathbf{U}_{\ell_k}\Bigg), \quad (52)$$

where (52) comes from the fact that $\|\mathbf{U}_{\ell_k}\|_F^2=1,\forall\ell_k$. From (52), the Lagrange dual function $\mathcal{G}(\{\lambda_{\ell_k}\})$ of (8) is given by

$$\mathcal{G}(\{\lambda_{\ell_k}\}) = \min_{\{\alpha_{\ell_k}\},\{\mathbf{U}_{\ell_k}\}}\mathcal{L}\left(\{\lambda_{\ell_k}\},\{\alpha_{\ell_k}\},\{\mathbf{U}_{\ell_k}\}\right), \quad (53)$$

which is always upper bounded by $\sum_{\ell_k}\alpha_{\ell_k}$, i.e.,

$$\mathcal{G}(\{\lambda_{\ell_k}\}) \leq \sum_{\ell_k}\alpha_{\ell_k}. \quad (54)$$

Next, we show that the KKT conditions of (8), which are necessary, are indeed optimal for (8) if the primal problem (8) is feasible. The KKT conditions of (8) are given by:

1) First order necessary condition for $\mathbf{U}_{\ell_k},\forall\ell_k$:

$$2\lambda_{\ell_k}\Bigg(\sum_{i_j\neq\ell_k}\alpha_{i_j}\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\mathbf{V}_{i_j}^*\mathbf{H}_{\ell_k,i}^*+\sigma^2\mathbf{I}_M - \frac{\alpha_{\ell_k}}{\xi_{\ell_k}}\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*\Bigg)\mathbf{U}_{\ell_k}=\mathbf{0}_{M\times d}. \quad (55)$$

2) First order necessary condition for $\alpha_{\ell_k},\forall\ell_k$:

$$1+\lambda_{\ell_k}\,\mathrm{tr}\left(\mathbf{U}_{\ell_k}^*(-\tfrac{1}{\xi_{\ell_k}}\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_K}^*)\mathbf{U}_{\ell_k}\right) + \sum_{i_j\neq\ell_k}\lambda_{i_j}\,\mathrm{tr}\left(\mathbf{U}_{i_j}^*\mathbf{H}_{i_j,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{i_j,\ell}^*\mathbf{U}_{i_j}\right)=0. \quad (56)$$

3) Complementary slackness, $\forall\ell_k$:

$$\lambda_{\ell_k}\,\mathrm{tr}\left(\mathbf{U}_{\ell_k}^*\left(\sum_{i_j\neq\ell_k}\alpha_{i_j}\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\mathbf{V}_{i_j}^*\mathbf{H}_{\ell_k,i}^*+\sigma^2\mathbf{I}_M - \frac{\alpha_{\ell_k}}{\xi_{\ell_k}}\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*\right)\mathbf{U}_{\ell_k}\right)=0. \quad (57)$$

The Lagrangian multiplier λ_{ℓ_k} must be $\lambda_{\ell_k}\neq 0,\forall\ell_k$, because otherwise equality in (56) does not hold. By the feasibility of (8), there exists a tuple $(\{\mathbf{U}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\},\{\lambda_{\ell_k}^\star\})$ satisfying the KKT conditions, such that

$$\mathcal{G}(\{\lambda_{\ell_k}^\star\}) = \sum_{\ell_k}\lambda_{\ell_k}^\star\sigma^2+\sum_{\ell_k}\lambda_{\ell_k}^\star\,\mathrm{tr}\Bigg(\mathbf{U}_{\ell_k}^{\star\,*}\Big(\frac{\alpha_{\ell_k}^\star}{\lambda_{\ell_k}^\star}\mathbf{I}_M + \sum_{i_j\neq\ell_k}\alpha_{i_j}^\star\mathbf{H}_{\ell_k,i}\mathbf{V}_{i_j}\mathbf{V}_{i_j}^*\mathbf{H}_{\ell_k,i}^*-\frac{\alpha_{\ell_k}^\star}{\xi_{\ell_k}}\mathbf{H}_{\ell_k,\ell}\mathbf{V}_{\ell_k}\mathbf{V}_{\ell_k}^*\mathbf{H}_{\ell_k,\ell}^*\Big)\mathbf{U}_{\ell_k}^\star\Bigg). \quad (58)$$

The duality gap is zero when $\alpha_{\ell_k}^\star=\lambda_{\ell_k}^\star\sigma^2,\forall\ell_k$, i.e., $\mathcal{G}(\{\lambda_{\ell_k}^\star\})=\sum_{\ell_k}\alpha_{\ell_k}^\star$, because the second term on the r.h.s. of (58) becomes zero due to (57). Therefore, strong duality of (8) holds and the tuple $(\{\mathbf{U}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\},\{\lambda_{\ell_k}^\star=\alpha_{\ell_k}^\star/\sigma^2\})$ are the primal and dual solutions. Because under strong duality, any pair of primal and dual optimal points must satisfy the KKT conditions [47][Chapter 5.5.3], the stationary condition in (55) and the complementary slackness in (57) are optimality conditions that imply (9) and (10), respectively. This completes the proof. ∎

Appendix C
Proof of Theorem 3

We first show the existence of a joint optimal solution of the SGPM subproblems in (8) and (11). Then, the proof for the coincidence between the joint optimal solution and the solution of (7) follows.

1) Existence of a joint optimal solution: If the original SGPM problem (7) is feasible and has an optimal tuple $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$, substituting it into the forward

SGPM subproblem (8) will make it feasible. Now, to show the feasibility of the backward subproblem (11), we fix $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\}\right)$ and formulate the forward and backward power allocation problems:

$$\min_{\{\alpha_{\ell_k}\}} \sum_{\ell_k} \alpha_{\ell_k} \;\text{ subject to }\; \gamma_{\ell_k}^{(DL)}\left(\{\alpha_{\ell_k}\}\right) \geq \xi_{\ell_k},\ \forall \ell_k, \quad (59)$$

and

$$\min_{\{\omega_{\ell_k}\}} \sum_{\ell_k} \omega_{\ell_k} \;\text{ subject to }\; \gamma_{\ell_k}^{(UL)}\left(\{\omega_{\ell_k}\}\right) \geq \xi_{\ell_k},\ \forall \ell_k, \quad (60)$$

respectively. Due to the UD duality between (59) and (60) [17], if we let $\{\omega_{\ell_k}^\star\}$ be the optimal solution of (60) we have the following equality

$$\sum_{\ell_k} \alpha_{\ell_k}^\star = \sum_{\ell_k} \omega_{\ell_k}^\star. \quad (61)$$

This concludes that the backward SGPM problem (11) is also feasible for the tuple $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\omega_{\ell_k}^\star\}\right)$. Hence there exists a joint optimal solution for the forward and backward subproblems.

2) Coincidence between the optimal solution of (7) *and the joint optimal solution of* (8) *and* (11)*:* First, we claim that the tuple $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$ optimal to (7) also jointly optimizes (8) and (11). This claim can be shown by contradiction. If the tuple $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$ does not jointly optimize (8) and (11), there exists at least another tuple, for example, $\left(\{\acute{\mathbf{U}}_{\ell_k}\},\{\mathbf{V}_{\ell_k}^\star\},\{\acute{\alpha}_{\ell_k}\}\right)$ that lowers the total transmit power of (8), i.e., $\sum_{\ell_k}\acute{\alpha}_{\ell_k} \leq \sum_{\ell_k}\alpha_{\ell_k}^\star$, meaning that the total transmit power of (7) can be further lowered. This contradicts the assumption that $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$ is optimal for (7).

Conversely, we claim that the joint optimal solution of (8) and (11), denoted as $\left(\{\mathbf{U}_{\ell_k}^o\},\{\mathbf{V}_{\ell_k}^o\},\{\alpha_{\ell_k}^o\}\right)$, also optimizes (7). First, assume the optimal solution of (7) is the tuple $\left(\{\mathbf{V}_{\ell_k}^\star\},\{\mathbf{U}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$, which is different from $\left(\{\mathbf{V}_{\ell_k}^o\},\{\mathbf{U}_{\ell_k}^o\},\{\alpha_{\ell_k}^o\}\right)$. This gives $\sum_{\ell_k}\alpha_{\ell_k}^\star \leq \sum_{\ell_k}\alpha_{\ell_k}^o$ for (7). If we insert the tuple $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\},\{\alpha_{\ell_k}^\star\}\right)$ into (8) the total transmit power will be decreased to $\sum_{\ell_k}\alpha_{\ell_k}^\star \leq \sum_{\ell_k}\alpha_{\ell_k}^o$. Moreover, fixing $\left(\{\mathbf{U}_{\ell_k}^\star\},\{\mathbf{V}_{\ell_k}^\star\}\right)$ in (11) results in $\sum_{\ell_k}\omega_{\ell_k}^\star = \sum_{\ell_k}\alpha_{\ell_k}^\star \leq \sum_{\ell_k}\omega_{\ell_k}^o = \sum_{\ell_k}\alpha_{\ell_k}^o$ due to (61). This contradicts the fact that the tuple $\left(\{\mathbf{V}_{\ell_k}^o\},\{\mathbf{U}_{\ell_k}^o\},\{\alpha_{\ell_k}^o\}\right)$ is optimal for (8) and (11).

Therefore, the joint optimal solution of (8) and (11) is equivalent to the optimal solution of (7). ∎

Appendix D
Proof of Lemma 2

Decomposing $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(n)} = \mathbf{L}^*\mathbf{L}$ and reformulating the constraint in (14), the optimal $\mathbf{U}_{\ell_k}^{(n)^\star}$ must satisfy

$$\mathbf{L}^*\Big(\mathbf{I}_M - \frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}\Big)\mathbf{L}\mathbf{U}_{\ell_k}^{(n)^\star} = \mathbf{0}_{M\times d}. \quad (62)$$

Because $\mathbf{U}_{\ell_k}^{(n)^\star}$ and $\mathbf{L}$ are full rank, (62) is true if and only if $\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}$ has at least d unit-eigenvalues. Because the rank of $\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}$ is d, the optimal solution must obey

$$\mathbf{L}\mathbf{U}_{\ell_k}^{(n)^\star} = a\ \nu_{1:d}\left(\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(n)}\mathbf{L}^{-1}\right), \quad (63)$$

where $a\in\mathbb{R}$ is a normalization constant. Taking $\mathbf{L}^{-1}$ of both sides of (63) and determining a to meet $\|\mathbf{U}_{\ell_k}^{(n)}\|_F = 1$ results in (15). ∎

Appendix E
Proof of Theorem 5

The proof consists of three constituent parts: (i) convergence of the objective, (ii) convergence of the parameters to a limit point, and (iii) stationary point. We first prove the convergence of the total transmit power value.

1) Convergence of the objective: We claim that (F1) and (P1) in the forward direction of the nth FB iteration do not increase the transmit power budget,

$$\sum_{\ell_k}\alpha_{\ell_k}^{(n-1)} \geq \sum_{\ell_k}\alpha_{\ell_k}^{(n)},\ \forall n>1. \quad (64)$$

This claim can be shown as follows. After the filter optimization (F1), the target SINR values satisfy $\xi_{\ell_k}\in\mathcal{S}_{\ell_k}^{(n)}$, $\forall\ell_k$, i.e., Remark 1, and make the initial power allocation $\mathbf{a}_{(0)}^{(n)} = \mathbf{1}_{LK}$ feasible, i.e.,

$$\mathbf{A}_{(n-1)}^{(n)}\mathbf{1}_{LK} \succeq \mathbf{q}. \quad (65)$$

From (26), we have for $m=1$,

$$\mathbf{a}_{(1)}^{(n)} = \Big(\mathbf{I}_{LK} - \Big(\mathbf{D}_{(n-1)}^{(n)}\Big)^{-1}\mathbf{A}_{(n-1)}^{(n)}\Big)\mathbf{1}_{LK} + \Big(\mathbf{D}_{(n-1)}^{(n)}\Big)^{-1}\mathbf{q}. \quad (66)$$

Rewriting the inequality in (65) as $\mathbf{1}_{LK} \succeq \big(\mathbf{I}_{LK} - \big(\mathbf{D}_{(n-1)}^{(n)}\big)^{-1}\mathbf{A}_{(n-1)}^{(n)}\big)\mathbf{1}_{LK} + \big(\mathbf{D}_{(n-1)}^{(n)}\big)^{-1}\mathbf{q}$ leads to $\mathbf{1}_{LK} \succeq \mathbf{a}_{(1)}^{(n)}$ due to (66). Because $\big(\mathbf{I}_{LK} - \big(\mathbf{D}_{(n-1)}^{(n)}\big)^{-1}\mathbf{A}_{(n-1)}^{(n)}\big) \succeq \mathbf{0}_{LK\times LK}$, iterating the Jacobi recursion in (26) results in $\mathbf{1}_{LK} \succeq \mathbf{a}_{(m)}^{(n)}$ for $m\geq 1$. Therefore, Algorithm 1 converges to a solution

$$\mathbf{0}_{LK\times 1} \preceq \mathbf{a}^{(n)} \preceq \mathbf{1}_{LK}. \quad (67)$$

Then Step 8 of Algorithm 1 gives

$$\Big(\boldsymbol{\alpha}^{(n-1)}\Big)^T\mathbf{1}_{LK} \geq \Big(\boldsymbol{\alpha}^{(n-1)}\Big)^T\mathbf{a}^{(n)} = \Big(\boldsymbol{\alpha}^{(n)}\Big)^T\mathbf{1}_{LK}, \quad (68)$$

where $\boldsymbol{\alpha}^{(n)} = [\alpha_{1_1}^{(n)},\ldots,\alpha_{L_K}^{(n)}]^T$ and the inequality in (68) is due to (67). This proves (64).

The backward iteration shares the same target SINR values and the same feasibility as the forward iteration. Hence the precoder design (F2) does not alter its feasibility due to (31). Moreover, the uplink power allocation will not increase the total uplink power budget due to the UD duality in (61). Because the total transmit power is bounded above 0, the objective converges.

2) Convergence of the parameters to a limit point: Suppose the objective has converged to the total transmit power value $P^\star$ at the lth FB iteration. We let the achieved solution at the lth FB iteration be

$\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$. Because the proposed Algorithm 3 guarantees that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l-1)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ satisfies (10), $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ must also satisfy (10); otherwise, the precoder update from $\{\mathbf{V}_{\ell_k}^{(l-1)}\}$ to $\{\mathbf{V}_{\ell_k}^{(l)}\}$ leads to a reduced total transmit power value, which contradicts the fact that the total transmit power has converged to $P^\star$. Due to the same reason, $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ satisfies (13). Provided the latter setting, we show that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\} = \{\{\mathbf{U}_{\ell_k}^{(l+1)}\},\{\mathbf{V}_{\ell_k}^{(l+1)}\},\{\alpha_{\ell_k}^{(l+1)}\},\{\omega_{\ell_k}^{(l+1)}\}\}$ below.

At the $(l+1)$th forward iteration, the combiner is updated to $\mathbf{U}_{\ell_k}^{(l+1)}$ according to (15). Due to the convergence of the total transmit power, the updated $\mathbf{U}_{\ell_k}^{(l+1)}$ leads to the strict equality $\mathbf{A}_{(l)}^{(l+1)}\mathbf{1}_{LK}=\mathbf{q}$, i.e., (65). Otherwise, the proposed algorithm reduces total transmit power by (68), which is a contradiction. Therefore, the equality $\mathbf{A}_{(l)}^{(l+1)}\mathbf{1}_{LK}=\mathbf{q}$ implies,

$$\mathrm{tr}\left((\mathbf{U}_{\ell_k}^{(l+1)})^*\Big(\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l+1)}-\frac{1}{\xi_{\ell_k}}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\Big)\mathbf{U}_{\ell_k}^{(l+1)}\right)=0,\forall\ell_k, \quad (69)$$

where $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l+1)}\triangleq\sum_{i_j\neq\ell_k}\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(l)}\big(\mathbf{H}_{\ell_k,i}\widetilde{\mathbf{V}}_{i_j}^{(l)}\big)^*+\sigma^2\mathbf{I}_M$ and $\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\triangleq\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}^{(l)}\big(\mathbf{H}_{\ell_k,\ell}\widetilde{\mathbf{V}}_{\ell_k}^{(l)}\big)^*$, which have been defined in Lemma 2. Given the Cholesky decomposition $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l+1)}=\mathbf{L}^*\mathbf{L}$, plugging $\mathbf{U}_{\ell_k}^{(l+1)}$ in (15) into (69) gives

$$\mathrm{tr}\left(\mathbf{\Gamma}^*\Big(\mathbf{I}_{M\times M}-\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}\Big)\mathbf{\Gamma}\right)=0, \quad (70)$$

where $\mathbf{\Gamma}\triangleq\nu_{1:d}\left(\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}\right)\in\mathbb{C}^{M\times d}$. Denoting the d-dominant eigenvalues of $\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}$ as $(\epsilon_1,\epsilon_2,\ldots,\epsilon_d)$ in descending order, the equality in (70) implies that the summation of the d-dominant eigenvalues of $\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}$ is equal to d, i.e., $\sum_{i=1}^d\epsilon_i=d$. Because $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ satisfies (10) for $\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}$ and $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l+1)}$, we have

$$\mathrm{tr}\left((\mathbf{U}_{\ell_k}^{(l)})^*\mathbf{L}^*\Big(\mathbf{I}_{M\times M}-\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}\Big)\mathbf{L}\mathbf{U}_{\ell_k}^{(l)}\right)=0. \quad (71)$$

Then, (71) holds only if $\mathbf{U}_{\ell_k}^{(l)}=\mathbf{U}_{\ell_k}^{(l+1)}=\frac{\mathbf{L}^{-1}\nu_{1:d}\left(\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}\right)}{\left\|\mathbf{L}^{-1}\nu_{1:d}\left(\frac{1}{\xi_{\ell_k}}(\mathbf{L}^*)^{-1}\overrightarrow{\mathbf{R}}_{\ell_k}^{(l+1)}\mathbf{L}^{-1}\right)\right\|_F}$, $\forall\ell_k$. Hence, after the combiner update (F1) at the $(l+1)$th forward iteration, the following holds $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}=\{\{\mathbf{U}_{\ell_k}^{(l+1)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$. Moreover, due to $\mathbf{U}_{\ell_k}^{(l)}=\mathbf{U}_{\ell_k}^{(l+1)}$, $\forall\ell_k$, the power update (P1) at the $(l+1)$th forward iteration must satisfy $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}=\{\{\mathbf{U}_{\ell_k}^{(l+1)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l+1)}\}\}$, i.e., $\alpha_{\ell_k}^{(l)}=\alpha_{\ell_k}^{(l+1)}$. The proof for $\mathbf{V}_{\ell_k}^{(l+1)}=\mathbf{V}_{\ell_k}^{(l)}$ and $\omega_{\ell_k}^{(l+1)}=\omega_{\ell_k}^{(l)}$, $\forall\ell_k$ in the backward iteration follows the exact same procedure as the forward iteration, so we omit it. Therefore, the solution $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ is a limit point.

3) Stationary point: We need to show that the limit point $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ satisfies the first-order necessary conditions in (9)-(10) and (12)-(13). We have already proved that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ and $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ satisfy, respectively, (10) and (13) in Appendix E-2. In what follows, we shown that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ satisfies (9) and $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ satisfies (12). The former is shown first.

Multiplying $\big(\mathbf{U}_{\ell_k}^{(l)}\big)^*$ to both sides of (9) gives

$$\mathbf{I}_d-\frac{1}{\xi_{\ell_k}}\mathbf{\Lambda}_{d\times d}=\mathbf{0}_{d\times d}. \quad (72)$$

The equality in (72) holds because $\mathbf{U}_{\ell_k}^{(l)}$ in (15) diagonalizes $\overrightarrow{\mathbf{R}}_{\ell_k}^{(l)}$ to $\big(\mathbf{U}_{\ell_k}^{(l)}\big)^*\overrightarrow{\mathbf{R}}_{\ell_k}^{(l)}\mathbf{U}_{\ell_k}^{(l)}=\mathbf{\Lambda}_{d\times d}$ and $\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l)}$ to $\big(\mathbf{U}_{\ell_k}^{(l)}\big)^*\overrightarrow{\mathbf{Q}}_{\ell_k}^{(l)}\mathbf{U}_{\ell_k}^{(l)}=\mathbf{I}_d$, respectively. For $d=1$, $\frac{1}{\xi_{\ell_k}}\mathbf{\Lambda}_{1\times1}=1$ because $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ satisfies (10), which is equal to (72). For $d>1$, we prove that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ satisfies (9) by contradiction. If $\frac{1}{\xi_{\ell_k}}\mathbf{\Lambda}_{d\times d}$ is not equal to $\mathbf{I}_d$, it means that $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ is not a solution to (9). Then, in the next FB iteration, after filter updates, one can decrease the total transmit power to $P'\leq P^\star$ by (64). This contradict with the fact that $P^\star$ is a limit point. Hence, $\frac{1}{\xi_{\ell_k}}\mathbf{\Lambda}_{d\times d}=\mathbf{I}_d$ and $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\alpha_{\ell_k}^{(l)}\}\}$ is the solution to (9). Because the same proof can be extended to show that the $\{\{\mathbf{U}_{\ell_k}^{(l)}\},\{\mathbf{V}_{\ell_k}^{(l)}\},\{\omega_{\ell_k}^{(l)}\}\}$ is the solution to (12), it is omitted. Therefore, the proposed algorithm converges to a stationary point. This completes the proof. ■

ACKNOWLEDGEMENT

We are deeply indebted to the reviewers, whose invaluable and consistent comments greatly improved the manuscript.

REFERENCES

[1] G. Xiong, T. Kim, and D. J. Love, "Distributed filter design and power allocation for small-cell MIMO networks," *IEEE Vehical Technology Conference-Fall*, pp. 1–5, Sep 2017.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] S. Samarakoon, M. Bennis, W. Saad, M. Debbah, and M. Latva-aho, "Ultra dense small cell networks: Turning density into energy efficiency," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1267–1280, May 2016.

[4] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, February 2014.

[5] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 GHz for ultra-dense 5G wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.

[6] *Initial Report on Horizontal Topics, First Results and 5G System Concept*. document METIS D6.2, March 2014.

[7] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Sum-rate maximization in sub-28-GHz millimeter-wave MIMO interfering networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1649–1662, July 2017.

[8] M. K. Samimi and T. S. Rappaport, *"Characterization of the 28GHz millimeter-wave dense urban channel for future 5G mobile cellular"*. Tech. Rep. TR 2014-04, March 2014.

[9] M. Dong and T. Kim, "Interference analysis for millimeter-wave networks with geometry-dependent first-order reflections," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12 404–12 409, Dec 2018.

[10] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[11] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, *Weighted Sum-Rate Maximization in Wireless Networks: A Review*. Now, 2012. [Online]. Available: https://ieeexplore.ieee.org/document/8186919

[12] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex MIMO interference channels," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 801–815, 2015.

[13] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Comparison of distributed beamforming algorithms for MIMO interference networks," *IEEE Trans. on Signal Processing*, vol. 61, pp. 3476–3489, July 2013.

[14] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Distributed low-overhead schemes for multi-stream MIMO interference channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, April 2015.

[15] M. Bengtsson, "A pragmatic approach to multi-user spatial multiplexing," *Sensor Array and Multichannel Signal Processing Workshop Proceedings*, pp. 130 – 134, 2002.

[16] D. P. Palomar, M. A. Lagunas, and J. M.Cioffi, "Optimum linear joint transmit-receive processiong for MIMO cahnnels with QoS constraints," *IEEE Transactions on Signal Processing*, May 2004.

[17] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, Jan 2006.

[18] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[19] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4639–4655, Sept 2007.

[20] J. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint transmitter receiver diversity for efficient space division multiaccess," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 16–27, Jan 2002.

[21] B. Song, R. L. Cruz, and B. D. Rao, "Network duality for multiuser MIMO beamforming networks and applications," *IEEE Transactions on Communications*, vol. 55, no. 3, pp. 618–630, March 2007.

[22] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *1999 IEEE 49th Vehicular Technology Conference*, vol. 1, May 1999, pp. 851–856 vol.1.

[23] Q. Shi, M. Razaviyayn, M. Hong, and Z. Q. Luo, "SINR constrained beamforming for a MIMO multi-user downlink system: Algorithms and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2920–2933, June 2016.

[24] H. Pennanen, A. Tolli, J. Kaleva, P. Komulainen, and M. Latva-aho, "Decentralized linear transceiver design and signaling strategies for sum power minimization in multi-cell MIMO systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1729–1743, April 2016.

[25] H. Pennanen, A. Tolli, and M. Latva-aho, "Decentralized coordinated downlink beamforming via primal decomposition," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 647–650, Nov 2011.

[26] A. Tolli, H. Pennanen, and P. Komulainen, "Distributed coordinated multi-cell transmission based on dual decomposition," in *2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–6.

[27] C. Shen, T. Chang, K. Wang, Z. Qiu, and C. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An admm approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, June 2012.

[28] K. Gomadam, V. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3309–3322, June 2011.

[29] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, Sep 2011.

[30] R. Brandt and M. Bengtsson, "Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2890–2906, May 2016.

[31] D. Kincaid and W. Cheney, *Numerical Analysis*, 2nd ed. Brooks/Cole Publishing Company, 1996.

[32] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, 1994. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611971262

[33] J.E.Peris and B. Subiza, "A characterization of weak-monotone matrices," *Linear Algebra and its Applications*, vol. 166, pp. 167–184, March 1992.

[34] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 652–664, Feb 2016.

[35] C. Hellings, F. Askerbeyli, and W. Utschick, "Two-user SIMO interference channel with treating interference as noise: Improper signaling versus time-sharing," *IEEE Transactions on Signal Processing*, pp. 1–1, 2020.

[36] C. Hellings and W. Utschick, "Improper signaling versus time-sharing in the two-user gaussian interference channel with TIN," *IEEE Transactions on Information Theory*, vol. 66, no. 5, pp. 2988–2999, 2020.

[37] H. Joudeh and B. Clerckx, "Rate-splitting for max-min fair multi-group multi-cast beamforming in overloaded systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7276–7289, 2017.

[38] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8754–8770, 2019.

[39] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Communications Letters*, vol. 9, no. 3, pp. 349–353, 2020.

[40] R. J. Plemmons, "M-matrices leading to semiconvergent splittings," *Linear Algebra and its Applications*, vol. 15, no. 3, pp. 243–252, 1976.

[41] C. D. Meyer, Jr, "The role of the group generalized inverse in the theory of finite Markov chains," *SIAM Review*, vol. 17, no. 3, pp. 443–464, 1975.

[42] G. Golub and C. V. Loan, *Matrix Computations*, 4th ed. The Johns Hopkins University Press, 2014.

[43] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. USA: Academic Press, Inc., 2011.

[44] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, p. 3, 2009.

[45] D. J. Love and R. W. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3674–3687, 2005.

[46] T. Liu and C. Yang, "On the feasibility of linear interference alignment for MIMO interference broadcast channels with constant coefficients," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2178–2191, May 2013.

[47] S. Boyd and L. Vandenderghe, *Convex Optimization*, 1st ed. Cambridge University Press, 2004.