Multi-Cell Multi-User MIMO Imperfect CSI Transceiver Design with Power Method Generalized Eigenvectors

Amel Tibhirt[1,2], Dirk Slock[1] and Yi Yuan-Wu[2]
[1]Communication Systems Department, EURECOM, France
[2]Orange Labs, Châtillon, France,
{amel.tibhirt,dirk.slock}@eurecom.fr, yi.yuan@orange.com

***Abstract*—In this paper, we consider the problem of user rate balancing in the downlink of multi-cell multi-user (MU) Multiple-Input-Multiple-Output (MIMO) systems with imperfect Channel State Information at the Transmitter (CSIT). We linearize the problem by introducing a rate minorizer and by formulating the balancing operation as constraints leading to a Lagrangian, allowing to transform rate balancing into weighted sum Mean Squared Error (MSE) or Interference Plus Noise (IPN) power minimization with Perron Frobenius theory. We introduce two imperfect CSIT formulations. One is based on the expected rate vs. Expected MSE (EMSE) relation, the other involves an original rate minorizer in terms of the received IPN covariance matrix, in the imperfect CSIT case applied to the Expected Signal and Interference Power (ESIP) rate. The main contribution here is another minorization step via an extended Rayleigh quotient which leads to a principled approach for introducing power method iterations replacing explicit generalized eigenvector computations. This allows to bring down the complexity per iteration of the better ESIP approaches to that of the EMSE based approaches. But we further reduce complexity by introducing also a (large) matrix inverse free power method.**

***Index Terms*—Multi-User MIMO, Rate Balancing, Imperfect CSIT, Power Method, Matrix-Inverse Free**

I. Introduction

Multi-user multiple-input multiple-output (MIMO) systems are considered as a promising technique for next generation cellular networks for their great potential to achieve high throughput [1]. In downlink communications, when a certain knowledge of the Channel State Information (CSI) at the transmitter is available, the system throughput can be maximized. Obtaining CSI at the receiver is easily possible via training, whereas CSI at the transmitter (CSIT) acquires reciprocity or feedback from the receiver. Therefore, many works address the problem of optimizing the performance of MIMO systems with the presence of CSIT uncertainties, better known as partial CSIT. Among the different optimization criteria, we distinguish the transmit power minimization, and the max-min/min-max problems w.r.t. either signal-to-interference-plus-noise ratio (SINR) [2]–[6], Mean Square Error (MSE) [7]–[9] or user rate. The latter is the focus of this work. In particular, we study Multi-Cell MIMO User Rate Balancing with Partial CSIT.

In this work, we focus on ergodic user rate balancing, which corresponds to maximizing the minimum (weighted) per user expected rate in the network. We consider a multi-cell multi-user MIMO system with imperfect CSIT, which combines both channel estimates and channel (error) covariance information. In particular, we introduce a novel [10] extension of [11] to imperfect CSIT, maximizing an expected rate lower bound in terms of expected MSE. Furthermore, we introduce a second algorithm by exploiting a better approximation of the expected rate as the Expected Signal and Interference Power (ESIP) rate. Whereas we have considered the ESIP approach in previous sum utility optimization work, the algorithm here is based on an original minorizer for every individual rate term, different from existing DC programming approaches in sum utility optimization. Both algorithms are based on a Lagrangian formulation introduced in [11] for perfect CSIT, in which utility balancing gets transformed into a weighted sum utility with known optimal beamformers.

A. Contributions

The main contribution then is another minorization step via an extended Rayleigh quotient which leads to a principled approach for introducing power method iterations replacing explicit generalized eigenvector computations. This allows to bring down the complexity per iteration of the better ESIP approaches to that of the EMSE based approaches. Namely, the updated generalized eigenvector(s) are obtained by solving a linear system of equations, i.e. as a matrix inverse times a (block) vector. In the case of perfect CSIT, the power method iterations for the transmit (Tx) beamformers (BF) lead to an algorithm that is very related to the approach in which receivers (Rx) are introduced and the algorithm alternates between updating transmitters and receivers. In the (ESIP) imperfect CSIT case however, the introduction of receivers and alternating optimization does not allow to reach something closely related to the proposed power method iterations on the transmitters (only). Although the power method replaces eigenvectors with LMMSE (Linear Minimum MSE) type filters, these still require the inversion of large matrices. Hence we also introduce matrix-inverse free methods that are related to polynomial expansion or Jacobi methods, but with combination coefficients optimizing the desired utility function.

B. Related Work

The large matrix inversions for LMMSE in Massive MIMO are the main motivation behind searching for low complexity solutions with close to optimal performance. [12] proposes truncated polynomial expansion (PE) for reducing precoder complexity. [13] uses the Approximate Message Passing (AMP) algorithm for LMMSE and introduces a non-parametric algorithm called NOPE that does not require any knowledge of the signal and noise powers. The authors also prove that in the large system limit, NOPE achieves the same performance as that of the LMMSE equalizer. [14] showed that the design of all variants of linear precoder/combiners for the downlink (DL) and uplink (UL) can be posed as the solution of a set of linear equations. Furthermore, this is solved using Kaczmarz method, which is essentially the Normalized LMS algorithm from adaptive filtering, applied to a randomized selection of the normal equations to be satisfied. In [15] we propose variational Bayesian learning (VBL) techniques to acquire it assuming TDD channel reciprocity. In particular a Space Alternating version of Variational Estimation (SAVE) allows a well founded alternative to AMP based techniques while being of similar complexity. Furthermore the resulting posterior parameter distributions allow to express covariance CSIT imprecision. The SAVE techniques can also be applied to obtain reduced complexity iterative techniques for determining the transmit/receive signals or beamformers themselves. The SAVE recursions are similar to PE. However, PE (Jacobi) only converges in case of sufficient diagonal dominance of the matrix to be inverted, whereas SAVE (Gauss-Seidel) is guaranteed to converge, since it corresponds to alternating minimization of a quadratic cost function. Other related work appears in [16], [17] that is more of the PE nature.

II. System Model

We consider a MIMO system with C cells. Each cell c has one base station (BS) of M_c transmit antennas serving K_c users, with total number of users $\sum_c K_c = K$. We refer to the BS of user $k \in \{1,\ldots,K\}$ by b_k. Each user has N_k antennas. The channel between the kth user and the BS in cell c is denoted by $\boldsymbol{H}_{k,c} \in \mathbb{C}^{N_k \times M_c}$. We consider zero-mean white Gaussian noise $\boldsymbol{n}_k \in \mathbb{C}^{N_k \times 1}$ with distribution $\mathcal{CN}(0, \sigma_n^2 \boldsymbol{I})$ at the kth user.

We assume independent unity-power transmit symbols $\mathbf{s}_c = [\mathbf{s}^{\mathrm{T}}_{K_{1:c-1}+1} \ldots \mathbf{s}^{\mathrm{T}}_{K_{1:c}}]^{\mathrm{T}}$, i.e., $\mathbb{E}[\mathbf{s}_c \mathbf{s}_c^{\mathrm{H}}] = \boldsymbol{I}$, where $\mathbf{s}_k \in \mathbb{C}^{d_k \times 1}$ is the data vector to be transmitted to the kth user, with d_k being the number of streams allowed by user k and $K_{1:c} = \sum_{i=1}^{c} K_i$. The latter is transmitted using the transmit filtering matrix $\boldsymbol{\mathcal{G}}_c = [\boldsymbol{\mathcal{G}}_{K_{1:c-1}+1} \ldots \boldsymbol{\mathcal{G}}_{K_{1:c}}] \in \mathbb{C}^{M_c \times N_c}$, with $\boldsymbol{\mathcal{G}}_k = p_k^{1/2} \boldsymbol{G}_k$, $\boldsymbol{G}_k$ being the (unit Frobenius norm) beamforming matrix, p_k is non-negative downlink power allocation of user k and $N_c = \sum_{k:b_k=c} d_k$ is the total number of streams in cell c. Each cell is constrained with $P_{\max,c}$, i.e., the total transmit power in c is limitted such that $\sum_{k:b_k=c} p_k \leq P_{\max,c}$. The received signal at user k in cell b_k is

$$\boldsymbol{y}_k = \underbrace{\boldsymbol{H}_{k,b_k} \boldsymbol{\mathcal{G}}_k \boldsymbol{s}_k}_{\text{signal}} + \underbrace{\sum_{\substack{i \neq k \\ b_i = b_k}} \boldsymbol{H}_{k,b_k} \boldsymbol{\mathcal{G}}_i \boldsymbol{s}_i}_{\text{intracel interf.}} + \underbrace{\sum_{j \neq b_k} \sum_{i: b_i = j} \boldsymbol{H}_{k,j} \boldsymbol{\mathcal{G}}_i \boldsymbol{s}_i}_{\text{intercell interf.}} + \boldsymbol{n}_k \tag{1}$$

Similarly, the receive filtering matrix for each user k is defined as $\boldsymbol{\mathcal{F}}_k^{\mathrm{H}} = p_k^{-1/2} \boldsymbol{F}_k^{\mathrm{H}} \in \mathbb{C}^{d_k \times N_k}$, composed of beamforming matrix $\boldsymbol{F}_k^{\mathrm{H}} \in \mathbb{C}^{d_k \times N_k}$. The received filter output is $\hat{\boldsymbol{s}}_k = \boldsymbol{\mathcal{F}}_k^{\mathrm{H}} \boldsymbol{y}_k$. For details about the (prior) separable channel correlation model and it's impact on the posterior channel model, please see [10]. It leads to e.g. (for arbitrary $\boldsymbol{P}$, $\boldsymbol{Q}$)

$$\mathrm{E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}_d} \boldsymbol{H}^H \boldsymbol{Q} \boldsymbol{H} = \widehat{\boldsymbol{H}}^H \boldsymbol{Q} \widehat{\boldsymbol{H}} + \mathrm{tr}\{\boldsymbol{C}_r \boldsymbol{Q}\} \boldsymbol{C}_p \tag{2}$$

$$\text{and } \mathrm{E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}_d} \boldsymbol{H} \boldsymbol{P} \boldsymbol{H}^H = \widehat{\boldsymbol{H}} \boldsymbol{P} \widehat{\boldsymbol{H}}^H + \mathrm{tr}\{\boldsymbol{C}_p \boldsymbol{P}\} \boldsymbol{C}_r \,. \tag{3}$$

Note that $\rho_P = \frac{\mathrm{tr}\{\widehat{\boldsymbol{H}}^H \widehat{\boldsymbol{H}}\}}{\mathrm{tr}\{\boldsymbol{C}_r\} \mathrm{tr}\{\boldsymbol{C}_p\}}$ is a form of Ricean factor that represents posterior channel estimation quality. It depends on the deterministic channel estimation quality $\rho_D = 1/\sigma^2_{\widetilde{\boldsymbol{H}}}$, which results from the training. Below we consider $\boldsymbol{C}_r = \boldsymbol{I}$ (to keep the posterior covariance separable), and the only covariance $\boldsymbol{C}$ we shall need is the Tx side posterior $\boldsymbol{C}_p$, hence we drop the subscript p. Perfect CSIT algorithms can be obtained by setting $\sigma^2_{\widetilde{\boldsymbol{H}}} = 0$, leading to $\widehat{\boldsymbol{H}} = \boldsymbol{H}$ and $\boldsymbol{C}_p = 0$.

III. Expected Rate Balancing Problem

In this work, we aim to solve the weighted user-rate max-min optimization problem under per cell total transmit power constraint, i.e., the user rate balancing problem expressed as follows

$$\begin{aligned} &\max_{\boldsymbol{G},p} \min_k \; r_k / r_k^{\circ} \\ &\text{s.t.} \sum_{k:b_k=c} p_k \leq P_{\max,c}, 1 \leq c \leq C \end{aligned} \tag{4}$$

where r_k is the kth user-rate

$$r_k = \ln\det\Big(\boldsymbol{I} + \boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{H}_{k,b_k} \boldsymbol{\mathcal{G}}_k \boldsymbol{\mathcal{G}}_k^{\mathrm{H}} \boldsymbol{H}_{k,b_k}^{\mathrm{H}}\Big) = \ln\det\Big(\boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{R}_k\Big), \tag{5}$$

$$\boldsymbol{R}_{\overline{k}} = \sigma_n^2 \boldsymbol{I} + \sum_{l \neq k} \boldsymbol{H}_{k,b_l} \boldsymbol{\mathcal{G}}_l \boldsymbol{\mathcal{G}}_l^{\mathrm{H}} \boldsymbol{H}_{k,b_l}^{\mathrm{H}}, \tag{6}$$

$$\boldsymbol{R}_k = \boldsymbol{R}_{\overline{k}} + \boldsymbol{H}_{k,b_k} \boldsymbol{\mathcal{G}}_k \boldsymbol{\mathcal{G}}_k^{\mathrm{H}} \boldsymbol{H}_{k,b_k}^{\mathrm{H}}, \tag{7}$$

$\boldsymbol{R}_{\overline{k}}$ and $\boldsymbol{R}_k$ are the interference plus noise and total received signal covariances, and r_k° is the rate priority (weight) for user k. Actually, in the presence of partial CSIT, we shall be interested in balancing the expected (or ergodic) rates

$$\begin{aligned} &\max_{\boldsymbol{G},p} \min_k \; \overline{r}_k / r_k^{\circ} \\ &\text{s.t.} \sum_{k:b_k=c} p_k \leq P_{\max,c}, \; c = 1,\ldots,C \end{aligned} \tag{8}$$

where $\overline{r}_k = \mathrm{E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}} r_k$. We shall need

$$\overline{\boldsymbol{S}}_{k,i} = \widehat{\boldsymbol{H}}_{k,b_i} \boldsymbol{G}_i \boldsymbol{G}_i^{\mathrm{H}} \widehat{\boldsymbol{H}}_{k,b_i}^{\mathrm{H}} + \mathrm{tr}\{\boldsymbol{G}_i^{\mathrm{H}} \boldsymbol{C}_{k,b_i} \boldsymbol{G}_i\} \boldsymbol{I}, \; \overline{\boldsymbol{S}}_k = \overline{\boldsymbol{S}}_{k,k} \tag{9}$$

$$\overline{\boldsymbol{R}}_{\overline{k}} = \mathrm{E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}} \boldsymbol{R}_{\overline{k}} = \sigma_n^2 \boldsymbol{I} + \sum_{i \neq k} p_i \overline{\boldsymbol{S}}_{k,i}, \;\; \overline{\boldsymbol{R}}_k = \overline{\boldsymbol{R}}_{\overline{k}} + p_k \overline{\boldsymbol{S}}_k \tag{10}$$

However, the problem presented in (8) is complex and can not be solved directly.

IV. Weighted User EMSE Balancing

Lemma 1. *The user k rate in (5) is lower bounded as [18]*

$$\overline{r}_k = E_{\boldsymbol{H}|\widehat{\boldsymbol{H}}} \max_{\boldsymbol{W}_k, \boldsymbol{\mathcal{F}}_k} \big[\ln\det\big(\boldsymbol{W}_k\big) - \mathrm{tr}\big(\boldsymbol{W}_k \mathbf{E}_k\big) + d_k\big] \tag{11}$$

$$\geq \overline{r}_k^{l} = \max_{\boldsymbol{W}_k, \boldsymbol{\mathcal{F}}_k} \underline{f}_k^{l}, \;\; \underline{f}_k^{l} = \ln\det\big(\boldsymbol{W}_k\big) - \mathrm{tr}\big(\boldsymbol{W}_k \overline{\boldsymbol{E}}_k\big) + d_k \tag{12}$$

where $\underline{f}_k^{l} = \underline{f}_k^{l}(\boldsymbol{W}_k, \boldsymbol{\mathcal{F}}_k)$, $\overline{\boldsymbol{E}}_k = \mathbb{E}\Big[(\hat{\mathbf{s}}_k - \mathbf{s}_k)(\hat{\mathbf{s}}_k - \mathbf{s}_k)^{\mathrm{H}}\Big]$

$$\begin{aligned} &= \boldsymbol{I} - \boldsymbol{\mathcal{F}}_k^{\mathrm{H}} \widehat{\boldsymbol{H}}_{k,b_k} \boldsymbol{\mathcal{G}}_k - \boldsymbol{\mathcal{G}}_k^{\mathrm{H}} \widehat{\boldsymbol{H}}_{k,b_k}^{\mathrm{H}} \boldsymbol{\mathcal{F}}_k + \sigma_n^2 \boldsymbol{\mathcal{F}}_k^{\mathrm{H}} \boldsymbol{\mathcal{F}}_k \\ &+ \sum_{l=1}^{K} \boldsymbol{\mathcal{F}}_k^{\mathrm{H}} \big(\widehat{\boldsymbol{H}}_{k,b_l} \boldsymbol{\mathcal{G}}_l \boldsymbol{\mathcal{G}}_l^{\mathrm{H}} \widehat{\boldsymbol{H}}_{k,b_l}^{\mathrm{H}} + \mathrm{tr}\{\boldsymbol{\mathcal{G}}_l^{\mathrm{H}} \boldsymbol{C}_{k,b_l} \boldsymbol{\mathcal{G}}_l\} \boldsymbol{I}\big) \boldsymbol{\mathcal{F}}_k \end{aligned} \tag{13}$$

is the k^{th} user downlink Expected MSE (EMSE) matrix between the decision variable $\hat{\mathrm{s}}_k$ and the transmit signal $\mathbf{s}_k$, and $\{\boldsymbol{W}_k\}_{1\leq k\leq K}$ are auxiliary weight matrix variables with optimal solution $\boldsymbol{W}_k^{\rm opt}=\overline{\boldsymbol{E}}_k^{-1}$ and the optimal receivers are

$$\boldsymbol{\mathcal{F}}_k=\overline{\boldsymbol{R}}_k^{-1}\widehat{\boldsymbol{H}}_{k,b_k}\boldsymbol{\mathcal{G}}_k. \quad (14)$$

with rate lower bound

$$\overline{r}_k^l=-\ln\det(\boldsymbol{I}-\boldsymbol{\mathcal{G}}_k^{\rm H}\widehat{\boldsymbol{H}}_{k,b_k}^{\rm H}\overline{\boldsymbol{R}}_k^{-1}\widehat{\boldsymbol{H}}_{k,b_k}\boldsymbol{\mathcal{G}}_k)\,. \quad (15)$$

The rest of this EMSE based approach is elaborated in [10]. The main issue that interests us here is that the Lagrangian of the expected rate balancing leads to a Weighted Sum EMSE minimization for the Tx filters with solution

$$\boldsymbol{\mathcal{G}}_k'=(\sum_{l=1}^K(\widehat{\boldsymbol{H}}_{l,b_k}^{\rm H}\boldsymbol{\mathcal{F}}_l\boldsymbol{W}_l'\boldsymbol{\mathcal{F}}_l^{\rm H}\widehat{\boldsymbol{H}}_{l,b_k}+{\rm tr}\{\boldsymbol{\mathcal{F}}_l\boldsymbol{W}_l'\boldsymbol{\mathcal{F}}_l^{\rm H}\}\boldsymbol{C}_{l,b_k})+\mu_{b_k}\boldsymbol{I})^{-1}$$
$$\times\,\widehat{\boldsymbol{H}}_{k,b_k}^{\rm H}\boldsymbol{\mathcal{F}}_k\boldsymbol{W}_k'\,,\;\boldsymbol{\mathcal{G}}_k=\sqrt{p_k}\,\boldsymbol{G}_k\,,\;\boldsymbol{G}_k=\frac{1}{\sqrt{{\rm tr}\{\boldsymbol{\mathcal{G}}_k'^{\rm H}\boldsymbol{\mathcal{G}}_k'\}}}\boldsymbol{\mathcal{G}}_k' \quad (16)$$

where $\boldsymbol{W}_k'=\lambda_k/\xi_k\boldsymbol{W}_k$. This is a Linear MMSE filter in a dual uplink, and the imperfect CSIT aspect has no effect on this nature of the solution. But WSEMSE corresponds to an ergodic rate lower bound. Better ergodic rate approximations are based on rate minorizers leading to generalized eigenvector beamformers. Note that when one substitutes the Rx $\boldsymbol{\mathcal{F}}_k$ from (14) in (16), then one can identify a power method iteration for $\boldsymbol{\mathcal{G}}_k$, which is fine in the perfect CSIT case, but suboptimal in the imperfect CSIT case, because in the (direct link) signal power, the channel estimation error covariance matrix is missing.

V. ESIP Rate Balancing

Now we follow another approximation of the expected rate expression The following approach will use a rate minorizer for every r_k, similar but not identical to what is used as in the DC programming approach which for the optimization of $\boldsymbol{G}_k$ keeps r_k and linearizes the $r_{\overline{k}}$. The approach does not require the introduction of Rxs. We consider again the (ergodic) rate balancing problem (8) where $\overline{r}_k=\;{\rm E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}}r_k$ is now approximated by the Expected Signal and Interference Power (ESIP) rate

$$\begin{aligned}\overline{r}_k=&\;{\rm E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}}\,{\rm lndet}\Big(\boldsymbol{I}+p_k\,\boldsymbol{G}_k^{\rm H}\,\boldsymbol{H}_{k,b_k}^{\rm H}\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{H}_{k,b_k}\boldsymbol{G}_k\Big)\\ \approx&\;{\rm lndet}\Big(\boldsymbol{I}+p_k\,\boldsymbol{G}_k^{\rm H}\;{\rm E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}}\{\boldsymbol{H}_{k,b_k}^{\rm H}(\,{\rm E}_{\boldsymbol{H}|\widehat{\boldsymbol{H}}}\boldsymbol{R}_{\overline{k}})^{-1}\boldsymbol{H}_{k,b_k}\}\boldsymbol{G}_k\Big)\\ =&\;\overline{r}_k^s=f_k^s(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})=\;{\rm lndet}\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\,\overline{\boldsymbol{B}}_k(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})\,\boldsymbol{G}_k\Big)\,, \quad (17)\\ \overline{\boldsymbol{B}}_k(\overline{\boldsymbol{T}}_k)=&\;\widehat{\boldsymbol{H}}_{k,b_k}^{\rm H}\overline{\boldsymbol{T}}_k^{-1}\,\widehat{\boldsymbol{H}}_{k,b_k}+{\rm tr}\{\overline{\boldsymbol{T}}_k^{-1}\}\,\boldsymbol{C}_{k,b_k} \quad (18)\end{aligned}$$

where the $\overline{r}_k$ approximation $\overline{r}_k^s$ in (17) in general is neither an upper nor lower bound but in the Massive MIMO limit becomes a tight upper bound.

Lemma 2. *The approximate $\overline{r}_k$, $\overline{r}_k^s$, can be obtained as $f_k^s(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})=\min_{\overline{\boldsymbol{T}}_k}\underline{f}_k^s(\overline{\boldsymbol{T}}_k,\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})$, with $\underline{f}_k^s(\overline{\boldsymbol{T}}_k,\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})$:*

$$\underline{f}_k^s={\rm lndet}\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\overline{\boldsymbol{B}}_k(\overline{\boldsymbol{T}}_k)\,\boldsymbol{G}_k\Big)+{\rm tr}\{\breve{\boldsymbol{W}}_k(\overline{\boldsymbol{T}}_k-\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})\} \quad (19)$$

where

$$\breve{\boldsymbol{W}}_k=\overline{\boldsymbol{T}}_k^{-1}\big(\widehat{\boldsymbol{H}}_{k,b_k}\boldsymbol{X}_k\,\widehat{\boldsymbol{H}}_{k,b_k}^{\rm H}+{\rm tr}\{\boldsymbol{X}_k\,\boldsymbol{C}_{k,b_k}\}\boldsymbol{I}\big)\overline{\boldsymbol{T}}_k^{-1} \quad (20)$$
$$\textit{with}\;\boldsymbol{X}_k=\boldsymbol{G}_k\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\overline{\boldsymbol{B}}_k(\overline{\boldsymbol{T}}_k)\,\boldsymbol{G}_k\Big)^{-1}\boldsymbol{G}_k^{\rm H} \quad (21)$$

The optimizer is $\overline{\boldsymbol{T}}_k=\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}}$. Also, $\underline{f}_k^s$ is a minorizer for $f_k^s(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})$ as a function of $\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}}$.

Indeed, since $f_k^s(.)$ is a convex function, it gets minorized by its tangent at any point:

$$f_k^s(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}})\geq\underline{f}_k^s=f_k^s(\overline{\boldsymbol{T}}_k)+{\rm tr}\{\frac{\partial f_k^s(\overline{\boldsymbol{T}}_k)}{\partial\overline{\boldsymbol{T}}_k}(\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}}-\overline{\boldsymbol{T}}_k)\} \quad (22)$$

and $\breve{\boldsymbol{W}}_k=-\frac{\partial f_k^s(\overline{\boldsymbol{T}}_k)}{\partial\overline{\boldsymbol{T}}_k}$. Note that for the Perron-Frobenius theory, we need a function that is linear in $\frac{\boldsymbol{p}_{\overline{k}}}{p_k}$, hence we need to work with $\frac{1}{p_k}\overline{\boldsymbol{R}}_{\overline{k}}$ instead of $\overline{\boldsymbol{R}}_{\overline{k}}$.

The Lagrangian formulation in [10] now leads to

$$\begin{aligned}&\sum_k\breve{\lambda}_k'(t\,r_k^o-\underline{f}_k^s)\\ &=-\sum_k\breve{\lambda}_k'\big(\,{\rm lndet}\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\overline{\boldsymbol{B}}_k\,\boldsymbol{G}_k\Big)-\frac{1}{p_k}\,{\rm tr}\{\breve{\boldsymbol{W}}_k\overline{\boldsymbol{R}}_{\overline{k}}\} \quad (23)\\ &+{\rm tr}\{\breve{\boldsymbol{W}}_k\overline{\boldsymbol{T}}_k\}-t\,r_k^o\big)=\sum_k\breve{\lambda}_k(\frac{1}{p_k\,\breve{\xi}_k}\,{\rm tr}\{\breve{\boldsymbol{W}}_k\overline{\boldsymbol{R}}_{\overline{k}}\}-1) \quad (24)\end{aligned}$$

$$\text{where }\breve{\xi}_k={\rm tr}\{\breve{\boldsymbol{W}}_k\overline{\boldsymbol{T}}_k\}+{\rm lndet}\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\overline{\boldsymbol{B}}_k\,\boldsymbol{G}_k\Big)-t\,r_k^o \quad (25)$$

and $\breve{\lambda}_k'=\breve{\lambda}_k/\breve{\xi}_k$, $\overline{\boldsymbol{B}}_k=\overline{\boldsymbol{B}}_k(\overline{\boldsymbol{T}}_k)$. The balancing of the rates in (8) or equivalently the weighted interference plus noise powers in (23) now finds the user powers as the Perron-Frobenius eigenvector of the matrix (see [10])

$$\breve{\boldsymbol{\Lambda}}=\breve{\boldsymbol{\xi}}^{-1}\breve{\boldsymbol{\Psi}}+\frac{1}{\boldsymbol{\theta}^{\rm T}\boldsymbol{p}_{\max}}\breve{\boldsymbol{\xi}}^{-1}\breve{\boldsymbol{\sigma}}\boldsymbol{\theta}^{\rm T}\boldsymbol{C}_C^{\rm T}\text{ with} \quad (26)$$

$$[\breve{\boldsymbol{\Psi}}]_{ij}=\begin{cases}{\rm tr}\{\breve{\boldsymbol{W}}_i(\widehat{\boldsymbol{H}}_{i,b_j}\boldsymbol{G}_j\boldsymbol{G}_j^{\rm H}\widehat{\boldsymbol{H}}_{i,b_j}^{\rm H}+{\rm tr}\{\boldsymbol{G}_j^{\rm H}\boldsymbol{C}_{i,b_j}\boldsymbol{G}_j\}\boldsymbol{I})\}, & i\neq j\\ 0, & i=j\end{cases} \quad (27)$$

$$\boldsymbol{\sigma}_i=\sigma_n^2\,{\rm tr}\{\breve{\boldsymbol{W}}_i\},\breve{\boldsymbol{\xi}}={\rm diag}(\breve{\xi}_1,\ldots,\breve{\xi}_K)\,. \quad (28)$$

The Tx BF and stream power optimization will be based on $\sum_i\frac{\breve{\lambda}_i}{\breve{\xi}_i}\underline{f}_i^s$, which from (23) becomes (apart from noise terms)

$$\sum_k\frac{\breve{\lambda}_k}{\breve{\xi}_k}\underline{f}_k^s=\sum_k\frac{\breve{\lambda}_k}{\breve{\xi}_k}\,{\rm lndet}\Big(\boldsymbol{I}+\boldsymbol{G}_k^{\rm H}\overline{\boldsymbol{B}}_k\,\boldsymbol{G}_k\Big)-\sum_k{\rm tr}\{p_k\boldsymbol{G}_k^H\overline{\boldsymbol{A}}_k\boldsymbol{G}_k\} \quad (29)$$
$$\text{with }\overline{\boldsymbol{A}}_k=\sum_{i\neq k}\frac{\breve{\lambda}_i}{p_i\,\breve{\xi}_i}\big(\widehat{\boldsymbol{H}}_{i,b_k}^{\rm H}\breve{\boldsymbol{W}}_i\widehat{\boldsymbol{H}}_{i,b_k}+{\rm tr}\{\breve{\boldsymbol{W}}_i\}\boldsymbol{C}_{i,b_k}\big)\,. \quad (30)$$

For the optimal Tx BF $\boldsymbol{G}_k$, the gradient of $\sum_i\frac{\breve{\lambda}_i}{\breve{\xi}_i}\underline{f}_i^s-\mu_{b_k}\sum_{i:b_i=b_k}p_i{\rm tr}\{\boldsymbol{G}_i^{\rm H}\boldsymbol{G}_i\}$ with (29) (or (17)) yields

$$\frac{\breve{\lambda}_k}{p_k\,\breve{\xi}_k}\overline{\boldsymbol{B}}_k\boldsymbol{G}_k\,(\boldsymbol{I}+\boldsymbol{G}_k^H\overline{\boldsymbol{B}}_k\boldsymbol{G}_k)^{-1}-(\overline{\boldsymbol{A}}_k+\mu_{b_k}\boldsymbol{I})\,\boldsymbol{G}_k=0\,. \quad (31)$$

The solution is the d_k maximal generalized eigen vectors

$$\boldsymbol{G}_k^{'}=V_{1:d_k}(\overline{\boldsymbol{B}}_k,\overline{\boldsymbol{A}}_k+\mu_{b_k}\boldsymbol{I}),\boldsymbol{G}_k=\boldsymbol{G}_k^{'}\overline{\boldsymbol{P}}_k^{1/2},\boldsymbol{\mathcal{G}}_k=\boldsymbol{G}_k\sqrt{p_k} \quad (32)$$

where the $\overline{\boldsymbol{P}}_k = \text{diag}(p_{k,1}, \ldots, p_{k,d_k})$, $\text{tr}\{\overline{\boldsymbol{P}}_k\} = 1$, are the relative stream powers. Indeed, (31) represents the definition of generalized eigen vectors. Consider

$$\Sigma_k^{(1)} = \boldsymbol{G}_k^{'\text{H}} \overline{\boldsymbol{B}}_k \boldsymbol{G}_k^{'},\ \Sigma_k^{(2)} = \boldsymbol{G}_k^{'\text{H}} \overline{\boldsymbol{A}}_k \boldsymbol{G}_k^{'} \tag{33}$$

then the generalized eigen vectors $\boldsymbol{G}_k^{'}$ of $\overline{\boldsymbol{B}}_k, \overline{\boldsymbol{A}}_k + \mu_{b_k} \boldsymbol{I}$ lead to diagonal matrices $\Sigma_k^{(1)}$, $\Sigma_k^{(2)} + \mu_{b_k} \boldsymbol{G}_k^{'\text{H}} \boldsymbol{G}_k^{'}$. Note that the normalized $\boldsymbol{G}_k^{'}$ are not orthogonal. Then (31) represents the generalized eigen vector condition with associated generalized eigen values in the diagonal matrix $\frac{p_k \breve{\xi}_k}{\breve{\lambda}_k}(\boldsymbol{I} + \Sigma_k^{(1)} \overline{\boldsymbol{P}}_k)$. Also, plugging in generalized eigen vectors into (29) reveals that one should choose the eigen vectors associated to d_k maximal eigen values to maximize (29). Now, premultiplying both sides of (31) by $p_k \boldsymbol{G}_k^{\text{H}}$, summing over all users $k : b_k = c$, taking trace and identifying the last term with $\sum_{k:b_k=c} p_k \text{tr}\{\boldsymbol{G}_k^{\text{H}} \boldsymbol{G}_k\} = P_{max,c}$ allows to solve for

$$\mu_c = \frac{1}{P_{max,c}} \left\lfloor \sum_{k:b_k=c} \text{tr}\{ \frac{\breve{\lambda}_k}{\breve{\xi}_k} \Sigma_k^{(1)} \overline{\boldsymbol{P}}_k (\boldsymbol{I} + \Sigma_k^{(1)} \overline{\boldsymbol{P}}_k)^{-1} - p_k \Sigma_k^{(2)} \overline{\boldsymbol{P}}_k \} \right\rfloor_+ . \tag{34}$$

The $\overline{\boldsymbol{P}}_k$ are themselves found from an interference leakage aware water filling (ILAWF) operation. Substituting $\boldsymbol{G}_k^{'}$ into term k of (29), dividing by p_k, and accounting for the constraint $\text{tr}\{\overline{\boldsymbol{P}}_k\} = 1$ by Lagrange multiplier ν_k, we get the Lagrangian

$$\begin{aligned} &\frac{\breve{\lambda}_k}{p_k \breve{\xi}_k} \ln\det\left(\boldsymbol{I} + \Sigma_k^{(1)} \overline{\boldsymbol{P}}_k\right) - \text{tr}\{(\Sigma_k^{(2)} + \nu_k \boldsymbol{I}) \overline{\boldsymbol{P}}_k\} = \\ &\frac{\breve{\lambda}_k}{p_k \breve{\xi}_k} \ln\det\left(\boldsymbol{I} + \Sigma_k^{(1)} \overline{\boldsymbol{P}}_k\right) - \text{tr}\{(\text{diag}(\Sigma_k^{(2)}) + \nu_k \boldsymbol{I}) \overline{\boldsymbol{P}}_k\}. \end{aligned} \tag{35}$$

Maximizing w.r.t. $\overline{\boldsymbol{P}}_k$ leads to the ILAWF solution

$$\overline{\boldsymbol{P}}_k = \left\lfloor \frac{\breve{\lambda}_k}{p_k \breve{\xi}_k} (\text{diag}(\Sigma_k^{(2)}) + \nu_k \boldsymbol{I})^{-1} - \Sigma_k^{-(1)} \right\rfloor_+ \tag{36}$$

where the Lagrange multiplier ν_k is adjusted (e.g. by bisection) to satisfy $\text{tr}\{\overline{\boldsymbol{P}}_k\} = 1$. Elements in $\overline{\boldsymbol{P}}_k$ corresponding to zeros in $\Sigma_k^{(1)}$ should also be zero. This completes the ESIP rate balancing algorithm derivation (Table I).

VI. Introducing Power Method Iterations

Consider first a single stream scenario. A max generalized eigenvector optimizes a Rayleigh quotient:

$$\arg\max_{\boldsymbol{g}} \frac{\boldsymbol{g}^H \boldsymbol{B} \boldsymbol{g}}{\boldsymbol{g}^H \boldsymbol{A} \boldsymbol{g}} = V_{max}(\boldsymbol{B}, \boldsymbol{A}) \tag{37}$$

where $V_{max}(\boldsymbol{B}, \boldsymbol{A})$ denotes the generalized eigenvector of matrices $\boldsymbol{B}$, $\boldsymbol{A}$ corresponding to their maximum generalized eigenvalue. We observe that power method iterations for a generalized eigenvector BF can actually be found by optimizing an extended Rayleigh quotient:

$$\max_{\boldsymbol{g}} \frac{\boldsymbol{g}^H \boldsymbol{B} \boldsymbol{g}}{\boldsymbol{g}^H \boldsymbol{A} \boldsymbol{g}} = \max_{\boldsymbol{g}} \max_{\widetilde{\boldsymbol{g}}} \frac{|\boldsymbol{g}^H \boldsymbol{B} \widetilde{\boldsymbol{g}}|^2}{\widetilde{\boldsymbol{g}}^H \boldsymbol{B} \widetilde{\boldsymbol{g}}\, \boldsymbol{g}^H \boldsymbol{A} \boldsymbol{g}} \tag{38}$$

TABLE I: ESIPrate based User Rate Balancing

1. initialize: $\boldsymbol{G}_k^{(0,0)} = (\boldsymbol{I}_{d_k} : \boldsymbol{0})^{\text{T}}$, $\boldsymbol{p}_k^{(0,0)} = \boldsymbol{q}_k^{(0,0)} = \frac{P_{\max,c}}{K}$, $m = n = 0$ and fix $n_{\max}, m_{\max}$, r_k°, and $\breve{\boldsymbol{W}}_k^{(0)}$ from (20)
2. compute $\overline{r}_k^{s\,(0)} = \text{lndet}\left(\boldsymbol{I} + \boldsymbol{G}_k^{\text{H}} \overline{\boldsymbol{B}}_k (\frac{1}{p_k} \overline{\boldsymbol{R}}_{\overline{k}})\ \boldsymbol{G}_k\right)$, determine $t = \min_k \frac{\overline{r}_k^{s\,(0)}}{r_k^{\circ}}$, $r_k^{\circ(0)} = t\, r_k^{\circ}$, and $\breve{\xi}_k^{(0)}$ from (25)
3. **repeat**
   - 3.1. $m \leftarrow m + 1$
   - 3.2. **repeat**
     - $n \leftarrow n + 1$
     - *i* update $\overline{\boldsymbol{A}}_k$ from (30)
     - *ii* update μ_c and $\boldsymbol{G}_k^{'}$ from (32),(34)
     - *iii* update $\overline{\boldsymbol{P}}_k$ from (36)
     - *iv* update $\boldsymbol{p}$ and $\boldsymbol{q}$ as maximal eigen vectors of $\breve{\boldsymbol{\Lambda}}$ in (26)
   - 3.3 **until** required accuracy is reached or $n \geq n_{\max}$
   - 3.4 compute $\overline{\boldsymbol{B}}_k(\overline{\boldsymbol{T}}_k)$ and update $\breve{\boldsymbol{W}}_k$ from (20)
   - 3.5 compute $\overline{r}_k^{s\,(m)} = \text{lndet}\left(\boldsymbol{I} + \boldsymbol{G}_k^{\text{H}} \overline{\boldsymbol{B}}_k (\frac{1}{p_k} \overline{\boldsymbol{R}}_{\overline{k}})\ \boldsymbol{G}_k\right)$ and determine $t = \min_k \frac{\overline{r}_k^{s\,(m)}}{r_k^{\circ(m-1)}}$, $r_k^{\circ(m)} = t\, r_k^{\circ(m-1)}$, and update $\breve{\xi}_k$ from (25)
   - 3.6 set $n \leftarrow 0$ and set $(.)^{(n_{\max}, m-1)} \rightarrow (.)^{(0,m)}$ in order to re-enter the inner loop
4. **until** required accuracy is reached or $m \geq m_{\max}$

which involves a Cauchy-Schwartz inequality. The alternating optimization of (38) leads to

$$\boldsymbol{g} = \boldsymbol{A}^{-1} \boldsymbol{B}\, \widetilde{\boldsymbol{g}}\ ,\ \widetilde{\boldsymbol{g}} = \boldsymbol{g} \tag{39}$$

where scale factors in $\boldsymbol{g}$, $\widetilde{\boldsymbol{g}}$ can be adjusted. The alternating updates in (39) are called the power method, here for finding the maximal generalized eigenvector.

Now let's consider the multi-stream case. If we isolate the BF matrix optimization problem for a generic user, we get

$$\begin{aligned} &\max_{\boldsymbol{G}} \{\ln\det(\boldsymbol{I} + \boldsymbol{G}^H \boldsymbol{B} \boldsymbol{G}) - \text{tr}\{\boldsymbol{G}^H (\boldsymbol{A} + \mu \boldsymbol{I}) \boldsymbol{G}\}\} = \\ &\max_{\boldsymbol{G}} \{\max_{\widetilde{\boldsymbol{G}}} \ln\det(\boldsymbol{I} + \boldsymbol{X}) - \text{tr}\{\boldsymbol{G}^H (\boldsymbol{A} + \mu \boldsymbol{I}) \boldsymbol{G}\}\} \\ &\text{where } \boldsymbol{X} = \boldsymbol{G}^H \boldsymbol{B} \widetilde{\boldsymbol{G}} (\widetilde{\boldsymbol{G}}^H \boldsymbol{B} \widetilde{\boldsymbol{G}})^{-1} \widetilde{\boldsymbol{G}}^H \boldsymbol{B} \boldsymbol{G} \end{aligned} \tag{40}$$

The gradient w.r.t. $\boldsymbol{G}^*$ yields:

$$\begin{aligned} &\boldsymbol{B} \widetilde{\boldsymbol{G}} (\widetilde{\boldsymbol{G}}^H \boldsymbol{B} \widetilde{\boldsymbol{G}})^{-1} \widetilde{\boldsymbol{G}}^H \boldsymbol{B} \boldsymbol{G} (\boldsymbol{I} + \boldsymbol{X})^{-1} - (\boldsymbol{A} + \mu \boldsymbol{I}) \boldsymbol{G} = 0 \\ &\Rightarrow\ \boldsymbol{G} = (\boldsymbol{A} + \mu \boldsymbol{I})^{-1} \boldsymbol{B} \widetilde{\boldsymbol{G}} (\widetilde{\boldsymbol{G}}^H \boldsymbol{B} \widetilde{\boldsymbol{G}})^{-1} \widetilde{\boldsymbol{G}}^H \boldsymbol{B} \boldsymbol{G} (\boldsymbol{I} + \boldsymbol{X})^{-1} \\ &\Rightarrow\ \boldsymbol{G}^{'} = (\boldsymbol{A} + \mu \boldsymbol{I})^{-1} \boldsymbol{B} \widetilde{\boldsymbol{G}} \boldsymbol{Y}\ ,\ \boldsymbol{G} = \boldsymbol{G}^{'} \overline{\boldsymbol{P}}^{1/2} \end{aligned} \tag{41}$$

for some $d \times d$ unitary $\boldsymbol{Y}$ and diagonal $\overline{\boldsymbol{P}}$. Before simplifying $\boldsymbol{G}$, consider the solution for $\widetilde{\boldsymbol{G}}$:

$$\boldsymbol{G}^H \boldsymbol{B} \widetilde{\boldsymbol{G}} (\widetilde{\boldsymbol{G}}^H \boldsymbol{B} \widetilde{\boldsymbol{G}})^{-1} \widetilde{\boldsymbol{G}}^H \boldsymbol{B} \boldsymbol{G} \leq \boldsymbol{G}^H \boldsymbol{B} \boldsymbol{G} \tag{42}$$

which reaches equality iff $P_{\widetilde{\boldsymbol{G}}} \boldsymbol{G} = \boldsymbol{G}$ where $P_{\widetilde{\boldsymbol{G}}}$ is the projection matrix onto the column space of $\widetilde{\boldsymbol{G}}$. In other words, $\widetilde{\boldsymbol{G}}$ should have the same column space as $\boldsymbol{G}$. (42) is the Schur complement's lemma, which is the matrix version of the Cauchy-Schwartz inequality. Now, since $\widetilde{\boldsymbol{G}}$ is only determined up to square mixture, we can require $\widetilde{\boldsymbol{G}}$ to satisfy $\widetilde{\boldsymbol{G}}^H \boldsymbol{B} \widetilde{\boldsymbol{G}} = \boldsymbol{I}$, which leads to the solution

$$\widetilde{\boldsymbol{G}} = \boldsymbol{G}\, (\boldsymbol{G}^H \boldsymbol{B} \boldsymbol{G})^{-1/2} \tag{43}$$

in which the lowest complexity computation of the matrix square-root would be via a Cholesky decomposition. With $\widetilde{\boldsymbol{G}}^H\boldsymbol{B}\widetilde{\boldsymbol{G}} = \boldsymbol{I}$, and $\boldsymbol{G}$ as in (41), the crucial terms in (40) become

$$\begin{aligned}\boldsymbol{X} &= \overline{\boldsymbol{P}}^{1/2}\boldsymbol{Y}^H(\widetilde{\boldsymbol{G}}^H\boldsymbol{B}(\boldsymbol{A}+\mu\boldsymbol{I})^{-1}\boldsymbol{B}\widetilde{\boldsymbol{G}})^2\boldsymbol{Y}\overline{\boldsymbol{P}}^{1/2}\\ \boldsymbol{G}^H(\boldsymbol{A}+\mu\boldsymbol{I})\boldsymbol{G} &= \overline{\boldsymbol{P}}^{1/2}\boldsymbol{Y}^H\widetilde{\boldsymbol{G}}^H\boldsymbol{B}(\boldsymbol{A}+\mu\boldsymbol{I})^{-1}\boldsymbol{B}\widetilde{\boldsymbol{G}}\boldsymbol{Y}\overline{\boldsymbol{P}}^{1/2}\end{aligned} \tag{44}$$

which hence get diagonalized simultaneously. Introduce now the eigendecomposition

$$\widetilde{\boldsymbol{G}}^H\boldsymbol{B}(\boldsymbol{A}+\mu\boldsymbol{I})^{-1}\boldsymbol{B}\widetilde{\boldsymbol{G}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H \tag{45}$$

which leads us to choose $\boldsymbol{Y} = \boldsymbol{U}$, which in turn transforms (40) to

$$\max_{\overline{\boldsymbol{P}}}\{\ln\det(\boldsymbol{I}+\boldsymbol{\Sigma}^{(1)}\,\overline{\boldsymbol{P}}) - \mathrm{tr}\{\overline{\boldsymbol{P}}\,(\mathrm{diag}(\boldsymbol{\Sigma}^{(2)})+\nu\boldsymbol{I})\}\} \tag{46}$$

where $\boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^2$, $\boldsymbol{\Sigma}^{(2)} = \boldsymbol{\Sigma} - \mu\boldsymbol{G}'^H\boldsymbol{G}'$, from which the usual ILAWF solution follows. So in summary, the generalized eigenvector computations are replaced by the alternating updates (40), (43), which possibly could be iterated separately more than once, if multiple power method iterations per overal update are desired.

A. Inverse-Free Generalized Power Method

The power method update in (41) for the generalized eigenvector problem is still complex due to the matrix inverse. Since the whole beamformer design is iterative, we may as well do this matrix inversion iteratively also, as in polynomial expansion or the Multi-Stage Wiener Filter. In fact, (41) means that $\boldsymbol{G}'$ is found from a linear system of equations (in which we shall adjust right multiplying factors later)

$$(\boldsymbol{A}+\mu\boldsymbol{I})\,\boldsymbol{G}' = \boldsymbol{B}\,\widetilde{\boldsymbol{G}},\ \boldsymbol{A}+\mu\boldsymbol{I} = \boldsymbol{D}+\overline{\boldsymbol{D}} \tag{47}$$

where the additive decomposition of $\boldsymbol{A}+\mu\boldsymbol{I}$ is into its diagonal part $\boldsymbol{D}$ and offdiagonal part $\overline{\boldsymbol{D}}$ (more sophisticated preconditioning $\boldsymbol{D}$ could be considered, trading complexity for per iteration convergence speed; e.g. taking $\boldsymbol{D}$ to be triangular, which leads to the Gauss-Seidel method, or diagonal plus low rank). Then we get the polynomial expansion (Jacobi) iteration

$$\boldsymbol{G}' = -\boldsymbol{D}^{-1}\,\overline{\boldsymbol{D}}\,\boldsymbol{G}' + \boldsymbol{D}^{-1}\,\boldsymbol{B}\,\widetilde{\boldsymbol{G}} \approx -\boldsymbol{D}^{-1}\,\overline{\boldsymbol{D}}\,\widetilde{\boldsymbol{G}} + \boldsymbol{D}^{-1}\,\boldsymbol{B}\,\widetilde{\boldsymbol{G}}\,. \tag{48}$$

Now, to guarantee and accelerate convergence, we can throw in $d_k \times d_k$ scale factors $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, $\boldsymbol{Q}_3$ that we shall optimize with the same utility function. So we end up with

$$\boldsymbol{G}' = \left[\widetilde{\boldsymbol{G}}\ \ -\boldsymbol{D}^{-1}\overline{\boldsymbol{D}}\widetilde{\boldsymbol{G}}\ \ \boldsymbol{D}^{-1}\boldsymbol{B}\widetilde{\boldsymbol{G}}\right]\begin{bmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\\ \boldsymbol{Q}_3\end{bmatrix} = \boldsymbol{T}\,\boldsymbol{Q}\,. \tag{49}$$

where $\boldsymbol{Q}$, of size $3d_k \times d_k$, is to be optimized and $\widetilde{\boldsymbol{G}}$ is the $\boldsymbol{G}$ of the previous iteration. By substituting $\boldsymbol{G}$ in (40) by the $\boldsymbol{G}'$ expression in (49), the optimization w.r.t. $\boldsymbol{Q}$ leads to

$$\boldsymbol{Q} = V_{1:d_k}(\boldsymbol{T}^H\boldsymbol{B}\boldsymbol{T}, \boldsymbol{T}^H(\boldsymbol{A}+\mu\boldsymbol{I})\boldsymbol{T}) \tag{50}$$

where both matrices again get diagonalized by the small dimension generalized eigenvectors, and the optimization of the stream powers proceeds as usual. Though the complexity of (50) is now small, it could still be reduced by the power method discussed above, where now the matrix inverse should be of acceptable complexity.

VII. Simulation results

We have evaluated numerically the performance of the proposed algorithms by following [10]. At convergence, the same results are reproduced by replacing generalized eigenvalue computations by power method iterations and inverse-free variations proposed here.

Acknowledgements

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the projects EEMW4FIX and CellFree6G (French ANR), and 5G-OPERA (Franco-German 5G ecosystem project).

References

[1] A. Goldsmith, *Wireless Communications*. Cambridge Univ. Press, 2005.

[2] D. Cai, T. Quek, C. W. Tan, and S. H. Low, "Max-Min SINR Coordinated Multipoint Downlink Transmission—Duality and Algorithms," in *IEEE Transactions on Signal Processing*, vol. 60, no. 10, Oct 2012.

[3] M. Schubert and H. Boche, "Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints," *IEEE Trans. Vehic. Tech.*, Jun 2004.

[4] W. Yu and T. Lan, "Transmitter Optimization for the Multi-antenna Downlink with Per-antenna Power Constraints," in *IEEE Trans. Signal Processing*, June 2007.

[5] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC Duality with Multiple Transmit Covariance Constraints," *IEEE Trans. Inform. Theory*, Apr. 2012.

[6] K. Cumanan, L. Musavian, S. Lambotharan, and A. B. Gershman, "SINR Balancing Technique for Downlink Beamforming in Cognitive Radio Networks," *IEEE Signal Process. Lett.*, Feb. 2010.

[7] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: Duality and sum-MSE Minimization," in *IEEE Trans. Signal Process*, vol. 55, no. 11, Nov 2007.

[8] ——, "Capacity Balancing for Multiuser MIMO Systems," in *Proc. IEEE ICASSP*, Apr 2007.

[9] R. Hunger, M. Joham, and W. Utschick, "On the MSE-Duality of the Broadcast Channel and the Multiple Access Channel," in *IEEE Trans. Signal Processing*, vol. 57, no. 2, Feb 2009, p. 698–713.

[10] I. Ghamnia, D. Slock, and Y. Yuan-Wu, "Multi-cell mimo user rate balancing with partial CSIT," in *IEEE Vehic. Tech. Conf. (VTC-Spring)*, 2021.

[11] ——, "MIMO user rate balancing in multicell networks with per cell power constraints," in *IEEE Vehic. Tech. Conf. (VTC-Spring)*, 2020.

[12] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 861–875, 2014.

[13] R. Ghods, C. Jeon, G. Mirza, A. Maleki, and C. Studer, "Optimally-Tuned Nonparametric Linear Equalization for Massive MU-MIMO Systems," in *IEEE International Symposium on Information Theory*, 2017.

[14] M. N. Boroujerdi, S. Haghighatshoar, , and G. Caire, "Low-Complexity Statistically Robust Precoder/Detector Computation for Massive MIMO Systems," in *arXiv preprint arXiv:1711.11405v2*, December 2017.

[15] C. Thomas and D. Slock, "Variational Bayesian Learning for Channel Estimation and Transceiver Determination," in *IEEE Information Theory and Applications (ITA) Workshop*, 2018.

[16] L. Pellaco and M. Bengtsson and J. Jaldén, "Matrix-Inverse-Free Deep Unfolding of the Weighted MMSE Beamforming Algorithm," *IEEE Open Journal Comm's Society*, March 2021.

[17] L. Pellaco and J. Jaldén, "A Matrix-Inverse-Free Implementation of the MU-MIMO WMMSE Beamforming Algorithm," 2022, arXiv:2205.08877.

[18] F. Negro, I. Ghauri, and D. T. Slock, "Sum Rate Maximization in the Noisy MIMO Interfering Broadcast Channel with Partial CSIT via the Expected Weighted MSE," in *Proc. Int'l Symp. on Wireless Communication Systems (ISWCS)*, Paris, France, 2012.