On MSE Balancing in the MIMO Broadcast Channel with Unequal Targets

José P. González-Coma*, Andreas Gründinger†, Michael Joham†, Luis Castedo Ribas* and Wolfgang Utschick†
*Grupo de Tecnología Electrónica y Comunicaciones, Universidade da Coruña, Spain
Email: {jose.gcoma,luis.castedo}@udc.es
†Fachgebiet Methoden der Signalverarbeitung, Technische Universität München, Germany
Email: {gruendinger,joham,utschick}@tum.de

***Abstract*—Average *mean square error* (MSE) balancing in a *multiple-input multiple-output* (MIMO) scenario was studied for multiple streams per-user and even for various power limitations that bound the achievable performance, but usually with equal MSE targets. We now focus on the case where the user MSE targets are different. In this situation, only a subset of the receivers may be served at the solution. At low transmit power, users with low target values are strictly prioritized and transmission to users with large MSE targets can even be switched off. That way, MSE balancing distinguishes from *signal-to-interference-plus-noise* (SINR) and rate balancing optimizations, where transmission to all users is always active. The transmission to users dependent on the balancing level leads to several practical and theoretical questions that we address here.**

I. Introduction

Downlink balancing optimizations have been studied under perfect and imperfect *channel state information* (CSI) and different power restrictions [1]–[5]. The assumption of imperfect CSI at the transmitter is reasonable in practical systems with fading channels, where non-cooperative users detect their channels and send this information back to the transmitter, e.g., as in *frequency-division duplex* (FDD) systems.

Several metrics can be employed for balancing optimizations, viz., the users' MSEs, SINRs, or data rates. The rate is the most important of these metrics for *quality of service* (QoS) requirements in data links, but is difficult to handle if only imperfect CSI is available at the transmitter. Therefore, we exploit the fact that the MSE provides a lower bound for the rate, which was pointed out in several works for perfect and imperfect CSI (see e.g., [6], [7]), to ensure reliable data transmission.

Using the MSE metric, the balancing problem becomes (quasi-)convex if viewed isolated with respect to the transmit or receive filters. More specifically, users compute their equalizers while the precoders are calculated at the transmitter using the available CSI. Thus, an *alternate optimization* (AO) is typically employed for both perfect and imperfect CSI, where the well known *broadcast channel* (BC) to *multiple access channel* (MAC) dualities are exploited for the precoder update steps [8]–[11]. The former dualities, e.g. [11], are appropriate for sum-power limitations but do not apply for the generalized power constraints studied in this work. Hence, an extension for such dualities was proposed to contemplate real setups [12].

In this work we focus on the case of unequal MSE targets which allows a flexible prioritization of users compared to the scenario where the targets (and the number of streams) for all the users are the same. In the latter scenario with equal targets, the MSEs for all the users are equal at the optimum. A similar behavior is observed when the figures of merit are the SINR or the data rate, where the trivial solution corresponds to the zero value for the balancing level. However, as soon as the transmit power is larger than zero, transmission to all the users is active. Remarkably, this behaviour is true for SINR and rate balancing even with unequal targets (see e.g. [1], [13], [14]), and in sharp constrast to weighted MSE balancing with different MSE targets for the users.

If the MSE is the metric for the balancing problem and users have unequal targets, only a subset of users may actively be served at the optimum. Consider a subset of users with very low targets and large targets for the remaining users. Furthermore, let the transmit power be strictly limited. Then, the MSEs are balanced only for the subset of users with low targets. If the other users' MSEs would be balanced as well, their achieved values would lie above the trivial upper bound. In other words, transmission is activated only for the users with high performance demands.

We study this interesting effect of (soft) switching on (and off) the transmission to users dependent on the achievable MSEs in the remainder of this work. Note that this behavior may even be exploited for scheduling in higher layers. The prioritization of users allows to distinguish between primary users, i.e., those with low MSE targets, and secondary users with weak targets. Only if the primary served users achieve a certain threshold, which is defined by the ratios between the target MSEs, the transmission to secondary users is activated.

In particular, we address the following questions that arise because of the transmission deactivation to users with large target values for downlink min-max MSE optimization:

- How similar do we have to choose the MSE targets such that all users are served for limited transmit power? Conversely, what is the minimum required transmit power for an active transmission to all users?
- What is the impact of switching off users in the MSE domain and how does the MSE balancing curve look in the rate domain? In other words, how to choose the MSE targets when we actually aim at rate balancing?

- What is the influence of the multiple power constraints compared to a single sum power constraint within the MSE and rate region?

We answer these questions by simulations and theoretical considerations. For the simulations, we adopt the AO method in [4], [5] to account for unequal MSE target values. Since this solution approach consists of an iterative process, it is necessary to enable each step of the AO to activate and deactivate transmissions to users while minimizing the MSEs.

To address the second of above items, we consider the vector broadcast case with two single antenna receivers in Section IV. We depict the achieved MSE pairs of these users within the feasible region, either employing MSE balancing based on equal and unequal target MSEs, or rate balancing with unequal targets. This figure, together with an intuitive relation between the MMSE and the data rate, helps to understand the different effects for unequal rate and MSE targets.

II. System Model

We consider a MIMO BC where an N-antenna *base station* (BS) sends M_k streams to user $k \in \{1,\dots,K\}$, which is equipped with R_k antennas. The signal is sent over the channel $\boldsymbol{H}_k \in \mathbb{C}^{N\times R_k}$ and is perturbed by the AWGN $\boldsymbol{\eta}_k \in \mathbb{C}^{R_k}$. The estimated data signal at the k-th receiver reads as

$$\hat{\boldsymbol{s}}_k = \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{B}_i\boldsymbol{s}_i + \boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{\eta}_k, \tag{1}$$

where $\boldsymbol{F}_k \in \mathbb{C}^{R_k\times M_k}$, $\boldsymbol{B}_k \in \mathbb{C}^{N\times M_k}$ and $\boldsymbol{s}_k \in \mathbb{C}^{M_k}$, are the equalizer, precoder and data vector for the k-th user, respectively. The MSE between the transmitted and estimated data vectors, i.e., $\mathrm{MSE}_k = \mathbb{E}[\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|_2^2]$, reads as

$$\begin{aligned}\mathrm{MSE}_k =& M_k - 2\Re\left\{\mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{B}_k\right)\right\}\\ &+\sum_{i=1}^{K}\left\|\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{B}_i\right\|_{\mathrm{F}}^2 + \mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\boldsymbol{F}_k\right)\end{aligned} \tag{2}$$

for a given $\boldsymbol{H}_k$ and $\boldsymbol{C}_{\boldsymbol{\eta}_k}$ being the noise covaraince matrix.

Our assumption is that the users acquire full information about the channel and, on the contrary, the BS only knows statistical information about the channel, e.g., $\hat{\boldsymbol{H}}_k = \boldsymbol{H}_k + \boldsymbol{E}_k$, where $\boldsymbol{E}_k$ is the estimation error. Consequently, (2) is not appropriate and we consider the average MSE instead, $\mathbb{E}[\mathrm{MSE}_k]$. This expectation is taken over the channel realizations in contrast to that in the MSE definition (2), and will be denoted by $\overline{\mathrm{MSE}}_k^{\mathrm{DL}}$. Let us now define v_k as a factor scaling the average receive power for the k-th user, $\boldsymbol{F}_k = v_k\tilde{\boldsymbol{F}}_k$. Hence, (2) is rewritten as follows

$$\begin{aligned}\overline{\mathrm{MSE}}_k^{\mathrm{DL}} &= M_k - 2\Re\{v_k^*\,\mathrm{tr}\left(\mathbb{E}\left[\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\right]\boldsymbol{B}_k\right)\} \qquad (3)\\ &+|v_k|^2\Big(\mathbb{E}\Big[\sum_{i=1}^{K}\|\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{B}_i\|_{\mathrm{F}}^2 + \mathrm{tr}\left(\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\tilde{\boldsymbol{F}}_k\right)\Big]\Big).\end{aligned}$$

Since the CSI is perfect at the users, the receive filters are calculated minimizing the MSE, i.e.,

$$\boldsymbol{F}_k^{\mathrm{MMSE}} = \boldsymbol{W}_k^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{B}_k, \tag{4}$$

with $\boldsymbol{W}_k = \boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^{K}\boldsymbol{B}_i\boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}$

At the transmitter, L power restrictions are imposed. Using the following expression, various limitations could be considered, e.g., sum power, per-beam or per-antenna (see [5])

$$\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}\boldsymbol{B}_k\right) = \sum_{k=1}^{K}\left\|\boldsymbol{A}_{k,l}^{1/2}\boldsymbol{B}_k\right\|_{\mathrm{F}}^2 \le P_l, \tag{5}$$

with $l = 1,\dots,L$, where L is the number of restrictions, $\boldsymbol{A}_{k,l} \in \mathbb{C}^{N\times N} \succeq 0$ and $\mathrm{rank}(\sum_{l=1}^{L}\boldsymbol{A}_{k,l}) = N$.

III. Problem Formulation

The main goal is to minimize the maximum ratio over the users between $\overline{\mathrm{MSE}}_k^{\mathrm{DL}}$ and the target, ε_k, while fulfilling the L power restrictions, that is

$$\min_{\{\boldsymbol{F}_k,\boldsymbol{B}_k\}_{k=1}^K}\max_j\frac{\overline{\mathrm{MSE}}_j^{\mathrm{DL}}}{\varepsilon_j}\ \text{s.t.}\ \sum_{k=1}^{K}\left\|\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\right\|_{\mathrm{F}}^2 \le P_l,\ \forall l. \tag{6}$$

Substituting (4) into (3), we obtain

$$\overline{\mathrm{MSE}}_k^{\mathrm{DL}} = M_k - \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{W}_k^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{B}_k\right)\right]. \tag{7}$$

The average MSE is bounded by $0 < \overline{\mathrm{MSE}}_k^{\mathrm{DL}} \le M_k$, where the upper bound occurs when $\boldsymbol{B}_k = \boldsymbol{0}_{N\times M_k}$. Due to this bound, some of the MSE to target ratios $\overline{\mathrm{MSE}}_k^{\mathrm{DL}}/\varepsilon_k \le M_k/\varepsilon_k$ may fulfill the restriction with equality at the optimum. Notice that the optimum of (6) satisfies $\overline{\mathrm{MSE}}_j^{\mathrm{DL}} = \min\{M_j, \varepsilon^{\mathrm{opt}}\varepsilon_j\}, \forall j$, with the balancing level $\varepsilon^{\mathrm{opt}}$. In particular, only the users with sufficiently small targets ε_k are balanced while transmissions to users with too large ε_k are switched off. Assume $\frac{\varepsilon_1}{M_1} \le \frac{\varepsilon_2}{M_2} \le \ldots \le \frac{\varepsilon_K}{M_K}$ and let $\varepsilon^{\mathrm{opt}}$ be the optimum for (6). Moreover, let $\ell \le K$ be the lowest index with $\varepsilon^{\mathrm{opt}}\varepsilon_\ell \ge M_\ell$. Then, only the MSEs of users $1,\dots,\ell-1$ are balanced. This is in contrast to the related SINR and rate balancing, where all the users are active and balanced in the optimum, e.g., $\frac{R_1}{\varrho_1} = \ldots = \frac{R_K}{\varrho_K} > 0$ (e.g., see [15, Section III.] and [16, Theorem 1]). The same reasoning also applies for MSE balancing with equal targets.

Despite this knowledge, the balancing optimization itself is difficult to handle. Even though a closed form representation may be found for the expectation in (7) (e.g., see [17] for zero-mean Gaussian channels) it is still non-convex in the precoders. To overcome this difficulty, we use an AO process to find a local solution for (6). For this process, we split up the precoders into $\boldsymbol{B}_k = \sqrt{p_k}\tilde{\boldsymbol{B}}_k$, with p_k being the transmit power for user k and $\|\tilde{\boldsymbol{B}}_k\|_{\mathrm{F}}^2 = 1$, and define the expected values $\bar{\boldsymbol{H}}_k = \mathbb{E}[\boldsymbol{H}_k\tilde{\boldsymbol{F}}_k]$, $\boldsymbol{R}_k = \mathbb{E}[\boldsymbol{H}_k\tilde{\boldsymbol{F}}_k\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}]$ and $\sigma_k^2 = \mathrm{tr}(\mathbb{E}[\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{\eta}_k}\tilde{\boldsymbol{F}}_k])$. Accordingly, (3) reads as

$$\begin{aligned}\overline{\mathrm{MSE}}_k^{\mathrm{DL}} &= M_k - 2\Re\left\{\sqrt{p_k}v_k^*\,\mathrm{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\tilde{\boldsymbol{B}}_k\right)\right\}\\ &+|v_k|^2\sum_{i=1}^{K}p_i\,\mathrm{tr}\left(\tilde{\boldsymbol{B}}_i^{\mathrm{H}}\boldsymbol{R}_k\tilde{\boldsymbol{B}}_i\right) + |v_k|^2\sigma_k^2.\end{aligned} \tag{8}$$

The AO exploits that (8) is biconvex [18] in the precoders and the equalizer functions within the expectations. The following steps are repeated until convergence (cf. [9]):

1) The equalizer functions $\boldsymbol{F}_k^{\text{MMSE}}$ and powers p_k are first found based on (4) for fixed $\tilde{\boldsymbol{B}}_k$, $k = 1, \ldots, K$.
2) The expected values $\bar{\boldsymbol{H}}_k$, $\boldsymbol{R}_k$, and σ_k^2 are computed.
3) Then, the downlink precoders $\boldsymbol{B}_k$ are optimized as equalizers in the dual uplink, based on (8).

A. Power Allocation

The power allocation in step 1 is responsible for switching users on if ε becomes sufficiently small. The corresponding optimization can equivalently be formulated as

$$\min_{\varepsilon, \boldsymbol{p} \geq \boldsymbol{0}_K} \varepsilon \text{ s.t. } \boldsymbol{p} \geq \boldsymbol{I}(\varepsilon, \boldsymbol{p}), \ \tilde{\boldsymbol{A}}\boldsymbol{p} \leq \boldsymbol{1}, \tag{9}$$

where $\boldsymbol{p} = [p_1, \ldots, p_K]^{\mathrm{T}}$ is the power allocation, $\boldsymbol{1}$ the all-ones vector, and $\boldsymbol{I}(\varepsilon, \boldsymbol{p}) = \boldsymbol{\Gamma}(\varepsilon)\boldsymbol{Z}(\boldsymbol{p})$, with

$$\gamma_k(\varepsilon) = \max\{0, M_k - \varepsilon\varepsilon_k\},$$
$$\boldsymbol{\Gamma} = \operatorname{diag}(\gamma_1(\varepsilon), \ldots, \gamma_K(\varepsilon))$$

the function $\boldsymbol{Z}(\boldsymbol{p}) = [Z_1(\boldsymbol{p}), \ldots, Z_K(\boldsymbol{p})]^{\mathrm{T}}$ is given by

$$Z_k(\boldsymbol{p}) = \left(\mathbb{E}\left[\operatorname{tr}\left(\tilde{\boldsymbol{B}}_k^{\mathrm{H}}\boldsymbol{H}_k\boldsymbol{W}_k^{-1}(\boldsymbol{p})\boldsymbol{H}_k^{\mathrm{H}}\tilde{\boldsymbol{B}}_k\right)\right]\right)^{-1}$$

where $\boldsymbol{W}_k(\boldsymbol{p}) = \boldsymbol{H}_k^{\mathrm{H}}\sum_{i=1}^K \tilde{\boldsymbol{B}}_i\tilde{\boldsymbol{B}}_i^{\mathrm{H}} p_i \boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}$, and

$$[\tilde{\boldsymbol{A}}]_{\ell,k} = P_\ell^{-1}\|\tilde{\boldsymbol{B}}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\|_{\mathrm{F}}^2.$$

We remark that $\boldsymbol{Z}(\boldsymbol{p})$ satisfies the properties of standard interference functions. Therefore, the solution $(\varepsilon^\star, \boldsymbol{p}^\star)$ of (9) is uniquely characterized by the two properties

$$\begin{aligned}\boldsymbol{p}^\star &= \boldsymbol{I}(\varepsilon^\star, \boldsymbol{p}^\star) \\ \varepsilon^\star &= \min\left\{\varepsilon \in \mathbb{R}_+ : \tilde{\boldsymbol{A}}\boldsymbol{I}(\varepsilon, \boldsymbol{p}^\star) \leq \boldsymbol{1}\right\}.\end{aligned} \tag{10}$$

This fixed-point is found, for example, via various adaptations of Yates procedure in [19], Schubert and Boches approach in [1], or using a Newton like method.

Next, we show existence of a sequence $(\varepsilon^{(n)}, \boldsymbol{p}^{(n)})$ to $(\varepsilon^\star, \boldsymbol{p}^\star)$. Notice that for fixed ε, $\boldsymbol{I}(\varepsilon, \boldsymbol{p})$ satisfies the properties of standard interference functions. Therefore, if $\varepsilon^{(n)}$ converges to $\varepsilon^\star$, the sequence $(\varepsilon^{(n)}, \boldsymbol{p}^{(n)})$ converges to the unique optimum $(\varepsilon^\star, \boldsymbol{p}^\star)$.

Consider a feasible tuple $(\varepsilon^{(n)}, \boldsymbol{p}^{(n)})$, i.e., $\boldsymbol{p}^{(n)} \geq \boldsymbol{I}(\varepsilon^{(n)}, \boldsymbol{p}^{(n)})$ and $\tilde{\boldsymbol{A}}\boldsymbol{p}^{(n)} \leq \boldsymbol{1}$, with at least one equality. The minimum feasible balance for the given power allocation is

$$\varepsilon^{(n+1)} = \min\{\varepsilon \in \mathbb{R}_+ : \boldsymbol{I}(\varepsilon, \boldsymbol{p}^{(n)}) \leq \boldsymbol{p}^{(n)}\} \tag{11}$$

such that $\varepsilon^{(n+1)} \leq \varepsilon^{(n)}$. Thus, we obtain a new vector $\tilde{\boldsymbol{p}} = \boldsymbol{I}(\varepsilon^{(n+1)}, \boldsymbol{p}^{(n)})$. However, this update might lead to $\max_l(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{p}}) < 1$. Hence, the power allocation update is

$$\boldsymbol{p}^{(n+1)} = w^{(n+1)}\tilde{\boldsymbol{p}} \tag{12}$$

where $w^{(n+1)} = (\max_l [\tilde{\boldsymbol{A}}\tilde{\boldsymbol{p}}]_l)^{-1}$.

In conclusion, the sequence $\varepsilon^{(n)}$ is monotonically decreasing and lower bounded.

Given the solution $\boldsymbol{p} \geq \boldsymbol{0}_K$ of (9), with $p_k > 0$ for $k = 1, \ldots, \ell - 1$ and $p_k = 0$ for $k = \ell, \ldots, K$, we compute the MMSE receive filters and the required expectations for (8). To keep the flexibility for switching precoders $\boldsymbol{B}_k$ on (or off) within step 3 of the AO iteration, we also compute the receive filters $\boldsymbol{F}_k^{\text{MMSE}}$, $\bar{\boldsymbol{H}}_k$, $\boldsymbol{R}_k$, and σ_k^2, for users $k \geq \ell$, but under the assumption that $p_k = 1$. The objective of the precoder optimization in step 3 of the AO is to minimize the maximum ratios of the MSEs and targets, which inherently contains the decision whether either of the MSEs for users $k = \ell, \ldots, K$ are balanced as well. In particular, the following optimization problem is solved with given expectations based on the MMSE filters, $\boldsymbol{F}_k^{\text{MMSE}} = v_k\tilde{\boldsymbol{F}}_k^{\text{MMSE}}$:

$$\min_{\{v_k, \boldsymbol{B}_k\}_{k=1}^K} \max_j \frac{\overline{\text{MSE}}_j^{\text{DL}}}{\varepsilon_j} \text{ s.t. } \sum_{k=1}^K \left\|\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\right\|_{\mathrm{F}}^2 \leq P_l, \forall l. \tag{13}$$

The solution may be found via a sequence of convex power minimization problems, each of which defines a second order cone program, or alternatively, using uplink-downlink duality.

B. Precoder Update via Convex Optimization

In order to solve (13), we first find the scalar receive filter

$$v_k^{\text{MMSE}} = \frac{\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k\right)}{\sum_{i=1}^K \boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{R}_k\boldsymbol{B}_i + \sigma_k^2}, \tag{14}$$

from which we obtain the following MSE

$$\overline{\text{MSE}}_k^{\text{DL}} = M_k - \frac{\left|\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k\right)\right|^2}{\sum_{i=1}^K \operatorname{tr}\left(\boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{R}_k\boldsymbol{B}_i\right) + \sigma_k^2}. \tag{15}$$

Introducing the slack variable ε, we reformulate the problem as

$$\begin{aligned}&\min_{\{\varepsilon, \boldsymbol{B}_k\}_{k=1}^K} \varepsilon \quad \text{s.t.} \quad \sum_{k=1}^K \left\|\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\right\|_{\mathrm{F}}^2 \leq P_l, \forall l \\ &M_k - \frac{\left|\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k\right)\right|^2}{\sum_{i=1}^K \operatorname{tr}\left(\boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{R}_k\boldsymbol{B}_i\right) + \sigma_k^2} \leq \varepsilon_k\varepsilon, \forall k.\end{aligned} \tag{16}$$

For fixed ε and restricting to positive and real $\operatorname{tr}(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k)$, the constraints in (16) can be reformulated as second order cone restrictions. The latter restriction is without loss of optimality since the constraints of (16) are independent with respect to substituting $\boldsymbol{B}_k' = \mathrm{e}^{\mathrm{j}\,\phi_k}\,\boldsymbol{B}_k$, $k = 1, \ldots, K$. Therewith, problem (13) is quasiconvex on $\{\boldsymbol{B}_k\}_{k=1}^K$ [20]. The formulation with second order cone like constraints reads as

$$\begin{aligned}\varepsilon^{\text{opt}} = &\min_{\{\varepsilon, \boldsymbol{B}_k\}_{k=1}^K} \varepsilon \quad \text{s.t.} \quad \sqrt{\sum_{k=1}^K \left\|\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\right\|_{\mathrm{F}}^2} \leq \sqrt{P_l}, \forall l \\ &\sqrt{\sum_{i=1}^K \operatorname{tr}\left(\boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{R}_k\boldsymbol{B}_i\right) + \sigma_k^2} \leq \frac{\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k\right)}{\sqrt{M_k - \varepsilon_k\varepsilon}}, \forall k.\end{aligned} \tag{17}$$

This problem formulation can be solved via a bisection search on $\varepsilon \in [0, \max_j \frac{M_j}{\varepsilon_j}]$. In each iteration, the following power minimization is solved to test whether a given ε is achievable:

$$\begin{aligned}\alpha^{\text{opt},2} = &\min_{\{\alpha, \boldsymbol{B}_k\}_{k=1}^K} \alpha^2 \text{ s.t.} \sqrt{\sum_{k=1}^K \left\|\boldsymbol{B}_k^{\mathrm{H}}\boldsymbol{A}_{k,l}^{1/2}\right\|_{\mathrm{F}}^2} \leq \alpha\sqrt{P_l}, \forall l \\ &\sqrt{\sum_{i=1}^K \operatorname{tr}\left(\boldsymbol{B}_i^{\mathrm{H}}\boldsymbol{R}_k\boldsymbol{B}_i\right) + \sigma_k^2} \leq \frac{\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\mathrm{H}}\boldsymbol{B}_k\right)}{\sqrt{M_k - \varepsilon_k\varepsilon}}, \forall k.\end{aligned} \tag{18}$$

When $\alpha^{\text{opt},2} \leq 1$, ε is a feasible solution of (16), and for $\alpha^{\text{opt},2} = 1$, the optimum precoders $\{\boldsymbol{B}_k\}_{k=1}^K$ and balancing level ε^{opt} are obtained.

C. Lagrangian Dual Problem Formulation

The basis for the dual formulation and solution to the downlink precoder optimization in (16) is the dual problem of the power minimization in (18), which we derive next. The Lagrangian function of (18) as

$$L(\alpha, \{\boldsymbol{B}_k\}_{k=1}^K, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \alpha^2 \left(1 - \sum_{l=1}^{L} \mu_l P_l\right) + \sum_{k=1}^{K} \lambda_k \sigma_k^2 + \sum_{i=1}^{K} \boldsymbol{b}_i^{\text{H}} \left(\boldsymbol{X}_i - \frac{\lambda_i}{M_i - \varepsilon_i \varepsilon} \bar{\boldsymbol{h}}_i \bar{\boldsymbol{h}}_i^{\text{H}}\right) \boldsymbol{b}_i \tag{19}$$

where we introduced the auxiliary matrix $\boldsymbol{X}_i = \sum_{l=1}^{L} \mu_l \left(\mathbf{I}_{M_i} \otimes \boldsymbol{A}_{i,l}\right) + \sum_{k=1}^{K} \lambda_k \left(\mathbf{I}_{M_i} \otimes \boldsymbol{R}_k\right)$, and the vectors $\bar{\boldsymbol{h}}_k = \text{vec}(\bar{\boldsymbol{H}}_k)$ and $\boldsymbol{b}_k = \text{vec}(\boldsymbol{B}_k)$. Moreover, Lagrangian multipliers $\boldsymbol{u}$ and $\boldsymbol{\lambda}$ similar to those in [2] are employed, preserving the optimality and strong duality. Thus, we get the dual function $d(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\alpha, \{\boldsymbol{B}_k\}_{k=1}^K} L(\alpha, \{\boldsymbol{B}_k\}_{k=1}^K, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

To avoid that $d(\boldsymbol{\lambda}, \boldsymbol{\mu}) \to -\infty$, the Lagrangian multipliers need to fulfill the following constraints [cf. (19)]:

$$\sum_{l=1}^{L} \mu_l P_l \leq 1, \quad \boldsymbol{X}_i - \frac{\lambda_i}{M_i - \varepsilon_i \varepsilon} \bar{\boldsymbol{h}}_i \bar{\boldsymbol{h}}_i^{\text{H}} \succeq \mathbf{0}, \, \forall i.$$

The second restriction can be rewritten into the following form using the Schur's complement [3], [20] condition:

$$M_i - \varepsilon_i \varepsilon - \lambda_i \bar{\boldsymbol{h}}_i^{\text{H}} \boldsymbol{X}_i^{-1} \bar{\boldsymbol{h}}_i \geq 0.$$

The equivalence holds for positive definite matrices $\boldsymbol{X}_i$ and is also valid for singular matrices, using the pseudoinverse of $\boldsymbol{X}_i$ instead of the inverse, if the vector $\bar{\boldsymbol{h}}_i$ lies in the range space of $\boldsymbol{X}_i$. To prove the last statement, observe that the first order moment $\bar{\boldsymbol{H}}_i$ is in the range space of the second order moment $\boldsymbol{R}_i$. Moreover, for $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{\lambda} > \mathbf{0}$, $\boldsymbol{D}_i = \sum_{l=1}^{L} \mu_l \boldsymbol{A}_{i,l} + \sum_{j \neq i} \lambda_j \boldsymbol{R}_j \succeq \mathbf{0}$. Defining $\boldsymbol{C}_i = \boldsymbol{D}_i + \lambda_i \boldsymbol{R}_i$, $\text{rank}(\boldsymbol{C}_i) \geq \text{rank}(\boldsymbol{R}_i)$ since $\boldsymbol{D}_i$ and $\boldsymbol{R}_i$ are positive semidefinite matrices. Therefore, $\bar{\boldsymbol{h}}_i = \text{vec}(\bar{\boldsymbol{H}}_i)$ is in the range space of the blockdiagonal matrix $\boldsymbol{X}_i = (\mathbf{I}_{M_i} \otimes \boldsymbol{C}_i)$.

The dual problem is defined as $\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}} d(\boldsymbol{\lambda}, \boldsymbol{\mu})$, i.e.,[1]

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}} \sum_{k=1}^{K} \lambda_k \sigma_k^2 \quad \text{s.t.} \quad \sum_{l=1}^{L} \mu_l P_l \leq 1, \quad M_i - \varepsilon_i \varepsilon - \lambda_i \bar{\boldsymbol{h}}_i^{\text{H}} \boldsymbol{X}_i^{-1} \bar{\boldsymbol{h}}_i \geq 0, \, \forall i. \tag{20}$$

We alternatively write the latter restrictions on $\boldsymbol{\lambda}$ as

$$\lambda_i \leq \frac{M_i - \varepsilon_i \varepsilon}{\bar{\boldsymbol{h}}_i^{\text{H}} \boldsymbol{X}_i^{-1} \bar{\boldsymbol{h}}_i}, \quad \forall i \tag{21}$$

and remark that their right hand side defines a standard interference function in $\boldsymbol{\lambda}$. Hence, there is a unique $\boldsymbol{\lambda}$ that simultaneously satisfies these constraints with equality and maximizes the objective [19]. Moreover, this allows to minimize with respect to $\boldsymbol{\lambda}$ when inverting the inequality in (21). The dual problem is then given by the max-min formulation

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \sum_{k=1}^{K} \lambda_k \sigma_k^2 \quad \text{s.t.} \quad \sum_{l=1}^{L} \mu_l P_l \leq 1, \quad M_i - \lambda_i \, \text{tr}\left(\bar{\boldsymbol{H}}_i^{\text{H}} \boldsymbol{Y}_i^{-1} \bar{\boldsymbol{H}}_i\right) \leq \varepsilon_i \varepsilon, \, \forall i, \tag{22}$$

where $\boldsymbol{Y}_i = \sum_{l=1}^{L} \mu_l \boldsymbol{A}_{i,l} + \sum_{k=1}^{K} \lambda_k \boldsymbol{R}_k$.

D. Uplink MSE Interpretation

The dual problem (22) can be interpreted as an uplink power minimization, where the MSE for user k reads as

$$\overline{\text{MSE}}_k^{\text{UL}} = M_k - 2\Re\left\{\sqrt{\lambda_k} \, \text{tr}\left(\boldsymbol{G}_k^{\text{H}} \bar{\boldsymbol{H}}_k\right)\right\} + \text{tr}\left(\boldsymbol{G}_k^{\text{H}} \boldsymbol{Y}_k \boldsymbol{G}_k\right), \tag{23}$$

and $\boldsymbol{G}_k$ is the receive filter for user k in the uplink. From the former expression, we identify the interference of the users as $\sum_{i=1}^{K} \lambda_i \boldsymbol{R}_i$ whereas the noise covariance matrix is $\sum_{l=1}^{L} \mu_l \boldsymbol{A}_{k,l}$. The filters minimizing (23) are

$$\boldsymbol{G}_k^{\text{MMSE}} = \boldsymbol{Y}_k^{-1} \bar{\boldsymbol{H}}_k \sqrt{\lambda_k}, \tag{24}$$

with $\lambda_k = 1$ for inactive users. Strong duality and, as a consequence, zero duality gap holds for the power minimization in the downlink and the uplink. Hence, the optimum values of (22) and (18) are the same. Next, we can define the uplink MSE from (22) as the balancing problem[2]

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\{\boldsymbol{G}_k\}_{k=1}^K, \boldsymbol{\lambda} \geq \mathbf{0}} \max_j \frac{\overline{\text{MSE}}_j^{\text{UL}}}{\varepsilon_j} \quad \text{s.t.} \quad \sum_{k=1}^{K} \lambda_k \sigma_k^2 \leq 1, \quad \sum_{l=1}^{L} \mu_l P_l \leq 1. \tag{25}$$

In the following, we study the relationship between the optimum values of (25) and (13).

Consider that ε^{opt} is the optimum value for (13). This is equivalent to getting $\alpha^{\text{opt}} = 1$ as the optimum of (18) for $\varepsilon = \varepsilon^{\text{opt}}$. Moreover, due to strong duality, 1 is also the optimum for the power minimization in the uplink (22). Then, $\sum_{k=1}^{K} \lambda_k \sigma_k^2 = 1$ holds in the optimum of (22) and the MSE constraints are fulfilled with equality for active users. Since the MSE is decreasing when scaling $\boldsymbol{\lambda}$ with $1+\delta$ and $\delta > 0$, both constraints are also fulfilled with equality in the optimum of (25). Note that the restriction over $\boldsymbol{\mu}$ represents the worst case noise covariance matrix, which is obtained for $\sum_{l=1}^{L} \mu_l P_l = 1$. Let us define the optimal values of (22) for ε^{opt}, i.e., $\boldsymbol{\lambda}^{\text{opt}}$ and $\boldsymbol{\mu}^{\text{opt}}$. Consider now that we obtain ε' as the optimal value of (25) for $\boldsymbol{\lambda}'$ and $\boldsymbol{\mu}'$. Since the optimal $\boldsymbol{\lambda}$ is unique for given

[1] Note also that the subsequent derivation can be generalized with the pseudoinverse of $\boldsymbol{X}_i$.

[2] Note that the constraints can be rewritten as $\sum_{k=1}^{K} \lambda_k \sigma_k^2 \leq \sum_{l=1}^{L} \mu_l P_l$ since the restriction is fulfilled with equality in the optimum and $\min_{\boldsymbol{G}_j} \text{MSE}_j$ is independent with respect to a common scaling of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$.

$\boldsymbol{\mu}$, for $\boldsymbol{\mu}' = \boldsymbol{\mu}^{\text{opt}}$ we also obtain $\boldsymbol{\lambda}' = \boldsymbol{\lambda}^{\text{opt}}$. On the contrary, if there is a feasible $\boldsymbol{\mu}' \neq \boldsymbol{\mu}$, two possible cases arise. Either $\varepsilon' < \varepsilon^{\text{opt}}$, due to suboptimal $\boldsymbol{\mu}'$, or $\varepsilon' > \varepsilon^{\text{opt}}$, which would contradict optimality of ε^{opt} and thus also of $\boldsymbol{\mu}^{\text{opt}}$. In summary, the optimum of (25) is ε^{opt}, which is obtained for $\boldsymbol{\lambda}^{\text{opt}}$ and $\boldsymbol{\mu}^{\text{opt}}$, the values achieved at the optimum of (22).

E. Iterative Uplink MSE Balancing

Consider the optimization problem in (25). The inner minimization finds the optimum filters and power allocation, whereas the outer maximization searches for the noise covariance matrix. Notice that the uplink MSE in (23) is affine in $\boldsymbol{\mu}$, as well as the constraints are affine in $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. Consequently, the maximization of (25) is over a concave function with affine constraints [20].

The solution for $\boldsymbol{\lambda}$ is found by the fixed point iteration

$$\lambda_k^{(n+1)} = \frac{\max\{0, M_k - \varepsilon\varepsilon_k\}}{\operatorname{tr}\left(\bar{\boldsymbol{H}}_k^{\text{H}} \boldsymbol{Y}_k(\boldsymbol{\lambda}^{(n)}, \boldsymbol{\mu})^{-1} \bar{\boldsymbol{H}}_k\right)} \tag{26}$$

where $\boldsymbol{Y}_k(\boldsymbol{\lambda}^{(n)}, \boldsymbol{\mu}) = \sum_{l=1}^{L} \mu_l \boldsymbol{A}_{k,l} + \sum_{i=1}^{K} \lambda_i^{(n)} \boldsymbol{R}_i$. From this expression, it is clear that λ_k decreases for increasing balance level ε and we ensure that it is bounded below by 0. Therefore, users with small ratios ε_k/M_k are preferred over those with larger ones because $\lambda_\ell^{(n+1)} = 0$ if $\varepsilon\varepsilon_\ell/M_\ell > 1$. Let these ratios be increasingly ordered in k, i.e., $\varepsilon_1/M_1 \leq \ldots \leq \varepsilon_K/M_K$ and only ℓ be the smallest index with $\varepsilon\varepsilon_\ell/M_\ell > 1$. As a consequence, to guarantee that the power constraint $\sum_{k=1}^{\ell-1} \lambda_k^{(n+1)} \sigma_k^2 \leq 1$ is fulfilled with equality, we compute $\varepsilon^{(n+1)}$ accordingly. Due to the maximization in (26), such an update can be performed by a bisection search for example.

The outer optimization in (25) can be performed via a projected gradient method. First, we compute the gradient step

$$\tilde{\mu}_l = \mu_l^{(n)} + s\delta_l, \tag{27}$$

where $\delta_l = -P_l + \sum_{i=1}^{K} \operatorname{tr}(\boldsymbol{B}_i^{\text{H}} \boldsymbol{A}_{i,l} \boldsymbol{B}_i)$ and s is the step size. Next, $\tilde{\mu}_l$ is projected to the set of feasible values

$$\mu_l^{(n+1)} = \kappa \max\left(0, \tilde{\mu}_l\right), \tag{28}$$

with the scalar factor κ to ensure $\sum_{l=1}^{L} \mu_l^{(n+1)} P_l = 1$.

F. Uplink-Downlink Duality

Knowing the solution vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ in the uplink, we obtain the corresponding downlink precoding vectors via the duality relationships (cf. [10])

$$v_k = \frac{1}{\alpha_k}\sqrt{\lambda_k} \quad \boldsymbol{B}_k = \alpha_k \boldsymbol{G}_k. \tag{29}$$

Observe that we require v_k to be real here, which is in accordance to the restriction to positive and real $\operatorname{tr}(\bar{\boldsymbol{H}}_k^{\text{H}} \boldsymbol{B}_k)$ below (16). Substituting (29) into the MSEs in (8) and equating it to (23) we arrive at

$$\sum_{i\neq k} \frac{\alpha_i^2}{\alpha_k^2} \operatorname{tr}\left(\boldsymbol{G}_i^{\text{H}} \lambda_k \boldsymbol{R}_k \boldsymbol{G}_i\right) + \frac{\lambda_k}{\alpha_k^2}\sigma_k^2 = \sum_{l=1}^{L} \operatorname{tr}\left(\boldsymbol{G}_k^{\text{H}} \mu_l \boldsymbol{A}_{k,l} \boldsymbol{G}_k\right) + \sum_{i\neq k} \operatorname{tr}\left(\boldsymbol{G}_k^{\text{H}} \lambda_i \boldsymbol{R}_i \boldsymbol{G}_k\right), \forall k.$$

Rewriting this set of equalities in matrix notation, we obtain $\boldsymbol{\Gamma}\boldsymbol{\alpha} = \boldsymbol{\Lambda}\boldsymbol{\sigma}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_{\ell-1}, 1, \ldots, 1)$, $\boldsymbol{\sigma} = [\sigma_1^2, \ldots, \sigma_K^2]^{\text{T}}$, $\boldsymbol{\alpha} = [\alpha_1^2, \ldots, \alpha_K^2]^{\text{T}}$ and

$$[\boldsymbol{\Gamma}]_{k,j} = \begin{cases} \sum_{l=1, i\neq k}^{L,K} \operatorname{tr}\left(\boldsymbol{G}_k^{\text{H}} \left(\mu_l \boldsymbol{A}_{k,l} + \lambda_i \boldsymbol{R}_i\right) \boldsymbol{G}_k\right) & j = k \\ -\operatorname{tr}\left(\boldsymbol{G}_j^{\text{H}} \lambda_k \boldsymbol{R}_k \boldsymbol{G}_j\right) & j \neq k, k < \ell \\ -\operatorname{tr}\left(\boldsymbol{G}_j^{\text{H}} \boldsymbol{R}_k \boldsymbol{G}_j\right) & j \neq k, k \geq \ell \\ 0 & j \neq k, j \geq \ell \end{cases}$$

Note that $\boldsymbol{\Gamma}$ is a column-wise diagonal dominant (M-matrix), i.e., $\boldsymbol{\Gamma}^{-1}$ exists and has only non-negative entries.

IV. MMSE and Rate Regions

We next discuss the effect of MSE balancing via a graphical example. To this end, we consider a *multiple-input single-output* (MISO) BC, where the BS is equipped with $N = 2$ antennas and sends data to $K = 2$ single-antenna receivers, i.e., $M_1 = M_2 = 1$. Due to imperfect CSI at the BS, the sum-MSE (SMSE) is lower bounded by $\text{MSE}_1 + \text{MSE}_2 \geq 1$. We sketch this MSE attainable region for unbounded transmit power, $\mathcal{M}$, in Fig. 1. The MSE region was studied in several works where a lower bound for the SMSE [21] or an upper bound for the minimum SINR [2] are computed. To find these bounds, asymptotically high transmit power is considered. The observation in both cases is that the bound depends on the rank of the channel matrix and the number of users.

Empirical regions under different power constraints have been calculated for the Rayleigh channel. The bounds of these regions are sketched via sample points in Fig. 1. We consider the regions under total transmit power for SNR values of 5 dB and 15 dB, represented by $\partial\mathcal{M}_{T5}$ and $\partial\mathcal{M}_{T15}$, respectively. Moreover, the bounds of the regions corresponding to per-antenna and per-user restrictions at 5 dB are also shown and denoted by $\partial\mathcal{M}_{A5}$ and $\partial\mathcal{M}_{U5}$.

For the MSE balancing formulation in (6), we move along a straight line with 45 degree slope by varying the balance factor ε within $(0, 1]$ if $\varepsilon_1 = \varepsilon_2$. In the case of asymmetric targets, e.g., $\varepsilon_2 = 2\varepsilon_1 = 0.5$, transmission to user 2 is inactive for $\varepsilon \geq 2$ since $\varepsilon\varepsilon_2 \geq M_2 = 1$. In this case, the achieved MSE for transmission to user 1 moves along the horizontal line on top of the figure as still $\text{MSE}_1 = \varepsilon_1\varepsilon \leq M_1$. This behavior is only possible when balancing is performed using the MSE metric. The obtained curve for balancing the ratios of the rates R_i over the targets $\varrho_i = -\log_2(\varepsilon_i)$, $i = 1, 2$ is also depicted in Fig. 1. As pointed out previously, transmission to all users is active in the optimal point.

The corresponding rate region and balancing curves to the considered setup is depicted in Fig. 2. The attainable rate region $\mathcal{R}$ is shown, whose upper right bound is reached asymptotically for unconstrained transmit power. The rate region boundaries for sum power restrictions at 5 dB and 15 dB, are $\partial\mathcal{R}_{T5}$ and $\partial\mathcal{R}_{T15}$, respectively. Furthermore, the boundaries corresponding to per-antenna and per-user restrictions at 5 dB have the labels $\partial\mathcal{R}_{A5}$ and $\partial\mathcal{R}_{U5}$. The MSE balancing formulation yields a horizontal line at the R_1 axis as long as transmission to user 2 is inactive, i.e.,

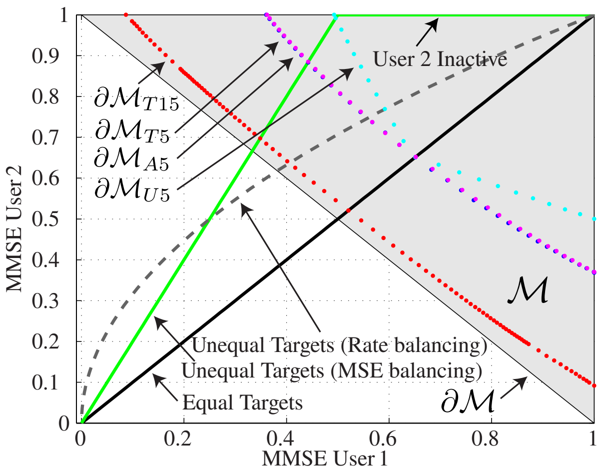

Fig. 1. MSE Region

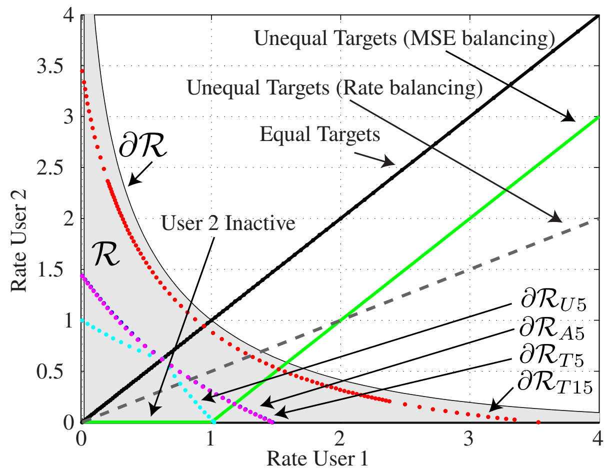

Fig. 2. Rate Region

for $-\log_2(\varepsilon_2\varepsilon) \leq -\log_2(M_2) = 0$ as $M_1 = M_2 = 1$. When $-\log_2(\varepsilon_2\varepsilon) \geq -\log_2(M_2) = 0$, the MSE balancing solution forms a straight line with 45 degree slope starting at $(-\log_2(\frac{\varepsilon_1/M_1}{\varepsilon_2/M_2}), 0)$ for $\varepsilon_1/M_1 \leq \varepsilon_2/M_2$. In contrast, the rate balancing line for the different targets $\varrho_i = -\log_2(\varepsilon_i)$, $i = 1, 2$ starts at the origin and has the slope ϱ_2/ϱ_1. This means that transmission to all users is active when optimizing with respect to the rates.

If perfect CSI is available at the transmitter, reducing the number of users does not affect the MSE feasible region for if the number of antennas at the transmitter is large enough, i.e. $N \geq K$, and the users' channels are linearly independent. Otherwise, the lower bound is reduced [21]. In the following, we study the behavior of the bound when the CSI is imperfect. Taking into account the duality of the MSE regions in the BC and the MAC, the SMSE lower bound reads as

$$\overline{\mathrm{SMSE}} \geq K - \mathrm{tr}\left(\boldsymbol{X}\right) \tag{30}$$

with the matrices $\boldsymbol{X} = \mathbb{E}\left[\boldsymbol{Y}^{\mathrm{H}}\right]\left(\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}\right] + \sigma^2\mathbf{I}\right)^{-1}\mathbb{E}\left[\boldsymbol{Y}\right]$ and $\boldsymbol{Y}$ containing the product of the precoders and the channels in the dual MAC (see [22]). Our assumption is $M_k = 1, \forall k$ since the same analysis applies if we consider every stream as an individual virtual user.

Assume that a certain user is deactivated. Therefore, $K' = K - 1$ and $\mathrm{tr}(\boldsymbol{X}') \leq \mathrm{tr}(\boldsymbol{X})$. Let us rewrite $\boldsymbol{X}'$ in terms of the matrix $\boldsymbol{C_Y} = \mathbb{E}[(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]^{\mathrm{H}})]$ as

$$\boldsymbol{X}' = \mathbb{E}\left[\boldsymbol{Y}^{\mathrm{H}}\right]\left(\boldsymbol{C_Y} + \mathbb{E}\left[\boldsymbol{Y}\right]\mathbb{E}\left[\boldsymbol{Y}^{\mathrm{H}}\right] + \sigma^2\mathbf{I}\right)^{-1}\mathbb{E}\left[\boldsymbol{Y}\right].$$

Therwith, we see that (30) decreases with $\boldsymbol{C_Y}$. In other words, the perfect CSI scenario with $\boldsymbol{C_Y} = \mathbf{0}_N$ and $\mathbb{E}\left[\boldsymbol{Y}\right] = \boldsymbol{Y}$ establishes a strict upper bound for $\mathrm{tr}(\boldsymbol{X}')$. Moreover, an additional upper bound for asymptotically high SNR leads to

$$\overline{\mathrm{SMSE}} \geq K' - \mathrm{tr}\left(\boldsymbol{X}'\right) > K' - \mathrm{rank}(\boldsymbol{Y}),$$

where $\mathrm{rank}(\boldsymbol{Y}) = \min\left(N, K'\right)$. Accordingly, the decrease in the lower bound caused by an inactive user is given by $K' - \min(N, K') - (K - \min(N, K)) = -1 + \beta$, with $\beta \in \{0, 1\}$. That is, the bound decreases or remains the same when a user is switched off. Note that in the asymptotically high SNR regime, any power constraint collapses to the sum transmit power restriction. Hence, the former reasoning applies for general power constraints.

A. *Ergodic Rate Balancing via MSE Minimization*

Data rate is one of the most common metrics to evaluate a system performance. However, rate based problem formulations are complicated to deal with. Hence, it is useful to implement the rate balancing through the MSE domain. Consider a system where only one stream is allocated to each user, $M_k = 1$, $\forall k$. The relationship between average rate and average MSE is (cf. [6], [7])

$$\mathbb{E}[R_k] \geq -\log_2\left(\mathbb{E}\left[\mathrm{MMSE}_k\right]\right). \tag{31}$$

Rate balancing is performed by adapting ϱ such that the rate targets are given by $\varrho\varrho_k$. This corresponds to a straight line starting at the origin in Fig. 2. In the MSE domain, however, the corresponding targets read as $2^{-\varrho\varrho_k}$. Accordingly, to perform a rate balancing in the MSE domain, the MSE targets are set to $\varepsilon_k = 2^{-\varrho_k}$, while the balance scaling in the rate domain translates into the exponent ε_k^{ϱ} for the MSE targets. For the example shown in Fig. 1, this results in a curve that starts at the point $(1, 1)$, where zero power is assigned to the users, and asymptotically reaches the origin for unlimited transmit power.

The scenario with several streams per user is different from single-stream transmission. Indeed, consider the matrix $\boldsymbol{\Sigma}_k$ such that the average MSE for the k-th user is $\mathbb{E}[\mathrm{MSE}] = \mathrm{tr}(\boldsymbol{\Sigma}_k)$, where the filter in (4) is employed. Using spatial decorrelation precoders, the matrix $\boldsymbol{\Sigma}_k$ becomes diagonal with positive entries since it is positive-definite. Such precoders do not modify the MSE nor the rate expression, and can be assumed without loss of generality [23]. Moreover, spatial decorrelation precoders are transparent to the power restrictions in (5). Thus, a lower bound for the average rate is calculated moving the expectation operator inside the concave function $\log_2 \det(\cdot)$, for positive-definite matrices, to get $\mathbb{E}[R_k] \geq \log_2 \det(\boldsymbol{\Sigma}_k^{-1})$ (cf. [23]). To guarantee the rate targets and perform rate balancing it is thereby necessary to satisfy the following condition

$$\varrho\varrho_k = -\log_2 \det\left(\boldsymbol{\Sigma}_k\right) = -\sum_{i=1}^{M_k} \log_2\left(\mathbb{E}\left[\mathrm{MSE}_{k,i}\right]\right). \qquad (32)$$

From the expression above it is clear that no direct (one-to-one) relation holds between the per-user MSEs and rates for this lower bound. To get a tight lower bound for the rate, the product of the per-stream MSEs has to fulfill

$$\prod_{i=1}^{M_k} \mathbb{E}\left[\mathrm{MSE}_{k,i}\right] = 2^{-\varrho\varrho_k}. \qquad (33)$$

V. Numerical Results

In this section, we numerically evaluate the proposed AO for the MSE balancing with imperfect transmitter CSI. In order to reduce the computational complexity, we remove the power update in AO step 1) since it is not necessary for the overall convergence. In the update of $\boldsymbol{\lambda}$ in the dual domain, users are switched on and off to fulfill the MSE restrictions and power constraints. This flexibility enables us to start the algorithm with a single active user for example.

The system setup consists of a $N = 4$ antenna BS allocating $M_k = 2$ streams to each of the $K = 2$ users that are both equipped with $R_k = 2$ antennas. We consider a Rician fading model, where $\boldsymbol{H}_k = \sqrt{\kappa(\kappa+1)^{-1}}\boldsymbol{\Theta}_{\boldsymbol{H}_k} + \sqrt{(\kappa+1)^{-1}}\tilde{\boldsymbol{H}}_k$ with i.i.d. complex Gaussian entries in $\tilde{\boldsymbol{H}}_k$, and a Rician factor $\kappa = 5$ dB. Different Frobenius norms for each user mean are assigned. In particular, we set $\|\boldsymbol{\Theta}_{\boldsymbol{H}_1}\|_{\mathrm{F}}^2 = 2\|\boldsymbol{\Theta}_{\boldsymbol{H}_2}\|_{\mathrm{F}}^2$. The noise covariance matrix is fixed to $\boldsymbol{C}_{\boldsymbol{\eta}_k} = \mathbf{I}_{R_k}, \forall k$ and the MSE targets are $\varepsilon_1 = 1$ and $\varepsilon_2 = 0.5$, respectively. We generate 1000 channel means $\boldsymbol{\Theta}_{\boldsymbol{H}_k}$ and 1000 channel realizations for each of the means.

Figures 3 and 4 illustrate the numerical results obtained with the proposed methods in the above described setup. We obtain the minimum average MSEs $\varepsilon\varepsilon_k$ considering sum, per-user and per-antenna power restrictions. In our model, this sum power constraint implies $\boldsymbol{A}_{k,l} = \mathbf{I}_N \, \forall k,l$, $L = 1$, and $P_1 = 1$. Per-user power constraints are given by $\boldsymbol{A}_{k,l} = \mathbf{I}_N$ for $k = l$ and $\boldsymbol{A}_{k,l} = \mathbf{0}_N$ otherwise, for $L = K$ and $P_l = 0.5, \forall l$. Finally, per-antenna constraints corresponds to $\boldsymbol{A}_{k,l} = \boldsymbol{e}_l\boldsymbol{e}_l^{\mathrm{T}}, \forall k$, $L = N$ and $P_l = 0.25, \forall l$. Moreover, the maximum powers P_ℓ are scaled by a common factor to obtain SNR values from 0 to 30 dB.

The MSE balancing values obtained from different SNRs are shown in Fig. 3. The performance of the sum and per-antenna constraints is very close to each other, whereas a significant gap appears with respect to the per-user restriction in the low SNR regime. Since the number of per-antenna restrictions, 4, is larger than the number of per-user constraints, which is 2, this result could be counterintuitive. However, the setup considered involves a highly unbalanced scenario, where

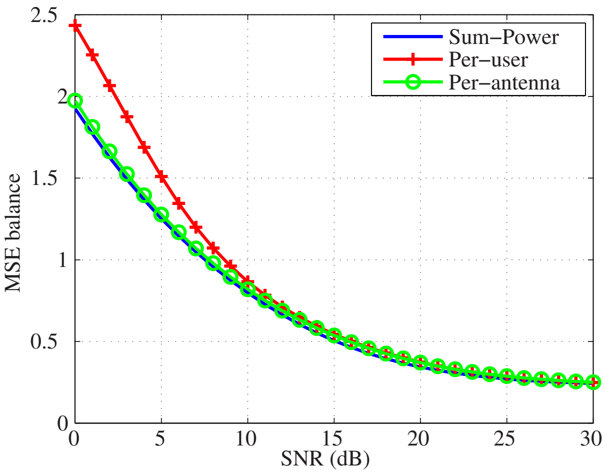

Fig. 3. MSE Balance ε vs SNR

a large portion of the available transmit power is consumed by transmission to user 2.

It is also interesting to observe the low SNR regime, where the user with the smallest priority, that is, user 1, is assigned zero power since $\varepsilon\varepsilon_1 \geq 2$. Then, user 2 demands all the available transmit power to miminize its MSE. Observe that there is no constraint violation when assigning all the power to a certain user for the sum and per-antenna transmit power restrictions, whereas only a half of the available power can be assigned to a user for per-user power limitations.

Empirical CDFs are depicted in Fig. 4 to show the largest per-antenna and per-user powers obtained under the sum power constraint, i.e. $\max_{n\in\{1,\ldots,4\}} p_n$ with $p_n = (\|\boldsymbol{e}_n\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{B}_1\|_{\mathrm{F}}^2 + \|\boldsymbol{e}_n\boldsymbol{e}_n^{\mathrm{T}}\boldsymbol{B}_2\|_{\mathrm{F}}^2)/(\|\boldsymbol{B}_1\|_{\mathrm{F}}^2 + \|\boldsymbol{B}_2\|_{\mathrm{F}}^2)$ and $\max_{k\in\{1,2\}} p_k$ with $p_k = \|\boldsymbol{B}_k\|_{\mathrm{F}}^2/(\|\boldsymbol{B}_1\|_{\mathrm{F}}^2 + \|\boldsymbol{B}_2\|_{\mathrm{F}}^2)$, respectively. The values are normalized with respect to the total power. As can be seen, balanced per-user power allocations ($\max_{k\in\{1,2\}} p_k \approx 0.5$) in the per-user power curve constitute a small portion of the channel realizations, e.g., user 2 power allocations greater than 60% of the total power surpass 95% of the cases. This fact is in accordance with the results obtained in Fig. 3.

VI. Conclusions

The balancing problem under imperfect CSI and different power constraints has been considered. The unbalanced targets scenario in the MSE is emphasized since it makes possible to allocate power to a subset of the users. An AO method was proposed to find the minimum balance level fulfilling the power restrictions via an uplink-downlink duality. The power allocation step allows the AO process to update the filters for both active and inactive users. Moreover, the MSE and the corresponding rate regions are studied via simulation experiments considering different power limitations.

References

[1] M. Schubert and H. Boche, "Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, January 2004.

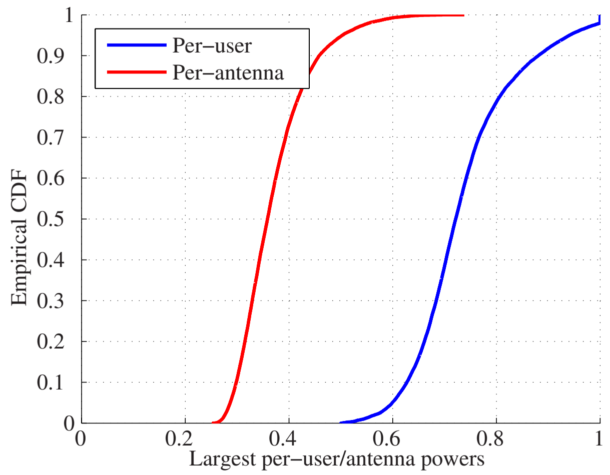

Fig. 4. Empirical CDF of largest per-user and per-antenna powers

[2] 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, January 2006.

[3] Wei Yu and Tian Lan, "Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[4] A. Gründinger, M. Joham, A. Barthelme, and W. Utschick, "Mean square error beamforming in SatCom: Uplink-downlink duality with per-feed constraints," in *Proc. International Symposium on Wireless Communications Systems (ISWCS)*, August 2014, pp. 600–605.

[5] A. Gründinger, M. Joham, J. González-Coma, L. Castedo, and W. Utschick, "Average Sum MSE Minimization in the Multi-User Downlink With Multiple Power Constraints," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, November 2014, pp. 1279–1285.

[6] Petre Stoica, Yi Jiang, and Jian Li, "On MIMO channel capacity: an intuitive discussion," *IEEE Signal Processing Magazine*, vol. 22, no. 3, pp. 83–84, May 2005.

[7] J.P. González-Coma, M. Joham, P.M. Castro, and L. Castedo, "Power minimization in the multiuser downlink under user rate constraints and imperfect transmitter CSI," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 4863–4867.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2658–2668, Oct 2003.

[9] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: Duality and Sum-MSE Minimization," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5436–5446, November 2007.

[10] R. Hunger, M. Joham, and W. Utschick, "On the MSE Duality of the Broadcast Channel and the Multiple Access Channel," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 698–713, February 2009.

[11] M. Joham, M. Vonbun, and W. Utschick, "MIMO BC/MAC MSE Duality with Imperfect Transmitter and Perfect Receiver CSI," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2010, pp. 1–5.

[12] T.E. Bogale, B.K. Chalise, and L. Vandendorpe, "Robust Transceiver Optimization for Downlink Multiuser MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 446–453, January 2011.

[13] F. Boccardi, H. Huang, and M. Trivellato, "Multiuser eigenmode transmission for mimo broadcast channels with limited feedback," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2007, pp. 1–5.

[14] M. Kobayashi and G. Caire, "Joint Beamforming and Scheduling for a Multi-Antenna Downlink with Imperfect Transmitter Channel Knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468–1477, September 2007.

[15] C. Hellings, M. Joham, and W. Utschick, "Gradient-Based Rate Balancing for MIMO Broadcast Channels with Linear Precoding," in *Proc. ITG Workshop on Smart Antennas (WSA)*, Februar 2011.

[16] E. Jorswieck and H. Boche, "Rate Balancing for the Multi-Antenna Gaussian Broadcast Channel," in *Proc. IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications*, 2002, vol. 2, pp. 545–549.

[17] A. Gründinger, M. Joham, and W. Utschick, "Stochastic Transceiver Design in Multi-Antenna Channels with Statistical Channel State Information," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.

[18] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex Sets and Optimization with Biconvex Functions: A Survey and Extensions," *Mathematical Methods of Operations Research*, vol. 66, no. 3, pp. 373–407, May 2007.

[19] R.D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, September 1995.

[20] S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[21] R. Hunger and M. Joham, "A Complete Description of the QoS Feasibility Region in the Vector Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3870 –3878, July 2010.

[22] J.P. González-Coma, M. Joham, P.M. Castro, and L. Castedo, "Power minimization and QoS feasibility region in the multiuser MIMO broadcast channel with imperfect CSI," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2013, pp. 619–623.

[23] J.P. González-Coma, M. Joham, P.M. Castro, and L. Castedo, "Power Minimization in the Multiple Stream MIMO Broadcast Channel with Imperfect CSI," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 165–168.