QoS Constrained Power Minimization in the Multiple Stream MIMO Broadcast Channel

José P. González-Coma[a,*], Michael Joham[b], Paula M. Castro[a], Luis Castedo[a]

[a] *Universidade da Coruña, CITIC, Campus Elviña s/n, A Coruña, 15071, Spain.*
[b] *Technische Universität München, Deparment of Electrical and Computer Engineering, Theresienstrasse 90 München, 80333, Germany.*

Abstract

This work addresses the design of optimal linear transmit filters for the *Multiple Input-Multiple Output* (MIMO) *Broadcast Channel* (BC) when several spatial streams are allocated to each user. We further consider that the *Channel State Information* (CSI) is perfect at the receivers but is only partial at the transmitter. A statistical model for the partial CSI is assumed and exploited for the filter design. The relationship between average rate and average *Mean Square Error* (MSE) is studied to determine the optimal way to distribute the per-user rates among the streams. Finally, the feasible average sum-MSE (sMSE) region is studied and the impact of the CSI uncertainty over the overall system performance is evaluated.

Keywords: Multiple-Input Single-Output, Broadcast Channel, imperfect CSI, QoS constraints, Multiple Stream

---

*Corresponding author

Email addresses: jose.gcoma@udc.es (José P. González-Coma), joham@tum.de (Michael Joham), pcastro@udc.es (Paula M. Castro), luis.castedo@udc.es (Luis Castedo)

1. Introduction

This work focuses on power minimization in the *Multiple Input-Multiple Output* (MIMO) *Broadcast Channel* (BC) when several streams are allocated to each user. Considering more than a single stream per user takes advantage of the spatial multiplexing gain of MIMO systems to increase the communication rate. Furthermore, multiple data streams fit into current scenarios where users connect more than one device, or where different and simultaneous data streams are required for a certain number of applications to simultaneously run at one device.

Our goal is to minimize the total amount of power needed to fulfill certain per-user *Quality-of-Service* (QoS) restrictions taking into account the flexibility of distributing the rate constraints between the different per-user streams. By imposing these restrictions, we ensure that all users achieve a certain level of data rate. This is in contrast to sum-rate maximizations [1] where users with poor channels obtain low data rates, or max-min formulations [2] where these users act as overall performance bottlenecks. Note that the *Base Station* (BS) usually has more degrees of freedom than the individual receivers. Therefore, precoding at the transmitter is helpful to mitigate the interferences between users.

Regarding *Channel State Information* (CSI), we consider that perfect knowledge is available at the receivers but only partial knowledge is available at the BS. This makes a difference with respect to previous works [3, 4, 5, 6, 7, 8] where perfect knowledge of the CSI at both ends of the BC is assumed. Moreover, we do not rely on error models as in [9, 10, 11, 1, 12] to design robust precoders, but leverage statistical knowledge of the channel responses at the BS by means of a *probability density function* (pdf).

Similar assumptions were considered in [13, 14, 15] but for a *Multiple Input-Single Output* (MISO) BC where only one stream per user is considered. The possibility of allocating more than one stream to each user has a profound impact on the problem formulation. Indeed, we show that the *Minimum Mean Square Error* (MMSE)-based lower bound employed in previous works, e.g. [15], is loose in the considered scenario. An accurate approach leads to an additional complexity since the designer has to decide between different per-stream rate targets fulfilling the per-user restrictions [16, 17]. Contrary to [16, 17], a low-complexity update is derived and the possibility of switching on and off some of the per-user streams is also considered.

Finally, the average sum-MSE (sMSE) feasible region is shown to be a generalization of the single stream case. Moreover, we compute a lower bound for the average sMSE which provides insight about the performance loss due to CSI uncertainty. The relationship between average MSE and average rate enables us to obtain conclusions in terms of average rate from the aforementioned lower bound.

The main contributions of this work are the following:

- The relationship between the per-user average MMSE and the per-user average rate is studied for the multiple stream case.
- A new algorithm, which reduces the computational complexity, is proposed to jointly find the filters and update the per-stream targets.
- The algorithm is adapted in such a way that users are allowed to switch on and off some of their data streams.

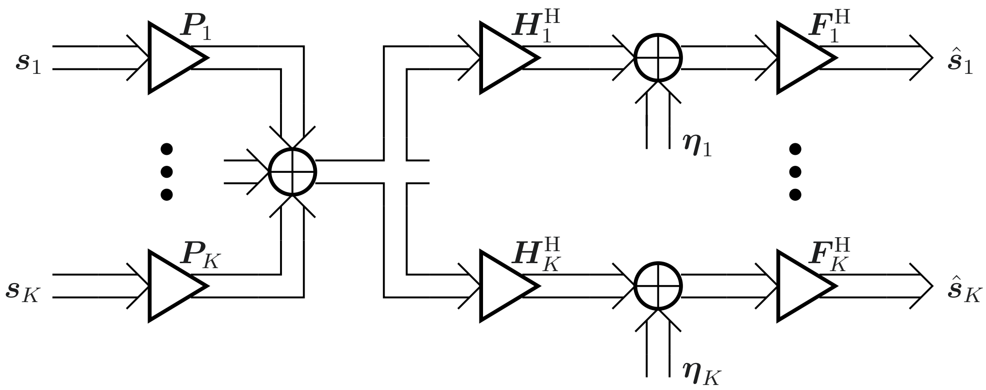


Figure 1: Multiple Stream MIMO BC System Model.

- The impact of CSI uncertainty on the overall system performance is evaluated in a realistic setup, via a sMSE lower bound.

The following notation is employed. Matrices and column vectors are written using upper an lower boldface characters, respectively. By $[\boldsymbol{X}]_{j,k}$, we denote the element in row j and column k of the matrix $\boldsymbol{X}$; $\operatorname{diag}(x_i)$ represents a diagonal matrix whose ith diagonal element is x_i; $\mathbf{I}_N$ stands for the $N \times N$ identity matrix, and $\boldsymbol{e}_i$ represents the canonical vector. The superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$ denote the complex conjugate, transpose, and Hermitian. $\Re\{\cdot\}$ represents the real part operator. Finally, $\mathbb{E}[\cdot]$ stands for statistical expectation, $\operatorname{tr}(\cdot)$ and $\det(\cdot)$ denote the trace and determinant operators, and $|\cdot|$, $\|\cdot\|_2$, $\|\cdot\|_{\mathrm{F}}$ stand for the absolute value, the Euclidean norm, and the Frobenius norm, respectively.

2. System Model

Fig. 1 shows the block diagram of a multiple stream MIMO BC. K users, with R antennas each, receive the information sent from a BS with

N antennas. The data symbols are represented by the vectors $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$ comprising the d_k data streams transmitted to the kth user, $k \in \{1, \ldots, K\}$. Such data vectors are considered to be zero-mean Gaussian with a covariance matrix $\mathbb{E}[\boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{H}}] = \mathbf{I}_{d_k}$, i.e. $\boldsymbol{s}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_{d_k})$, and independent among users, i.e. $\mathbb{E}[\boldsymbol{s}_k \boldsymbol{s}_l^{\mathrm{H}}] = \mathbf{0}$ for $l \neq k$. Prior to be transmitted, the data vectors are precoded with $\boldsymbol{P}_k \in \mathbb{C}^{N \times d_k}$ to produce the signal that propagates over the MIMO channel $\boldsymbol{H}_k \in \mathbb{C}^{N \times R}$. Next, it is perturbed by zero-mean additive Gaussian noise $\boldsymbol{\eta}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_k})$. We assume a block fading-channel that remains constant for a packet of symbols. According to such system model, the k-th user data rate is given by

$$R_k = \log_2 \det \left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k \boldsymbol{P}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{X}_k^{-1} \right), \tag{1}$$

where $\boldsymbol{X}_k = \boldsymbol{C}_{\boldsymbol{\eta}_k} + \boldsymbol{H}_k^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{P}_i \boldsymbol{P}_i^{\mathrm{H}} \boldsymbol{H}_k$ represents the interference from the other users and the noise. The total transmit power is $P_T = \sum_{k=1}^{K} ||\boldsymbol{P}_k||_{\mathrm{F}}^2$.

In this work, we consider that the CSI at the BS, v, is partial and available through the conditional pdfs $f_{\boldsymbol{H}_k|v}(\boldsymbol{H}_k|v)$. This information is used to design the precoders for many packets of a block-fading channel. In other words, v contains the statistical information available at the transmitter to design precoders for multiple channel realizations. A practical example is to model the channel uncertainty as a statistical error and assume that the channel is constant for one symbol transmission. Hence, a channel realization reads as

$$\boldsymbol{H}_k = \hat{\boldsymbol{H}}_k + \tilde{\boldsymbol{H}}_k, \tag{2}$$

with the expectation $\hat{\boldsymbol{H}}_k = \mathbb{E}[\boldsymbol{H}_k | \, v]$ and $\tilde{\boldsymbol{H}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{C}_{\tilde{\boldsymbol{H}}_k})$ being the imperfect CSI error. Thus, the partial CSI available at the transmitter are $\hat{\boldsymbol{H}}_k$ and the error covariance matrix $\boldsymbol{C}_{\tilde{\boldsymbol{H}}_k} = \mathbb{E}[(\boldsymbol{H}_k - \hat{\boldsymbol{H}}_k)(\boldsymbol{H}_k - \hat{\boldsymbol{H}}_k)^{\mathrm{H}} | \, v]$.

On the contrary, users are assumed to perfectly know the channel. Hence, the problem at hand consists on finding the mappings between the precoding matrices and v according to some metric. This difficult problem will be solved point-wise for a given v and, in accordance with the system description, the QoS metric is given by the following kth user conditional average rate

$$\mathbb{E}[R_k|\, v] = \mathbb{E}\left[\log_2 \det\left(\mathbf{I}_R + \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k \boldsymbol{P}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{X}_k^{-1}\right) |\, v\right], \tag{3}$$

which should be larger than a given value ρ_k, thus leading to the formulation

$$\min_{\{\boldsymbol{P}_k(v)\}_{k=1}^K} P_T = \sum_{k=1}^K \|\boldsymbol{P}_k(v)\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \mathbb{E}\left[R_k|\, v\right] \geq \rho_k \,\forall k, \tag{4}$$

where we have remarked the dependency of $\boldsymbol{P}_k$ on v. The optimization problem (4) is difficult to solve due to the QoS restrictions. In the ensuing section, we propose more manageable MMSE based constraints. A similar strategy was employed in [13, 15]. However, the multiple data streams make it necessary a new approach.

3. MMSE-Based Problem Formulation

Let us introduce the receive filter $\boldsymbol{F}_k \in \mathbb{C}^{R \times d_k}$, which is applied to the received signal to obtain the data estimates for user k

$$\hat{\boldsymbol{s}}_k = \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum\nolimits_{i=1}^K \boldsymbol{P}_i \boldsymbol{s}_i + \boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{\eta}_k. \tag{5}$$

Note that linear filters consider the inter-user interference as noise. Accordingly, the multiple-stream BC MSE, is $\mathrm{MSE}_k^{\mathrm{BC}} = \mathbb{E}[\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|_2^2]$, i.e.,

$$\mathrm{MSE}_k^{\mathrm{BC}} = \mathrm{tr}\left(\mathbf{I}_{d_k} - 2\Re\left\{\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k\right\}\right) + \sum_{i=1}^K \left\|\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_i\right\|_{\mathrm{F}}^2 + \mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_k} \boldsymbol{F}_k\right).$$

Since the CSI is imperfect at the BS, the appropriate MSE metric is the conditional average $\mathbb{E}[\text{MSE}|\, v] = \overline{\text{MSE}}_k^{\text{BC}}(v)$. This is in accordance with the rate in (3). Recall, however, that CSI is perfect at the receiver side and hence the MMSE receive filters for given precoders $\boldsymbol{P}_k(v)$ can be determined as

$$\boldsymbol{F}_k^{\text{MMSE}} = \left(\boldsymbol{H}_k^{\text{H}}\boldsymbol{P}_k(v)\boldsymbol{P}_k^{\text{H}}(v)\boldsymbol{H}_k + \boldsymbol{X}_k\right)^{-1}\boldsymbol{H}_k^{\text{H}}\boldsymbol{P}_k(v), \tag{6}$$

with $\boldsymbol{X}_k = \boldsymbol{H}_k^{\text{H}}\sum_{i\neq k}\boldsymbol{P}_i(v)\boldsymbol{P}_i^{\text{H}}(v)\boldsymbol{H}_k + \boldsymbol{C}_{\boldsymbol{\eta}_k}$. Plugging (6) into MSE_k^{BC} yields the following expression for the k-th user average minimum MSE

$$\overline{\text{MMSE}}_k^{\text{BC}}(v) = \mathbb{E}\left[\text{tr}\left(\boldsymbol{\Sigma}_k(v)\right)|\, v\right], \tag{7}$$

where $\boldsymbol{\Sigma}_k(v) = (\mathbf{I}_{d_k} + \boldsymbol{P}_k^{\text{H}}(v)\boldsymbol{H}_k\boldsymbol{X}_k^{-1}\boldsymbol{H}_k^{\text{H}}\boldsymbol{P}_k(v))^{-1}$. Observe from (3) that the average rate can be written as a function of $\boldsymbol{\Sigma}_k(v)$, as follows

$$\mathbb{E}\left[R_k|\, v\right] = \mathbb{E}\left[\log_2\det\left(\boldsymbol{\Sigma}_k^{-1}(v)\right)\,|\, v\right]. \tag{8}$$

Notice that $\boldsymbol{\Sigma}_k(v)$ in (7) and (8), as well as $\mathbb{E}\left[\boldsymbol{\Sigma}_k(v)|\, v\right]$, are symmetric positive-semidefinite. We thereby introduce the eigenvalue decomposition $\mathbb{E}[\boldsymbol{\Sigma}_k(v)|\, v] = \boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{U}_k^{\text{H}}$, with the unitary matrix $\boldsymbol{U}_k$ and the diagonal matrix $\boldsymbol{\Lambda}_k = \text{diag}(\lambda_{k,1},\ldots,\lambda_{k,d_k})$, where $\lambda_{k,i}\geq 0$, $\forall k,i$, are the eigenvalues.

The columns of $\boldsymbol{U}_k$ form a basis that enables to introduce the spatial decorrelation precoders $\boldsymbol{P}'_k(v) = \boldsymbol{P}_k(v)\boldsymbol{U}_k$. Such precoders remove the off-diagonal elements of $\mathbb{E}[\boldsymbol{\Sigma}_k(v)|\, v]$, for all k, without changing the total transmit power $\sum_{k=1}^{K}\|\boldsymbol{P}'_k(v)\|_{\text{F}}^2 = \sum_{k=1}^{K}\|\boldsymbol{P}_k(v)\|_{\text{F}}^2$, nor the expressions of the average rate (3) and the average MMSE (7). We henceforth consider that the spatial decorrelation precoders $\boldsymbol{P}'_k$ are employed from now on. Thus, the per-user average MMSE in the BC is

$$\overline{\text{MMSE}}_k^{\text{BC}}(v) = \text{tr}\left(\mathbb{E}\left[\boldsymbol{\Sigma}_k(v)|v\right]\right) = \sum_{i=1}^{d_k}\lambda_{k,i}. \tag{9}$$

Notice that $\lambda_{k,i}$ can be interpreted as the k-th user i-th stream average MMSE, i.e. $\lambda_{k,i} = \overline{\text{MMSE}}_{k,i}^{\text{BC}}(v)$, and the average MMSE in (9) corresponds to the sum of all of them, i.e., $\overline{\text{MMSE}}_{k}^{\text{BC}}(v) = \sum_{i=1}^{d_k} \overline{\text{MMSE}}_{k,i}^{\text{BC}}(v)$.

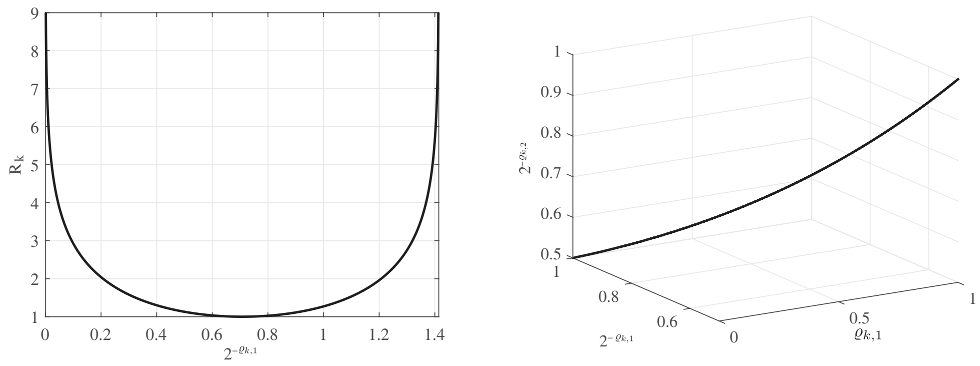
We now study the relationship between the per-user average rate and the per-user average MMSE. The average rate from (8) yields

$$\begin{aligned}\mathbb{E}[R_k|\, v] &\overset{\text{(a)}}{\geq} -\log_2 \det\left(\mathbb{E}\left[\boldsymbol{\Sigma}_k(v)|\, v\right]\right) \overset{\text{(b)}}{=} -\sum_{i=1}^{d_k} \log_2(\lambda_{k,i}) \\ &= -\log_2\left(\prod_{i=1}^{d_k} \lambda_{k,i}\right) \overset{\text{(c)}}{\geq} -d_k \log_2\left(\frac{\overline{\text{MMSE}}_k}{d_k}\right), \qquad (10)\end{aligned}$$

where (a) comes from Jensen's inequality since $f(\boldsymbol{A}) = -\log(\det(\boldsymbol{A}))$, with $\boldsymbol{A}$ being positive semidefinite, is convex [18]. However, contrary to the single stream scenario [15], a lower bound based on the per-user average MMSE is not tight in general. Equality (b) holds because we are using the aforementioned spatial decorrelation precoders, i.e., the streams can be separately encoded without loss of optimality [cf. (9), [19]]. Finally, (c) results from the inequality between arithmetic and geometric means, and equality holds for $d_k = 1$ or when $\lambda_{k,1} = \lambda_{k,2} = \ldots = \lambda_{k,d_k}$. For such cases, the original constraints are ensured if $\overline{\text{MMSE}}_k \leq d_k 2^{-\rho_k/d_k}$. Since equal per-stream MMSEs do not lead to the smallest transmit power in general, we propose to introduce per-stream MMSE targets by imposing per-stream rate constraints $\varrho_{k,i}$ to obtain an accurate bound for the average rate. Indeed, the per-user average rate requirements are satisfied when

$$\overline{\text{MMSE}}_{k,i}^{\text{BC}}(v) = \lambda_{k,i} \leq 2^{-\varrho_{k,i}}, \qquad (11)$$

where $\sum_{i=1}^{d_k} \varrho_{k,i} = \rho_k \, \forall k$ or, equivalently, $\prod_{i=1}^{d_k} \lambda_{k,i} \leq 2^{-\rho_k} \, \forall k$.



(a) (b)

Figure 2: (a) Per-User Rate for per-User MMSE Target of 0.5; (b) MMSE Targets satisfying per-User Rate Requirement.

The lack of precision of the per-user MMSE targets as lower bounds of the per-user rate targets for the multiple stream scenario is illustrated by the perfect CSI example in Fig. 2a. Such figure shows the rates achieved for user k, which allocates two streams, for given per-user MMSE target of $\sqrt{2}$. We vary the per-stream MMSE targets $2^{-\varrho_{k,1}}$ and $2^{-\varrho_{k,2}}$ between 0 and $\sqrt{2}$, such that $2^{-\varrho_{k,1}} + 2^{-\varrho_{k,2}} = \sqrt{2}$. Observe that $R_k = \rho_k = 1$ for $2^{-\varrho_{k,1}} = 2^{-\varrho_{k,2}}$.

Fig. 2b depicts the curve corresponding to the pairs of per-stream MMSE targets $(2^{-\varrho_{k,1}}, 2^{-\varrho_{k,2}})$ ensuring the per-user rate $\rho_k = 1$. The per-user rate target $\varrho_{k,1}$ goes from 0 to 1 while $\varrho_{k,2} = 1 - \varrho_{k,1}$.

We next find the targets leading to the smallest transmit power. Towards this aim we resort to a nested optimization procedure to solve (4). The outer procedure finds the optimum way to split the target rate ρ_k into the d_k per-stream target rates $\varrho_{k,i}$, i.e.

$$\min_{\{\boldsymbol{\varrho}_k\}_{k=1}^K} P_T(\boldsymbol{\varrho}) \text{ s.t. } \mathbf{1}^{\mathrm{T}}\boldsymbol{\varrho}_k = \rho_k, \text{ and } \boldsymbol{\varrho}_k \geq \mathbf{0}\ \forall k, \tag{12}$$

with $\boldsymbol{\varrho} = [\boldsymbol{\varrho}_1^{\mathrm{T}}, \ldots, \boldsymbol{\varrho}_K^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{\varrho}_k = [\varrho_{k,1}, \ldots, \varrho_{k,d_k}]^{\mathrm{T}}$. The inner optimization determines the minimum transmit power for given per-stream average rate targets $\boldsymbol{\varrho}$, that is, $P_T(\boldsymbol{\varrho})$ is the solution to the following variational problem

$$\min_{\{\boldsymbol{P}_k(v), \boldsymbol{F}_k\}_{k=1}^K} \sum_{k=1}^{K} \|\boldsymbol{P}_k(v)\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \overline{\mathrm{MSE}}_{k,i}^{\mathrm{BC}}(v) \leq 2^{-\varrho_{k,i}}\, \forall k, i. \tag{13}$$

This strategy is similar to the so-called "margin adaptive" rate allocation, proposed for multiuser DSL scenarios in [20]. Note that this new formulation allows us to treat each user's streams as virtual users [see (9)]. Thus, we solve (13) point-wise for each v as done in [15]. Notice, however, that this per-stream constrained optimization is more stringent than the original one.

4. Projected generalized gradient

Similarly to [21], the optimization problem (12) is solved in the dual MAC. To this end, we introduce $\boldsymbol{t}_{k,i}$, $\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2}$, $\boldsymbol{g}_{k,i}$ and $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I}_N)$ to represent the precoders, the channel, the equalizer and the noise, respectively. We now define the average transmit power for the k-th user i-th stream $\xi_{k,i} = \mathbb{E}[\|\boldsymbol{t}_{k,i}\|_2^2 | \, v]$, the normalized precoders $\boldsymbol{\tau}_{k,i} = \xi_{k,i}^{-1} \boldsymbol{t}_{k,i}$, and the expectations $\boldsymbol{\mu}_{k,i} = \mathbb{E}[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{\tau}_{k,i} | \, v]$ and $\boldsymbol{\Theta}_{k,i} = \mathbb{E}[\boldsymbol{H}_k \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{\tau}_{k,i} \boldsymbol{\tau}_{k,i}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-1/2} \boldsymbol{H}_k^{\mathrm{H}} | \, v]$. Let us introduce the scalar equalizers $r_{k,i}$ with $\boldsymbol{g}_{k,i} = r_{k,i} \tilde{\boldsymbol{g}}_{k,i}$. Note that the precoders in the dual MAC are functions of the channel whereas the receivers depend on the imperfect CSI v. Hence the average MSE reads as

$$\begin{aligned} \overline{\mathrm{MSE}}_{k,i}^{\mathrm{MAC}}(v) = {} & 1 - 2\Re\left\{ r_{k,i}^* \tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \boldsymbol{\mu}_{k,i} \sqrt{\xi_{k,i}} \right\} \\ & + |r_{k,i}|^2 \left(\tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \sum\nolimits_{l=1}^{K} \sum\nolimits_{j=1}^{d_l} \xi_{l,j} \boldsymbol{\Theta}_{l,j} \tilde{\boldsymbol{g}}_{k,i} + \|\tilde{\boldsymbol{g}}_{k,i}\|_2^2 \right), \end{aligned} \tag{14}$$

and the minimum average MSE is

$$\overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v) = 1 - \xi_{k,i} \left| \tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \boldsymbol{\mu}_{k,i} \right|^2 y_{k,i}^{-1}, \tag{15}$$

with the scalar $y_{k,i} = \tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \sum_{l=1}^{K} \sum_{j=1}^{d_l} \xi_{l,j} \boldsymbol{\Theta}_{l,j} \tilde{\boldsymbol{g}}_{k,i} + \|\tilde{\boldsymbol{g}}_{k,i}\|_2^2$.

Consequently, we rewrite the optimization problem (13) as

$$P_T(\boldsymbol{\varrho}) = \min_{\left\{\boldsymbol{\tau}_{k,i}, \tilde{\boldsymbol{g}}_{k,i}, \xi_{k,i}\right\}_{k,i}^{K,d_k}} \sum_{m=1}^{K} \sum_{n=1}^{d_m} \xi_{m,n} \text{ s.t. } \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v) \leq 2^{-\varrho_{k,i}} \, \forall k, \forall i. \tag{16}$$

Remember that this latter optimization problem can be solved in a way similar to the case of a single stream per user (see [15]).

In order to solve (12), we propose a generalized gradient-projection algorithm. In such algorithm, the direction of the generalized gradient is followed but it is projected onto the set of values fulfilling the original per-user restrictions. Observe that the objective function of (12) is the solution of (16) and the derivative might not exist. The generalized gradient, however, extends the gradient definition to the scenario where the function is possibly non differentiable but locally Lipschitz [22]. This is appropriate because the latter condition is satisfied for any feasible targets. Furthermore, the generalized gradient is a subgradient and, if the function is differentiable at a certain point, it is equivalent to the actual gradient. Indeed, we update the per-stream average rate targets as

$$\varrho\prime_{k,i}^{(\ell+1)} = \varrho_{k,i}^{(\ell)} - s^{(\ell)} \frac{\partial P_T(\boldsymbol{\varrho}^{(\ell)})}{\partial \varrho_{k,i}^{(\ell)}}, \tag{17}$$

with the diminishing step size $s^{(\ell)} \to 0$ for the ℓ-th iteration, such that $\sum_{\ell=1}^{\infty} s^{(\ell)} = \infty$. Note that the conventional notation for the partial derivative is used for the generalized gradient. In the ensuing sections we propose two approximations to determine the generalized gradient in (17).

4.1. Jacobian matrix approach

To compute the generalized gradient in (17), we start deriving (15) with respect to the power allocation elements $\xi_{m,n}$. For the cases: $m=k$, $n=i$, and $m \neq k$ or $n \neq i$, we get

$$\frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{k,i}} = -\frac{\left|\tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}}\boldsymbol{\mu}_{k,i}\right|^2}{y_{k,i}^2}\left(y_{k,i} - \xi_{k,i}\tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}}\boldsymbol{\Theta}_{k,i}\tilde{\boldsymbol{g}}_{k,i}\right), \text{and} \tag{18}$$

$$\frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{m,n}} = \frac{\xi_{k,i}\left|\tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}}\boldsymbol{\mu}_{k,i}\right|^2\tilde{\boldsymbol{g}}_{m,n}^{\mathrm{H}}\boldsymbol{\Theta}_{m,n}\tilde{\boldsymbol{g}}_{m,n}}{y_{k,i}^2},$$

respectively. Taking into account the dependency of the transmit power $P_T(\boldsymbol{\varrho})$ with respect to the per-stream targets and that the equality $\overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}} = 2^{-\varrho_{k,i}}$ holds in the solution of (16), we get that the generalized gradient

$$\frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \varrho_{l,j}} = \sum_{m=1}^{K}\sum_{n=1}^{d_m}\frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{m,n}}\frac{\partial \xi_{m,n}}{\partial \varrho_{l,j}}, \tag{19}$$

is equal to $-\ln(2)2^{-\varrho_{k,i}}$ for $k,i=l,j$, and 0 for $k,i \neq l,j$. The Jacobian matrix of $\boldsymbol{f}(\boldsymbol{\xi}) = [\overline{\mathrm{MMSEs}}_{1,1}^{\mathrm{MAC}}(v),\ldots,\overline{\mathrm{MMSEs}}_{K,d_K}^{\mathrm{MAC}}(v)]^{\mathrm{T}}$ reads as

$$[\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})]_{a,b} = \frac{\partial \overline{\mathrm{MMSE}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{l,j}}, \tag{20}$$

where $a = \sum_{m=1}^{k-1} d_m + i$, and $b = \sum_{m=1}^{l-1} d_m + j$. Similarly, we define the matrix $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{\varrho}) = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\varrho}^{\mathrm{T}}}$. Hence, we rewrite (19) as

$$\frac{\partial \overline{\mathrm{MMSE}}_{k,i}^{\mathrm{MAC}}}{\partial \varrho_{l,j}} = \left[\boldsymbol{J}_{\boldsymbol{f}}\left(\boldsymbol{\xi}\right)\boldsymbol{J}_{\boldsymbol{\xi}}\left(\boldsymbol{\varrho}\right)\right]_{a,b} = -\ln(2)\left[\boldsymbol{W}\right]_{a,b}, \tag{21}$$

with $\boldsymbol{W} = \mathrm{diag}(2^{\varrho_{1,1}}, \dots, 2^{\varrho_{1,d_1}}, \dots, 2^{\varrho_{K,d_K}})$. Hence, $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{\varrho})$, which contains the entries necessary to compute the update in (17), is obtained by left multiplying (21) times the inverse of $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})$, that is

$$\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{\varrho}) = -\ln(2)\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})^{-1}\boldsymbol{W}. \tag{22}$$

Therefore, the update for the k-th user i-th stream reads as

$$\frac{\partial P_T(\boldsymbol{\varrho})}{\partial \varrho_{k,i}} = -\ln(2)\mathbf{1}^{\mathrm{T}}\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})^{-1}\boldsymbol{W}\boldsymbol{e}_{\sum_{m=1}^{k-1} d_m + i}. \tag{23}$$

We now prove that $-\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})$ is a Z-matrix, i.e., the diagonal elements are positive and the off-diagonal ones are negative. Indeed, let us define the diagonal matrix $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{\xi})$. Observe that the following inequality $\sum_{b \neq a} |[-\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})\boldsymbol{D}]_{a,b}| < [-\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})\boldsymbol{D}]_{a,a}$ holds for every row $a = \sum_{j=1}^{k-1} d_j + i$, corresponding to the k-th user i-th stream. Hence, $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})\boldsymbol{D}$ is strictly diagonally dominant and $-\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})$ is a non-singular M-matrix with positive inverse [23]. This result aligns with the intuition that a lower target rate $\varrho_{l,j}$ also leads to a lower transmit power $P_T(\boldsymbol{\varrho})$.

4.2. Lagrangian multiplier approach

Observe that the update step given by (23) requires to compute the inverse of the Jacobian matrix. The dimensions of such matrix depend on the number of users and streams allocated by the BS. For example, a BS serving to $K = 10$ users allocating 4 streams each, leads to a 40×40 matrix. We thereby provide an alternative to avoid this computationally costly calculation.

Let us define the Lagrangian function corresponding to the i-th stream of user k of the inner optimization problem (16) as

$$L(\boldsymbol{\xi}, \beta_{k,i}) = \psi(\boldsymbol{\xi}) + \beta_{k,i}\nu_{k,i}(\boldsymbol{\xi}), \tag{24}$$

with the objective function $\psi(\boldsymbol{\xi}) = \sum_{k=1}^{K}\sum_{i=1}^{d_k}\xi_{k,i}$, the constraint $\nu_{k,i}(\boldsymbol{\xi}) = \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v) - 2^{-\varrho_{k,i}}$, and $\beta_{k,i} \geq 0$. The *Karush-Kuhn-Tucker* (KKT) necessary conditions for a local optimum [24, Theorem 4.2.13] state that

$$\frac{\partial L(\boldsymbol{\xi}^{\mathrm{opt}}, \beta_{k,i}^{\mathrm{opt}})}{\partial \xi_{k,i}} = \frac{\partial \psi(\boldsymbol{\xi}^{\mathrm{opt}})}{\partial \xi_{k,i}} + \beta_{k,i}^{\mathrm{opt}} \frac{\partial \nu_{k,i}(\boldsymbol{\xi}^{\mathrm{opt}})}{\partial \xi_{k,i}} = 0. \tag{25}$$

Equating the two terms in (25) we readily obtain

$$\frac{\partial \psi(\boldsymbol{\xi}^{\mathrm{opt}})}{\partial \xi_{k,i}} = -\beta_{k,i}^{\mathrm{opt}} \frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{k,i}}. \tag{26}$$

Thus, taking into account that $\overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v) = 2^{-\varrho_{k,i}}$ in the optimum, making explicit the dependency on the targets $\xi_{k,i}(\boldsymbol{\varrho})$, and using the chain rule we arrive at

$$\frac{\partial \psi(\boldsymbol{\xi}^{\mathrm{opt}}(\boldsymbol{\varrho}))}{\partial \varrho_{k,i}} \frac{\partial \varrho_{k,i}}{\partial \xi_{k,i}} = \beta_{k,i}^{\mathrm{opt}} \ln(2) 2^{-\varrho_{k,i}} \frac{\partial \varrho_{k,i}}{\partial \xi_{k,i}}. \tag{27}$$

This interpretation is similar to that in [18, Sec. 5.6] for convex problems.

Finally, in order to calculate the Lagrangian multiplier, we compute the derivative in (25) using (18), leading to

$$\beta_{k,i} = -\left(\frac{\partial \overline{\mathrm{MMSEs}}_{k,i}^{\mathrm{MAC}}(v)}{\partial \xi_{k,i}}\right)^{-1} = \frac{y_{k,i}^2}{\left|\tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \boldsymbol{\mu}_{k,i}\right|^2 \left(y_{k,i} - \xi_{k,i} \tilde{\boldsymbol{g}}_{k,i}^{\mathrm{H}} \boldsymbol{\Theta}_{k,i} \tilde{\boldsymbol{g}}_{k,i}\right)}. \tag{28}$$

Note that (27) is positive. Hence, the gradient step (17) reduces the targets.

4.3. Projection

It is important to note that after the target update (17), the per-stream targets do not fulfill the original constraints $\sum_{i=1}^{d_k} \varrho'_{k,i} = \rho_k$. Therefore, we propose to perform a projection onto the set of feasible target rates of the k-th user by minimizing the following Euclidean distance

$$\min_{\varrho_{k,i} \geq 0} \sum_{i=1}^{d_k} (\varrho_{k,i} - \varrho'_{k,i})^2 \quad \text{s.t.} \quad \sum_{i=1}^{d_k} \varrho_{k,i} = \rho_k. \tag{29}$$

The KKT conditions of (29) lead to the following projection

$$\varrho_{k,i} = \max\left\{\varrho'_{k,i} - \mu_k, 0\right\}, \quad \mu_k = \frac{1}{d_k}\left(\sum_{i=1}^{d_k} \varrho'_{k,i} - \rho_k\right). \tag{30}$$

Note that some of the k-th user per-stream targets could be switched off (i.e. $\varrho_{k,i} = 0$) after the projection. In such a case, the power assigned to that user is $\xi_{k,i} = 0$, and the corresponding generalized gradient is also zero. This way the stream will not be switched on again. Such behavior, observed in [17], is avoided by using "dummy" filters. That is, the updates of $\boldsymbol{\tau}_{k,i}, \tilde{\boldsymbol{g}}_{k,i}$ in (16) are performed with $\xi_{k,i} = 1$ (cf. [15], see also step 9 of Algorithm 1). Consequently, "dummy" filters do not affect the entries of $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{\xi})$ for active streams while the entries of the inactive streams are forced to be non-zero.

4.4. Proposed Algorithm

Algorithm 1 implements the solution previously described. After the initialization in line 1, the power minimization (13) is solved via the methods proposed in [15] (see line 2). Next, the generalized gradient is computed in line 5, and the step size is set to the initial value s_0. Line 7 updates the per-stream target rates $\varrho_{k,i}$ according to (17), while the projection in (30) is implemented in line 8. Next, the power minimization is updated (see line 9). Then, if the BC total power is smaller than that achieved in the previous iteration, the per-stream target rates and the corresponding transmit and receive filters are updated. If not, the step size s is reduced in line 13. If the initial QoS constraints are feasible, this simple line search guarantees convergence to a local minimum since the generalized gradient leads to either a projected gradient or a projected subgradient (see [25, Prop. 2.2.1] and [25,

Algorithm 1 Power Minimization Algorithm

1: $\ell \leftarrow 0$, initialization: $\boldsymbol{P}_k$ and $\varrho_{k,i}^{(0)}$, $\boldsymbol{P}_k \leftarrow \boldsymbol{P}'_k, \forall k$

2: Solve (13), i.e. $\boldsymbol{t}_{k,i}^{(0)}$, $\boldsymbol{g}_{k,i}^{(0)}$, and $\xi_{k,i}^{(0)}\, \forall k, i$.

3: **repeat**

4: $\quad \ell \leftarrow \ell + 1$

5: $\quad \delta_{k,i}^{(\ell)} \leftarrow \frac{\partial P_T(\boldsymbol{\varrho}^{(\ell-1)})}{\partial \varrho_{k,i}^{(\ell-1)}}$, $\forall k, i$, $b_{\text{exit}} \leftarrow 0$, $s^{(\ell)} \leftarrow s_0$

6: $\quad$ **repeat**

7: $\qquad \varrho\prime_{k,i}^{(\ell)} \leftarrow \varrho_{k,i}^{(\ell-1)} - s^{(\ell)} \delta_{k,i}^{(\ell)}, \forall k, i$. Gradient step (17)

8: $\qquad \varrho_{k,i}^{(\ell)} \leftarrow \max\{\varrho\prime_{k,i}^{(\ell)} - \mu_k^{(\ell)}, 0\}\ \forall k, i$. Projection

9: $\qquad$ Solve (13), i.e. $\boldsymbol{t}_{k,i}^{(\ell)}$, $\boldsymbol{g}_{k,i}^{(\ell)}$, and $\xi_{k,i}^{(\ell)}$, $\forall k, i$

10: $\qquad$ **if** $P_T^{(\ell-1)} - P_T^{(\ell)} > 0$ **then**

11: $\qquad\quad b_{\text{exit}} \leftarrow 1$

12: $\qquad$ **else**

13: $\qquad\quad s^{(\ell)} \leftarrow \frac{s^{(\ell)}}{2}$. Step size update

14: $\qquad$ **end if**

15: $\quad$ **until** b_{exit}

16: **until** $\sum_{k=1}^{K} \sum_{i=1}^{d_k} \xi_{k,i}^{(\ell-1)} - \sum_{k=1}^{K} \sum_{i=1}^{d_k} \xi_{k,i}^{(\ell)} \leq \gamma$

Section 6.3.1]). Finally, we set the threshold γ in line 16 to check the accuracy of the solution.

5. Problem Feasibility and sMSE Lower Bound

For feasibility testing, we generalize the single-stream vector channel procedure employed in [26] to the multiple-stream MIMO channel. We thereby obtain the matrix

$$\boldsymbol{E} = \mathbf{I}_d - \mathbb{E}\left[\boldsymbol{\Upsilon}^{\mathrm{H}}|v\right]\left(\mathbb{E}\left[\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\mathrm{H}}|v\right] + \sigma^2\mathbf{I}_N\right)^{-1}\mathbb{E}\left[\boldsymbol{\Upsilon}|v\right], \tag{31}$$

using $\boldsymbol{T}_l = [\boldsymbol{t}_{l,1}, \ldots, \boldsymbol{t}_{l,d_l}]$, $\boldsymbol{\Upsilon}_l = \boldsymbol{H}_l\boldsymbol{T}_l \in \mathbb{C}^{N\times d_l}$ and $\boldsymbol{\Upsilon} = [\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_K] \in \mathbb{C}^{N\times d}$, with the total number of data streams $d = \sum_{k=1}^{K} d_k$. Note that $\mathrm{tr}(\boldsymbol{E})$ is the average sum-MMSE, and $\boldsymbol{E}$ contains the average MMSEs for the k-th user i-th stream, $\overline{\mathrm{MMSE}}_{k,i}$, in the entry $[\boldsymbol{E}]_{a,a}$, with $a = \sum_{l=1}^{k-1} d_l + i$. Accordingly, the k-th user average MMSE, $\overline{\mathrm{MMSE}}_k$, corresponds to $\mathrm{tr}([\boldsymbol{E}]_{b:c,b:c})$, with $b = \sum_{l=1}^{k-1} d_l + 1$ and $c = \sum_{l=1}^{k} d_l$. When setting $\sigma^2 = 0$, $\mathrm{tr}(\boldsymbol{E})$ gives the sum-MMSE lower bound for the set of precoders $\{\boldsymbol{T}_k\}_{k=1}^{K}$, i.e., feasible MMSE targets have to fulfill

$$\sum_{k=1}^{K}\sum_{i=1}^{d_k} 2^{-\varrho_{k,i}} \geq d - \mathrm{tr}(\mathbb{E}\left[\boldsymbol{\Upsilon}^{\mathrm{H}}|v\right]\left(\mathbb{E}\left[\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{\mathrm{H}}|v\right]\right)^{-1}\mathbb{E}\left[\boldsymbol{\Upsilon}|v\right]). \tag{32}$$

Note that if $\boldsymbol{\varrho}$ is feasible, any distribution among the streams $\boldsymbol{\varrho}'$ such that $\sum_{k=1}^{K}\sum_{i=1}^{d_k} 2^{-\varrho'_{k,i}} = \sum_{k=1}^{K}\sum_{i=1}^{d_k} 2^{-\varrho_{k,i}}$ satisfies the inequality (32).

5.1. Sum-MSE Lower Bound

This section studies the bound in (32) to provide more insight about the impact of imperfect CSI in the overall BC performance. Unlike for the power minimization problem, the distribution of the per-user average MSE among the streams is irrelevant for the sum-MSE lower bound. Accordingly, we can assume $\overline{\mathrm{MMSE}}_k = d_k\lambda_k$, $\forall k$ in (10) to obtain a tight bound of the per-user average rate. Moreover, since (31) turns out to be a direct extension of the single-stream MIMO case, we will focus on the latter for simplicity.

First, we consider perfect CSI at the BS. Hence, the expectations in (32) can be removed and the lower bound is 0 for $N \geq d$ (cf. [27]). However, under imperfect CSI assumption, the lower bound in (32) depends on the MAC precoders $\{\boldsymbol{T}_k\}_{k=1}^K$. Therefore, we need to provide transmit and receive filters minimizing the average sum-MSE lower bound to investigate the performance loss caused by the uncertainty of the statistical CSI available at the BS.

Consider the sMSE expression in the MAC for zero-power noise, i.e.,

$$\mathrm{sMSE}^{\mathrm{MAC}} = K - 2\Re\left\{\sum_{k=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{t}_k\right\} + \sum_{k=1}^{K}\sum_{i=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_i \boldsymbol{t}_i \boldsymbol{t}_i^{\mathrm{H}} \boldsymbol{H}_i^{\mathrm{H}} \boldsymbol{g}_k. \tag{33}$$

Recall that we have considered perfect CSI at the receivers in the BC. Therefore, the filters $\boldsymbol{t}_k$ minimizing the sum-MSE are readily obtained as the MMSE MAC precoders $\boldsymbol{t}_k = (\boldsymbol{H}_k^{\mathrm{H}} \sum_{i=1}^K \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{H}_k)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{g}_k$, to get

$$\mathrm{sMSE}^{\mathrm{MAC}} = K - \sum\nolimits_{k=1}^{K} \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{H}_k \big(\boldsymbol{H}_k^{\mathrm{H}} \sum\nolimits_{i=1}^{K} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \boldsymbol{H}_k\big)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{g}_k. \tag{34}$$

Thus, using $\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K]$ and our imperfect CSI assumption we get

$$\overline{\mathrm{sMSE}}^{\mathrm{MAC}}(v) = K - \sum_{k=1}^{K} \boldsymbol{e}_k^{\mathrm{T}}\, \mathbb{E}\left[\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k\right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G} \middle| \, v\right] \boldsymbol{e}_k. \tag{35}$$

To find an analytical solution of the receive filters $\boldsymbol{G}$ is difficult due to the non-convexity of (35) and the unknown closed-form expression of the expectation. Hence, we propose to use a steepest descent algorithm. Let us start defining the gradient for user k as

$$\begin{aligned}\boldsymbol{\delta}_k = \frac{\partial \overline{\mathrm{sMSE}}^{\mathrm{MAC}}(v)}{\partial \boldsymbol{G}^*} \boldsymbol{e}_k = -\,\mathbb{E}\left[\boldsymbol{H}_k \left(\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_k\right)^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{G} \middle| \, v\right] \boldsymbol{e}_k \\ + \sum_{m=1}^{K} \mathbb{E}\left[\boldsymbol{H}_m \left(\boldsymbol{H}_m^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_m\right)^{-1} \boldsymbol{A}_m \left(\boldsymbol{H}_m^{\mathrm{H}} \boldsymbol{G}\boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_m\right)^{-1} \boldsymbol{H}_m^{\mathrm{H}} \boldsymbol{G} \middle| \, v\right] \boldsymbol{e}_k,\end{aligned} \tag{36}$$

with $\boldsymbol{A}_m = \boldsymbol{H}_m^{\mathrm{H}} \boldsymbol{G} \boldsymbol{e}_m \boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{G}^{\mathrm{H}} \boldsymbol{H}_m$ (see Appendix A). The receive filters at iteration ℓ are updated as

$$\boldsymbol{G}^{(\ell)} \boldsymbol{e}_k = \boldsymbol{G}^{(\ell-1)} \boldsymbol{e}_k + s \boldsymbol{\delta}_k^{(\ell)}, \tag{37}$$

with the step size s. Note that $\overline{\mathrm{sMSE}}^{\mathrm{MAC}}(v)$ is reduced at every iteration. This fact, together with the lower bound $\overline{\mathrm{sMSE}}^{\mathrm{MAC}}(v) \geq 0$, guarantees convergence of the proposed method to a local optimum.

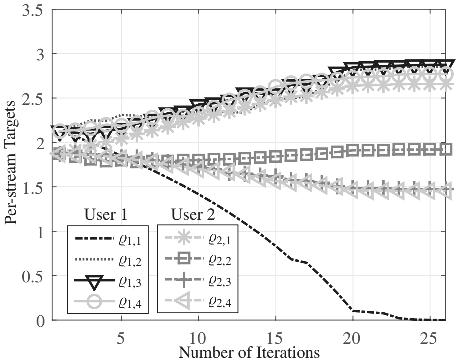
6. Results and Discussion

The performance of Algorithm 1 is illustrated here by considering a MIMO BC with $K = 2$ users, $R = 6$ receive antennas per user and $N = 8$ transmit antennas. Notice that d_k sets an upper bound for the number of streams that might be allocated to user k since Algorithm 1 can switch some streams off but not include additional ones. Taking into account that there is no prior information regarding users' channel qualities, we start with the fair criterion of $d_1 = d_2 = 4$. The AWGN is zero-mean with $\boldsymbol{C}_{\boldsymbol{\eta}} = \mathbf{I}_R$, and the per-user target rates are set to $\rho_1 = 8.5$ and $\rho_2 = 7.5$ bits per channel use. The error model in (2) is assumed, and we consider first and second order moments $[\mathbb{E}[\boldsymbol{H}_k|v]]_{1:N,r} = \boldsymbol{u}_{k,r}$, for each $r \in \{1, \ldots, R\}$ with $u_{k,r,n} = \mathrm{e}^{j(n-1)\varphi_k}$ and $\varphi_k \sim \mathcal{U}(0, 2\pi)$, and $\boldsymbol{C}_{\tilde{\boldsymbol{H}}_k} = R\mathbf{I}_N$, $\forall k$. Recall that no closed-form expressions for the expectations in (15) have been found. Therefore, we employ Monte Carlo numerical integration with $M = 1\,000$ channel realizations. This way, we calculate

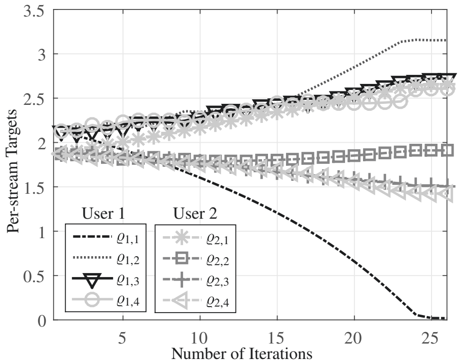
$$\boldsymbol{\mu}_{k,i} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{H}_k^{(m)} \boldsymbol{C}_{\boldsymbol{\eta}_k} \boldsymbol{\tau}_{k,i}^{(m)}$$

$$\boldsymbol{\Theta}_{k,i} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{H}_k^{(m)} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\mathrm{H}/2} \boldsymbol{\tau}_{k,i}^{(m)} \boldsymbol{\tau}_{k,i}^{(m),\mathrm{H}} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{-\frac{1}{2}} \boldsymbol{H}_k^{(m),\mathrm{H}}$$

where $\boldsymbol{\tau}_{k,i}^{(m)}$ is the normalized MAC precoder for the m-th channel realization. In our first experiment, we perform a comparison between the two approaches



(a)



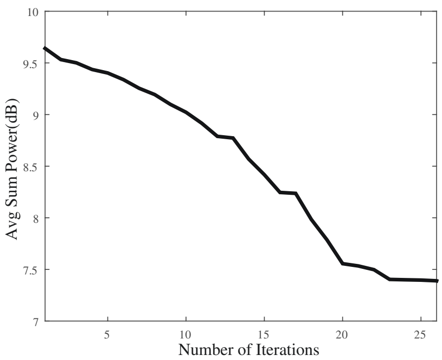
(b)

Figure 3: Performance of Algorithm 1 using two approaches: (a) Targets using Matrix Inversion in (23); (b) Targets using Lagrangian Multipliers in (27)

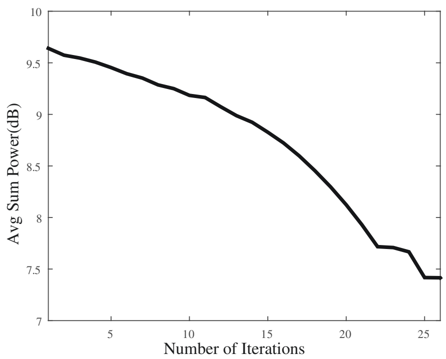
proposed in Section 4. Since both of them converge to a local optimum, it is not possible to know a priori which one performs better in a particular scenario. In this example we start with equal per-stream targets, which is a good initial candidate in terms of power consumption since none of the MMSE targets is too small. Fig. 3 depicts the target updates for both approaches. Similar values are reached for targets $\varrho_{1,1}$, $\varrho_{2,2}$, $\varrho_{2,3}$ and $\varrho_{2,4}$, whereas larger differences can be noticed for the rest of them. Recall that, at each iteration, the sum of the k-th user per-stream targets is equal to ρ_k.

Fig. 4 shows the gradual reduction of the transmit powers. After convergence, similar results are achieved for both update steps. However, this is not true in general due to the local optimality of the solutions. The performance of the two steps highly depends on the initial targets, the particular channel realizations, and the starting precoding matrices.

For the second setup, we generate 1 000 channel realizations according to (2) and initial random per-stream targets. The evolution of such targets can



(a)



(b)

Figure 4: Performance of Algorithm 1 using two approaches. (a) Power using Matrix Inversion in (23). (b) Power using Lagrangian Multipliers in (27).

be observed in Fig. 5. Observe that stream 1 of user 1 is deactivated and afterwards activated (see iterations 3 and 4). The total transmit power is shown in Fig. 6. Observe that both the per-stream targets and the total transmit power converge at about 15 iterations and, after iteration 9, the power reduction is negligible.

6.1. Sum-MSE Lower Bound

In this section we empirically evaluate the performance degradation, in terms of sMSE, which results from the CSI uncertainty. The setup consists of one BS with $N = 4$ antennas and $K = 4$ users with $R = 2$ antennas. We define the vector $\boldsymbol{h}_k[q]$ containing the stacked columns of the k-th user Gaussian channel, $\boldsymbol{H}_k[q]$, where q is the time slot, and $\boldsymbol{C}_{\boldsymbol{h}_k} = \mathbf{I}_{NR}$ is the slow-variation channel covariance.

In this example we consider that the users receive a training sequence. Afterwards, the received signal is processed to reduce the number of bits fed back to the BS. Indeed, a rank reduction is performed by truncating the

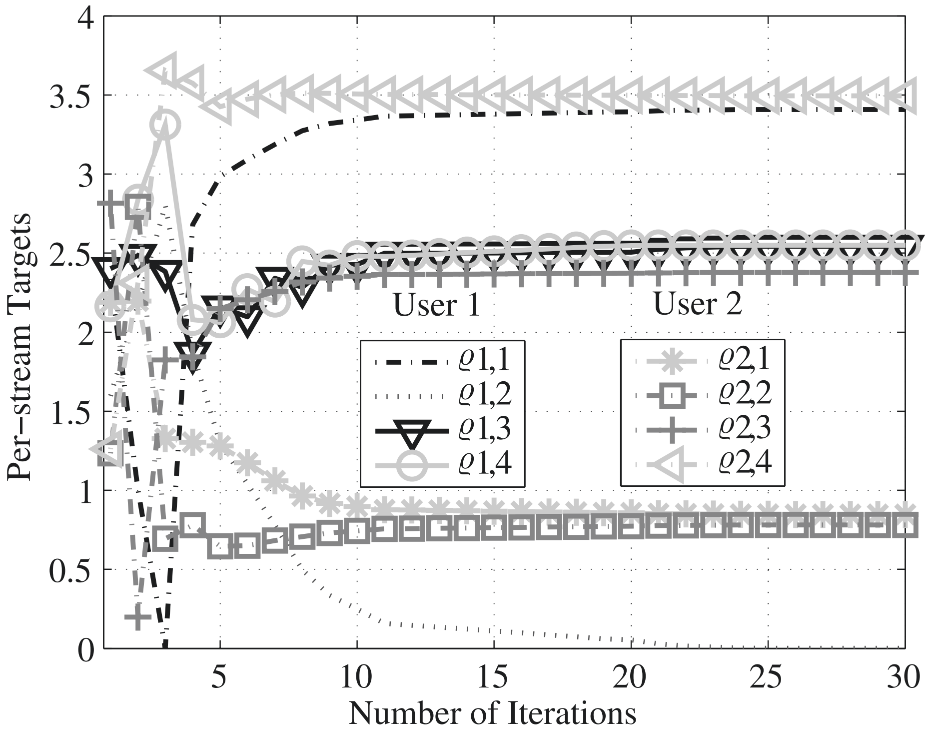


Figure 5: Per-Stream Targets vs. Number of Iterations.

Karhunen-Loève transform. Next, the coefficients are quantized using scalar quantization. The CSI at the users, $\hat{\boldsymbol{h}}_k[q] \in \mathbb{C}^n$, reads as [28]

$$\hat{\boldsymbol{h}}_k[q] = \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{S} \boldsymbol{h}_k[q] + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{n} + \boldsymbol{n}_{Q,k}, \tag{38}$$

where $\boldsymbol{V}_k \in \mathbb{C}^{NR \times n}$ collects $n < NR$ eigenvectors of $\boldsymbol{C}_{\boldsymbol{h}_k}$, $\boldsymbol{S} \in \mathbb{C}^{NR \times NR}$ such that $\boldsymbol{S}^{\mathrm{H}} \boldsymbol{S} = \mathbf{I}_{NR}$ is the training sequence, $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{NR}, \boldsymbol{C}_{\boldsymbol{n}})$ is the noise during the training stage, and $\boldsymbol{n}_{Q,k} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_n, \frac{\gamma^2}{6} \mathbf{I}_n)$ is the quantization error, with $\gamma = \frac{2\sqrt{2}}{2^b}$ and 2^b being the number of quantization levels. For simplicity, the channel, the noise, and the quantization error are assumed to be mutually independent and Gaussian.

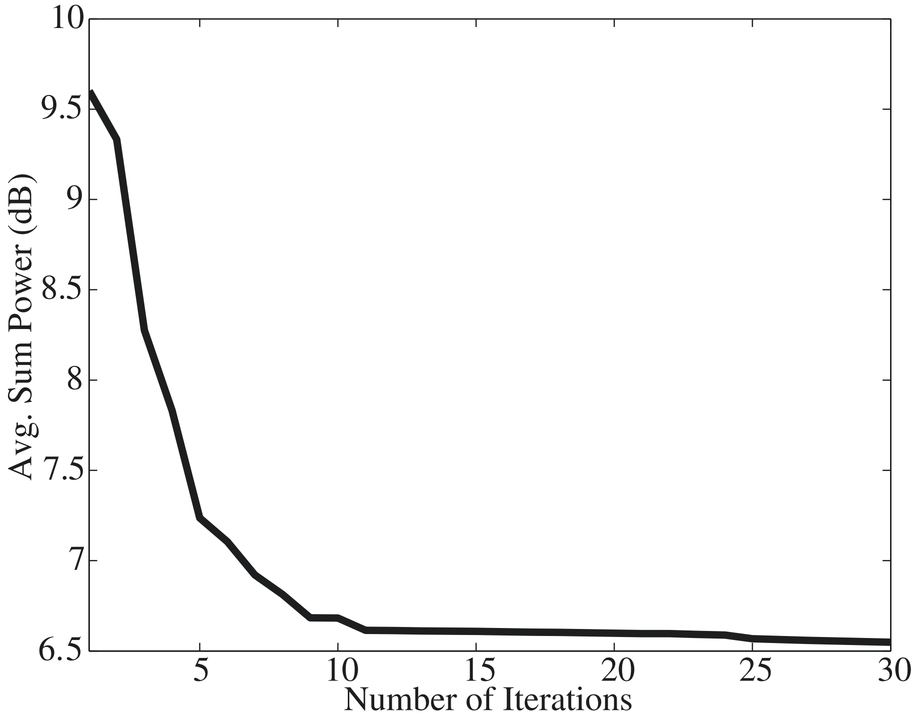


Figure 6: Average Sum Power (dB) vs. Number of Iterations.

Each user selects an entry from its 2^{bn} codebook entries and sends the bn bits to the BS. We consider a feedback delay of D time slots for acquiring the CSI, and temporal correlations of the channel according to Jakes model (see [28] and references therein). Then, the statistical CSI at the BS is given by the first and second order moments of $\boldsymbol{h}_k[q]|\hat{\boldsymbol{h}}_k[q-D]$ as follows [28, 29]

$$\mathbb{E}[\boldsymbol{h}_k[q]|\hat{\boldsymbol{h}}_k[q-D]] = \boldsymbol{C}^{\mathrm{H}}_{\hat{\boldsymbol{h}}_k\boldsymbol{h}_k}\boldsymbol{C}^{-1}_{\hat{\boldsymbol{h}}_k}\hat{\boldsymbol{h}}_k[q-D], \tag{39}$$

$$\mathbb{E}[\boldsymbol{h}_k[q]\boldsymbol{h}_k^{\mathrm{H}}[q]|\hat{\boldsymbol{h}}_k[q-D]] = \boldsymbol{C}_{\boldsymbol{h}_k} - \boldsymbol{C}^{\mathrm{H}}_{\hat{\boldsymbol{h}}_k\boldsymbol{h}_k}\boldsymbol{C}^{-1}_{\hat{\boldsymbol{h}}_k}\boldsymbol{C}_{\hat{\boldsymbol{h}}_k\boldsymbol{h}_k}, \tag{40}$$

where $\boldsymbol{C}_{\hat{\boldsymbol{h}}_k,\boldsymbol{h}_k} = \mathbb{E}[\hat{\boldsymbol{h}}_k[q-D]\boldsymbol{h}_k^{\mathrm{H}}[q]] = \boldsymbol{V}_k^{\mathrm{H}}\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_k}J_0(\alpha_k D)$, $\boldsymbol{C}_{\hat{\boldsymbol{h}}_k} = \mathbb{E}[\hat{\boldsymbol{h}}_k\hat{\boldsymbol{h}}_k^{\mathrm{H}}] = \boldsymbol{V}_k^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{h}_k}\boldsymbol{S}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{n}})\boldsymbol{V}_k + \frac{\gamma^2}{6}\mathbf{I}_n$, with $J_0(\cdot)$ being the zero-order Bessel function

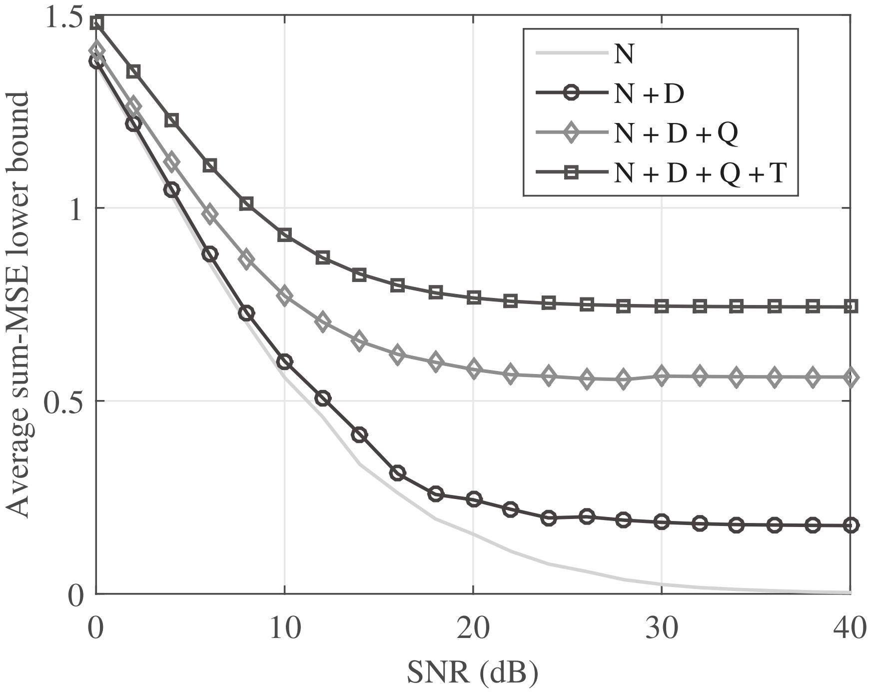


Figure 7: Average sMSE Lower Bound under CSI Uncertainties.

of the first kind, and $\alpha_k = 2\pi f_D / f_s$.

In order to compute the average sMSE in (35), we generate 2 000 channel realizations according to the statistics in (39) and (40), using the locally optimum MAC receive filters $\boldsymbol{G}$, obtained after convergence of the iteration in (37). In particular, we show the bounds considering the following CSI uncertainties: (N) channel noise, with $\boldsymbol{C_n} = \sigma^2 \mathbf{I}_{NR}$ and SNR $\in [0, 40]$ dB; (D) $D = 2$ time slots delay, $f_D = 18.5185$ Hz, and $f_s = 1500$ Hz; (Q) quantization with 2 bits per channel coefficient (12 per user); and (T) rank reduction with $n = 6$ eigenvectors.

Figure 7 depicts the severe system performance degradation due to CSI uncertainty. Theoretical studies can be found in the literature, e.g. [30], but assuming Zero-Forcing precoders. This example exhibits the trade-off between minimum average MMSE (or maximum average rate) and the CSI accuracy at the BS. Note that the feedback rate is constant for all the SNR values, and the impact on performance dramatically increases for high SNRs. More sophisticated quantizer designs could be used but these are out of the scope of this work.

7. Conclusions

This work addresses the power minimization of the multiple-stream MIMO BC subject to per-user average rate restrictions. Moreover, the practical assumption of imperfect CSI at the transmitter is considered, leading to a complicated problem formulation. To tackle with this difficulty, we have investigated the relationship with the average MMSE and proposed to reformulate the problem by introducing average MMSE-based per-stream restrictions. Accordingly, a strategy to distribute rates among the streams via a projected gradient method is performed. The flexibility of the proposed algorithm is improved by allowing the streams to switch on and off for convenience. Additionally, we study the feasible region and the average sMSE lower bound considering different CSI errors.

8. Acknowledgment

This work was funded by Xunta de Galicia (ED431C 2016-045, ED341D R2016/012, ED431G/01), AEI of Spain (TEC2013-47141-C4-1-R, TEC2015-69648-REDC, TEC2016-75067-C4-1-R), and ERDF funds (AEI/FEDER, EU).

Appendix A.

In order to obtain the gradient in (36), we compute the derivative of $\overline{\mathrm{sMSE}}^{\mathrm{MAC}}(v)$ in (35) with respect to $\boldsymbol{G}$ entry-wise as

$$\begin{aligned}\delta_{i,j} &= \frac{\partial}{\partial[\boldsymbol{G}^*]_{i,j}} \sum_{k=1}^{K} \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\left(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k\right)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\middle|\, v\right]\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\right) \qquad \text{(A.1)}\\ &= \boldsymbol{e}_i^{\mathrm{T}}\,\mathbb{E}\left[\boldsymbol{H}_j\boldsymbol{B}_j^{-1}\boldsymbol{H}_j^{\mathrm{H}}\boldsymbol{G}\middle|\, v\right]\boldsymbol{e}_j - \boldsymbol{e}_i^{\mathrm{T}}\sum_{k=1}^{K}\mathbb{E}\left[\boldsymbol{H}_k\boldsymbol{B}_k^{-1}\boldsymbol{A}_k\boldsymbol{B}_k^{-1}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\middle|\, v\right]\boldsymbol{e}_j\end{aligned}$$

where we used the matrices $\boldsymbol{A}_k = \boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k$ and $\boldsymbol{B}_k = \boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}_k$. Stacking the N elements corresponding to the user j, $\boldsymbol{\delta}_j = [\delta_{1,j}, \ldots, \delta_{N,j}]^{\mathrm{T}}$, we arrive at the gradient (36).

References

[1] F. Negro, I. Ghauri, D. T. M. Slock, Sum Rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE, in: International Symposium on Wireless Communication Systems (ISWCS), 2012, pp. 576–580. doi:10.1109/ISWCS.2012.6328433.

[2] S. Shi, M. Schubert, H. Boche, Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: MMSE Balancing, IEEE Transactions on Signal Processing 56 (8) (2008) 3702–3712.

[3] M. Kobayashi, G. Caire, A Practical Approach for Weighted Rate Sum Maximization in MIMO-OFDM Broadcast Channels, in: Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC), 2007, pp. 1591–1595.

[4] M. Codreanu, A. Tolli, M. Juntti, M. Latva-aho, Joint Design of Tx-Rx Beamformers in MIMO Downlink Channel, IEEE Transactions on Signal Processing 55 (9) (2007) 4639–4655.

[5] W. Yu, T. Lan, Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints, IEEE Transactions on Signal Processing 55 (6) (2007) 2646–2660.

[6] S. Vishwanath, N. Jindal, A. Goldsmith, Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels, Information Theory, IEEE Transactions on 49 (10) (2003) 2658–2668. doi:10.1109/TIT.2003.817421.

[7] S. Shi, M. Schubert, H. Boche, Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: Duality and Sum-MSE Minimization, IEEE Transactions on Signal Processing 55 (11) (2007) 5436 –5446. doi:10.1109/TSP.2007.899283.

[8] C. Hellings, M. Joham, W. Utschick, Gradient-based power minimization in mimo broadcast channels with linear precoding, Signal Processing, IEEE Transactions on 60 (2) (2012) 877–890.

[9] M. B. Shenouda, T. N. Davidson, Convex Conic Formulations of Robust Downlink Precoder Designs With Quality of Service Constraints, IEEE Journal on Selected Areas in Signal Processing 1 (4) (2007) 714–724.

[10] A. Mutapcic, S. Kim, S. Boyd, A Tractable Method for Robust Downlink Beamforming in Wireless Communications, in: Proc. Asilomar Con-

ference on Signals, Systems and Computers (ACSSC), 2007, pp. 1224–1228.

[11] G. Zheng, K. Wong, B. Ottersten, Robust Cognitive Beamforming With Bounded Channel Uncertainties, IEEE Transactions on Signal Processing 57 (12) (2009) 4871–4881.

[12] M. Razaviyayn, M. Boroujeni, Z.-Q. Luo, A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel, in: IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2013, pp. 325–329.

[13] H. Joudeh, B. Clerckx, AMMSE optimization for multiuser MISO systems with imperfect CSIT and perfect CSIR, in: Proc. IEEE Global Communications Conference (GLOBECOM), 2014, pp. 3308–3313. doi:10.1109/GLOCOM.2014.7037317.

[14] J. González-Coma, M. Joham, P. Castro, 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), 2013, pp. 4863–4867. doi:10.1109/ICASSP.2013.6638585.

[15] J. González-Coma, M. Joham, P. Castro, L. Castedo, QoS constrained power minimization in the MISO broadcast channel with imperfect CSI, Signal Processing 131 (2017) 447 – 455. doi:10.1016/j.sigpro.2016.09.007.

[16] J. González-Coma, M. Joham, P. M. Castro, L. Castedo, Power Minimization in the Multiuser MIMO-OFDM Broadcast Channel with Im-

perfect CSI, in: Proc. European Signal Processing Conference (EUSIPCO), 2014.

[17] J. González-Coma, M. Joham, P. Castro, L. Castedo, Power minimization in the multiple stream MIMO Broadcast Channel with imperfect CSI, in: IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 165–168. doi:10.1109/SAM.2014.6882366.

[18] S. P. Boyd, L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

[19] R. Hunger, M. Joham, A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel, in: Proc. IEEE Global Telecommunications Conference, 2008, pp. 1–5.

[20] P. S. Chow, J. M. Cioffi, J. A. C. Bingham, A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels, IEEE Transactions on Communications 43 (2/3/4) (1995) 773–775. doi:10.1109/26.380108.

[21] M. Joham, M. Vonbun, 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), 2010, pp. 1–5. doi:10.1109/SPAWC.2010.5670866.

[22] F. Clarke, Optimization and Nonsmooth Analysis, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1990.

[23] A. Berman, R. Plemmons, Nonnegative matrices in the mathematical sciences , New York: Academic Press, 1979.

[24] M. S. Bazaraa, H. D. Sherali, C. M. Shetty, Nonlinear programming: theory and algorithms, John Wiley & Sons, 2006.

[25] D. Bertsekas, Nonlinear programming, Athena Scientific, Belmont, Massachusetts, 1999.

[26] J. González-Coma, M. Joham, P. Castro, L. Castedo, Power Minimization and QoS Feasibility Region in the Multiuser MIMO Broadcast Channel with Imperfect CSI, in: Proc. IEEE Workshop onSignal Processing Advances in Wireless Communications (SPAWC), 2013, pp. 619–623. doi:10.1109/SPAWC.2013.6612124.

[27] R. Hunger, M. Joham, A Complete Description of the QoS Feasibility Region in the Vector Broadcast Channel, IEEE Transactions on Signal Processing 58 (7) (2010) 3870 –3878.

[28] M. Joham, P. M. Castro, L. Castedo, W. Utschick, Robust Precoding With Bayesian Error Modeling for Limited Feedback MU-MISO Systems, IEEE Transactions on Signal Processing 58 (9) (2010) 4954–4960. doi:10.1109/TSP.2010.2052046.

[29] S. M. Kay, Fundamentals of Statistical Signal Processing, Prentice Hall, 1993.

[30] N. Jindal, MIMO Broadcast Channels With Finite-Rate Feedback, IEEE Transactions on Information Theory 52 (11) (2006) 5045–5060. doi:10.1109/TIT.2006.883550.