INTERFERENCE MITIGATION FOR COOPERATIVE MIMO CHANNELS WITH ASYMMETRIC FEEDBACK

Lorenzo Miretti and David Gesbert

EURECOM, Sophia Antipolis, France

ABSTRACT

Current cooperative transmission strategies for distributed MIMO systems are typically designed by assuming perfect, or at least perfectly shared, channel state information at the transmitters (CSIT). However, when this assumption is not met, the naïve application of existing schemes encounters severe performance degradation. Recently, an information theoretical result unveiled the intriguing phenomenon that, for distributed MIMO systems with asymmetric CSIT, optimal encoding may require the transmission of a number of data streams that goes beyond the classical upper bound predicted by the number of antennas in the system. In this work, we explore the implications of the aforementioned result from precoder design point of view. More precisely, we propose a method for distributed precoding design which optimally minimizes the expected interference. As in the information theoretical result, the key idea lies in allowing the transmission of additional data streams. We show that this trick helps in transforming an otherwise difficult robust optimization problem into a convex formulation. Numerical simulations corroborate the robustness gain in terms of achievable sum-rate of the proposed method over traditional schemes.

1. INTRODUCTION

Network-wide interference management via cooperation among geographically distributed transmitters (TXs) is a well-established paradigm, also known as *multi-cell* multiple-input multiple-output (MIMO) or *network MIMO*, with the potential of overcoming current cellular systems limitations [1, 2]. However, the immense capacity gains promised by this paradigm are mostly achieved under the assumption of perfect, or at least perfectly shared, channel state information at the transmitters (CSIT). Although there exist commercial-grade systems implementing the original network MIMO idea (see, e.g., [3]), the aforementioned assumption has been questioned for several practical scenarios where more stringent backhaul and/or feedback constraints severely limit accurate CSIT distribution across the network, giving rise to a so-called *distributed CSIT* configuration [4, 5]. As an extreme case, the most recent embodiment of network MIMO called *cell-free massive MIMO* [6] often advocates transmission strategies based on local CSIT only, hence renouncing to network-wide interference cancellation.

Despite its relevance, research on the distributed CSIT assumption is far from being mature. Most of the available insights are based on asymptotic analysis [5, 7], while many problems remain open for finite signal-to-noise ratio (SNR) [4]. In this work we focus on this last category of problems and study the simple network in Fig. 1, where two TXs wish to cooperatively serve two RXs on the basis of CSIT acquired via over-the-uplink feedback links. The salient feature of the consider model is that, because of possibly different feedback rates, the TXs may not share the same CSIT. We

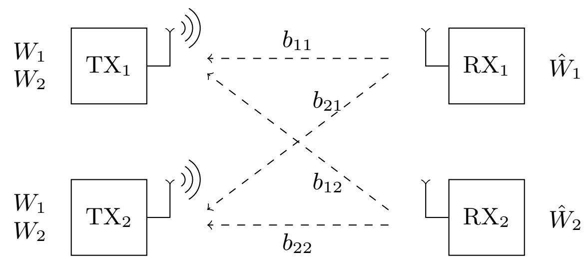

Fig. 1. Cooperative MIMO channel with asymmetric feedback.

refer to this specific cooperation regime as cooperative MIMO with *asymmetric feedback*, a particular case of distributed CSIT. Note that settings with MIMO cooperation under message sharing but asymmetric CSIT are particularly relevant to use cases where data content can be pre-stored at the transmitters (as in caching applications) while feedback rates are subject to resource constraints that vary as a function of devices nature/grade level.

The main contribution of this paper is a novel precoding design technique that optimally mitigates interference in the aforementioned channel, building upon recent information theoretical insights given by [8, 9]. Specifically, by focusing on the single RX case, [9] proves that optimal transmission under asymmetric feedback may require *additional* data streams w.r.t. the classical symmetric feedback setup. This previously overlooked design parameter is here exploited to transform an otherwise difficult robust optimization problem into a convex formulation, which can be solved by off-the-shelf tools. Moreover, we show that if the CSIT distribution exhibits some favourable structure (in particular, a so-called *common information*), the complexity of the proposed technique can be sensibly reduced. Finally, numerical simulations demonstrate the effectiveness of the proposed method over traditional schemes. Interestingly, our experiments and preliminary theoretical insights suggest that, although very useful for the precoding design phase, the use of additional data streams seems not crucial in the transmission phase, which is a desirable property for low-complexity decoders.

Notation: we use a, $\mathbf{a}$, and $\mathbf{A}$ to denote respectively scalars, column vectors, and matrices. The entry of $\mathbf{A}$ in the ith column and jth row is denoted by $[\mathbf{A}]_{i,j}$. The operators $(\cdot)^{\mathsf{H}}$ and $\mathrm{tr}\{\cdot\}$ denote respectively the Hermitian transpose and the trace, and $\|\cdot\|$ is the Euclidean norm. We use $\mathrm{diag}(\mathbf{a})$ to denote a diagonal matrix with $\mathbf{a}$ on the main diagonal. By $\mathbb{S}_+^n$ we denote the set of Hermitian positive semidefinite matrices of dimension n. Given a random variable x, $\mathbb{E}[x]$ denotes its expected value, and $p_x(x)$ its distribution. In this work we do not typographically differentiate random variables from their realizations.

2. SYSTEM MODEL AND PROBLEM STATEMENT

2.1. Cooperative MIMO with asymmetric feedback

Consider a wireless system composed by 2 TXs and 2 RXs, each of them equipped with a single antenna, and governed by the following MIMO fading channel law

$$y_l = \mathbf{h}_l^{\mathsf{H}}\mathbf{x} + z_l = \mathbf{h}_l^{\mathsf{H}}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + z_l, \quad l = 1, 2$$

where $y_l \in \mathbb{C}$ is the received signal at the l-th RX, $x_k \in \mathbb{C}$ is the transmitted signal at the k-th TX subject to a power constraint $\mathbb{E}[|x_k|^2] \leq P_k$, $\mathbf{h}_l \in \mathbb{C}^2$ is an arbitrarily distributed channel vector, and $z_l \sim \mathcal{CN}(0,1)$. We assume the channel vectors $\mathbf{h}_l$ to be mutually independent, which can be easily justified for geographically spaced RXs.

We consider a cooperative setup where all TXs have access to the full message set (W_1, W_2), where $W_l \in \{1, \ldots, 2^{\lceil nR_l \rceil}\}$ denotes the independently and uniformly distributed message of rate $R_l \geq 0$ intended for RX l. Furthermore, we assume perfect CSIR and we let the CSIT available at the k-th TX to be a quantized representation of the true state, i.e. we assume the k-th TX to causally observe integer valued signals s_{kl} given by

$$s_{kl} = q_{kl}(\mathbf{h}_l), \quad q_{kl} : \mathbb{C}^2 \to \mathcal{S}_{kl} := \{1, \ldots, 2^{b_{kl}}\},$$

where b_{kl} denotes the feedback rate from RX l towards TX k. The full CSIT at TX k is denoted by $\mathbf{s}_k := (s_{k1}, s_{k2})$. As already mentioned, in this work we are mostly interested in the asymmetric feedback regime, i.e. where the quantizers $q_{kl}(\cdot)$ are different across k. We remark that asymmetric feedback makes the above model formally (and, as we will see in the following, also practically) different from the classical 2×2 MIMO broadcast channel (BC), where full cooperation among the TXs is assumed.

2.2. Distributed linear precoding

In this section we describe a simple achievable scheme obtained by extending the classical concept of linear precoding of Gaussian data streams (i.e., a form of superposition coding) to distributed settings. Let

$$x_k = \mathbf{g}_k^{\mathsf{H}}(\mathbf{s}_k)\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \sum_{l=1}^{2} \mathbf{g}_{kl}^{\mathsf{H}}(\mathbf{s}_k)\mathbf{u}_l, \quad \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{2d}), \tag{1}$$

where $\mathbf{u}_l \in \mathbb{C}^d$ is a vector of independent Gaussian coded data symbols intended for the l-th RX, and where $\mathbf{g}_{kl}(\mathbf{s}_k) \in \mathbb{C}^d$ is a linear precoder applied to $\mathbf{u}_l$ at the k-th TX based on the local CSIT $\mathbf{s}_k$. We denote this scheme by *distributed* linear precoding.

By treating interference as noise, and by allowing coding over multiple fading realizations, it can be shown by classical arguments [10] that distributed linear precoding achieves the rate pair (R_1, R_2) given by

$$R_l = \mathbb{E}\left[\log_2\left(1 + \frac{\mathbf{h}_l^{\mathsf{H}}\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)\mathbf{h}_l}{1 + \mathbf{h}_l^{\mathsf{H}}\mathbf{\Sigma}_{\bar{l}}(\mathbf{s}_1,\mathbf{s}_2)\mathbf{h}_l}\right)\right], \quad \bar{l} \neq l, \tag{2}$$

where we defined the conditional input covariance matrices

$$\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2) := \begin{bmatrix} \mathbf{g}_{1l}^{\mathsf{H}}(\mathbf{s}_1) \\ \mathbf{g}_{2l}^{\mathsf{H}}(\mathbf{s}_2) \end{bmatrix}\begin{bmatrix} \mathbf{g}_{1l}(\mathbf{s}_1) & \mathbf{g}_{2l}(\mathbf{s}_2) \end{bmatrix} \in \mathbb{S}_+^2. \tag{3}$$

The goal of this work is to design the distributed precoders $\mathbf{g}_k(s_k)$ such that the sum-rate $R_{\text{sum}} := R_1 + R_2$ is maximized, under peak-power constraints $\|\mathbf{g}_k(\mathbf{s}_k)\|^2 \leq P_k$ and feedback strategy (q_1, q_2).

At this stage it is important to stress two major differences between precoder design in (1) and classical [11, 12, 13] based on centralized CSIT (or *symmetric* feedback). First let us note how $\mathbf{g}_{kl}(\mathbf{s}_k) \in \mathbb{C}^d$, which is a linear precoder applied to $\mathbf{u}_l$ at the k-th TX, depends solely on the local CSIT $\mathbf{s}_k$, hence fulfilling the distributed nature of the design. Secondly, in contrast with classical schemes, the number of data streams for each RX d is here allowed to take arbitrary values. While the first aspect was already considered for example in [4, 5], the exploration of the second aspect in the context of interference management is the main focus of this work, and it will be detailed in the following sections.

2.3. Multi-stream precoding

Traditionally, precoding design for the 2×2 MIMO BC focuses on single-stream transmission, i.e., $d = 1$ Gaussian symbol per user. More generally, although less commonly done in practice since it requires multi-stream decoding [10] (e.g., successive interference cancellation), an optimal design choice is to bound the number of data streams d by the *total* number of TX antennas, i.e. by letting $d \leq 2$ in the considered antenna setup.

As a matter of fact, such bound is also optimal for the special case of the cooperative MIMO system at hand for $\mathbf{s}_1 = \mathbf{s}_2 =: \mathbf{s}$, i.e. where the system boils down to a *virtually* centralized MIMO BC. In fact, in such case, it is easy to see that every conditional input covariance $\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2) = \mathbf{\Sigma}_l(\mathbf{s})$ is achievable by letting

$$\begin{bmatrix} \mathbf{g}_{1l}^{\mathsf{H}}(\mathbf{s}_1) \\ \mathbf{g}_{2l}^{\mathsf{H}}(\mathbf{s}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{1l}^{\mathsf{H}}(\mathbf{s}) \\ \mathbf{g}_{2l}^{\mathsf{H}}(\mathbf{s}) \end{bmatrix} = \mathbf{\Sigma}_l(\mathbf{s})^{\frac{1}{2}} \in \mathbb{C}^{2\times 2}.$$

In contrast, such approach is not valid for systems with asymmetric feedback, as taking the matrix square-root of $\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)$ may violate the functional dependencies of $\mathbf{g}_k(\mathbf{s}_k)$.

Surprisingly, by studying the single RX case, in [8, 9] we show that increasing the number of data streams d beyond classical design choices, i.e. by letting $d > 2$, may be beneficial. In particular, we show that it allows distributed linear precoding to span the whole set of feasible conditional input covariance matrices $\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)$, and, as a byproduct, to achieve capacity. On the other hand, it is also shown that $d \leq 2$ allows only for a *proper* subset of the feasible conditional input covariances, and, for some fading distributions, this leads to *strictly* suboptimal rates. Furthermore, computational advantages of multi-stream transmission for the single-RX precoding design problem are also reported in [9, 14].

In this work we extend the aforementioned works by studying how this previously overlooked design parameter can be exploited for distributed precoding desing in systems with interference.

3. DISTRIBUTED PRECODING VIA EXPECTED INTERFERENCE MINIMIZATION

3.1. Expected interference minimization

As a heuristics for precoding design, in this section we propose to solve the following optimization problem

$$\begin{aligned} \underset{\mathbf{g}_{kl}(\mathbf{s}_k)\in\mathbb{C}^d}{\text{minimize}} \quad & \mathbb{E}\left[\mathbf{h}_{\bar{l}}^{\mathsf{H}}\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)\mathbf{h}_{\bar{l}}\right], \quad \bar{l} \neq l \\ \text{subject to} \quad & \|\mathbf{g}_{kl}(\mathbf{s}_k)\|^2 \leq \frac{P_k}{2}, \quad k = 1, 2 \\ & \mathbb{E}\left[\operatorname{tr}\{\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)\}\right] = \frac{\alpha}{2}(P_1 + P_2) \end{aligned} \tag{4}$$

where the objective $I_l := \mathbb{E}\left[\mathbf{h}_{\bar{l}}^{\mathsf{H}}\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)\mathbf{h}_{\bar{l}}\right]$ denotes the expected interference caused by the transmission of message W_l with input covariance $\mathbf{\Sigma}_l(\mathbf{s}_1,\mathbf{s}_2)$ given by (3), and where the equality constraint is introduced to avoid the trivial solution $\mathbf{g}_{kl}(\mathbf{s}_k)=\mathbf{0}$, by imposing an average signal strength equal to a tunable fraction $\alpha\in(0,1]$ of the per-user total peak power $(P_1+P_2)/2$. Note that, as often done in the literature on linear precoding with limited feeedback [15], we simplified the power allocation problem across users (i.e. across W_l) by splitting the available power P_k equally.

A direct relation between the optimum of Problem (4) and the optimum of the direct optimization of (2) is not surprisingly hard to establish. Despite this limitation, we will show in the following that Problem (4) has a particularly favourable structure for the asymmetric feedback setup. We point out that asymmetry of information makes Problem (4) to fall into the category of *Team Decision* problems [16], for which no optimal and efficient solution method is known in general.

Before presenting the main result of this work, we observe that, by the mutual independence of $\mathbf{h}_1$ and $\mathbf{h}_2$, Problem (4) is equivalent to

$$\begin{aligned}\underset{\mathbf{g}_{kl}(s_{k\bar{l}})\in\mathbb{C}^d}{\text{minimize}}\quad & \mathbb{E}\left[\mathbf{h}_{\bar{l}}^{\mathsf{H}}\mathbf{\Sigma}_l(s_{1\bar{l}},s_{2\bar{l}})\mathbf{h}_{\bar{l}}\right],\quad \bar{l}\neq l\\ \text{subject to}\quad & \|\mathbf{g}_{kl}(s_{k\bar{l}})\|^2\leq\frac{P_k}{2},\quad k=1,2 \\ & \mathbb{E}\left[\mathrm{tr}\{\mathbf{\Sigma}_l(s_{1\bar{l}},s_{2\bar{l}})\}\right]=\frac{\alpha}{2}(P_1+P_2).\end{aligned}\tag{5}$$

That is, the optimal $\mathbf{g}_{kl}(\mathbf{s}_k)$ for Problem (4) needs not to depend on s_{kl}. This is reminescent of the fact that, in classical zero-forcing precoding [11, 15], interference can be avoided by simply selecting a precoding vector in the null-space of the interfering channel $\mathbf{h}_{\bar{l}}$, i.e. information about the direct channel $\mathbf{h}_l$ is not required.

Finally, to avoid cumbersome notation, in the following we omit the RX subscripts $l,\bar{l}$ and focus on the following archetypal problem

$$\begin{aligned}\underset{\mathbf{g}_{k}(s_{k})\in\mathbb{C}^d}{\text{minimize}}\quad & \mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1,s_2)\mathbf{h}\right],\\ \text{subject to}\quad & \|\mathbf{g}_{k}(s_{k})\|^2\leq\frac{P_k}{2},\quad k=1,2 \\ & \mathbb{E}\left[\mathrm{tr}\{\mathbf{\Sigma}(s_1,s_2)\}\right]=\frac{\alpha}{2}(P_1+P_2),\end{aligned}\tag{6}$$

for some triple of random variables $(s_1,s_2,\mathbf{h})$ taking values in some sets $\mathcal{S}_1\times\mathcal{S}_2\times\mathbb{C}^2$, $\mathcal{S}_k:=\{1,\dots,d_k\}$, and where

$$\mathbf{\Sigma}(s_1,s_2):=\begin{bmatrix}\mathbf{g}_1^{\mathsf{H}}(s_1)\\ \mathbf{g}_2^{\mathsf{H}}(s_2)\end{bmatrix}\begin{bmatrix}\mathbf{g}_1(s_1) & \mathbf{g}_2(s_2)\end{bmatrix}.$$

Clearly, (4) and (5) can be simply mapped into Problem (6) by letting $(s_1,s_2,\mathbf{h})\sim(s_{1\bar{l}},s_{2\bar{l}},\mathbf{h}_l)$.

3.2. Convex formulation via multi-stream precoding

In this section we describe the main result of this work, which states that by letting the dimension d of the precoders to grow large (beyond classical design), it is possible to recast Problem (4) as an equivalent convex problem. To this end, let us first define the diagonal matrix

$$\mathbf{\Pi}:=\begin{bmatrix}\mathbf{\Pi}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Pi}_2\end{bmatrix},\quad \mathbf{\Pi}_k\in\mathbb{C}^{d_k\times d_k},$$
$$[\mathbf{\Pi}_k]_{i,j}:=\begin{cases}p_{s_k}(i) & i=j\\ 0 & i\neq j\end{cases},\quad k=1,2,$$

and the covariance matrix $\mathbf{\Psi}:=\mathbb{E}[\mathbf{h}_{\mathrm{eq}}\mathbf{h}_{\mathrm{eq}}^{\mathsf{H}}]$ of an equivalent $d_{\max}$-dimensional channel

$$\mathbf{h}_{\mathrm{eq}}:=\mathbf{E}(s_1,s_2)\mathbf{h},$$
$$\mathbf{E}(i,j):=\begin{bmatrix}\mathbf{e}_i & \mathbf{0}\\ \mathbf{0} & \mathbf{e}_j\end{bmatrix}\in\{0,1\}^{d_{\max}\times 2},$$

where $\mathbf{e}_i\in\{0,1\}^{d_1}$ (resp. $\mathbf{e}_j\in\{0,1\}^{d_2}$) denotes a standard column selector, i.e., with the i-th entry (resp. j-th) set to 1, and all the other entries set to 0. With these definitions in hand, we obtain the following result:

Proposition 1. *By letting $d\leq d_{\max}$, where*

$$d_{\max}:=d_1+d_2,$$

Problem (6) is equivalent to the following convex problem

$$\begin{aligned}\underset{\mathbf{Q}\in\mathbb{S}_+^{d_{\max}}}{\textit{minimize}}\quad & \mathrm{tr}\{\mathbf{\Psi}\mathbf{Q}\}\\ \textit{subject to}\quad & [\mathbf{Q}]_{i,i}\leq\frac{P_1}{2},\quad 1\leq i\leq d_1\\ & [\mathbf{Q}]_{i,i}\leq\frac{P_2}{2},\quad d_1<i\leq d_{\max}\\ & \mathrm{tr}\{\mathbf{\Pi}\mathbf{Q}\}=\frac{\alpha}{2}(P_1+P_2)\end{aligned}\tag{7}$$

Proof. Let us define the codebook matrix $\mathbf{F}\in\mathbb{C}^{d\times d_{\max}}$ given by

$$\mathbf{F}:=\begin{bmatrix}\mathbf{g}_1(1) & \dots & \mathbf{g}_1(2^{b_1}) & \mathbf{g}_2(1) & \dots & \mathbf{g}_2(2^{b_2})\end{bmatrix},$$

which is obtained by stacking in the given order all the $d_{\max}:=2^{b_1}+2^{b_2}$ possible values assumed by the precoders $\mathbf{g}_k(s_k)$. It is easy to verify that, for a given CSIT realization (s_1,s_2), the distributed precoders are given by

$$\begin{bmatrix}\mathbf{g}_1(s_1) & \mathbf{g}_2(s_2)\end{bmatrix}=\mathbf{F}\mathbf{E}(s_1,s_2),$$

We can now rewrite the objective $I:=\mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1,s_2)\mathbf{h}\right]$ of Problem (6) as

$$\begin{aligned}I&=\mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{E}(s_1,s_2)^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{F}\mathbf{E}(s_1,s_2)\mathbf{h}_2\right]\\ &=\mathbb{E}\left[\mathbf{h}_{\mathrm{eq}}^{\mathsf{H}}\mathbf{F}^{\mathsf{H}}\mathbf{F}\mathbf{h}_{\mathrm{eq}}\right]\\ &=\mathbb{E}\left[\mathbf{h}_{\mathrm{eq}}^{\mathsf{H}}\mathbf{Q}\mathbf{h}_{\mathrm{eq}}\right]\\ &=\mathrm{tr}\left\{\mathbb{E}\left[\mathbf{h}_{\mathrm{eq}}\mathbf{h}_{\mathrm{eq}}^{\mathsf{H}}\right]\mathbf{Q}\right\}\\ &=\mathrm{tr}\left\{\mathbf{\Psi}\mathbf{Q}\right\},\end{aligned}$$

where $\mathbf{Q}:=\mathbf{F}^{\mathsf{H}}\mathbf{F}$ is a positive semidefinite matrix of dimension $d_{\max}$ and maximum rank d. The power constraints can be trivially reformulated by following similar steps. Finally, by removing the rank constraint on $\mathbf{Q}$, i.e. by allowing the precoders dimension d to grow up to $d_{\max}$, we obtain the convex optimization problem (7). □

Problem (4) can be solved by computing the optimal solution $\mathbf{Q}_l^\star$ of an instance of Problem (7) obtained for $(s_1,s_2,\mathbf{h})\sim(s_{1\bar{l}},s_{2\bar{l}},\mathbf{h}_l)$. The optimal distributed precoders can be then recovered from $\mathbf{Q}_l^\star$ by letting

$$\begin{bmatrix}\mathbf{g}_{1l}(\mathbf{s}_1) & \mathbf{g}_{2l}(\mathbf{s}_2)\end{bmatrix}=(\mathbf{Q}_l^\star)^{\frac{1}{2}}\mathbf{E}(s_{1\bar{l}},s_{2\bar{l}})\in\mathbb{C}^{d_{\max}\times 2},$$

which theoretically implies the encoding of W_l into up to

$$d_{\max} = 2^{b_{1\bar{l}}} + 2^{b_{2\bar{l}}}$$

data streams. However, note that if $\mathrm{rank}(\mathbf{Q}_l^\star) = r < d_{\max}$, we can reduce without loss of performance the dimension of $\mathbf{g}_{kl}(\mathbf{s}_k)$ down to r.

3.3. Exploiting common information for reducing complexity

One of the major drawbacks of the method proposed in Sect. 3.2 is that it suffers from scalability issues, since the dimension $d_{\max}$ of the precoders optimization problem scales exponentially with the number of feedback bits. However, the proposed method is based on a worst-case assumption on the distribution of the asymmetric feedback, which is in practice too restrictive. In fact, in many applications, CSIT is strongly correlated across the TXs, and this correlation can be intuitively exploited for precoders optimization. In light of this intuition, in this section we describe an approach based on the concept of *common information*, whose definition will be detailed in the following. Interestingly, we show that the availability of common information allows for a dramatic reduction in complexity, with zero or little loss of optimality.

Let us focus on Problem (6). Consider a random variable $z \in \mathcal{Z}$, where $\mathcal{Z} := \{1, \dots, Z\}$ is some alphabet of finite cardinality Z, such that there exist two functions $f_k : \mathcal{S}_k \to \mathcal{Z}$, $k = 1, 2$, satisfying

$$z = f_1(s_1) = f_2(s_2) \quad \text{a.s.}$$

We refer to such z as a *common information* between s_1 and s_2. Note that for the same pair of random variables (s_1, s_2) there may be multiple common informations satisfying the above definition (in particular, with different cardinalities Z). For example, we can always define at least a trivial common information for $Z = 1$.

For a given choice of common information, we can rewrite the objective of Problem (6) as

$$\begin{aligned} I &:= \mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1, s_2)\mathbf{h}\right] \\ &= \sum_{z=1}^{Z} \mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1, s_2)\mathbf{h}|z\right] p_z(z) \\ &= \sum_{z=1}^{Z} I^{(z)} p_z(z), \end{aligned}$$

where $I^{(z)} := \mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1, s_2)\mathbf{h}|z\right]$ can be interpreted as the average interference conditioned on a given realization of the common information z. The key idea is to optimize each $I^{(z)}$ disjointly, i.e., to decompose the optimization of I into Z subproblems of reduced complexity. The reduction in complexity is explained as follows. Intuitively, we can bijectively map s_k into a pair $\left(z, s_k^{(z)}\right)$, where $z = f_k(s_k)$ is a common information between s_1 and s_2, and where $s_k^{(z)} \in \mathcal{S}_k^{(z)} := \{1, \dots, d_k^{(z)}\}$, $d_k^{(z)} \ll d_k$, is an additional index corresponding to the residual local information. Hence, conditioned on z, the optimization of $I^{(z)}$ needs only to consider the joint distribution of the residual local informations $(s_1^{(z)}, s_2^{(z)})$ instead of the full CSIT pair (s_1, s_2).

More formally, we denote the pre-image of z under f_k by $f_k^{(-1)}(z) \subseteq \mathcal{S}_k$. The sets $\{f_k^{(-1)}(z)\}_{z=1}^{Z}$ form a partition of $\mathcal{S}_k$. For every z, we further index the elements of $f_k^{(-1)}(z)$ by means of a bijective map $i_k^{(z)} : f_k^{(-1)}(z) \to \mathcal{S}_k^{(z)}$, where $d_k^{(z)} := |f_k^{(-1)}(z)|$. Then, we define the triple of random variables

$$\left(s_1^{(z)}, s_2^{(z)}, \mathbf{h}^{(z)}\right) \in \mathcal{S}_1^{(z)} \times \mathcal{S}_2^{(z)} \times \mathbb{C}^2$$

distributed according to

$$\begin{aligned} &p_{s_1^{(z)}, s_2^{(z)}, \mathbf{h}^{(z)}}\left(s_1^{(z)}, s_2^{(z)}, \mathbf{h}\right) \\ &= \sum_{s_1, s_2} p_{s_1^{(z)}, s_2^{(z)}, s_1, s_2, \mathbf{h}|z}\left(s_1^{(z)}, s_1^{(z)}, s_1, s_2, \mathbf{h}\middle|z\right), \end{aligned}$$

and the precoding vectors $\mathbf{g}_k^{(z)}(s_k^{(z)}) \in \mathbb{C}^d$, one-to-one mapped to $\mathbf{g}_k(s_k)$ according to

$$\mathbf{g}_k^{(f_k(s_k))}\left(i_k^{(f_k(s_k))}(s_k)\right) = \mathbf{g}_k(s_k).$$

Under the above definitions, we obtain

$$\begin{aligned} I^{(z)} &= \mathbb{E}\left[\mathbf{h}^{\mathsf{H}}\mathbf{\Sigma}(s_1, s_2)\mathbf{h}|z\right] \\ &= \mathbb{E}\left[\left(\mathbf{h}^{(z)}\right)^{\mathsf{H}}\mathbf{\Sigma}^{(z)}(s_1^{(z)}, s_2^{(z)})\mathbf{h}^{(z)}\right], \end{aligned}$$

where

$$\mathbf{\Sigma}^{(z)}(s_1^{(z)}, s_2^{(z)}) := \begin{bmatrix} \mathbf{g}_1^{\mathsf{H}(z)}(s_1^{(z)}) \\ \mathbf{g}_2^{\mathsf{H}(z)}(s_2^{(z)}) \end{bmatrix} \begin{bmatrix} \mathbf{g}_1^{(z)}(s_1^{(z)}) & \mathbf{g}_2^{(z)}(s_2^{(z)}) \end{bmatrix}.$$

The conditional interference I_l can be (disjointly) optimized by solving an instance of Problem (7) obtained by replacing I with $I^{(z)}$ and $(s_1, s_2, \mathbf{h})$ with $\left(s_1^{(z)}, s_2^{(z)}, \mathbf{h}^{(z)}\right)$. The resulting precoding vectors $\mathbf{g}_k^{(z)}(s_k^{(z)})$ are obtained similarly to Sect. 3.2. They imply precoding of a number of data streams bounded by $d_1^{(z)} + d_2^{(z)}$. Note that this upper bound may vary according to the realization of the common information. Overall, the proposed scheme considers the transmission of up to $\max_z(d_1^{(z)} + d_2^{(z)})$ data streams, which can be much lower than the upper bound $d_1 + d_2 = \sum_z(d_1^{(z)} + d_2^{(z)})$ predicted without (or by neglecting) common information.

We conclude by poiting out that the proposed disjoint optimization is not exactly equivalent to Problem (6): the main difference lies in the average sum-power constraint, which is here satisfied for every realization of the common information, while the original constraint is looser. Hence, the proposed method is formally a suboptimal solution to Problem (6). However, we recall that the aforementioned constraint is mainly introduced to avoid trivial solutions, hence the loss in performance is expected to be small.

4. NUMERICAL EXPERIMENTS AND CONCLUDING REMARKS

As an example, we consider the following asymmetric version of the classical random vector quantization feedback scheme

$$q_{kl}(\mathbf{h}_l) \in \arg\max_{i \in \{1, \dots, 2^{b_k}\}} |\mathbf{v}_i^{\mathsf{H}}\mathbf{h}_l|^2, \quad l = 1, 2, \quad k = 1, 2,$$

where $\mathbf{v}_i$ is a unit-norm vector belonging to a codebook $\mathcal{V}_k$ of 2^{b_k} randomly drawn channel directions. Note that the adopted simplification $b_k = b_{k1} = b_{k2}$ still captures the asymmetry of information across k, i.e., across the TXs. In the following experiments, we let $P_1 = P_2 = \mathrm{SNR}$, and we estimate the input matrices $\mathbf{\Psi}$ and $\mathbf{\Pi}$ capturing the effect of the CSIT distribution by sample averaging. Furthermore, we simulate the performance of the proposed method for different values of $\alpha \in (0, 1]$, and we keep the choice delivering the highest sum-rate.

4.1. Performance comparison

In Fig. 2 we compare the performance of the proposed method against time-division multiplexing (TDM), and regularized zero-forcing (RZF) precoding [15] naïvely transposed to the considered asymmetric feedback setup as described in [4]. As an ideal benchmark, we also show the performance of the same RZF precoder applied to a symmetric setup where the highest rate CSIT is available at both TXs. As we can observe, careful precoding design tailored to the asymmetric feedback assumption is absolutely necessary for meaningful spatial multiplexing, as TDM outperforms traditional RZF at all SNR levels. Note that TDM eventually outperforms all the considered spatial multiplexing schemes at sufficiently high SNR, but this issue could be solved by combining the proposed method with *rate-splitting* approaches with common message decoding [17].

In Fig. 3 we repeat the above experiments for the following *hierarchical* quantized feedback scheme:

$$s_{1l} = q_{1l}(\mathbf{h}_l), \quad s_{2l} = q_{2l}(q_{1l}(\mathbf{h}_l)), \quad l = 1, 2,$$

where q_{1l} and q_{2l} are the same as in the previous experiment. The above hierarchical quantizers consume the same feedback resources as their non-hierarchical counterparts, but they induce a common information $z_k = s_{2k}$. Hence, we use the technique described in Sect. 3.3 and exploit z_k for complexity reduction. In the example in Fig. 3, it means decoupling the problem into $2^{b_2} = 128$ semidefinite programs of dimension $\approx (2^{b_1} + 1)/2^{b_2} \approx 8$ in place of a unique semidefinite program of dimension $2^{b_1} + 2^{b_2} = 1152$, hence dramatically reducing the optimization effort[1].

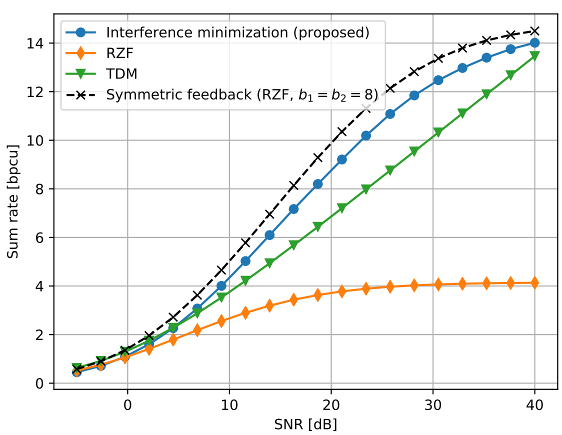

Fig. 2. Performance evaluation for random vector quantizers with asymmetric rates $(b_1, b_2) = (8, 3)$.

4.2. Optimal number of data streams

As a final important observation, by inspecting the optimized precoders, $d = 1$ turns out to be optimal in almost all our experiments. Moreover, in the very few cases where $d > 1$ is obtained, mostly happening at low SNR, it corresponds to adding 1 or 2 additional streams with very low power, hence with negligible impact on the end performance. Although we do not have a complete explanation for this observation, some preliminary insights are given by the following proposition.

[1]Semidefinite programs are known to scale badly with the dimension of the matrices involved.

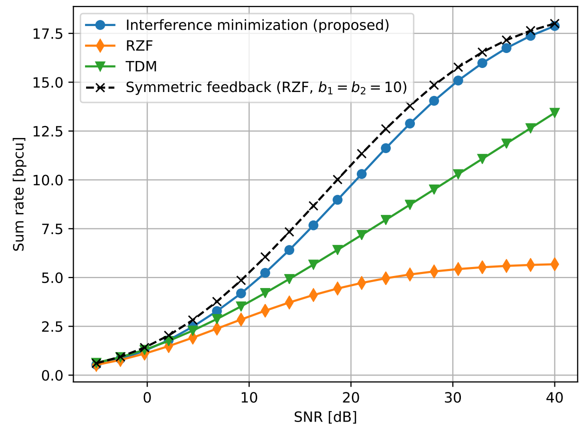

Fig. 3. Performance evaluation for hierarchical random vector quantizers with asymmetric rates $(b_1, b_2) = (10, 7)$.

Proposition 2. *In Problem* (4), *if the tunable fraction* $\alpha \in (0, 1]$ *of the total peak-power is small enough such that the peak-power constraints are inactive at the optimum, then* $d = 1$ *data stream is optimal.*

Proof. Consider the equivalent problem (7). By applying the change of variable $\tilde{\mathbf{Q}} := \mathbf{\Pi}^{\frac{1}{2}}\mathbf{Q}\mathbf{\Pi}^{\frac{1}{2}}\frac{2}{\alpha(P_1+P_2)}$, we can consider its normalized form

$$\begin{aligned} \underset{\tilde{\mathbf{Q}} \in \mathbb{S}_+^{d_{\max}}}{\text{minimize}} \quad & \mathrm{tr}\{\tilde{\mathbf{\Psi}}\tilde{\mathbf{Q}}\} \\ \text{subject to} \quad & [\tilde{\mathbf{Q}}]_{i,i} \leq b_i/\alpha, \\ & \mathrm{tr}\{\tilde{\mathbf{Q}}\} = 1, \end{aligned}$$

where $b_i := [\Pi]_{i,i}\frac{P_1}{P_1+P_2}$ for $1 \leq i \leq d_1$, $b_i := [\Pi]_{i,i}\frac{P_2}{P_1+P_2}$ for $d_1 < i \leq d_{\max}$, and $\tilde{\mathbf{\Psi}} := \mathbf{\Pi}^{-\frac{1}{2}}\mathbf{\Psi}\mathbf{\Pi}^{-\frac{1}{2}}$. If we remove the inequality constraints, standard results [18] show that, at the optimum, $r = \mathrm{rank}(\tilde{\mathbf{Q}}^\star) = \mathrm{rank}(\mathbf{Q}^\star) = 1$. Clearly, this also holds if α is small enough such that the inequality constraints become redundant. Finally, as discussed in Sect. 3.2, we recall that precoding using $d = r$ data streams incurs no loss of optimality. □

The above proposition states that $d > 1$ is not necessary for optimal interference mitigation if we allow for sufficient power underconsumption, i.e., if α is small enough. Note that power consumption is not the main bottleneck in the high SNR regime and indeeed this is reflected by our experiments. Moreover, the (almost) optimality of $d = 1$ could also be an artifact of the chosen design metric, since it is not directly related to the sum-rate. These heuristic explanations do not collide with the results in [9] for the single RX case, where instead efficient power allocation is important and where we work directly on the achievable rate.

However, further work is needed to fully understand the role of the additional data streams in systems with multiple RXs, that is, if they provide substantial rate gains for some settings on top of being a useful optimization trick.

5. REFERENCES

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, 2006.

[2] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[3] A. Forenza, S. Perlman, F. Saibi, M. Di Dio, R. van der Laan, and G. Caire, ," in *Proc. Asilomar Conf. Signals, Systems and Computers*. 2015, pp. 286–293, IEEE.

[4] D. Gesbert and P. de Kerret, "Team methods for device cooperation in wireless networks," in *Cooperative and Graph Signal Processing*, pp. 469–487. Elsevier, 2018.

[5] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Info. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[6] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[7] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "On the degrees-of-freedom of the k-user distributed broadcast channel," *IEEE Trans. Info. Theory*, vol. 66, no. 9, pp. 5642–5659, 2020.

[8] L. Miretti, P. de Kerret, and D. Gesbert, "On the fundamental limits of cooperative multiple-access channels with distributed CSIT," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2019.

[9] L. Miretti, M. Kobayashi, D. Gesbert, and P. de Kerret, "Cooperative multiple-access channels with distributed state information," *Submitted to IEEE Trans. Info. Theory*, 2019, preprint http://www.eurecom.fr/publication/6128.

[10] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge university press, 2011.

[11] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Info. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[12] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, 2009.

[13] F. Negro, I. Ghauri, and D. T. M. Slock, "Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *9th Int. Symp. on Wireless Commun. Syst., (ISWCS)*. 2012, IEEE.

[14] Lorenzo Miretti, Mari Kobayashi, and David Gesbert, "Precoding for cooperative MIMO channels with asymmetric feedback," in *Proc. IEEE Int. Conf. Communications (ICC)*, 2020.

[15] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Info. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[16] J. Marschak and R. Radner, *Economic Theory of Teams*, Yale University Press, 1972.

[17] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser miso systems with partial csit: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, 2016.

[18] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.