Precoding for Cooperative MIMO Channels with Asymmetric Feedback

Lorenzo Miretti†, Mari Kobayashi⋆, and David Gesbert†
†Communication Systems Department, EURECOM, Sophia Antipolis, France
⋆Technical University of Munich, Munich, Germany
{miretti,gesbert}@eurecom.fr, mari.kobayashi@tum.de

***Abstract*—The problem of optimally precoding over cooperative MIMO channels when the transmitters are endowed with different noisy channel state information is a long standing and challenging open problem. Recently an information theoretic result was obtained which characterized the common message capacity of a channel with two transmitters and a single receiver with such distributed channel state information (D-CSIT) generated from different feedback links. While classical MIMO precoding with centralized CSIT implies the transmission of a number of spatial streams bounded by the number of transmit and receiver antennas, the above result suggests that, surprisingly, the transmission of *additional* streams may be beneficial. In this work, we explore the operational implications of the above intuition to optimally tackle the problem of ergodic rate optimization under distributed feedback. In particular, we propose a method for joint distributed precoding and feedback design under asymmetric feedback rate constraints. In doing so, we also optimize the number of spatial data streams under practical complexity constraints. Finally, we provide numerical simulations and illustrate the performance gains compared to conventional precoder design.**

***Index Terms*—Cooperative communication, MIMO, ergodic rate, distributed CSIT, precoding.**

I. Introduction

Wireless communication networks can substantially benefit from transmitter (TX) cooperation, especially in interference dominated scenarios. Traditionally, transmission schemes and performance analysis have been mostly derived by assuming perfect, or at least perfectly shared, i.e. centralized, channel state information at the transmitters (CSIT) [1]–[3].

Recently, several practical scenarios have put the centralized CSIT assumption under question, notably in high mobility scenarios when the channel coherence time is short. In such case any attempt to centralize CSIT via an information exchange mechanism among transmitters will induce additional delays which in turn create transmitter-specific outdating of the exchanged CSI elements. In fact, in scenarios where the delay sensitivity of the data payloads is low or moderate (e.g. popular data contents that are amenable to caching [4]), it is reasonable to assume that data can be prestored at multiple devices whereas timely exchange of highly time-sensitive CSIT across such devices is difficult to achieve. In this case, the devices (cooperating transmitters) are endowed with different noisy versions of the same underlying CSIT, a setting that is commonly referred to as a *distributed* CSIT (D-CSIT) [5].

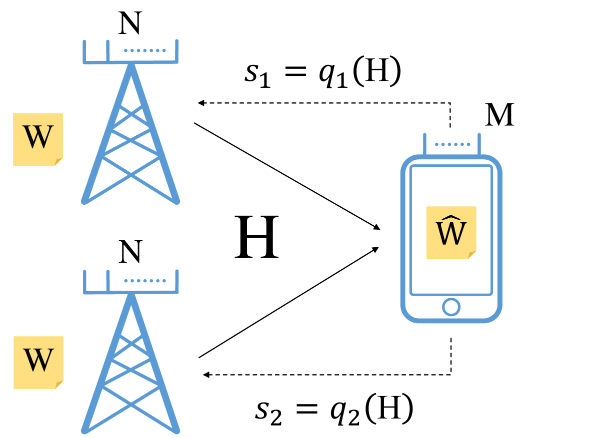


Fig. 1: Illustration of a cooperative MIMO channel with D-CSIT obtained from different error-free feedback links.

Both information theoretical limits and practical precoding designs for cooperative networks with D-CSIT are still far from being understood. Partial results on capacity analysis have been derived e.g. in [6], [7] by focusing on the asymptotic high-SNR regime, but very little is known for finite SNR.

In this paper, we consider the cooperative MIMO channel illustrated in Fig. 1, where the TXs acquire the CSIT through different feedback links of limited rate from the receiver (RX), a setting we also refer to as "asymmetric feedback". Each TX is equipped with N antennas, while the RX is equipped with M antennas. For a special case of this setup with $N = 1$ and $M = 2$, the ergodic capacity of the channel at hand was characterized in a recent work [8] building on distributed linear precoding over Gaussian codewords. Surprisingly, the proof involves the unconventional choice of letting the number of data streams d to grow large, and in particular larger than the classical bound $d \leq \min(2N, M)$.

In this work, we make use of the above information theoretical insight to address the problem of optimal distributed precoding design for the channel in Figure 1. This problem has not been tackled by [8] due to the non-convexity of the

This work was in part supported by the French-German Academy towards Industry 4.0 (SeCIF project) under Institut Mines-Telecom.

L. Miretti and D. Gesbert are also supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Agreement no. 670896).

optimization problem involved in the original formulation.

Specifically, by recasting the ergodic rate optimization problem as an equivalent convex problem similarly to the recent result given by [9] for the single TX antenna case, this paper provides the following novel contributions:

- We propose a method for jointly designing a distributed precoding and feedback strategy which maximizes the ergodic rate under possibly asymmetric feedback rate constraints. The key idea is a non-trivial adaptation of the celebrated *generalized Lloyd algorithm* [10], [11] for vector quantization to jointly optimized distributed quantizers.
- The optimal precoding design considers $d_{\max}$ spatial streams, where $d_{\max}$ depends on the feedback rates. Since the optimal $d = d_{\max}$ can be very large in theory, we study how to constrain d to practical values while minimizing the performance loss.

In Sect. II, we provide the system model and review existing results on the considered problem. In Sect. III, we detail the proposed algorithms. Finally, in Sect. IV we illustrate the above results through numerical simulations, and we compare the performance of the proposed algorithms with techniques derived from traditional centralized design.

Notation: We use boldface to denote vectors and matrices, and calligraphic uppercase to denote sets. $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote respectively the transpose and Hermitian transpose, and $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm. The set of Hermitian positive-semidefinite matrices of dimension n is denoted by $\mathbb{S}_+^n$. The shorthand $\mathbf{A}_{[i:j,k:l]}$ denotes the sub-matrix of $\mathbf{A}$ corresponding to the $(i,\dots,j)$-th rows and $(k,\dots,l)$-th columns. $\otimes$ denotes the Kronecker product. Probability mass/density functions and expectations are denoted respectively by $p(\cdot)$ and $\mathbb{E}[\cdot]$.

II. System Model and Preliminaries

We consider a flat-fading cooperative MIMO channel with 2 TXs equipped with N antennas and a single RX equipped with M antennas, described by the following memoryless input/output relation:

$$\mathbf{y} = \mathbf{H}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^M$ is the RX signal, $\mathbf{H} \in \mathbb{C}^{M\times 2N}$ is an arbitrarily distributed matrix of fast-fading coefficients, $\mathbf{x}_k \in \mathbb{C}^N$ is the signal transmitted at the k-th TX, and where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$.

Furthermore, we assume perfect CSIR, and imperfect CSIT obtained via error-free feedback links of limited and, most importantly, possibly different rate. More precisely, we assume the k-th TX to have access to an integer-valued signal $s_k \in \mathcal{S}_k := \{1,\dots,2^{\beta_k}\}$, where β_k denotes the number of feedback bits per fading realization, given by

$$s_k = q_k(\mathbf{H}), \quad q_k : \mathbb{C}^{M\times N} \to \mathcal{S}_k, \tag{2}$$

where q_k models for example a deterministic mapping from $\mathbf{H}$ to the index of a codebook of quantized representations $\hat{\mathbf{H}}_k$. The joint probability distribution

$$p(\mathbf{H}, s_1, s_2) = \mathbb{1}[s_1 = q_1(\mathbf{H})]\mathbb{1}[s_2 = q_2(\mathbf{H})]p(\mathbf{H}) \tag{3}$$

fully describes the given CSI structure, where $\mathbb{1}[\cdot]$ denotes an indicator function. When $q_1 = q_2$, the system boils down to a virtually centralized $2N \times M$ MIMO channel, that can be analyzed by using classical techniques [12]. On the other hand, when $q_1 \neq q_2$, we fall into what we call a *distributed CSIT* (D-CSIT) configuration, i.e. when the TXs do not share the same CSIT.

We finally assume that both TXs have access to a uniformly distributed message W. In this work, we focus on the *ergodic* rate of the considered channel, i.e. the rate R at which W can be reliably transmitted by coding over multiple i.i.d. fading realizations [13].

In what follows we briefly review existing approaches and recent information theoretical results in the considered scenario.

A. Distributed Linear Precoding

Inspired by classical centralized MIMO systems, where linear precoding over Gaussian codewords is known to be capacity achieving [12], a reasonable approach is to extend this concept to the considered channel by letting

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} = \begin{bmatrix}\mathbf{G}_1^{\mathsf{H}}(s_1)\\ \mathbf{G}_2^{\mathsf{H}}(s_2)\end{bmatrix}\mathbf{u}, \quad \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_d), \tag{4}$$

where $\mathbf{G}_k(s_k) \in \mathbb{C}^{d\times N}$, $k = 1,2$, are *distributed* linear precoding matrix of dimension $d \times N$, depending only on the local CSIT s_k, and where d denotes the number of independent Gaussian streams over which W is encoded.

From classical arguments [12], [13], the scheme in (4) can be shown to achieve the rate

$$R(\mathbf{G}_1, \mathbf{G}_2) := \mathbb{E}\left[\log\det\left(\mathbf{I} + \mathbf{H}\mathbf{\Sigma}(s_1,s_2)\mathbf{H}^{\mathsf{H}}\right)\right], \tag{5}$$

where $\mathbf{\Sigma}(s_1,s_2) \in \mathbb{S}_+^{2N}$ is the conditional input covariance matrix given by

$$\begin{aligned}\mathbf{\Sigma}(s_1,s_2) :=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{H}} | s_1, s_2\right]\\ =& \begin{bmatrix}\mathbf{G}_1^{\mathsf{H}}(s_1)\\ \mathbf{G}_2^{\mathsf{H}}(s_2)\end{bmatrix}\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix}.\end{aligned} \tag{6}$$

Clearly, such approach is optimal when $q_1 = q_2$, i.e. when the network behaves like a virtually centralized MIMO system.

In this work, we focus on the individual power constraint

$$\mathbb{E}[\|\mathbf{x}_k\|^2 | s_k] = \|\mathbf{G}_k(s_k)\|_{\mathsf{F}}^2 \leq P_k, \ \forall s_k \in \mathcal{S}_k. \tag{7}$$

B. Ergodic Capacity

It has been recently shown in [8, Theorem 3] that, for $N = 1$ and $M = 2$, the distributed linear precoding scheme described in Sect. II-A is indeed capacity achieving. More precisely, we have the following result:

Theorem 1. *For $N = 1$, $M = 2$, a given $p(\mathbf{H}, s_2, s_2)$, and power constraints (P_1, P_2), the ergodic capacity is given by*

$$C = \max_{\substack{\mathbf{G}_k(s_k)\in\mathbb{C}^{d\times 1}\\ \|\mathbf{G}_k(s_k)\|_2^2 \leq P_k}} R(\mathbf{G}_1, \mathbf{G}_2), \tag{8}$$

where $d \leq d_{\max} := |\mathcal{S}_1| + |\mathcal{S}_2|$.

Proof: The proof is given in [8, Sect. III], by trivially adapting the average power constraint to (7). ∎

The main characteristic of Theorem 1 is that the underlying achievable scheme exploits the unconventional choice of letting the number of precoded Gaussian streams d to grow large, up to a given upper bound $d_{\max}$.

This surprising condition is in sharp contrast to the traditional design choice $d \leq \min(2N, M) = 2$. Indeed, for classical centralized 2×2 MIMO systems with CSIT s and per-antenna power constraints, such choice is optimal, as the capacity

$$C_{\text{centralized}} = \max_{\substack{\mathbf{\Sigma}(s) \in \mathbb{S}_+^2 \\ \mathbf{\Sigma}_{kk}(s) \leq P_k}} \mathbb{E}\left[\log\det\left(\mathbf{I} + \mathbf{H}\mathbf{\Sigma}(s)\mathbf{H}^{\mathsf{H}}\right)\right], \quad (9)$$

is always achievable by letting

$$\mathbf{x} = \mathbf{G}^{\mathsf{H}}(s)\mathbf{u}, \quad \mathbf{G}(s) = \mathbf{\Sigma}^{\frac{1}{2}}(s) \in \mathbb{C}^{2\times 2}, \quad \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_2). \quad (10)$$

However, such an approach cannot be in general applied to systems with D-CSIT, as taking any arbitrarily rotated matrix square-root of $\mathbf{\Sigma}(s_1, s_2)$ may generally violate the functional dependencies $(s_k, \mathbf{u}) \mapsto x_k$ at the distributed TXs.

Being an achievable scheme, the use of a number of data streams $d > 2$ for approaching the ergodic capacity in (8) is only a sufficient condition. However, this result is further enhanced in [9], where it is shown that the use of $d > 2$ is also necessary for some $p(\mathbf{H}, s_1, s_2)$.

In this work, we generalize the information theoretical insight given by [8] to arbitrary number of antennas N and M, by considering the following ergodic rate maximization problem:

$$\begin{aligned} \underset{\mathbf{G}_k(s_k)\in\mathbb{C}^{d\times N}}{\text{maximize}} \quad & R(\mathbf{G}_1, \mathbf{G}_2) \\ \text{subject to} \quad & \|\mathbf{G}_1(s_1)\|_{\mathsf{F}}^2 \leq P_1, \\ & \|\mathbf{G}_2(s_2)\|_{\mathsf{F}}^2 \leq P_2, \\ & d \leq d_{\max} := N(|\mathcal{S}_1| + |\mathcal{S}_2|). \end{aligned} \quad (11)$$

In the following sections we will see that by letting $d = d_{\max}$ is sufficient to exhaust the set of ergodic rates achievable via distributed linear precoders. Note that, for the same reasons explained above, for centralized systems a sufficient condition for optimality is instead $d = 2N$.

Remark 1. *By extending the proof in [8, Theorem 3], it is not difficult to show that the optimum in* (11) *is indeed the ergodic capacity, that is, distributed linear precoding is capacity achieving for arbitrary N and M. However, the proof of this statement is out of the scope of this work.*

III. Precoding Design with Asymmetric Feedback

A. Optimal Distributed Precoders Design

The distributed linear precoding design problem (11) is an instance of a well-known class of non-convex optimization problems called *static team decision* problems [5], [14], for which efficient optimal approaches are available only for very particular cases.

However, by letting $d = d_{\max}$, Problem (8) can be recast as an equivalent convex optimization problem. We point out that, conversely, with the traditional choice $d \leq 2N$ this was instead not possible in general.

Proposition 1. *Problem* (11) *is equivalent to the following convex problem*

$$\begin{aligned} \underset{\mathbf{Q}\in\mathbb{S}_+^{d_{\max}}}{\textit{maximize}} \quad & \mathbb{E}\left[\log\det\left(\mathbf{I} + \mathbf{H}_{\text{eq}}\mathbf{Q}\mathbf{H}_{\text{eq}}^{\mathsf{H}}\right)\right] \\ \textit{subject to} \quad & \operatorname{tr}\{\mathbf{Q}_{[N(i-1)+1:Ni]}\} \leq P_1, \\ & \operatorname{tr}\{\mathbf{Q}_{[N(j-1)+1:Nj]}\} \leq P_2, \\ & i = 1, \ldots, |\mathcal{S}_1|, \\ & j = (|\mathcal{S}_1| + 1), \ldots, d_{\max}, \end{aligned} \quad (12)$$

where we defined $\mathbf{H}_{\text{eq}} := \mathbf{H}\mathbf{E}^{\mathsf{H}}(s_1, s_2) \in \mathbb{C}^{M\times d_{\max}}$*, and*

$$\mathbf{E}(i,j) := \begin{bmatrix} \mathbf{e}_i \otimes \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_j \otimes \mathbf{I}_N \end{bmatrix} \in \mathbb{C}^{d_{\max}\times 2N}, \quad (13)$$

where $\mathbf{e}_i \in \{0,1\}^{|\mathcal{S}_1|}$ *(resp.* $\mathbf{e}_j \in \{0,1\}^{|\mathcal{S}_2|}$*) is a standard column selector, i.e. with the i-th (resp. j-th) element set to* 1 *and all the other elements set to* 0*.*

Proof: The proof follows the one given by [9] for the special case $N = 1$, $M = 2$.

Let us consider a given choice of distributed linear precoders $\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix} \in \mathbb{C}^{d\times 2N}$, collected in a codebook $\mathbf{F} \in \mathbb{C}^{d\times d_{\max}}$ ordered as follows

$$\mathbf{F} = \begin{bmatrix}\mathbf{G}_1(1) & \ldots & \mathbf{G}_1(|\mathcal{S}_1|) & \mathbf{G}_2(1) & \ldots & \mathbf{G}_2(|\mathcal{S}_2|)\end{bmatrix}.$$

The first $|\mathcal{S}_1|$ submatrices of $\mathbf{F}$ are the possible precoder choices for TX 1, while the remaining $|\mathcal{S}_2|$ submatrices are for TX 2. With these definitions, the distributed precoders matrix $\mathbf{G}_k(s_k)$ can be obtained from the codebook $\mathbf{F}$ as

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

where $\mathbf{E}(s_1, s_2)$ is defined in (13), and hence

$$\begin{aligned} \mathbf{\Sigma}(s_1, s_2) &= \begin{bmatrix}\mathbf{G}_1^{\mathsf{H}}(s_1) \\ \mathbf{G}_2^{\mathsf{H}}(s_2)\end{bmatrix}\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix} \\ &= \mathbf{E}^{\mathsf{H}}(s_1, s_2)\mathbf{F}^{\mathsf{H}}\mathbf{F}\mathbf{E}(s_1, s_2) \\ &= \mathbf{E}^{\mathsf{H}}(s_1, s_2)\mathbf{Q}\mathbf{E}(s_1, s_2), \end{aligned} \quad (14)$$

where $\mathbf{Q} := \mathbf{F}^{\mathsf{H}}\mathbf{F} \in \mathbb{S}_+^{d_{\max}}$ is a (fixed) positive semi-definite matrix with rank $d \leq d_{\max}$. We can now rewrite the rate achieved by $\mathbf{G}_k(s_k)$ as

$$\begin{aligned} R(\mathbf{G}_1, \mathbf{G}_2) &= \mathbb{E}\left[\log\det\left(\mathbf{I} + \mathbf{H}\mathbf{\Sigma}(s_1, s_2)\mathbf{H}^{\mathsf{H}}\right)\right] \\ &= \mathbb{E}\left[\log\det\left(\mathbf{I} + \mathbf{H}_{\text{eq}}\mathbf{Q}\mathbf{H}_{\text{eq}}^{\mathsf{H}}\right)\right] =: R_{\text{eq}}(\mathbf{Q}), \end{aligned}$$

where we defined the equivalent channel

$$\mathbf{H}_{\text{eq}} := \mathbf{H}\mathbf{E}^{\mathsf{H}}(s_1, s_2) = \mathbf{H}\mathbf{E}^{\mathsf{H}}(q_1(\mathbf{H}), q_2(\mathbf{H})) \in \mathbb{C}^{M\times d_{\max}}.$$

By removing the rank constraint $\operatorname{rank}(\mathbf{Q}) \leq d$, i.e. by letting $d = d_{\max}$, the function $R_{\text{eq}}(\mathbf{Q})$ becomes convex in $\mathbf{Q}$. The proof is completed by writing the power constraints in terms of linear constraints on $\mathbf{Q}$. ■

We notice that Problem (12) corresponds to the capacity of an equivalent $d_{\max} \times 2$ MIMO channel with state $\mathbf{H}_{\text{eq}}$, perfect CSIR, no CSIT, and fixed transmit covariance $\mathbf{Q}$. The optimal distributed precoders for the original channel can be then designed from the optimal $\mathbf{Q}^\star$ as follows

$$\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix} = (\mathbf{Q}^\star)^{\frac{1}{2}}\mathbf{E}(s_1, s_2) \in \mathbb{C}^{d_{\max}\times 2N}. \quad (15)$$

Note that, although the optimal $\mathbf{Q}^\star$ is unique because the objective of (12) is strictly convex in the feasible set, the optimal distributed precoders are not necessarily unique, as one can equivalently consider any arbitrary rotation of $(\mathbf{Q}^\star)^{\frac{1}{2}}$.

B. Joint Feedback and Precoders Design

In this section we show how problem (12) can be used to design the quantizers (q_1, q_2) at the RX, given constraints β_k on the feedback rates towards TX k.

More precisely, we are interested in solving the following optimization problem:

$$\max_{\substack{\mathbf{G}_k(s_k)\in\mathbb{C}^{d_{\max}\times N},\ \|\mathbf{G}_k(s_2)\|_{\mathsf{F}}^2\leq P_k\\ q_k(\mathbf{H})\in\mathcal{S}_k,\ |\mathcal{S}_k|=2^{\beta_k}}} R(\mathbf{G}_1,\mathbf{G}_2), \tag{16}$$

where the ergodic rate is maximized by optimizing both precoders $\mathbf{G}_k(s_k)$ under individual power constraints and vector quantizer functions $q_k(\mathbf{H})$ with feedback rate constraints.

To address the above problem, which is non-convex, we propose a suboptimal approach based on the *generalized Lloyd algorithm* [10] for vector quantization, similarly to [11] for centralized MIMO channels.

In particular, similarly to the classical Lloyd algorithm, we use a converging alternating optimization procedure composed by the following two steps:

1) Quantizers update step:
$$(q_1^\star(\mathbf{H}), q_2^\star(\mathbf{H})) \in \arg\max_{(s_1,s_2)\in\mathcal{S}_1\times\mathcal{S}_2} R(\mathbf{H},\mathbf{Q}), \tag{17}$$
2) Precoders update step:
$$\mathbf{Q}^\star = \arg\max_{\substack{\mathbf{Q}\in\mathbb{S}_+^{d_{\max}}\\ \mathbf{Q}_{ii}\leq P_1,\ i=1,\ldots,|\mathcal{S}_1|\\ \mathbf{Q}_{jj}\leq P_2,\ j=(|\mathcal{S}_1|+1),\ldots,d_{\max}}} \mathbb{E}\left[R(\mathbf{H},\mathbf{Q})\right], \tag{18}$$

where we defined
$$R(\mathbf{H},\mathbf{Q}) := \log\det\left(\mathbf{I}+\mathbf{H}_{\mathrm{eq}}\mathbf{Q}\mathbf{H}_{\mathrm{eq}}^{\mathsf{H}}\right),$$
where $\mathbf{H}_{\mathrm{eq}} = \mathbf{H}\mathbf{E}^{\mathsf{H}}(s_1,s_2)$ is given by (12). Note that the precoders update step corresponds to Problem (12), hence it is a convex problem.

The two above steps are respectively similar to the *quantization regions* update rule and the *centroid* update rule given by the Lloyd algorithm, under a modified distortion measure $-R(\mathbf{H},\mathbf{Q})$. However, there is a non-trivial differences between the proposed algorithm and the ones in [10], [11]. In particular, because of the D-CSIT assumption, the centroids must be jointly optimized in a unique step, and not disjointly as in the classical Lloyd algorithm.

Similarly to [10], to coope with the expectation in the precoders update step (or equivalently, the expectation in problem (12)), we approximate $p(\mathbf{H})$ by its empirical distribution $\hat{p}(\mathbf{H}) = \frac{1}{L}\sum_{i=1}^{L}\mathbb{1}[\mathbf{H}=\mathbf{H}_i]$ obtained from L training samples $\{\mathbf{H}_i\}_{i=1}^{L}$ generated i.i.d. according to $p(\mathbf{H})$. This allows us to replace the expectation in (12) with a finite sum of L convex functions.

Note that, for an arbitrary initialization, the above procedure converges to a local optimum of (16). To avoid bad local optima, classical multi-start methods (e.g. random sampling) may be applied. However, in this work we omit this step.

Finally, we point out that although during the design phase the quantizers are optimized over the samples $\mathbf{H}_i$, the quantizers for arbitrary input $\mathbf{H}$ are simply given by (17).

C. Constrained Number of Data Streams

Although letting $d = d_{\max}$ is capacity achieving, practical systems may be constrained to use $d \ll d_{\max}$ data streams. For example, if the RX adopts a non-ideal successive interference cancellation decoder, a lower d typically results in lower decoding complexity and smaller error propagation.

A constraint d on the number of data streams can be imposed by constraining the rank of $\mathbf{Q}$ in (12) (see (14)), leading to the non-convex optimization problem

$$\begin{aligned}\underset{\mathbf{Q}\in\mathbb{S}_+^{d_{\max}}}{\text{maximize}}\quad & \mathbb{E}\left[\log\det\left(\mathbf{I}+\mathbf{H}_{\mathrm{eq}}\mathbf{Q}\mathbf{H}_{\mathrm{eq}}^{\mathsf{H}}\right)\right]\\ \text{subject to}\quad & \mathrm{tr}\{\mathbf{Q}_{[N(i-1)+1:Ni]}\}\leq P_1,\\ & \mathrm{tr}\{\mathbf{Q}_{[N(j-1)+1:Nj]}\}\leq P_2,\\ & i=1,\ldots,|\mathcal{S}_1|,\\ & j=(|\mathcal{S}_1|+1),\ldots,d_{\max},\\ & \mathrm{rank}(\mathbf{Q})\leq d\end{aligned} \tag{19}$$

Finding a good solution for rank constrained problems similar to (19) is an interesting open problem that is out of the scope of this work. In what follows, we propose a simple yet effective sub-optimal approach for solving (19) based on creating a feasible solution from the optimal $\mathbf{Q}^\star$ of the unconstrained problem (12).

Let $\mathbf{Q}_r = \arg\min_{\mathbf{Q}\in\mathcal{Q}_r}\|\mathbf{Q}-\mathbf{Q}^\star\|_{\mathrm{F}}$ be the projection of $\mathbf{Q}^\star$ onto the set of rank constrained positive semidefinite matrices $\mathcal{Q}_r := \{\mathbf{Q}\in\mathbb{S}_+^{d_{\max}} \mid \mathrm{rank}(\mathbf{Q})\leq r\}$. This can be readily obtained by letting

$$\mathbf{Q}_r = \mathbf{V}_r\mathbf{\Delta}_r\mathbf{V}_r^{\mathsf{H}},\quad \mathbf{V}_r\in\mathbb{C}^{d_{\max}\times r},\quad \mathbf{\Delta}_r\in\mathbb{C}^{r\times r},$$

where $\mathbf{\Delta}_r$ is a diagonal matrix containing the r largest eigenvalues of $\mathbf{Q}^\star$, and the columns of $\mathbf{V}_r$ are the corresponding eigenvectors. Note that if $\mathbf{Q}^\star\in\mathcal{Q}_d$, then $\mathbf{Q}_d = \mathbf{Q}^\star$ is clearly also optimal for (19), and no particular processing is required. In such case, the optimal precoders are simply given by

$$\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix} = \mathbf{\Delta}_d^{\frac{1}{2}}\mathbf{V}_d^{\mathsf{H}}\mathbf{E}(s_1,s_2)\in\mathbb{C}^{d\times 2}.$$

On the other hand, if $\mathbf{Q}^\star\notin\mathcal{Q}_d$, the suboptimal solution $\mathbf{Q}_d$ may incur some power loss. To handle this problem, we consider instead $\mathbf{Q}_{d-2N}$, and exploit the remaining power to transmit $2N$ data streams, each of them available at one TX only. More precisely, we let

$$[\mathbf{G}_1'(s_1)\ \mathbf{G}_2'(s_2)] = \mathbf{\Delta}_{d-2N}^{\frac{1}{2}}\mathbf{V}_{d-2N}^{\mathsf{H}}\mathbf{E}(s_1,s_2)\in\mathbb{C}^{(d-2N)\times 2N},$$

and

$$\begin{bmatrix}\mathbf{G}_1(s_1) & \mathbf{G}_2(s_2)\end{bmatrix} = \begin{bmatrix}\mathbf{G}_1'(s_1) & \mathbf{G}_2'(s_2)\\ \mathbf{G}_1''(s_1) & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_2''(s_2)\end{bmatrix}\in\mathbb{C}^{d\times 2},$$

where $\mathbf{G}_k''(s_k)\in\mathbb{C}^{N\times N}$ are two disjoint local precoders (similar to Sect. IV-B). Note that the above precoders have the desired dimension. Furthermore, after some manipulations it can be shown that the resulting $\mathbf{Q}$, defined as in (14), satisfies

$$\mathbf{E}^{\mathsf{H}}(s_1,s_2)\mathbf{Q}\mathbf{E}(s_1,s_2) = \mathbf{E}^{\mathsf{H}}(s_1,s_2)(\mathbf{Q}_{d-2N}+\mathbf{\Gamma})\mathbf{E}(s_1,s_2),$$

where

$$\mathbf{\Gamma} := \mathrm{blkdiag}(\mathbf{\Gamma}_1(1),\ldots,\mathbf{\Gamma}_1(|\mathcal{S}_1|),\mathbf{\Gamma}_2(1),\ldots,\mathbf{\Gamma}_2(|\mathcal{S}_2|)),$$
$$\mathbf{\Gamma}_k(s_k) := (\mathbf{G}_k''(s_k))^{\mathsf{H}}\mathbf{G}_k''(s_k).$$

Thus, the disjoint local precoders can be optimized by solving the following convex problem

$$\begin{aligned} \underset{\mathbf{\Gamma}\in\mathbb{S}_+^{d_{\max}}}{\text{maximize}} \quad & \mathbb{E}\left[\log\det\left(\mathbf{I}+\mathbf{H}_{\text{eq}}(\mathbf{Q}_{d-2N}+\mathbf{\Gamma})\mathbf{H}_{\text{eq}}^{\mathsf{H}}\right)\right] \\ \text{subject to} \quad & \operatorname{tr}\{\mathbf{\Gamma}_{[N(i-1)+1:Ni]}\} \leq P_1 - \|\mathbf{G}_1'(i)\|_{\mathsf{F}}^2, \\ & \operatorname{tr}\{\mathbf{\Gamma}_{[N(j-1)+1:Nj]}\} \leq P_2 - \|\mathbf{G}_2'(j-|\mathcal{S}_1|)\|_{\mathsf{F}}^2, \\ & i = 1,\ldots,|\mathcal{S}_1|, \\ & j = (|\mathcal{S}_1|+1),\ldots,d_{\max}, \end{aligned} \tag{20}$$

Note that the suboptimal scheme described cannot handle well very low values of d, as it converges to a suboptimal disjoint local precoding scheme for $d = 2N$. Furthermore, the above scheme is not defined for $d \leq 2N$.

To conclude, we point out that the algorithm described in this section can be applied as a final step of the method given by Sect. III-B to obtain precoders of practical dimension.

IV. Performance Evaluation

A. Simulation Setup

We consider two different antenna configurations $(N, M) = (2, 2)$ and $(N, M) = (1, 1)$, denoted respectively by *distributed MIMO* and *distributed MISO*. We consider i.i.d. Rayleigh fading, i.e. we let the elements of $\mathbf{H}$ to be i.i.d. distributed according to $\mathcal{CN}(0, 1)$. For simplicity, we consider equal power constraints $P_1 = P_2 =: \text{SNR}$.

The feedback links rates are constrained to $(\beta_1, \beta_2) = (4, 3)$ for the cooperative MIMO case, and to $(\beta_1, \beta_2) = (2, 1)$ for the cooperative MISO case. This implies that optimal precoders are designed by using respectively up to $d_{\max} = 48$ and $d_{\max} = 6$ spatial streams for the two cases.

For all the optimization problems involved, we approximate the statistical averages by their empirical averages obtained from $L = 100$ i.i.d. training samples $\{\mathbf{H}_i\}_{i=1}^{L}$, as described in Sect. III-B, and we use a numerical convex solver. The performance are then evaluated over a different i.i.d. test set of $L_{\text{test}} = 10000$ samples.

B. Precoding From Conventional Centralized Design

In this section we briefly review available suboptimal distributed precoding techniques derived from centralized design, used here for performance comparison.

- *Statistical* precoding (no CSIT): by neglecting the feedback, solve

$$\max_{\substack{\mathbf{\Sigma}\in\mathbb{S}_+^{2N} \\ \operatorname{tr}\{\mathbf{\Sigma}_{[1:N,1:N]}\}\leq P_1 \\ \operatorname{tr}\{\mathbf{\Sigma}_{[N+1:2N,N+1:2N]}\}\leq P_2}} \mathbb{E}\left[\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{\Sigma}\mathbf{H}^{\mathsf{H}}\right)\right],$$

  and let $[\mathbf{G}_1(s_1)\ \mathbf{G}_2(s_2)] = \mathbf{\Sigma}^{\frac{1}{2}}$.
- *Local* precoding: each TX is treated disjointly, as if it is the only TX in the system. Hence, for each TX k, we run a centralized version of the alternating optimization procedure in Sect III-B similar to [11], given by the following two steps:

$$\max_{\substack{\mathbf{\Sigma}(s_k)\in\mathbb{S}_+^{N} \\ \operatorname{tr}\{\mathbf{\Sigma}(s_k)\}\leq P_k,}} \mathbb{E}\left[\log\det\left(\mathbf{I}+\mathbf{H}^{(k)}\mathbf{\Sigma}(s_k)(\mathbf{H}^{(k)})^{\mathsf{H}}\right)\right],$$

$$q_k(\mathbf{H}^{(k)}) \in \arg\max_{s_k\in\mathcal{S}_k} \log\det\left(\mathbf{I}+\mathbf{H}^{(k)}\mathbf{\Sigma}(s_k)(\mathbf{H}^{(k)})^{\mathsf{H}}\right),$$

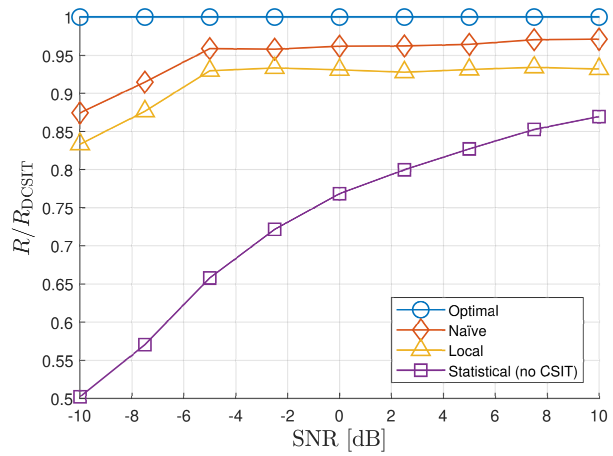
  where $\mathbf{H}^{(k)} \in \mathbb{C}^{N\times N}$ is the channel submatrix corresponding to TX k, i.e. $[\mathbf{H}^{(1)}\ \mathbf{H}^{(2)}] =: \mathbf{H}$.
  We then let $\mathbf{G}_k(s_k) = \mathbf{\Sigma}^{\frac{1}{2}}(s_k)$.
- *(Robust) Naïve* precoding [5]: each TX is optimized by assuming that the other TX shares the same CSIT. Hence, for each TX k, we run an instance of the alternating optimization procedure given by the following two steps:

$$\max_{\substack{\mathbf{\Sigma}\in\mathbb{S}_+^{2N} \\ \operatorname{tr}\{\mathbf{\Sigma}_{[1:N,1:N]}(s_1)\}\leq P_1 \\ \operatorname{tr}\{\mathbf{\Sigma}_{[N+1:2N,N+1:2N]}(s_2)\}\leq P_2}} \mathbb{E}\left[\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{\Sigma}(s_k)\mathbf{H}^{\mathsf{H}}\right)\right],$$

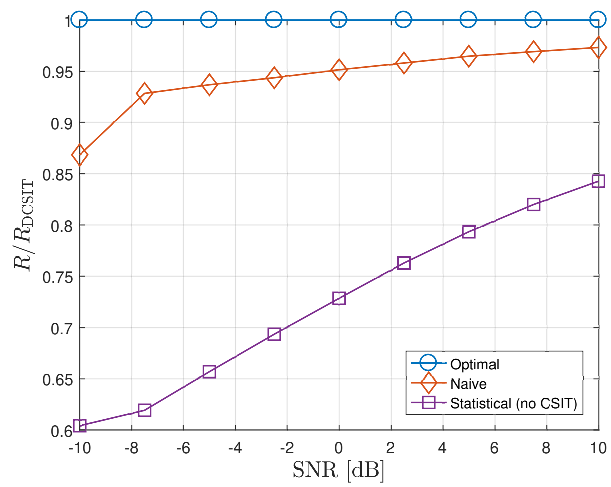
$$q_k(\mathbf{H}) \in \arg\max_{s_k\in\mathcal{S}_k} \log\det\left(\mathbf{I}+\mathbf{H}\mathbf{\Sigma}(s_k)\mathbf{H}^{\mathsf{H}}\right).$$

  We then let $\mathbf{G}_k(s_k) = \mathbf{\Sigma}^{\frac{1}{2}}(s_k)(\mathbf{e}_k\otimes\mathbf{I}_N)$. Note that each TX computes both precoders, but only the solution for its own precoder is kept.

For a fair comparison, although for the design phase the local

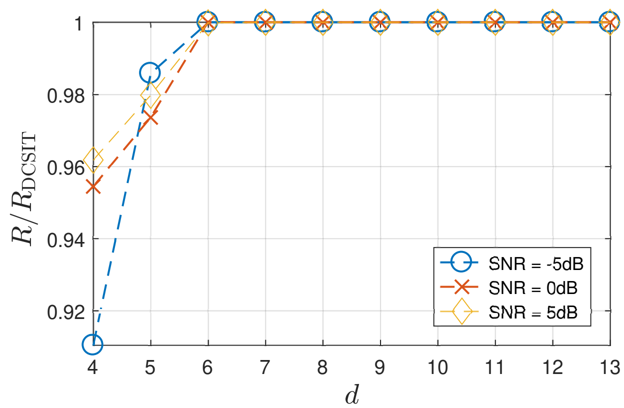


(a) 4×2 Cooperative MIMO

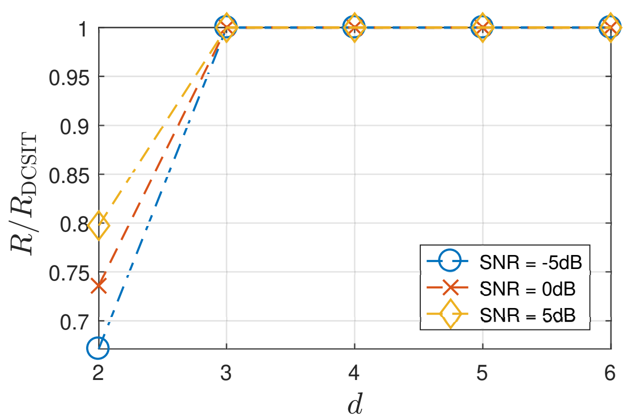


(b) 2×1 Cooperative MISO

Fig. 2: Performance comparison vs SNR for (a) $(\beta_1, \beta_2) = (4, 3)$ and (b) $(\beta_1, \beta_2) = (2, 1)$ feedback bits.



(a) 4×2 Cooperative MIMO



(b) 2×1 Cooperative MISO

Fig. 3: Relative rate loss vs constraint $d \leq d_{\max}$ on the number of spatial streams, for (a) $d_{\max} = 48$ and (b) $d_{\max} = 6$.

and the naïve precoding algorithms assume the RX to adopt a suboptimal feedback strategy, in the performance evaluation phase we use the rate optimal strategy given by (17).

C. Simulation Results

In Fig. 2 we plot the performance comparison versus SNR for the aforementioned setup, normalized by the rate R_{DCSIT} achieved by the proposed joint feedback and precoders design given in Sect. III-B. Note that, for the cooperative MISO case, the local precoding is omitted since it corresponds to the disjoint optimization of two SISO links, where CSIT is useless (we don't consider power allocation over time, see power constraint (7)).

We observe that, although designed over a training set of relatively small size, the proposed precoding and feedback strategy exhibits strictly better performance than the competing approaches also over the test set. However, this comes at the price of using $d = d_{\max}$ spatial streams.

In Fig. 3 we plot instead the normalized performance loss of the proposed algorithm for constraining the maximum number of data streams given by Sect. III-C. We recall that the proposed algorithm is suboptimal, especially for low values of d. However, we observe that, for the considered setting, it is able to reduce the number of data streams from the theoretical upper bound $d_{\max}$ down to respectively $d = 6$ and $d = 3$ for the MIMO and for the MISO case, with negligible performance loss.

V. Conclusion

Inspired by the recent information theoretical findings in [8], we derived practical guidelines for precoding and feedback design for cooperative MIMO channels with asymmetric feedback rates. The benefits of the proposed methods are assessed via numerical simulations. Future works include the extension of the ideas exploited in this paper to the more interesting scenario of systems with multiple receivers.

References

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

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

[3] A. G. Davoodi and S. A. Jafar, "Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5603–5618, Nov. 2016.

[4] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

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

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

[7] A. Bazco Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving vanishing rate loss in decentralized network MIMO," *Proc. IEEE Int. Symp. Inf. Theory*, 2019.

[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*, 2019.

[9] L. Miretti, M. Kobayashi, D. Gesbert, and P. de Kerret, "Cooperative multiple-access channels with distributed channel state information," *available on arXiv*, 2019.

[10] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[11] V. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

[12] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[13] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

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