Capacity Bounds for Gaussian MIMO relay channel with Channel State Information

Sebastien Simoens, Olga Muñoz[1], Josep Vidal[1], Aitor Del Coso[2]

Motorola Labs, Paris, France. E-mail : simoens@motorola.com

[1]Universitat Politècnica de Catalunya (UPC), Signal Processing and Communications group, Barcelona, Spain

[2]Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Casteldefells, Barcelona, Spain

Abstract—**In this paper, source and relay precoders are derived which optimize upper and lower bounds on the Gaussian MIMO relay channel capacity. First, the prior art on the cut-set upper-bound on capacity is extended by showing that the optimization of the source and relay codebooks can be formulated as a convex problem without having to introduce a scalar parameter that captures their cross-correlation. Both the Full-Duplex and Time Division Duplex (TDD) relay channels are addressed, assuming perfect knowledge of all channels, and two procedures are proposed which solve the problem efficiently by relying on analytical expressions of gradients, subgradients and projection operators: the first one solves the dual problem while the second one applies the barrier method. Similar techniques are then used to maximize the achievable rate of Decode-and-Forward (DF) TDD MIMO relaying strategies with either partial or full decoding at the relay. Sub-optimum precoders are also proposed which have a closed-form expression that can be obtained from the KKT conditions, thus reducing the computational complexity at the expense of a lower rate. Simulations in a cellular downlink scenario show that the partial DF strategy can achieve a rate very close to capacity for realistic values of the Source to Relay signal-to-noise ratio. Finally, the availability of Channel State Information (CSI) in a real system is discussed.**

Index Terms—**cooperative, relaying, decode, forward, CSI, MIMO**

I. INTRODUCTION

Cooperative Relaying is regarded as a promising technique for next generation cellular wireless systems (e.g. [1]) and is being considered by several standardization groups (e.g. IEEE802.16j). This paper focuses on the one-way single relay channel, where the Relay (R) cooperates with the Source (S) and the Destination (D) to maximize the information rate from S to D. Although full-duplex relaying is briefly addressed, most of the paper considers a two-phase TDD relaying[1] protocol with a relay-receive slot of relative duration denoted by $t \in [0;1]$ and a relay-transmit slot of duration $1-t$. The number of antennas at S, R and D are respectively denoted by N_S, N_R and N_D and each one may be greater than 1. The channels from S to D, S to R and R to D are all assumed static and are denoted respectively $\mathbf{H}_0$, $\mathbf{H}_1$ and $\mathbf{H}_2$. Full knowledge of all channels is assumed to be available at all nodes, except in section IV where this assumption, called "full CSI" in the rest of the paper, is challenged.

The maximization of the Gaussian MIMO channel capacity with perfect CSI at the source is solved by Telatar in [2]. In the landmark paper [3], the capacity of the full-duplex single-relay channel is studied: the Decode-and-Forward (DF) (Theorem 1) and Compress-and-Forward (CF) (Theorem 6) coding strategies are introduced and an upper-bound on the relay channel capacity (a.k.a. cut-set bound) is provided. In DF the relay decodes at least a part of the message transmitted by the source, whereas in CF the relay observation is compressed and forwarded to the destination as a new codeword. In this paper, we focus on DF. Neither DF nor CF is capacity-achieving in the general relay channel, but DF meets the cut-set bound when R gets infinitely close to S, while CF outperforms DF when R gets close to D, ultimately achieving the capacity when their distance goes to zero (see e.g.[4]). Other strategies exist like Amplify-and-Forward (AF) which can be viewed as a sub-case of Linear Relaying (LR). One advantage of LR is processing simplicity and we refer the reader to e.g. [5] for the optimization of linear relaying with full MIMO CSI. A drawback of LR in the single TDD relaying case is that the time-sharing is constrained to $t = 1/2$, which severely limits the achievable rate. This effect can be alleviated by two-path relaying (when multiple relays are available) or two-way relaying (which applies to bidirectional traffic) for which precoder optimization can also be performed [8].

In [6], the cut-set bound and the achievable rate of a partial DF strategy are expressed in the case of Gaussian TDD relaying, assuming single antenna devices. In [7], an upper-bound and several lower-bounds are derived for the full-duplex MIMO relay channel with full CSI. In [9], two strategies are introduced which improve the achievable rate on the full-duplex MIMO relay channel. The present paper aims at contributing to the study of MIMO relaying as follows:

- Two alternative convex optimization procedures are presented in section II, which compute the cut-set upper-bound on both full-duplex and TDD MIMO relay channels without requiring the introduction of a scalar parameter to capture the cross-correlation as in [7].

- The achievable rates of DF strategies with either partial or full decoding at the relay are computed for the TDD MIMO relay channel in section III. Moreover, sub-

---

[1] Half-Duplex relaying lends itself more easily to practical implementation, by removing the need to physically isolate the relay transmit and receive chains

optimum source and relay precoders are proposed with analytical expressions derived from KKT conditions.

- The upper and lower bounds are benchmarked by simulations in a cellular downlink scenario (section IV) before discussing on CSI availability in real systems.

II. THE CUT-SET BOUND WITH FULL MIMO CSI

We start by introducing the notations and formulating the cut-set bound as the solution of a convex optimization problem in the full-duplex and TDD relaying cases, before proposing methods to solve it.

A. Formulation of the problem

In full-duplex MIMO relaying, the signals received at the relay and destination can be written as in [7]:

$$\begin{aligned} \mathbf{y}_R &= \mathbf{H}_1 \mathbf{x}_S + \mathbf{n}_R \\ \mathbf{y}_D &= \mathbf{H}_0 \mathbf{x}_S + \mathbf{H}_2 \mathbf{x}_R + \mathbf{n}_D \end{aligned} \tag{1}$$

where circularly symmetric complex white Gaussian noise of unit variance is assumed at the relay and destination, i.e. $\mathbf{n}_R \sim \mathbb{CN}(\mathbf{0}, \mathbf{I}_{N_R})$ and $\mathbf{n}_D \sim \mathbb{CN}(\mathbf{0}, \mathbf{I}_{N_D})$. The expression of the cut-set bound is given by equation (3) in [7]:

$$C_{FD} \le \max_{p(\mathbf{x}_S, \mathbf{x}_R)} \min(I(\mathbf{x}_S; \mathbf{y}_D, \mathbf{y}_R | \mathbf{x}_R), I(\mathbf{x}_S, \mathbf{x}_R; \mathbf{y}_D)) \tag{2}$$

where the maximization is performed over the distribution of the source and relay codebooks $p(\mathbf{x}_S, \mathbf{x}_R)$. The authors in [7] show that Gaussian codebooks are optimum and conclude that the optimization of (2) must be carried on w.r.t. three matrices $\mathbf{R}_S \triangleq E[\mathbf{x}_S \mathbf{x}_S^H]$, $\mathbf{R}_R \triangleq E[\mathbf{x}_R \mathbf{x}_R^H]$ and the cross-correlation $E[\mathbf{x}_S \mathbf{x}_R^H]$. This optimization seems highly non-trivial and the authors resort to introducing a scalar parameter ρ that captures the cross-correlation in order to finally obtain (Theorem 3.1) an upper-bound which involves a maximization only over $\mathbf{R}_S$, $\mathbf{R}_R$ and ρ. This bound is however not evaluated in the fixed channel case in [7] and its numerical computation seems complex because evaluating the objective requires to solve an optimization problem (equation (9) in [7]). However, let define the joint covariance matrix:

$$\mathbf{R}_{SR} \triangleq \begin{bmatrix} \mathbf{R}_S & E[\mathbf{x}_S \mathbf{x}_R^H] \\ E[\mathbf{x}_R \mathbf{x}_S^H] & \mathbf{R}_R \end{bmatrix} \tag{3}$$

Let also define the following matrices:

$$\mathbf{D}_S \triangleq [\mathbf{I}_{N_S} \ \mathbf{0}_{N_S \times N_R}] \text{ and } \mathbf{D}_R \triangleq [\mathbf{0}_{N_R \times N_S} \ \mathbf{I}_{N_R}] \tag{4}$$

From (3) and (4), the following relationships hold:

$$\mathbf{R}_S = \mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H \text{ and } \mathbf{R}_R = \mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H \tag{5}$$

Note that if $\mathbf{R}_{SR}$ is Positive Semidefinite (PSD), then $\mathbf{R}_S$ and $\mathbf{R}_R$ are PSD too:

$$\mathbf{R}_{SR} \succeq \mathbf{0} \Rightarrow \mathbf{R}_S \succeq \mathbf{0} \text{ and } \mathbf{R}_R \succeq \mathbf{0} \tag{6}$$

Indeed, for any two vectors $\mathbf{x} \in \mathbb{C}^{N_S}$ and $\mathbf{y} \in \mathbb{C}^{N_R}$, defining $\tilde{\mathbf{x}} \triangleq \mathbf{D}_S^H \mathbf{x}$ and $\tilde{\mathbf{y}} \triangleq \mathbf{D}_R^H \mathbf{y}$, the following holds:

$$\begin{aligned} \mathbf{x}^H \mathbf{R}_S \mathbf{x} &\overset{(a)}{=} \mathbf{x}^H \mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H \mathbf{x} = \tilde{\mathbf{x}}^H \mathbf{R}_{SR} \tilde{\mathbf{x}} \overset{(b)}{\ge} 0 \\ \mathbf{y}^H \mathbf{R}_R \mathbf{y} &\overset{(a)}{=} \mathbf{y}^H \mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H \mathbf{y} = \tilde{\mathbf{y}}^H \mathbf{R}_{SR} \tilde{\mathbf{y}} \overset{(b)}{\ge} 0 \end{aligned} \tag{7}$$

where (a) comes from equation (5) and (b) from the positive semi-definiteness of $\mathbf{R}_{SR}$.

The cut-set bound (2) can be expressed as:

$$\begin{aligned} C_{FD} &\le \max_{\mathbf{R}_{SR} \succeq 0} \min(C_A, C_B) \\ C_A &= \log_2 \left| \mathbf{I}_{N_S} + \mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H (\mathbf{H}_0^H \mathbf{H}_0 + \mathbf{H}_1^H \mathbf{H}_1) \right| \\ C_B &= \log_2 \left| \mathbf{I}_{N_D} + [\mathbf{H}_0 \ \mathbf{H}_2] \mathbf{R}_{SR} [\mathbf{H}_0 \ \mathbf{H}_2]^H \right| \\ \text{s.t. } &\operatorname{tr}(\mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H) \le P_S \text{ and } \operatorname{tr}(\mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H) \le P_R \end{aligned} \tag{8}$$

where P_S and P_R are the maximum transmit powers at S and R. The objective is concave on the PSD cone, because it is the pointwise minimum of two concave functions [10]. Therefore the problem is convex with respect to a unique variable $\mathbf{R}_{SR}$.

In the case of TDD relaying, the cut-set bound can be expressed as follows (see equation (77) in [6]):

$$C_{TDD} \le \max_{\substack{0 \le t \le 1 \\ p(\mathbf{x}_S^{(1)}, \mathbf{x}_S^{(2)}, \mathbf{x}_R^{(2)})}} \min \begin{Bmatrix} t I(\mathbf{x}_S^{(1)}; \mathbf{y}_R^{(1)}, \mathbf{y}_D^{(1)} | \mathbf{x}_R^{(1)} = \mathbf{0}) + (1-t) I(\mathbf{x}_S^{(2)}; \mathbf{y}_D^{(2)} | \mathbf{x}_R^{(2)}), \\ t I(\mathbf{x}_S^{(1)}; \mathbf{y}_D^{(1)} | \mathbf{x}_R^{(1)} = \mathbf{0}) + (1-t) I(\mathbf{x}_S^{(2)}, \mathbf{x}_R^{(2)}; \mathbf{y}_D^{(2)}) \end{Bmatrix} \tag{9}$$

where the superscript $^{(i)}$ indicates the slot during which the signal was transmitted or received. Using similar arguments as in the full-duplex case, the cut-set bound for the TDD MIMO relay channel can be expressed as:

$$C_{TDD} \le \max_{\substack{0 \le t \le 1 \\ \mathbf{R}_S^{(1)} \succeq 0, \mathbf{R}_{SR}^{(2)} \succeq 0}} \min\{C_A, C_B\}$$

$$C_A = t \log_2 \left| \mathbf{I}_{N_S} + \mathbf{R}_S^{(1)} (\mathbf{H}_0^H \mathbf{H}_0 + \mathbf{H}_1^H \mathbf{H}_1) \right| + (1-t) \log_2 \left| \mathbf{I}_{N_D} + \mathbf{H}_0 \mathbf{D}_S \mathbf{R}_{SR}^{(2)} \mathbf{D}_S^H \mathbf{H}_0^H \right|$$

$$C_B = t \log_2 \left| \mathbf{I}_{N_D} + \mathbf{H}_0 \mathbf{R}_S^{(1)} \mathbf{H}_0^H \right| + (1-t) \log_2 \left| \mathbf{I}_{N_D} + [\mathbf{H}_0 \ \mathbf{H}_2] \mathbf{R}_{SR}^{(2)} [\mathbf{H}_0 \ \mathbf{H}_2]^H \right|$$

$$\text{s.t. } \operatorname{tr}(\mathbf{R}_S^{(1)}) \le P_S \ ; \ \operatorname{tr}(\mathbf{D}_S \mathbf{R}_{SR}^{(2)} \mathbf{D}_S^H) \le P_S \text{ and } \operatorname{tr}(\mathbf{D}_R \mathbf{R}_{SR}^{(2)} \mathbf{D}_R^H) \le P_R \tag{10}$$

which is convex in $(\mathbf{R}_S^{(1)}, \mathbf{R}_{SR}^{(2)})$ for a given t.

B. Solution of the problems

The procedures presented in this section aim at solving (8) and (10) efficiently by relying on the convex optimization literature [10][11]. Let first focus on (8). It can be turned into a differentiable problem by writing the equivalent epigraph [10] form of the problem:

$$\begin{aligned} C_{FD} &\le -\min_{\mathbf{R}_{SR} \succeq 0}(-\varepsilon) \\ \text{s.t. } &\underbrace{\varepsilon - C_A}_{\triangleq f_1(\varepsilon, \mathbf{R}_{SR})} \le 0 \ ; \ \underbrace{\varepsilon - C_B}_{\triangleq f_2(\varepsilon, \mathbf{R}_{SR})} \le 0 \\ &\underbrace{\operatorname{tr}(\mathbf{D}_S \mathbf{R}_{SR} \mathbf{D}_S^H) - P_S}_{\triangleq f_3(\varepsilon, \mathbf{R}_{SR})} \le 0 \text{ and } \underbrace{\operatorname{tr}(\mathbf{D}_R \mathbf{R}_{SR} \mathbf{D}_R^H) - P_R}_{\triangleq f_4(\varepsilon, \mathbf{R}_{SR})} \le 0 \end{aligned} \tag{11}$$

This problem is convex in $(\varepsilon, \mathbf{R}_{SR})$ on $\mathbb{R}^+ \times \mathbb{S}^+$ where $\mathbb{S}^+$ is the cone of PSD matrices. For notational convenience we introduce the following column vector:

$$\mathbf{v} \triangleq [\varepsilon \ \operatorname{vec}(\mathbf{R}_{SR}^T)^T]^T \tag{12}$$

where vec(.) is the operator that stacks the colums of a matrix to generate a column vector.

Starting from (11), a first method to compute the cut-set bound consists in solving the dual problem:

$$\max_{\boldsymbol{\mu} \ge 0} g(\boldsymbol{\mu}) \text{ where } g(\boldsymbol{\mu}) \triangleq \inf_{\varepsilon, \mathbf{R}_{SR} \succeq 0} \left(\underbrace{-\varepsilon + \boldsymbol{\mu}^T \mathbf{f}(\varepsilon, \mathbf{R}_{SR})}_{\triangleq L(\mathbf{v}, \boldsymbol{\mu})} \right) \tag{13}$$

$L(\mathbf{v},\boldsymbol{\mu})$ denotes the Lagrangian and $\mathbf{f}$ is the column vector formed by stacking the 4 inequality constraints f_i. Note that from the weak Slater's conditions (5.27) in [10], strong duality (i.e. zero duality gap) holds in (11) and therefore solving (13) or (11) yields the same value for the cut-set bound. The computation of the dual function $g(\boldsymbol{\mu})$ at a given point $\boldsymbol{\mu}_0$ can be performed by the Gradient Projection Method (GPM, see section 2.3 in [11]). The gradient of the Lagrangian at a point $\mathbf{v}_0$ reads:

$$\nabla L|_{\mathbf{v}_0} \triangleq \left(\frac{\partial L}{\partial \varepsilon}\bigg|_{\mathbf{v}_0} \quad \text{vec}\left(2\frac{\partial L}{\partial \mathbf{R}_{SR}^*}\bigg|_{\mathbf{v}_0} \right)^T \right)^T \quad (14)$$

An analytical expression of the gradient can be obtained from the following formulas (see [12] and [13]):

$$\frac{\partial \log|\mathbf{A}+\mathbf{BXC}|}{\partial \mathbf{X}^*} = \left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^H \;;\; \frac{\partial \text{tr}(\mathbf{AXA}^H)}{\partial \mathbf{X}^*} = \mathbf{A}^H\mathbf{A} \quad (15)$$

The GPM is an iterative algorithm which computes at step k the following points:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \alpha^{(k)}\left(\overline{\mathbf{v}}^{(k)} - \mathbf{v}^{(k)}\right) \text{ and } \overline{\mathbf{v}}^{(k)} = \left[\mathbf{v}^{(k)} - s^{(k)}\nabla L|_{\mathbf{v}^{(k)}}\right]^+ \quad (16)$$

where $s^{(k)}$ is a positive scalar, $\alpha^{(k)} \in (0,1]$ is the step size and $[.]^+$ denotes the projection on the feasible set. Here, when minimizing the Lagrangian, a projection on the PSD cone is required. This projection is applied by forming a Hermitian (but not necessarily PSD) matrix denoted by $\mathbf{M}$ from the $(N_S+N_R)^2$ last coordinates of the candidate point $\mathbf{v}^{(k)} - s^{(k)}\nabla L|_{\mathbf{v}^{(k)}}$ then, considering its eigenvalue decomposition $\mathbf{M} = \mathbf{U}\text{diag}(\lambda_i)\mathbf{U}^H$, the projected matrix is ([13]):

$$\mathbf{M}^+ = \mathbf{U}\text{diag}(\max(\lambda_i,0))\mathbf{U}^H \quad (17)$$

There exist various methods to select parameters $s^{(k)}$ and $\alpha^{(k)}$ in order to ensure convergence. In our simulations we used the Armijo Rule along the feasible direction (sec. 2.3 in [11]).

In order to solve the dual problem, it remains to maximize the dual function with respect to $\boldsymbol{\mu}$. From (13), the dual function may not be differentiable. However, as shown below, an analytical expression of a subgradient can be found (see also sec. 6.3 of [11]). Since the dual function is concave in $\boldsymbol{\mu}$, a vector $\mathbf{h}$ is a subgradient of $\mathbf{g}$ at $\boldsymbol{\mu}_0$ if for all $\boldsymbol{\mu}_1$:

$$g(\boldsymbol{\mu}_1) \leq g(\boldsymbol{\mu}_0) + \mathbf{h}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \quad (18)$$

Let $\tilde{\mathbf{v}}_0$ and $\tilde{\mathbf{v}}_1$ be minimizers of the Lagrangian at respectively $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. Then

$$g(\boldsymbol{\mu}_1) \triangleq L(\tilde{\mathbf{v}}_1,\boldsymbol{\mu}_1) \leq L(\tilde{\mathbf{v}}_0,\boldsymbol{\mu}_1)$$
$$\Rightarrow g(\boldsymbol{\mu}_1) - g(\boldsymbol{\mu}_0) \leq L(\tilde{\mathbf{v}}_0,\boldsymbol{\mu}_1) - L(\tilde{\mathbf{v}}_0,\boldsymbol{\mu}_0) = \mathbf{f}(\tilde{\mathbf{v}}_0)^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) \quad (19)$$

where the constraints $\mathbf{f}$ were rewritten as a function of $\mathbf{v}$. From (19) and (18) it can be concluded that $\mathbf{f}(\tilde{\mathbf{v}}_0)$ is a subgradient of $\mathbf{g}$ at $\boldsymbol{\mu}_0$, and the dual can be solved efficiently by subgradient methods. We refer the reader to sec. 6.3 of [11] for step size selection strategies.

We have described a procedure that computes the dual using the GPM and maximizes it via a subgradient method. Another alternative is to solve the primal problem using the barrier method [10], which we only briefly describe here for lack of space. In this case, a (logarithmic) barrier of the form $-m\log(-f_i)$ is added to the objective for each inequality constraint before minimization, with m a positive parameter that is decreased at each outer iteration of the method. The constraint $\mathbf{R}_{SR} \succeq 0$ can be taken into account by adding the barrier $-m\log|\mathbf{R}_{SR}|$ (Theorem 5.1 in [14]). As before, an analytical expression of the gradient can be found.

The two previously described optimization procedures can be directly applied to compute C_{FD}, but computing C_{TDD} requires to run a one-dimensional maximization over t on top of the convex optimization. Fortunately, the search over t can be performed by efficient techniques such as the Golden Section Search. Indeed, it can be checked numerically that the following function of t is unimodal (i.e. monotonically increasing until a maximum, then monotonically decreasing):

$$\max_{\mathbf{R}_S^{(1)} \succeq 0, \mathbf{R}_{SR,0}^{(2)} \succeq 0} \min\{C_A,C_B\} \quad (20)$$

In this section, the cut-set upper-bound on capacity was expressed in the full-duplex and TDD MIMO relaying cases, and two numerical evaluation procedures were proposed. The next section will address the maximization of achievable rates.

III. DECODE-AND-FORWARD WITH FULL CSI

A. The Partial DF strategy

In a partial DF scheme [4][6], S transmits a first message ω_0 at rate R_0 using signal $\mathbf{x}_S^{(1)}(\omega_0)$ during the first slot. The relay (R) transmits $\mathbf{x}_R^{(2)}(\omega_0)$ during the second slot while S transmits $\mathbf{x}_S^{(2)}(\omega_0,\omega_1) = \mathbf{x}_{S,0}^{(2)}(\omega_0) + \mathbf{x}_{S,1}^{(2)}(\omega_1)$. This strategy is often called partial DF because R only decodes the part ω_0 of the message (ω_0,ω_1) to cooperate with S during the second slot. Because we assume a synchronized scenario, the signals $\mathbf{x}_R^{(2)}(\omega_0)$ and $\mathbf{x}_{S,0}^{(2)}(\omega_0)$ are correlated, whereas ω_1 is mapped onto an independent signal $\mathbf{x}_{S,1}^{(2)}(\omega_1)$ transmitted at rate R_1 using superposition coding. The destination successively decodes $\hat{\omega}_0$ and $\hat{\omega}_1$. The derivation of the achievable rate is a straightforward extension of the proof of Prop. 2 in [6] and leads to:

$$R_{\text{DF}} = \max_{t,\, p(\mathbf{x}_S^{(1)},\mathbf{x}_{S,0}^{(2)},\mathbf{x}_{S,1}^{(2)},\mathbf{x}_R^{(2)})} R_0 + R_1 = \max_{t,\, \mathbf{R}_S^{(1)} \succeq 0,\, \mathbf{R}_{S,1}^{(2)} \succeq 0,\, \mathbf{R}_{SR,0}^{(2)} \succeq 0} \min(R_A,R_B) \quad (21)$$

$$R_A \triangleq t\log_2\left|\mathbf{I}_{N_R} + \mathbf{H}_1\mathbf{R}_S^{(1)}\mathbf{H}_1^H\right| + (1-t)\log_2\left|\mathbf{I}_{N_D} + \mathbf{H}_0\mathbf{D}_S\mathbf{R}_{S,1}^{(2)}\mathbf{D}_S^H\mathbf{H}_0^H\right| \quad (22)$$

$$R_B \triangleq t\log_2\left|\mathbf{I}_{N_D} + \mathbf{H}_0\mathbf{R}_S^{(1)}\mathbf{H}_0^H\right|$$
$$+ (1-t)\log_2\left|\mathbf{I}_{N_D} + \mathbf{H}_0\mathbf{R}_{S,1}^{(2)}\mathbf{H}_0^H + [\mathbf{H}_0 \; \mathbf{H}_2]\mathbf{R}_{SR,0}^{(2)}[\mathbf{H}_0 \; \mathbf{H}_2]^H\right| \quad (23)$$

where $\mathbf{R}_S^{(1)} \triangleq E\left[\mathbf{x}_S^{(1)}(\mathbf{x}_S^{(1)})^H\right]$, $\mathbf{R}_{S,1}^{(2)} \triangleq E\left[\mathbf{x}_{S,1}^{(2)}(\mathbf{x}_{S,1}^{(2)})^H\right]$ and $\mathbf{R}_{SR,0}^{(2)} \triangleq E\left[\left[(\mathbf{x}_{S,0}^{(2)})^T \; (\mathbf{x}_R^{(2)})^T\right]^T\left[(\mathbf{x}_{S,0}^{(2)})^H \; (\mathbf{x}_R^{(2)})^H\right]\right]$ and the transmit power constraints can be written as:

$$\text{tr}(\mathbf{R}_S^{(1)}) \leq P_S \;;\; \text{tr}(\mathbf{D}_R\mathbf{R}_{SR,0}^{(2)}\mathbf{D}_R^H) \leq P_R \quad (24)$$

$$\text{tr}(\mathbf{R}_{S,1}^{(2)}) + \text{tr}(\mathbf{D}_S\mathbf{R}_{SR,0}^{(2)}\mathbf{D}_S^H) \leq P_S \quad (25)$$

The problem (21) is convex and the maximization of the achievable rate of the partial DF strategy can be carried on using the methods described in the previous section.

B. Sub-optimum strategies

We now consider a sub-optimum strategy, that we call full DF (FDF) which is defined as a variant of partial DF in which the relay decodes the whole message and the source does not

superimpose a new message during the second slot (i.e. $R_1=0$). In this case, the achievable rate simplifies as:

$$R_{\text{FDF}}=\max_{t,\mathbf{R}_S^{(1)}\geq0,\mathbf{R}_{SR}^{(2)}\geq0}\min(tR_A,tR_{B,1}+(1-t)R_{B,2}) \quad (26)$$

$$R_A\triangleq\log_2\left|\mathbf{I}_{N_R}+\mathbf{H}_1\mathbf{R}_S^{(1)}\mathbf{H}_1^H\right| \quad (27)$$

$$R_{B,1}\triangleq\log_2\left|\mathbf{I}_{N_D}+\mathbf{H}_0\mathbf{R}_S^{(1)}\mathbf{H}_0^H\right| \quad (28)$$

$$R_{B,2}\triangleq\log_2\left|\mathbf{I}_{N_D}+[\mathbf{H}_0\ \mathbf{H}_2]\mathbf{R}_{SR}^{(2)}[\mathbf{H}_0\ \mathbf{H}_2]^H\right| \quad (29)$$

s.t. $\text{tr}(\mathbf{R}_S^{(1)})\leq P_S$; $\text{tr}(\mathbf{D}_R\mathbf{R}_{SR}^{(2)}\mathbf{D}_R^H)\leq P_R$; $\text{tr}(\mathbf{D}_S\mathbf{R}_{SR}^{(2)}\mathbf{D}_S^H)\leq P_S$ (30)

For a given t, it can be noticed that R_{FDF} is an increasing function of $R_{B,2}$, which only depends on $\mathbf{R}_{SR}^{(2)}$. Therefore the optimization (26) can be carried on in two steps:

$$\tilde{R}_{B,2}\triangleq\max_{\mathbf{R}_{SR}^{(2)}\geq0}R_{B,2} \quad (31)$$

$$R_{\text{FDF}}=\max_{t,\mathbf{R}_S^{(1)}\geq0}\min(tR_A,tR_{B,1}+(1-t)\tilde{R}_{B,2}) \quad (32)$$

Although R_{FDF} can be solved using the procedures described in section II.B, we decide to evaluate sub-optimum precoders at the Source and Relay with a structure that further reduces the optimization complexity.

1) Sub-optimum Source precoder during 1st slot
Let first consider the optimization of $\mathbf{R}_S^{(1)}$ and introduce the SVD of $\mathbf{H}_0$ and $\mathbf{H}_1$:

$$\mathbf{H}_0=\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{V}_0^H \text{ and } \mathbf{H}_1=\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{V}_1^H \quad (33)$$

with $\mathbf{\Lambda}_0\triangleq\text{diag}(\lambda_{0,i})$ and $\mathbf{\Lambda}_1\triangleq\text{diag}(\lambda_{1,i})$. We arbitrarily impose the following structure to the source covariance matrix:

$$\mathbf{R}_S^{(1)}=\underbrace{\mathbf{V}_0\text{diag}(\mathbf{p}_0)\mathbf{V}_0^H}_{\triangleq\mathbf{R}_0}+\underbrace{\mathbf{V}_1\text{diag}(\mathbf{p}_1)\mathbf{V}_1^H}_{\triangleq\mathbf{R}_1} \quad (34)$$

The structure (34) stems from the intuition that the source shall transmit part of its power on the eigenmodes of the channel to the relay and the rest on the eigenmodes of the channel to the destination (note the similarity with the precoder optimization in [8]). Let L_0 and L_1 be the number of non-zero singular values of $\mathbf{H}_0$ and $\mathbf{H}_1$. Inserting (34) into (27) and (28) gives:

$$R_A=\log_2\left|\mathbf{I}_{N_R}+\mathbf{H}_1\mathbf{R}_0\mathbf{H}_1^H+\mathbf{H}_1\mathbf{R}_1\mathbf{H}_1^H\right|$$
$$\overset{(a)}{\geq}\log_2\left|\mathbf{I}_{N_R}+\mathbf{H}_1\mathbf{R}_1\mathbf{H}_1^H\right|\overset{(b)}{=}\sum_{i=1}^{L_1}\log_2\left(1+\lambda_{1,i}^2p_{1,i}\right)\triangleq J_1 \quad (35)$$
$$R_{B,1}\overset{(a)}{\geq}\log_2\left|\mathbf{I}_{N_D}+\mathbf{H}_0\mathbf{R}_0\mathbf{H}_0^H\right|\overset{(b)}{=}\sum_{i=1}^{L_0}\log_2\left(1+\lambda_{0,i}^2p_{0,i}\right)\triangleq J_0$$

Where (a) comes from the Minkowski determinant inequality and (b) comes from (34). Inserting these lower bounds on R_A and $R_{B,1}$ into (32) yields a lower-bound on R_{FDF} :

$$R_{FDF}\geq\max_{t,\mathbf{p}_0,\mathbf{p}_1}\min(tJ_1,tJ_0+(1-t)\tilde{R}_{B,2})$$
$$\text{s.t. } \mathbf{p}_0\geq0; \mathbf{p}_1\geq0; \mathbf{1}_{L_0}^T\mathbf{p}_0+\mathbf{1}_{L_1}^T\mathbf{p}_1=P_S \quad (36)$$

From the epigraph form of (36), writing the KKT conditions leads to:

$$p_{0,i}=(\alpha-1/\lambda_{0,i}^2)^+; p_{1,i}=(\beta-1/\lambda_{1,i}^2)^+ \quad (37)$$

The solution (37) can be obtained by a water-filling algorithm with two water levels α and β which are not independent due to the total source power constraint in (36). A fraction

$P_1\triangleq(\mathbf{1}^T\mathbf{p}_1)/P_S$ of it will be transmitted on the source-relay channel eigenmodes, while the rest $1-P_1$ will be transmitted on the eigenmodes of the source-destination channel. Finding α and β amounts to finding the optimum $\tilde{P}_1\in[0;1]$. It can be checked that J_1 is non-decreasing with P_1 while J_0 is non-increasing. Therefore, the optimum $\tilde{P}_1$ can be found by equating the left-hand and right-hand sides of the min(.) in (36), provided that $(1-t)\tilde{R}_{B,2}\leq tJ_1$ at $P_1=1$. Otherwise the solution is trivial (i.e. $\tilde{P}_1=0$ or $\tilde{P}_1=1$). In summary, we propose in this section a sub-optimum approach which has low complexity. It consists in transmitting a fraction of the source power on the eigenmodes of the channel to the relay and the rest on the eigenmodes of the channel to the destination, after determining the optimum balance. Note that equality occurs in (35) when the two channels turn out to be orthogonal, and in this case the solution (34) is optimum. Also note that optimality is approached anyway at large N_S because (not shown here for lack of space) assuming $\mathbf{H}_0$ and $\mathbf{H}_1$ have i.i.d. Rayleigh components, the distribution of the ratio $\|\mathbf{H}_1^H\mathbf{H}_0\|_F^2/(\|\mathbf{H}_0\|_F^2\|\mathbf{H}_1\|_F^2)$ concentrates around 0 as N_S grows.

2) Sub-optimum Source and Relay precoder during 2nd slot
Let now consider the optimization of $\mathbf{R}_{SR}^{(2)}$. Introducing the SVD $[\mathbf{H}_2\ \mathbf{H}_0]=\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ and the change of variable $\mathbf{R}\triangleq\mathbf{V}^H\mathbf{R}_{SR}^{(2)}\mathbf{V}$, the problem (31) can be rewritten as:

$$\tilde{R}_{B,2}=\max_{\mathbf{R}\geq0}\log_2\left|\mathbf{I}_{N_D}+\mathbf{\Lambda}\mathbf{R}\mathbf{\Lambda}^H\right|$$
$$\text{s.t. } \text{tr}(\mathbf{D}_S\mathbf{V}\mathbf{R}\mathbf{V}^H\mathbf{D}_S^H)\leq P_S ; \text{tr}(\mathbf{D}_R\mathbf{V}\mathbf{R}\mathbf{V}^H\mathbf{D}_R^H)\leq P_R \quad (38)$$

As in appendix B of [7], we arbitrarily enforce a diagonal structure $\mathbf{R}=\text{diag}(\mathbf{p}_2)$. This turns the matrix optimization problem (38) into the following vector optimization:

$$\max_{\mathbf{p}_2}\sum_{i=1}^{L}\log_2(1+\lambda_i^2p_{2,i})$$
$$\text{s.t. } \mathbf{a}^T\mathbf{p}_2\leq P_S ; \mathbf{b}^T\mathbf{p}_2\leq P_R ; \mathbf{p}_2\geq0 \quad (39)$$

where the components of vectors $\mathbf{a}$ and $\mathbf{b}$ can be computed as $a_i\triangleq\sum_{j=1}^{N_S}|V_{N_R+j,i}|^2$ and $b_i\triangleq\sum_{j=1}^{N_R}|V_{j,i}|^2$. The problem (39) can then be solved at a much lower computational cost than (31). Note that if the individual power constraints in (38) are replaced by a sum-power constraint $\text{tr}(\mathbf{R}_{SR}^{(2)})=\text{tr}(\mathbf{R})\leq P_S+P_R$, then Hadamard determinant inequality can be applied as in [2] to show that the optimum $\mathbf{R}$ is diagonal. In other words, under a total transmit power constraint at the source and relay, transmitting on the eigenmodes of the joint channel $[\mathbf{H}_2\ \mathbf{H}_0]$ is optimum.

IV. APPLICATION TO THE DOWNLINK OF CELLULAR SYSTEMS

In this section, we illustrate the application of the upper and lower bounds on the TDD MIMO relay channel in a cellular downlink scenario. Let assume a Base Station (BS) equipped with $N_S=2$ antennas per sector. The BS is in Line-Of-Sight (LOS) with a dual antenna $N_R=2$ fixed Relay Station (RS),

and both cooperate to transmit to a dual antenna Mobile Station (MS) $N_D = 2$. The MS is in NLOS from both the BS and the RS. The MIMO channels are modeled by i.i.d. complex gaussian components in the NLOS case, and a 10dB Rice factor is introduced in the LOS case. The SNR on the BS-MS link (SNR_0) and on the RS-MS link SNR_2 are assumed equal $SNR_0 = SNR_2 = 0dB$. On Fig. 1, the average achievable rate of the partial DF and full DF (FDF) strategies are plotted as a function of the SNR on the BS-RS link (SNR_1). The 2x2 MIMO channel capacity is also plotted for reference with (solid line) and without (dashed line) CSI at the transmitter side, in order to highlight the waterfilling gain. It can be checked that single-hop transmission outperforms non-cooperative DF relaying even at high SNR_1, but is outperformed by all cooperative DF strategies (even the sub-optimum FDF with vector optimization) when SNR_1 exceeds a threshold around 3 dB. As expected, the partial DF outperforms FDF and both achieve rates quite close to the cut-set bound at high SNR_1 (only 0.2 bit per channel use away at $SNR_1 = 20dB$). Also note that the cooperative DF strategies exploit a larger total transmit power during the 2nd slot. However, it can be verified (not plotted here) that cooperative DF still improves upon both single-hop and non-cooperative DF under the same sum-power constraint.
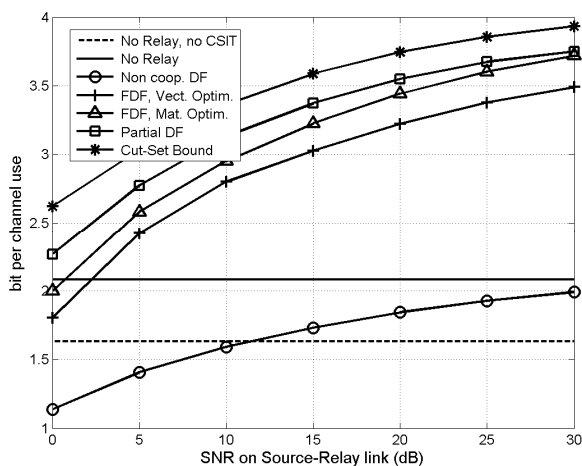

Figure 1: Upper and Lower bounds on TDD MIMO relaying channel capacity with Source and Relay precoder optimization in a 2x2x2 antenna configuration.

State-of-the-art TDD systems often rely on uplink/downlink channel reciprocity to enable CSIT exploitation without having to explictly feed back the channel coefficients. However, in cooperative relaying it is hard if not impossible to avoid explicit signaling because for instance the BS cannot estimate the RS-MS channel by means of reciprocity. Quantized codebooks [15] allow to reduce the amount of feedback. For instance, the MS can estimate $\mathbf{H}_0$ and $\mathbf{H}_2$ and feedback the index of the best codeword to the BS and RS. Some preliminary simulation results (not detailed here for lack of space) obtained with Grassmannian beamforming codebooks in the $N_S = 2; N_R = 2; N_D = 1$ case indicate that only 4 bits of feedback (i.e. a codebook of size 16) are required to

approach the average FDF achievable rate by less than 0.5 bit per channel use. The design of quantized codebooks for cooperative relaying thus seems a promising topic.

V. Conclusion

The cut-set bound on the MIMO relay channel capacity can be efficiently computed in the full-duplex and TDD cases by convex optimization techniques, assuming full knowledge of all channels. Moreover, the same techniques can be applied to derive the achievable rate of partial DF with source and relay precoders optimization. Sub-optimum strategies and precoders with lower complexity can also be considered at the expense of lower rate. Finally, the exploitation of partial and statistical channel knowledge and the application of quantized codebooks seem promising topics in order to approach in a real system the achievable rate gains derived in this paper.

References

[1] Nosratinia, T. E. Hunter, A. Hedayat, "Cooperative Communication in Wireless Networks", *IEEE Communications Magazine*, Oct 2004

[2] E. Telatar, "Capacity of multi-antenna gaussian channels" *European Trans. Telecommun.*, vol. 10, pp. 585--595, 1999

[3] T. M. Cover, A.A. El Gamal, "Capacity theorems for the relay channel", *IEEE Trans. on Information Theory*, vol. IT-25, No 5, Sep 1979

[4] G. Kramer, M. Gastpar, P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks", *IEEE Trans. on Information Theory*, Vol 51. No 9. Sep 2005

[5] O. Muñoz, J. Vidal, A. Agustin, "Linear Transceiver Design in Non-Regenerative Relays with Channel State Information", in *IEEE trans. on Signal Processing*, Vol. 55, No. 6, Jun. 2007

[6] A. Høst-Madsen, J. Zhang "Capacity Bounds and Power Allocation for Wireless Relay channels" *IEEE Transactions on Information Theory*, Vol 51, N°6, pp. 2020-2040, June 2005

[7] B. Wang, J. Zhang, A. Høst-Madsen "On the Capacity of MIMO Relay Channels" *IEEE Trans. Information Theory*, Vol. 51, No 1, Jan 2005.

[8] I. Hammerstrom, M. Kuhn, C. Esli, Z. Jihan, A. Wittneben, G. Bauch "MIMO two-way relaying with transmit CSI at the relay" *Proc. IEEE SPAWC'07*, June 2007.

[9] C.K. Lo, S. Vishwanath, R.W. Heath Jr. "Rate Bounds for MIMO Relay Channels Using Precoding" *Proc. IEEE Globecom'05*. Nov 2005.

[10] S. Boyd and L. Vandenberghe "Convex Optimization" Cambridge University Press, 2004. ISBN 0521833787. Available at: http://www.stanford.edu/~boyd/cvxbook/

[11] D. P. Bertsekas "Nonlinear Programming". Athena Scientific. ISBN 1-886529-00-0. 2nd edition, 1999.

[12] J. Liu, Y. T. Hou, H. D. Sherali "Conjugate Gradient Projection Approach for Multi-Antenna Gaussian Broadcast Channels" *Proc. IEEE International Symposium. on Information Theory. (ISIT'07)*. Jun 2007.

[13] S. Ye and R. Blum "Optimized Signaling for MIMO Interference Systems with Feedback" IEEE Trans. Signal Processing. Vol 51, No 11, Nov 2003.

[14] M.J. Todd "Semidefinite Optimization" in Acta Numerica 10, pp 515-560. Cambridge Univ. Press. 2001.

[15] D.J. Love, R.W. Heath Jr., T. Strohmer "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems" *IEEE Trans. on Information Theory*. Vol 49, No 10, Oct 2003.