Linear Precoding for Mutual Information Maximization in MIMO Systems

Meritxell Lamarca[1,2]

[1]*Signal Theory and Communications Department, Technical University of Catalonia, Barcelona (Spain)*
[2]*ECE Department, University of Delaware, Newark DE (USA)*
xell@gps.tsc.upc.edu

Abstract—**We study the design of linear precoders for maximization of the mutual information in MIMO systems with arbitrary constellations and with perfect channel state information at the transmitter. We derive the structure of the optimum precoder and we show that the mutual information is concave in a quadratic function of the precoder coefficients. An iterative algorithm is also proposed to find this optimum value.**

I. INTRODUCTION

The ultimate goal of employing a linear precoder in a MIMO system is to maximize the transmission rate that this system can achieve. Given a certain constellation to be employed, and assuming all symbols are equally likely, this rate is determined by the mutual information between the precoder input and the channel output. This is the parameter that defines the performance of capacity-approaching systems employed in modern digital communications and, therefore, it should be the driving criterion for precoder design.

Unfortunately, the task of mutual information maximization for arbitrary constellations encountered some difficulties, so in the literature some other alternative parameters were employed for optimization: the mutual information assuming Gaussian signals were transmitted [15], the linear minimum mean square error (MMSE) [1], the uncoded BER [6], etc. The only information-theoretic criterion proposed for quite some time was the channel cut-off rate [13], [14]. This situation changed when the relationship between MMSE and the derivative of mutual information in Gaussian channels was introduced in [3]. This relationship was exploited to propose an algorithm to allocate power to independent parallel subchannels by means of mercury waterfilling [5] and was subsequently exploited to provide the first analysis on MIMO precoder design for mutual information optimization [11], [12]. The authors in [12] employed the KKT conditions satisfied by all critical points of the optimization criterion to derive a gradient algorithm to find the precoder coefficients, this algorithm was quite slow and did not guarantee convergence. Recently, [10] showed that generally the mutual information is not a concave function of the precoder coefficients, although it is concave function of the power allocated to each subchannel when the MIMO channel matrix is diagonal.

In this work we derive the structure of the optimum precoder and we show that, even though the mutual information is not concave in the precoder coefficients, it is concave in a quadratic function of precoder coefficients, which could be very useful to cast the problem of mutual information maximization into a convex optimization framework. While the optimal structure applies to complex modulations, concavity is proven for real signals and channels, although it is expected to hold also in the complex case (this is a topic of current work). Additionally, we also present an iterative algorithm to find the optimum solution.

The following notation is employed in the paper: $(\mathbf{A})_{ij}$ stands for the coefficient in the i-th row, j-th column of matrix $\mathbf{A}$, $diag(\mathbf{A})$ is a column vector containing the diagonal elements of matrix $\mathbf{A}$, $Diag(\mathbf{a})$ is a diagonal matrix containing in the main diagonal the coefficients of vector $\mathbf{a}$, the superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian operations, $\odot$ and $tr\{\cdot\}$ correspond to the Schur product and the trace, $Re\{\cdot\}$ gets the real part, $vec(\mathbf{A})$ and $eig_i(\mathbf{A})$ represent the vector obtained stacking the columns of $\mathbf{A}$ and the i-th eigenvalue of $\mathbf{A}$, $\mathbf{e}_i$ and $\mathbf{1}$ stand for the column vector with all-zero entries except for $(\mathbf{e}_i)_{i1} = 1$ and for the vector with all entries equal to 1, $\mathbf{D_T}\boldsymbol{\Psi}$ and $\mathbf{H_T}\boldsymbol{\Psi}$ denote the Jacobian and the Hessian of the function $\boldsymbol{\Psi}$ with respect to parameter $\mathbf{T}$ [10], and $\mathbf{K}$ and $\mathbf{N}$ stand for the commutation and symmetrization matrices defined in [10, Appendix A].

II. PROBLEM STATEMENT

We consider the problem of transmission through a MIMO system with N_T transmit and N_R receive antennas[1]. The signal model is

$$\mathbf{y} = \mathbf{HPs} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ is the MIMO channel (a matrix of size $N_R \times N_T$ with complex entries), $\mathbf{s}$ stands for the transmitted symbols, which are assumed to belong to an arbitrary constellation of size C and have $E\{\mathbf{ss}^H\} = \mathbf{I}$, and $\mathbf{n}$ is the complex additive Gaussian noise component with zero mean and variance $\sigma_n^2\mathbf{I}$. The square matrix $\mathbf{P}$ stands for the linear precoder that must be optimized. Throughout the paper the following notation is used to denote the singular value decomposition (SVD) of the channel and precoder matrices:

$$\mathbf{H} = \mathbf{U_H}\boldsymbol{\Sigma_H}\mathbf{V_H}^H, \quad \boldsymbol{\Sigma_H} = Diag(\sigma_{H,1}, \ldots, \sigma_{H,N_T})$$
$$\mathbf{P} = \mathbf{U_P}\boldsymbol{\Sigma_P}\mathbf{V_P}^H, \quad \boldsymbol{\Sigma_P} = Diag(\sigma_{P,1}, \ldots, \sigma_{P,N_T}).$$

[1]In order to simplify notation, in the paper we focus in the case $N_R \geq N_T$, though analogous conclusions are obtained when $N_R < N_T$.

In this paper we assume that the transmitter has perfect knowledge of the channel status and we study the design of the linear precoder that maximizes the mutual information $I(\mathbf{s}; \mathbf{y})$ subject to an average power constraint:

$$P^\circ = \arg\max_{\mathbf{P}} I(\mathbf{s}; \mathbf{y}) \text{ subject to } tr\left\{\mathbf{P}^H\mathbf{P}\right\} \leq 1 \quad (2)$$

which according to [10] is a non-convex problem in $\mathbf{P}$.

III. OPTIMUM PRECODER STRUCTURE

A. Precoder structure

Denote the MMSE matrix as $\mathbf{E} = E\left\{(\mathbf{s}-\widehat{\mathbf{s}})(\mathbf{s}-\widehat{\mathbf{s}})^H\right\}$, $\widehat{\mathbf{s}} = E\{\mathbf{s}|\mathbf{y}\}$ and its eigendecomposition as $\mathbf{E} = \mathbf{Q_E}\mathbf{\Lambda_E}\mathbf{Q_E}^H$, $\mathbf{\Lambda_E} = diag(\lambda_{E1}, \ldots, \lambda_{E N_T})$. The following theorem gives the structure of the optimum precoder.

Theorem 1: The linear precoder $\mathbf{P}^\circ$ that maximizes the mutual information with an average power constraint can be written as

$$\mathbf{P}^\circ = \mathbf{V_H}\mathbf{\Pi}\,\overline{\mathbf{D}}\mathbf{Q_E}^H \quad (3)$$

or equivalently as

$$\mathbf{P}^\circ = \mathbf{V_H}\mathbf{D}\mathbf{\Pi}\mathbf{Q_E}^H \quad (4)$$

where $\mathbf{V_H}$ and $\mathbf{Q}_E$ are the right singular vectors of the channel matrix and the eigenvectors of the MMSE matrix respectively, $\mathbf{\Pi}$ is a permutation matrix, $\mathbf{D}$ is a diagonal matrix and $\overline{\mathbf{D}}$ the same matrix as $\mathbf{D}$ except for a permutation of the diagonal elements.

Therefore, for the optimum precoder[2] $\mathbf{U_P} = \mathbf{V_H}$, $\mathbf{\Sigma_P} = \mathbf{D}$, $\mathbf{V_P} = \mathbf{Q_E}\mathbf{\Pi}^T$, and to find the optimum solution we only need to find the values of $\mathbf{\Sigma_P}$ and $\mathbf{V_P}$.

Proof: Define $\mathbf{B} = \mathbf{V_H}^H\mathbf{P}^\circ\mathbf{Q_E}$. According to equation (4) in [11] the optimum solution (and all other critical points) of this optimization problem verify for a certain value of the constraint ν

$$\mathbf{P}^\circ = \nu^{-1}\mathbf{H}^H\mathbf{H}\mathbf{P}^\circ\mathbf{E} \quad (5)$$

so replacing $\mathbf{H}$ by its SVD and $\mathbf{B}$ by its equation we get

$$
\begin{aligned}
\mathbf{P}^\circ &= \nu^{-1}\mathbf{V_H}\mathbf{\Sigma_H}^T\mathbf{\Sigma_H}\mathbf{V_H}^H\mathbf{P}^\circ\mathbf{Q_E}\mathbf{\Lambda_E}\mathbf{Q_E}^H \\
\mathbf{V_H}^H\mathbf{P}^\circ\mathbf{Q_E} &= \nu^{-1}\mathbf{\Sigma_H}^T\mathbf{\Sigma_H}\mathbf{V_H}^H\mathbf{P}^\circ\mathbf{Q_E}\mathbf{\Lambda_E} \\
\mathbf{B} &= \nu^{-1}\mathbf{\Sigma_H}^T\mathbf{\Sigma_H}\mathbf{B}\mathbf{\Lambda_E} \\
&= \nu^{-1}\left(diag\left(\mathbf{\Sigma_H}^T\mathbf{\Sigma_H}\right)diag^T(\mathbf{\Lambda_E})\right)\odot\mathbf{B}
\end{aligned}
$$

The last equation can only be satisfied for the i-th row, j-th column if either $(\mathbf{B})_{ij} = 0$ or

$$1 = \nu^{-1}\sigma_{Hi}^2\lambda_{Ej} \quad (6)$$

If all channel singular values are different and all MMSE matrix eigenvalues are different too the last equation can only be true for at most as many (i, j) pairs as the matrix size N_T, each one corresponding to a different row and a different column. That is, all coefficients in $\mathbf{B}$ must be zero except for at most one coefficient per column or per row. Therefore,

[2]The fact that $\mathbf{U_P} = \mathbf{V_H}$ is mentioned in [10], where it is said that it can be proven following the same procedure as [7, Appendix A].

$\mathbf{B}$ can be written as $\mathbf{\Pi}\overline{\mathbf{D}}$ or $\mathbf{D}\mathbf{\Pi}$, being $\mathbf{D}$ and $\overline{\mathbf{D}}$ diagonal matrices and being $\mathbf{\Pi}$ a permutation matrix. Therefore, as $\mathbf{P}^\circ = \mathbf{V_H}\mathbf{B}\mathbf{Q_E}^H$ the results in (3) and (4) are found. ∎

Note that the structure in equations (3-4) is satisfied by the precoder values corresponding to all critical points of the optimization problem defined in (2).

B. Interpretation

The precoder structure in equation (4) allows for a simple interpretation of all its components.

First, notice that $\mathbf{U_P} = \mathbf{V_H}$ implies that we are transmitting through the channel singular vectors, a procedure that has been shown to be optimum for many other problems (e.g. [7]). The precoder dependency on the signal to noise ratio (SNR) and the constellation is included in $\overline{\mathbf{D}}$, $\mathbf{\Pi}$ and $\mathbf{Q_E}$. In the case of the permutation matrix $\mathbf{\Pi}$, equation (6) indicates that it is selected to match the strongest channel singular vectors (i.e. highest per-subchannel SNR) with the most reliable components of the rotated signal $\mathbf{Q_E}^H\mathbf{s}$ (where smallest MSE is committed in the estimation). The diagonal matrix $\overline{\mathbf{D}}$ does power allocation over the channel modes, a procedure that has also been shown to be optimal for many other problems.

Regarding matrix $\mathbf{Q_E}$, we can view it as the rotation matrix that transforms the constellation employed in $\mathbf{s}$ into another one whose MMSE matrix is diagonal (estimation errors are uncorrelated, as they would be if independent signals were transmitted through parallel subchannels). Then, after transforming the constellation symbols we transmit them through a diagonal channel. To show this fact we consider a rotated version of the received signal and the Gaussian noise $\mathbf{r} = \mathbf{U_H}^H\mathbf{y}$ and $\mathbf{w} = \mathbf{U_H}^H\mathbf{n}$, so we get the model:

$$\mathbf{r} = \mathbf{\Sigma_H}\overline{\mathbf{D}}\mathbf{\Pi}\mathbf{Q_E}^H\mathbf{s} + \mathbf{w} \quad (7)$$

we also define the rotated constellation symbol as $\mathbf{t} = \mathbf{\Pi}\mathbf{Q_E}^H\mathbf{s}$. Hence, substituting in (7) we get

$$\mathbf{r} = \mathbf{\Sigma_H}\overline{\mathbf{D}}\mathbf{t} + \mathbf{w}$$

which is a diagonal channel. Besides, the MMSE matrix for the estimation of $\mathbf{t}$ in this model is also diagonal: denoting the diagonal channel as $\mathbf{\Gamma} = \mathbf{\Sigma_H}\overline{\mathbf{D}}$ and $\Delta\overline{\mathbf{t}} = \mathbf{t} - \overline{\mathbf{t}}$, $\Delta\tilde{\mathbf{t}} = \mathbf{t} - \tilde{\mathbf{t}}$, $\Delta\overline{\mathbf{s}} = \mathbf{s} - \overline{\mathbf{s}}$, $\Delta\tilde{\mathbf{s}} = \mathbf{s} - \tilde{\mathbf{s}}$ we get

$$
\begin{aligned}
&E_{\mathbf{t},\mathbf{w}}\left\{(\mathbf{t} - E\{\mathbf{t}|\mathbf{r}\})(\mathbf{t} - E\{\mathbf{t}|\mathbf{r}\})^H\right\} \\
&= \sum_{\mathbf{t},\,\overline{\mathbf{t}},\,\tilde{\mathbf{t}}}\Delta\overline{\mathbf{t}}\Delta\tilde{\mathbf{t}}^H E_\mathbf{w}\left\{p(\overline{\mathbf{t}}|\mathbf{r} = \mathbf{\Gamma}\mathbf{t} + \mathbf{w})\,p(\tilde{\mathbf{t}}|\mathbf{r} = \mathbf{\Gamma}\mathbf{t} + \mathbf{w})\right\} \\
&= \sum_{\mathbf{s},\,\overline{\mathbf{s}},\,\tilde{\mathbf{s}}}\mathbf{\Pi}\mathbf{Q_E}^H\Delta\overline{\mathbf{s}}\Delta\tilde{\mathbf{s}}^H\mathbf{Q_E}\mathbf{\Pi}^T\cdot \\
&\quad E_\mathbf{w}\left\{\exp-\frac{1}{\sigma_n^2}\left(\left\|\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Q_E}^H\Delta\overline{\mathbf{s}} + \mathbf{w}\right\|^2 + \left\|\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Q_E}^H\Delta\tilde{\mathbf{s}} + \mathbf{w}\right\|^2\right)\right\} \\
&= \mathbf{\Pi}\mathbf{Q_E}^H\left[\sum_{\mathbf{s},\,\overline{\mathbf{s}},\,\tilde{\mathbf{s}}}\Delta\overline{\mathbf{s}}\Delta\tilde{\mathbf{s}}^H\cdot E_\mathbf{n}\left\{\exp-\frac{1}{\sigma_n^2}\left(\left\|\mathbf{U_H}\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Q_E}^H\Delta\overline{\mathbf{s}} + \mathbf{n}\right\|^2\right.\right.\right. \\
&\quad \left.\left.\left. + \left\|\mathbf{U_H}\mathbf{\Gamma}\mathbf{\Pi}\mathbf{Q_E}^H\Delta\tilde{\mathbf{s}} + \mathbf{n}\right\|^2\right)\right\}\right]\mathbf{Q_E}\mathbf{\Pi}^T \\
&= \mathbf{\Pi}\mathbf{Q_E}^H\mathbf{E}\mathbf{Q_E}\mathbf{\Pi}^\mathbf{T} = \mathbf{\Pi}\mathbf{\Lambda_E}\mathbf{\Pi}^\mathbf{T}
\end{aligned}
$$

where the summations run over all constellation symbols. Thus, we obtain that in the transformed domain the MMSE

matrix is a diagonal matrix with the eigenvalues of $\mathbf{E}$ in permuted order.

Finally, it is also worth mentioning that we can also provide an interpretation for the scaling constant ν in (5). Going back to equation (5), right-multiplying it by $\mathbf{P}^{\circ\mathbf{H}}$ and taking the trace we get

$$
\begin{aligned}
tr\left\{\mathbf{P}^\circ\mathbf{P}^{\circ\mathbf{H}}\right\} &= \nu^{-1}tr\left\{\mathbf{H}^H\mathbf{HP}^\circ\mathbf{EP}^{\circ\mathbf{H}}\right\} \\
&= \nu^{-1}tr\left\{\mathbf{HP}^\circ\mathbf{EP}^{\circ\mathbf{H}}\mathbf{H}^H\right\} \\
&= \frac{\nu^{-1}}{SNR}E\left\{\left\|\sqrt{SNR}\mathbf{HP}^\circ\left(\mathbf{s}-\widehat{\mathbf{s}}\right)\right\|^2\right\} \\
&= \frac{\nu^{-1}}{SNR}\frac{d}{d\,SNR}I\left(\mathbf{s};\mathbf{y}\right)=1 \\
\nu &= \frac{1}{SNR}\frac{d}{d\,SNR}I\left(\mathbf{s};\mathbf{y}\right)
\end{aligned}
$$

IV. CONCAVITY OF MUTUAL INFORMATION

In this section we prove that the mutual information is concave in the quadratic form $\mathbf{P^TH^THP}$ for the case where all the signals in equation (2) are all real, so σ_n^2 now stands for the real noise variance per component. This fact could be very useful to cast the problem of mutual information maximization into the convex optimization framework. While the generalization of Theorem 2 to complex signals is straightforward, the generalization of Theorem 3 is more involved and is object of current work, so we will formulate the whole section for real signals.

Theorem 2: The mutual information depends on the precoder $\mathbf{P}$ through the quadratic form $\mathbf{P^TH^THP}$.

Proof: Define the symbol log-likelihood ratios (LLR) for detection of the transmitted symbol and the vector that groups all LLRs for the same symbol $\mathbf{s}$ as follows:

$$
\begin{aligned}
LLR\left(\mathbf{s}_k\right) &= \log\frac{p\left(\mathbf{y}|\mathbf{s}=\mathbf{s}_k\right)}{p\left(\mathbf{y}|\mathbf{s}=\mathbf{s}_1\right)} \quad k=1...C \\
\mathbf{L} &= \left[LLR\left(\mathbf{s}_1\right)\,...\,LLR\left(\mathbf{s}_C\right)\right]^T
\end{aligned}
$$

being $\mathbf{s}_1$ an arbitrary constellation symbol taken as reference and C the constellation size. As the LLRs are a sufficient statistic for data detection

$$I\left(\mathbf{s};\mathbf{y}\right)=I\left(\mathbf{s};\mathbf{L}\right)$$

Assume that we transmit symbol $\mathbf{s}=\mathbf{s}_i$. Then the LLR's are:

$$
\begin{aligned}
LLR\left(\mathbf{s}_k\right) &= \log\frac{p\left(\mathbf{y}|\mathbf{s}=\mathbf{s}_k\right)}{p\left(\mathbf{y}|\mathbf{s}=\mathbf{s}_1\right)} \\
&= -\frac{1}{2\sigma_n^2}\left\|\mathbf{n}+\mathbf{HP}\left(\mathbf{s}_i-\mathbf{s}_k\right)\right\|^2+\frac{1}{2\sigma_n^2}\left\|\mathbf{n}+\mathbf{HP}\left(\mathbf{s}_i-\mathbf{s}_1\right)\right\|^2
\end{aligned}
$$

Therefore, it can be seen that when symbol $\mathbf{s}=\mathbf{s}_i$ is transmitted $\mathbf{L}$ is a Gaussian multivariate variable $N\left(\mathbf{m}_{Li},\mathbf{C}_L\right)$, whose mean and covariance are given by

$$
\begin{aligned}
\mathbf{m}_{Li} &= \frac{1}{2\sigma_n^2}\left(-diag\left(\mathbf{\Delta}_i^T\mathbf{P}^T\mathbf{H}^T\mathbf{HP}\mathbf{\Delta}_i\right)+\mathbf{1}\left\|\mathbf{HP}\left(\mathbf{s}_i-\mathbf{s}_1\right)\right\|^2\right) \\
\mathbf{C}_L &= \frac{1}{\sigma_n^2}\mathbf{\Delta}_1^T\mathbf{P}^T\mathbf{H}^T\mathbf{HP}\mathbf{\Delta}_1 \\
\mathbf{\Delta}_j &= \left[\begin{array}{ccc} \mathbf{s}_1-\mathbf{s}_j & ... & \mathbf{s}_C-\mathbf{s}_j \end{array}\right]
\end{aligned}
$$

(8)

being $\mathbf{\Delta}_j$ the matrix of size $N_T\times C$ that lists the symbol differences $\mathbf{s}_k-\mathbf{s}_j$. Therefore, as the probability density function

of the LLRs only depends on the precoder through $\mathbf{P^TH^THP}$ the mutual information has also the same dependency. ∎

Several comments can be made on Theorem 2:

Remark 1: The quadratic form $\mathbf{P^TH^THP}$ has been found to play a key role in many other precoder design criteria, since the cost function to be optimized also depends on the precoder through $\mathbf{P^TH^THP}$, e.g. optimization of the mutual information with Gaussian symbols, of the MMSE and bit error rate with linear receivers, and of the channel cut-off rate [9],[14].

Remark 2: For the optimum precoder $\mathbf{P^{\circ T}H^THP^\circ}=\mathbf{V_P\Sigma_P^T\Sigma_H^T\Sigma_H\Sigma_PV_P^T}$ so this quadratic form is a symmetric matrix that has eigenvalues $\sigma_{Hi}^2\cdot\sigma_{Pi}^2$ and its eigenvectors are the right singular values of $\mathbf{P}^\circ$. Thus, as we know that $\mathbf{U_P}=\mathbf{V_H}$, if we can find the eigenvalues and eigenvectors of $\mathbf{P^{\circ T}H^THP^\circ}$ then we can fully identify the precoder. This is the approach taken in section V.

Remark 3: The structure of $\mathbf{P}^\circ$ as $\mathbf{P}=\mathbf{U_H\Sigma_PV_P^H}$ can be obtained through the combination of Theorem 2 with the fact that the mutual information is a monotonically increasing function of the SNR and with the lemma in [9, Pag.436], which states that for a given matrix value $\mathbf{P}^T\mathbf{H}^T\mathbf{HP}$ the Frobenius norm of $\mathbf{P}$ is minimized by selecting $\mathbf{P}=\mathbf{U_H\Sigma_PV_P^H}$.

Remark 4: It is also worth mentioning that equations (8) can also be employed to draw a relevant result that to the best of the authors knowledge remained unknown, though it does not have a direct application for the problem under analysis: The mutual information achieved by a constellation is fully determined by the set of distances between triplets of constellation symbols $\left(\mathbf{s}_i,\mathbf{s}_j,\mathbf{s}_k\right)$ (i.e.$\left(\mathbf{s}_i-\mathbf{s}_j\right)^T\mathbf{P}^T\mathbf{H}^T\mathbf{HP}\left(\mathbf{s}_i-\mathbf{s}_k\right)$). While only pairwise distances $\left\|\mathbf{HP}\left(\mathbf{s}_i-\mathbf{s}_j\right)\right\|^2$ are required to compute the channel cut-off rate [14] and it is widely known that they are also useful to find bounds on the detection error probability (e.g. the union bound), equations in (8) indicate that the set of all triplets distances provides a complete characterization of constellation performance in terms of mutual information. That is, all constellations with the same set of triplets distances will yield identical mutual information in the MIMO Gaussian channel.

Theorem 3: The mutual information is a concave function of $\mathbf{P^TH^THP}$.

Proof: In the proof we employ some of the Jacobian properties indicated in [10], as well as some of the Jacobian expressions there. According to Theorem 2 the mutual information depends on the precoder $\mathbf{P}$ through the quadratic form $\mathbf{P^TH^THP}$. Therefore, the chain rule can be applied to find the Hessian of the mutual information with respect to $\mathbf{P^TH^THP}$. First the chain rule for the Jacobian yields

$$\mathbf{D_P}I\left(\mathbf{s};\mathbf{y}\right)=\mathbf{D_{P^TH^THP}}I\left(\mathbf{s};\mathbf{y}\right)\mathbf{D_p}\mathbf{P^TH^THP} \quad (9)$$

From the results in [10] it follows that $\mathbf{D_P}I\left(\mathbf{s};\mathbf{y}\right)=\frac{1}{\sigma_n^2}vec^T\left(\mathbf{H^THPE}\right)$ and using lemma B.7 in the same paper it can be proven that $\mathbf{D_p}\mathbf{P^TH^THP}=2\left(\mathbf{P^TH^TH}\otimes\mathbf{I}\right)\mathbf{K}$.

Isolating the first term in the right hand side of (9) we get that

$$\mathbf{D_{P^T H^T HP}} I\left(\mathbf{s};\mathbf{y}\right) = \frac{1}{2\sigma_n^2} vec^T\left(\mathbf{E}\right)$$

Furthermore, as $\mathbf{H_T}\Psi = \mathbf{D_T}\left[\mathbf{D_T}^T\Psi\right]$ for scalar function Ψ, we can apply the chain rule once more to obtain

$$\mathbf{D_P} vec\left(\mathbf{E}\right) = \mathbf{D_{P^T H^T HP}} vec\left(\mathbf{E}\right)\mathbf{D_p P^T H^T HP}$$

From [10] we know that

$$\mathbf{D_P} vec\left(\mathbf{E}\right) = -\frac{1}{\sigma_n^2}\mathbf{N}E\left\{\mathbf{\Phi}\left(y\right)\otimes\mathbf{\Phi}\left(y\right)\right\}\left(\mathbf{I}\otimes\mathbf{P^T H^T H}\right)$$

where $\mathbf{\Phi}\left(y\right)$ stands for the MMSE matrix conditioned on a specific realization of the output $\mathbf{y}$:

$$\mathbf{\Phi}\left(\mathbf{y}\right) = E\left\{\left|\left(\mathbf{s}-\widehat{\mathbf{s}}\right)\left(\mathbf{s}-\widehat{\mathbf{s}}\right)^T\right|\mathbf{y}\right\},\quad \widehat{\mathbf{s}} = E\left\{\mathbf{s}\,|\,\mathbf{y}\right\}$$

Combining the previous equations and using properties of matrix $\mathbf{N}$ [10] we obtain that the Hessian of the mutual information with respect to $\mathbf{P^T H^T HP}$ is

$$\mathbf{H_{P^T H^T HP}} I\left(\mathbf{s};\mathbf{y}\right) = \frac{-1}{2\sigma_n^4}\mathbf{N}E\left\{\mathbf{\Phi}\left(\mathbf{y}\right)\otimes\mathbf{\Phi}\left(\mathbf{y}\right)\right\}\mathbf{N^T}$$

Thus, as matrix $E\left\{\mathbf{\Phi}\left(\mathbf{y}\right)\otimes\mathbf{\Phi}\left(\mathbf{y}\right)\right\}$ is positive semi-definite, the Hessian is negative semi-definite and, therefore, the theorem follows. ∎

V. GRADIENT ALGORITHM

In this section we propose an iterative algorithm to find the precoder coefficients. As we know that the optimum precoder follows the structure in (4), we need to identify the optimum singular values and right singular vectors of $\mathbf{P}$, or equivalently, the optimum eigenvalues and eigenvectors of $\mathbf{P^H H^H HP}$. We propose a two stage approach where we update alternatively two parameters. First we update the estimate of the power allocated per subchannel taking into account the total power constraint and freezing the singular values of the precoder; afterwards we optimize the eigenvectors of $\mathbf{P^H H^H HP}$ by searching for the value of $\mathbf{P^H H^H HP}$ that optimizes the mutual information with the constraint that its eigenvalues are known.

A. Gradient updating of the power allocated per subchannel

In the first stage of the iterative procedure we optimize the problem

$$\max_{\left\{\sigma_{Pi}^2\right\}} I\left(\mathbf{s};\mathbf{y}\right) \quad s.t. \sum_i \sigma_{Pi}^2 = 1,\ \sigma_{Pi}^2 \geq 0 \quad (10)$$

assuming $\mathbf{V_P}$ is fixed. According to the results in [10], this is a convex problem if the precoder has the optimal structure described in section III and the remaining precoder parameters are fixed. We will try to simplify notation defining $\mathbf{s} = diag\left(\mathbf{\Sigma_P}\right)$, $\mathbf{p} = diag\left(\mathbf{\Sigma_P^T\Sigma_P}\right)$. To derive the updating equation we apply Lemma 3a and equation (22) in [8] to find that

$$\nabla_{\mathbf{P}} I\left(\mathbf{s};\mathbf{y}\right) = \mathbf{H^T HPE}$$
$$\nabla_{\mathbf{\Sigma_P}} I\left(\mathbf{s};\mathbf{y}\right) = \mathbf{U_P^T}\nabla_{\mathbf{P}} I\left(\mathbf{s};\mathbf{y}\right)\mathbf{V_P^T}$$

so
$$\nabla_{\mathbf{\Sigma_P}} I\left(\mathbf{s};\mathbf{y}\right) = \mathbf{\Sigma_P\Sigma_H^T\Sigma_H V_P^T EV_P}$$

and, taking into account that we are only interested in the diagonal values of $\mathbf{\Sigma_P}$, we get $\nabla_{\mathbf{s}} I\left(\mathbf{s};\mathbf{y}\right) = \mathbf{diag}\left(\mathbf{\Sigma_P\Sigma_H^T\Sigma_H V_P^T EV_P}\right)$. Besides, applying the chain rule to relate the derivative with respect to the precoder eigenvalues and its squared values we can easily find that $\nabla_{\mathbf{p}} I\left(\mathbf{s};\mathbf{y}\right) = \frac{1}{2}\mathbf{diag}\left(\mathbf{\Sigma_H^T\Sigma_H V_P^T EV_P}\right)$. If the only constraint on $\mathbf{p}$ were that the sum of its values should be constant, its optimum value could be found initializing $\widehat{\mathbf{p}}_1 = \mathbf{1}\frac{1}{N}$ and using the following gradient algorithm to update its estimates for the k-th iteration:

$$\widehat{\mathbf{p}}_k = \widehat{\mathbf{p}}_{k-1} + \mu_p\left(\nabla_{\mathbf{p}} I\left(\mathbf{s};\mathbf{y}\right) - \mathbf{1}\frac{1}{N}\left(\mathbf{1^T}\nabla_{\mathbf{p}} I\left(\mathbf{s};\mathbf{y}\right)\right)\right) \quad (11)$$

which corresponds to the gradient of the mutual information with the linear constraint. However, as $\widehat{\mathbf{p}}_k$ must have non-negative coefficients, we propose to correct the estimates obtained with this update by setting to zero the coefficients that become negative after updating with equation (11) and re-normalizing the resulting vector to satisfy the power constraint. After this normalization we can set $\mathbf{\Sigma_P} = \left(Diag(\widehat{\mathbf{p}}_k)\right)^{\frac{1}{2}}$.

B. Gradient updating of $\mathbf{P^H H^H HP}$

In the second stage of the iterative procedure we optimize the following problem

$$\max_{\mathbf{P^T H^T HP}} I\left(\mathbf{s};\mathbf{y}\right)$$
$$s.t.\ \text{eigenvalues}\left\{\mathbf{P}^T\mathbf{H}^T\mathbf{HP}\right\} = \mathbf{\Sigma}_P^T\mathbf{\Sigma}_H^T\mathbf{\Sigma}_H\mathbf{\Sigma}_P$$

where the estimates of $\mathbf{\Sigma_P}$ are the values obtained in the previous stage of the iterative process. We will try to simplify notation defining $\mathbf{M} = \mathbf{P}^T\mathbf{H}^T\mathbf{HP}$. As we know that $\nabla_{\mathbf{P^T H^T HP}} I\left(\mathbf{s};\mathbf{y}\right) = \frac{1}{\sigma_n^2}\mathbf{E}$, we propose to find this optimum value by updating the estimate of $\mathbf{M}$ as

$$\widehat{\mathbf{M}}_k = \widehat{\mathbf{M}}_{k-1} + \mu_M\mathbf{E}$$

and projecting the new update $\widehat{\mathbf{M}}_k$ into a matrix with the desired eigenvalues and close to $\widehat{\mathbf{M}}_k$ in Frobenius norm sense. While the closest matrix to $\widehat{\mathbf{M}}_k$ with the prescribed values can only be found numerically [2], if the eigenvalue decomposition of $\widehat{\mathbf{M}}_k$ is $\mathbf{Q_M\Lambda_M Q_M^T}$ and we write $\widehat{\mathbf{M}}_k = \mathbf{G_M G_M^T}$ with $\mathbf{G_M} = \mathbf{Q}_M\mathbf{\Lambda}_M^{1/2}$, then the orthogonal matrix $\mathbf{R}$ that minimizes the Frobenius norm of $\mathbf{G_M^T} - \mathbf{\Sigma}_H\mathbf{\Sigma}_P\mathbf{R}$ has a closed form solution [4]. If we denote the SVD of $\mathbf{\Sigma}_H\mathbf{\Sigma}_P\mathbf{G_M^T}$ as $\widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^T$ then we can obtain the desired projection as $\widehat{\mathbf{M}}_k = \widetilde{\mathbf{V}}\widetilde{\mathbf{U}}^T\mathbf{\Sigma}_P^T\mathbf{\Sigma}_H^T\mathbf{\Sigma}_H\mathbf{\Sigma}_P\widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T$, so at each iteration we will update the estimate of $\mathbf{P}^T\mathbf{H}^T\mathbf{HP}$ with this value and the estimate of the precoder right singular vectors with $\mathbf{V}_P^T = \widetilde{\mathbf{U}}\widetilde{\mathbf{V}}^T$.

C. Simulation results

In this section we illustrate the performance of the proposed algorithm by showing the convergence of the solution for a 4x4 MIMO channel that was randomly generated and has singular values $Diag([3.10\ 1.59\ 0.79\ 0.50])$. A BPSK signal was transmitted at SNR = 8 dB. For each iteration of the gradient algorithm the MMSE was updated with 100 new samples. The mutual information convergence is shown as a function of the number of gradient iterations for the algorithm proposed
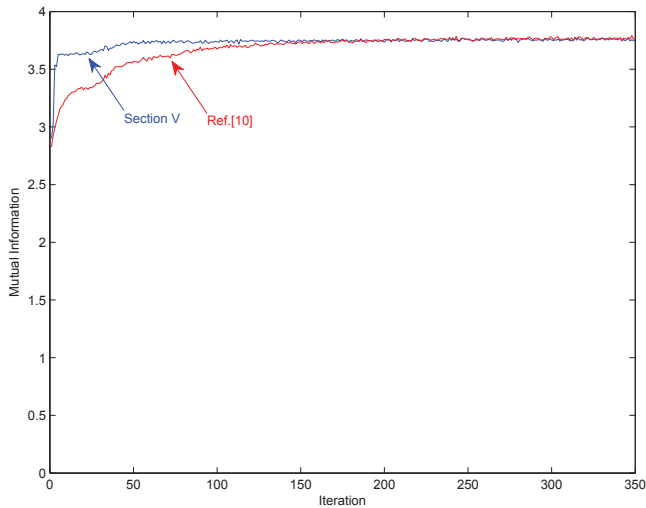
Figure 1. Mutual information evolution for the algorithm proposed in this paper and for the algorithm proposed in [12].

here as well as for that one employed in [12], which was an algorithm based on the gradient of the mutual information with respect to **P** without exploiting precoder structure. In both cases the same amount of data was used, so in the case of [12] the precoder estimate was updated every iteration, whereas in the algorithm proposed here the precoder singular values were updated in the even iterations and the right singular vectors in the odd iterations. In both cases the algorithm step size was selected to achieve fast convergence without any oscillation behavior. Note the improved convergence thanks to the exploitation of the precoder structure knowledge.

VI. CONCLUSION

In this paper we have derived the structure for the linear MIMO precoder that maximizes the mutual information and we have shown that the mutual information is concave in a quadratic function of precoder coefficients. We have also proposed an iterative gradient algorithm with increased convergence speed.

ACKNOWLEDGMENT

REFERENCES

[1] S. Barbarossa and A. Scaglione, "Time-varying channels," in *Signal Processing Advances in Wireless Communications*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000, vol. II, Trends in Single- and Multiuser Systems, ch. 1.

[2] M. T. Chu and K. R. Driessel, "The Projected Gradient Method for Least Squares Matrix Approximations with Spectral Constraints", SIAM Journal on Numerical Analysis, vol. 47, no. 4, pp.1050-1060, August 1990.

[3] D. Guo, S. Shamai, and s. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels", *IEEE Trans. on Information Theory*, vol. 51, no.4, pp.1261-1282, April 2005.

[4] N. J. Higham, "Matrix Nearness Problems and Applications", *Proc. of IMA Conference on Applications of Matrix Theory*, pp.1-27, Oxford University Press, 1989

[5] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum Power Allocation for Parallel Gaussian Channels with Arbitrary Input Distributions", *IEEE Trans. on Information Theory*, vol.52, no.7, pp.3033-3051, July 2006.

[6] E. N. Onggosanusi, A. M. Sayeed, and D. Van Veen, "Optimal antenna diversity signaling for wide-band systems utilizing channel side information," *EEE Trans. on Communications*, vol. 50, no. 2, pp. 341–353, Feb. 2002

[7] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO channels: a unified framework for convex optimization". *IEEE Trans. on Signal Processing*, vol.51, no.9, pp.2381-2401, September 2003.

[8] D. P. Palomar and S. Verdú, "Gradient of Mutual Information in Linear Vector Gaussian Channels", *IEEE Trans. on Information Theory*, vol.52, no.1, pp.141-154, January 2006.

[9] D. P. Palomar and Y. Jiang, "MIMO Transceiver Design via Majorization Theory", *Foundations and Trends in Communications and Information Theory*, Now Publishers, vol. 3, no. 4-5, pp. 331-551, 2006.

[10] M. Payaró and D. P. Palomar, "Hessian and Concavity of Mutual Information: Differential Entropy, and Entropy Power in Linear Vector Gaussian Channels", *accepted for publication at IEEE Trans. on Information Theory*.

[11] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "Optimal Linear Precoding for Multiple-Input Multiple-Output Gaussian Channels with Arbitrary Inputs", *submitted to IEEE Trans. on Information Theory*.

[12] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "Optimal Precoding for Digital Subscriber Lines", *Proc. of Intl. Communications Conference (ICC'08),* Beijing (China), May 2008.

[13] F. Rey, M. Lamarca, G. Vázquez, "A Transmitter Design for Coded Systems in the Presence of CSI Errors", *Proc. of 37th Asilomar Conf. on Signals, Systems and Computers,* Monterey (California), November 2003.

[14] F. Rey, M. Lamarca, and G. Vázquez, "Linear Precoder Design Through Cut-off Rate Maximization in MIMO-OFDM Coded Systems with Imperfect CSIT", *submitted to IEEE Trans. on Communications*.

[15] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels", *IEEE Trans. on Information Theory*, vol.45, no.3, pp.1019-1032, April 1999.