SPATIAL TRANSMIT PREFILTERING FOR FREQUENCY-FLAT MIMO TRANSMISSION WITH MEAN AND COVARIANCE INFORMATION

Ruben de Francisco, Dirk T.M. Slock

Institut Eurécom, 2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Tel: +33 4 9300 2916/2606; Fax: +33 4 9300 2627
e-mail:{defranci,slock}@eurecom.fr

ABSTRACT

In this paper techniques are proposed for combining information about the mean and the covariance of the channel for the purpose of MIMO transmission. (Partial) channel state information at the transmitter (CSIT) is typically used in MIMO systems for the design of spatial prefiltering and waterfilling. For the purpose of generating CSIT, the cases of mean or covariance information have only been solved separately in the literature. A Bayesian approach is presented here incorporating both pieces of information, but in which correlations are limited to the transmitter side. The approach yields the existing cases of mean or (transmit) covariance information as special instances. Various cases of mean and covariance information are discussed, including prior mean and covariance (Ricean channel distribution) and posterior mean and covariance (based on a noisy channel estimate and prior covariance information). The case of a singular covariance matrix is treated in detail also, allowing to treat zero covariance (known channel) as a special case.

1. INTRODUCTION

In practical wireless systems, training sequences or pilot symbols are incorporated in the transmitted signal to allow for channel estimation at the receiver. The density of training data needs to increase as the mobility and the channel variation increases. Nevertheless, even with training data available, the channel estimate can only be of limited quality, and the channel estimation errors reduce the channel capacity. Furthermore, the fact of substituting data to be transmitted by training data obviously also limits the capacity. All this means that the channel capacity degrades

with mobile speed and to minimize this decrease, all a priori information about the channel should be exploited for its estimation.

In order to exploit partial Channel State Information at the Transmitter (CSIT) in MIMO systems, most of the current precoding schemes exploit either mean [1] or covariance [2] information. A combination of the two can improve exploitation of channel knowledge by weighting them according to a certain criteria. However, mean and covariance information do not necessary have to correspond to actual channel mean and covariance, i.e. to prior distributions. They can be given by a Bayesian approach, with a certain posterior mean and covariance.

In our work, we present different techniques to combine mean and covariance information. We show how to exploit both sources of partial CSIT to optimize the error rate in MIMO systems, by performing linear precoding at the transmitter.

2. SYSTEM DESCRIPTION

We consider transmission over a flat-fading MIMO channel, described by a matrix $\mathbf{h}$ with random complex entries and dimension $N_R \text{x} N_T$, being N_T and N_R the number of transmit and receive antennas, respectively. Prior to transmission, the data codewords are prefiltered with the matrix $\mathbf{W}$, which takes into account mean and covariance information, as we detail in the following section. In the presence of additive white Gaussian noise, the received signal is given by

$$\mathbf{Y} = \mathbf{h}\,\mathbf{WC} + \mathbf{V} \tag{1}$$

where the noise power spectral density matrix is $S_{\boldsymbol{vv}}(z) = \sigma^2$. The transmitted data is recovered by a Maximum Likelihood receiver.

We assume that the channel impulse response matrix $\mathbf{h}$ has separable covariance structure. Hence

$$\mathrm{E}\,\mathbf{h}\mathbf{h}^H = \mathrm{tr}\{C_T\}\,C_R \tag{2}$$
$$\mathrm{E}\,\mathbf{h}^T\mathbf{h}^* = \mathrm{tr}\{C_R\}\,C_T$$

3. PARTIAL CSIT COMBINING MEAN AND COVARIANCE

When mean and covariance information are present at the transmitter side, this information can be of different nature. In the presence of a Line-of-Sight component between transmitter and receiver, the MIMO channel may be modeled as Ricean. In this first case of interest, the channel $\mathbf{h}$ has a prior distribution $\mathbf{h} \sim \mathcal{CN}(m_\mathbf{h}, C_\mathbf{hh})$, with mean $m_\mathbf{h}$ due to the LOS component. The Ricean channel can be modeled as $\mathbf{h} = m_\mathbf{h} + C_R^{1/2} \mathbf{h}_w C_T^{H/2}$ with $\mathbf{h}_w$ distributed as $\mathcal{CN}(0, I_{N_T N_R})$. A special case can be considered in the Ricean model when also the mean is separable. Thus, the vectorized mean can be represented as $m_\mathbf{h} = m_T^* \otimes m_R$ and matricially as $m_\mathbf{h} = m_R m_T^H$. A second source of partial CSIT is the case when mean information corresponds to the channel estimate and perfect covariance information is present at the transmitter. The channel in this case is modeled as Rayleigh with distribution $\mathbf{h} \sim \mathcal{CN}(0, C_\mathbf{hh})$. On the other hand, the channel estimates can be modeled as $\widehat{\mathbf{h}} = \mathbf{h} + \widetilde{\mathbf{h}}$, where $\widetilde{\mathbf{h}}$ follows a distribution $\mathcal{CN}(0, \sigma_{\widetilde{h}}^2 I_{N_T N_R})$ and could be due to a combination of the following sources of error: estimation noise, quantization noise and prediction noise. The combination of mean and covariance information leads to a Gaussian posterior distribution with posterior mean given by

$$\widehat{\underline{\mathbf{h}}} = (I_{N_R N_T} + \sigma_{\widetilde{h}}^2 C_T^{-1} \otimes C_R^{-1})^{-1} \widehat{\underline{\mathbf{h}}} \qquad (3)$$

and posterior covariance

$$\widetilde{C} = (\sigma_{\widetilde{h}}^{-2} I_{N_R N_T} + C_T^{-1} \otimes C_R^{-1})^{-1} \qquad (4)$$

If only posterior mean is present, it is due to noise-free channel estimation ($\sigma_{\widetilde{h}}^2 \to 0$) and thus $\|C\| \to 0$. On the other hand, only posterior covariance will be present if $\widehat{\underline{\mathbf{h}}} = \widehat{\mathbf{h}} = 0$ or the estimation noise tends to infinity. In addition, if a rich scattering environment is assumed at the Rx side, the covariance at the receiver can be modeled as identity. In this case, the posterior (vectorized) mean is given by

$$\widehat{\underline{\mathbf{h}}} = \widehat{\mathbf{h}}(I_{N_T} + \sigma_{\widetilde{h}}^2 C_T^{-1})^{-1} \qquad (5)$$

and the posterior covariance

$$\widetilde{C} = \mathbf{I}_{N_R} \otimes \widetilde{C}_T \qquad (6)$$

with posterior covariance seen from the transmitter

$$\widetilde{C}_T = (\sigma_{\widetilde{h}}^{-2} I_{N_T} + C_T^{-1})^{-1} \qquad (7)$$

In a simplified scenario with noisification of the mean, assume we only have access to $D\widehat{\underline{\mathbf{h}}}$ instead of having acess to $\widehat{\underline{\mathbf{h}}}$ directly, where the elements of the diagonal matrix D

are i.i.d. $\mathcal{CN}(0,1)$. Now the distribution becomes zero mean with transmit side covariance matrix $\widetilde{R}_T = \widehat{\widehat{\mathbf{h}}}^H \widehat{\widehat{\mathbf{h}}} + \widetilde{C}_T$. Under these circumstances, the mean information falls into the covariance information, and thus the correlation becomes the covariance. Hence, optimal MIMO transmission schemes with partial CSIT for the case of only covariance information will apply for the described model that combines mean and covariance infomation. If linear prefiltering is carried out at the transmitter side (after the ST encoding stage) to adapt the transmission to the channel knowledge, an optimal prefilter will lead to capacity maximization or tipically Pairwise Error Probability (PEP) minimization. If we assume $C_R = \mathbf{I}_{N_R}$, the optimal prefilter that maximizes capacity and minimizes PEP pours power along the eigenvectors of the posterior correlation matrix seen from the transmitter, following a waterfilling power allocation policy for minimum PEP [2] and possibly different weighting for the capacity maximization solution [6].

4. LINEAR PRECODING FOR ERROR RATE MINIMIZATION

In this section, we derive an optimal precoding strategy for error rate minimization in MIMO systems combining mean and covariance information at the transmitter. The source of mean and covariance information can be either prior or posterior, as described in the previous section. We optimize the performance of the proposed system in terms of PEP averaged over $\mathbf{h}$, prior or posterior, with a certain distribution $\mathcal{CN}(m_\mathbf{h}, C_\mathbf{hh})$. We assume identity covariance matrix at the receiver. In the analysis, we follow the work developed by *Jongren et al.* in [4].

The PEP is defined as the error probability of choosing the nearest distinct codeword $\mathbf{C}^j$ instead of $\mathbf{C}^i$. The code error matrix can be defined as $\widetilde{\mathbf{E}} := [\mathbf{C}^i - \mathbf{C}^j]$. In practice, the average PEP is limited by the minimum distance code error matrix, given by $\mathbf{E} = arg \min_{\widetilde{\mathbf{E}}(i,j)} det\left[\widetilde{\mathbf{E}}(i,j)\widetilde{\mathbf{E}}^H(i,j)\right]$.

The average PEP is given by

$$P(\mathbf{C}^i \to \mathbf{C}^j) = \int P(\mathbf{C}^i \to \mathbf{C}^j | \mathbf{h}) p_\mathbf{h}(\mathbf{h}) d\mathbf{h} \qquad (8)$$

where the complex Gaussian PDF $p_\mathbf{h}(\mathbf{h})$ is

$$p_\mathbf{h}(\mathbf{h}) = \frac{e^{-tr\left[(\mathbf{h}-m_\mathbf{h})^H C_\mathbf{hh}^{-1}(\mathbf{h}-m_\mathbf{h})\right]}}{\pi^{N_R N_T} \det(C_\mathbf{hh})^{N_R}} \qquad (9)$$

By applying the Chernoff bound and averaging over the distribution of $\mathbf{h}$, an upper bound on the average PEP is given by

$$P(\mathbf{C}^i \to \mathbf{C}^j) \leq \int e^{-d_{min}^2(\mathbf{C}^i, \mathbf{C}^j)/4} p_\mathbf{h}(\mathbf{h}) d\mathbf{h} \qquad (10)$$

When concatenating the Space-Time encoder at the transmitter with a linear prefilter to exploit partial CSIT, the minimum Euclidean distance is

$$d^2_{min}(\mathbf{C}^i, \mathbf{C}^j) = d^2(\mathbf{E}) = \frac{1}{\sigma^2}\|\mathbf{hWE}\|^2_F \qquad (11)$$

where $\mathbf{W}$ is the linear prefilter. On the other hand, it can be shown that if $\mathbf{EE}^H = \alpha\mathbf{I}$, the PEP is minimized at high SNR for a given optimal prefilter [3]. Thus, the system under consideration has $\mathbf{EE}^H = \alpha\mathbf{I}$, e.g. orthogonal ST block codes [7] (single stream) or ST spreading [5] (full stream). Introducing $\eta = \frac{\alpha}{4\sigma^2}$ and $\Psi = \mathbf{WW}^H$, the solution to (10) is given by

$$P(\mathbf{C}^i \rightarrow \mathbf{C}^j) \leq \frac{e^{\text{tr}\left[m_\mathbf{h} C^{-1}_\mathbf{hh}\left((\eta\Psi + C^{-1}_\mathbf{hh})^{-1} - C_\mathbf{hh}\right)C^{-1}_\mathbf{hh}m^H_\mathbf{h}\right]}}{\det(\eta\Psi + C^{-1}_\mathbf{hh})^{N_R}\det(C_\mathbf{hh})^{N_R}} \qquad (12)$$

The performance criterion can be expressed logarithmically (neglecting parameter-independent terms) as follows

$$J = \text{tr}\left[m_\mathbf{h} C^{-1}_\mathbf{hh}(\eta\Psi + C^{-1}_\mathbf{hh})^{-1}C^{-1}_\mathbf{hh}m^H_\mathbf{h}\right] - N_R\text{logdet}(\eta\Psi + C^{-1}_\mathbf{hh}) \qquad (13)$$

Assuming a normalized average power constraint, the optimal Ψ that minimizes the performance criterion in (13) is given by (see Appendix A)

$$\Psi = \left\{\frac{1}{2\mu}\left[N_R\mathbf{I}_{N_T} + \left(N_R^2\mathbf{I}_{N_T} + \frac{4\mu}{\eta}C^{-1}_\mathbf{hh}m^H_\mathbf{h}m_\mathbf{h}C^{-1}_\mathbf{hh}\right)^{\frac{1}{2}}\right] - \frac{1}{\eta}C^{-1}_\mathbf{hh}\right\}_+ \qquad (14)$$

where μ is the Lagrange multiplier associated with the power constraint and $\{\cdot\}_+$ takes the positive semidefinite part. It can be seen straightforward from (14) that as η tends to infinity (i.e. SNR tends to infinity), the optimal Ψ tends to $\Psi = \frac{1}{N_T}\mathbf{I}_{N_T}$, since in this particular case the value of the Lagrange multiplier is $\mu = N_R N_T$. This result is equivalent to transmission without CSIT, which shows that as the SNR increases the importance of CSIT gets reduced. Another solution assuming full-rank Ψ is provided, to have a more intuitive idea of the unequal power-loading policy at the transmitter. We define the eigenvalue decomposition $C^{-1}_\mathbf{hh}m^H_\mathbf{h}m_\mathbf{h}C^{-1}_\mathbf{hh} = \mathbf{U}\Sigma\mathbf{U}^H$ where $\mathbf{U}$ is a unitary matrix and $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_{N_T})$. The solution is given by

$$\Psi = \mathbf{U}\Lambda\mathbf{U}^H - \frac{1}{\eta}C^{-1}_\mathbf{hh} \qquad (15)$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{N_T})$. The elements in the matrix of eigenvalues Λ are given by (see Appendix B)

$$\lambda_i = \frac{N_R + \sqrt{N_R^2 + 4\frac{\mu\sigma_i}{\eta}}}{2\mu} \qquad (16)$$

To obtain the optimal precoder either from (14) or (15), let the eigenvalue decomposition of Ψ be $\Psi = \mathbf{V}_\Psi\Lambda_\Psi\mathbf{V}^H_\Psi$.

Since $\Psi = \mathbf{WW}^H$, the optimal precoder is $\mathbf{W} = \mathbf{V}_\Psi\Lambda^{1/2}_\Psi$. Thus, the optimal transmission strategy as reflected in the above equations corresponds to transmission along the eigenvectors of a combination of mean and covariance and a waterfilling power allocation policy. The first and second term in (15) are differently weighted depending on the SNR and the covariance information. In the remaining of this section we introduce some particular cases of special interest.

4.1. Zero mean information

When the mean information is zero, it can be seen from equation (14) that in this case Ψ becomes

$$\Psi = \left\{\frac{N_R}{\mu}\mathbf{I}_{N_T} - \frac{1}{\eta}C^{-1}_\mathbf{hh}\right\}_+ \qquad (17)$$

The value of the lagrange multiplier can be analitically expressed as

$$\mu = \frac{N_R N_T}{\left[1 + \frac{1}{\eta}tr(C^{-1}_\mathbf{hh})\right]} \qquad (18)$$

It is clear from (17) and (18) that as the SNR increases the covariance information becomes less important, and Ψ converges to a scaled identity matrix.

4.2. Unit rank mean

A particular case of interest is the case when the mean information has rank one. Since $m_\mathbf{h}$ is unit rank, also $C^{-1}_\mathbf{hh}m^H_\mathbf{h}m_\mathbf{h}C^{-1}_\mathbf{hh}$ becomes unit rank. The mean $m_\mathbf{h}$ can be represented as a combination of a pair of vectors s and t, $m_\mathbf{h} = s\ t^H$. The solution for Ψ in the case of unit rank mean derived from the full-rank solution in (15) is given by

$$\Psi = [u_1\mathbf{U}_2]\Lambda[u_1\mathbf{U}_2]^H - \frac{1}{\eta}C^{-1}_\mathbf{hh} \qquad (19)$$

where $\Lambda = diag(1 + \frac{tr\left[C^{-1}_\mathbf{hh}\right]}{\eta}, 0, \ldots, 0)$, u_1 is the eigenvector associated with the only non-zero eigenvalue and $\mathbf{U}_2$ are arbitrary vectors chosen such that the matrix $[u_1\mathbf{U}_2]$ forms an orthonormal basis. It can be seen from (19) that as the SNR increases the solution approaches to beamforming along a single direction, defined by $C^{-1}_\mathbf{hh}m^H_\mathbf{h}m_\mathbf{h}C^{-1}_\mathbf{hh}$, which is a combination of mean and covariance information.

4.3. Singular covariance information

When the covariance information is singular, it can be modeled as follows

$$C_\mathbf{hh} = \begin{bmatrix} X_{/\!/} X_\perp \end{bmatrix}\begin{bmatrix} \mathbf{A} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X_{/\!/} X_\perp \end{bmatrix}^H \qquad (20)$$

where $\perp$ and $\parallel$ represent singular and non-singular parts respectively. Let $\left[m_{\mathbf{h}\parallel} m_{\mathbf{h}\perp} \right] = m_{\mathbf{h}} \left[X_\parallel X_\perp \right]$ and $C_\parallel$ be the non-singular part of $C_{\mathbf{hh}}$. The optimization problem in this case becomes

$$\begin{cases} J = \min_{\Psi} tr \left[m_{\mathbf{h}\parallel} C_\parallel^{-1} \left(\eta \Psi_\parallel + C_\parallel^{-1} \right)^{-1} C_\parallel^{-1} m_{\mathbf{h}\parallel}^H \right] \\ \quad - N_R \log \det(\eta \Psi_\parallel + C_\parallel^{-1}) - \eta tr \left(m_{\mathbf{h}\perp} \Psi_\perp m_{\mathbf{h}\perp}^H \right) \\ s.t. \ tr(\Psi_\parallel + \Psi_\perp) = 1 \end{cases}$$
(21)

where $\Psi = [\Psi_\parallel \Psi_\perp]^T$. The objective function J can be divided in two optimization problems $J = J_\parallel + J_\perp$ minimized separately. The power constraints in both cases have to be adjusted so that $P_\parallel + P_\perp = 1$. Hence, each minimization problem has a different power constraint associated. The optimization problem for the singular part is given by

$$\begin{cases} J_\perp = \min_{\Psi_\perp} -\eta tr \left(m_{\mathbf{h}\perp} \Psi_\perp m_{\mathbf{h}\perp}^H \right) \\ s.t. \ tr(\Psi_\perp) = P_\perp \end{cases}$$
(22)

The solution for $\Psi_\perp$ is derived in Appendix C. The precoding solution corresponds to eigenbeamforming in the direction of the eigenvector of $m_{\mathbf{h}\perp}^H m_{\mathbf{h}\perp}$ associated with the largest eigenvalue $\lambda_{m_\perp,1}$. The result of the objective funtion is $J_\perp = -\eta P_\perp \lambda_{m_\perp,1}$. The remaining optimization problem for the non-singular part is given by

$$\begin{cases} J_\parallel = \min_{\Psi_\parallel} tr \left[m_{\mathbf{h}\parallel} C_\parallel^{-1} \left(\eta \Psi_\parallel + C_\parallel^{-1} \right)^{-1} C_\parallel^{-1} m_{\mathbf{h}\parallel}^H \right] \\ \quad - N_R \log \det(\eta \Psi_\parallel + C_\parallel^{-1}) - \eta P_\perp \lambda_{\perp 1} \\ s.t. \ tr(\Psi_\parallel) = P_\parallel = 1 - P_\perp \end{cases}$$
(23)

The solution for this part is equivalent to the general solution with waterfilling shown in (14), but with reduced dimension and power constraint due to singularities. On the other hand, an optimal power split solution exits $(P_\parallel, P_\perp)$ under certain circumstances such that $J(P_\parallel) = J_\parallel(P_\parallel) + J_\perp(1 - P_\parallel)$ is minimized. If $J(P_\parallel)$ has an absolute minimum $P_{\parallel | J_{min}}$, there are three different possibilities. If $0 < P_{\parallel | J_{min}} < 1$, the optimal power for the non-singular part is $P_{\parallel | opt} = P_{\parallel | J_{min}}$ and for the singular part $P_{\perp | opt} = 1 - P_{\parallel | opt}$. The solution is a combination of beamforming (in $\perp$ part) and waterfilling (in $\parallel$ part). If $P_{\parallel | J_{min}} \geq 1$ then $P_{\parallel | opt} = 1$ and $P_{\perp | opt} = 0$, and the solution is given by waterfilling in the non-singular part. Finally, if $P_{\parallel | J_{min}} \leq 0$ then $P_{\parallel | opt} = 0$ and $P_{\perp | opt} = 1$, and the solution is given by beamforming in $m_\perp^H m_\perp$.

5. SIMULATION RESULTS

The system considered is 2x2 O-STBC as described in [7] with QPSK modulation. The symbols are transmitted over a channel with an arbitrary mean and a correlation factor ρ

(cross-diagonal terms in $C_{\mathbf{hh}}$). Fig. 1 shows the gain in performance that can be obtained w.r.t. a non-precoded system by combining mean and covariance knowledge. We can observe a remarkable improvement in the simulated range of up to ≈ 2.5dB in SNR decrease for a given PEP. In this particular case, the performance is close to the one achived by an optimal prefilter with perfect CSIT. In Fig. 2 we compare the performance of a system with only covariance or mean knowledge at the transmitter and a system with both sources of CSIT. As we have shown, the cases of only mean or covariance CSIT can be considered special instances of a system that combines both. The contribution of the mean to the average channel power is denoted by γ.
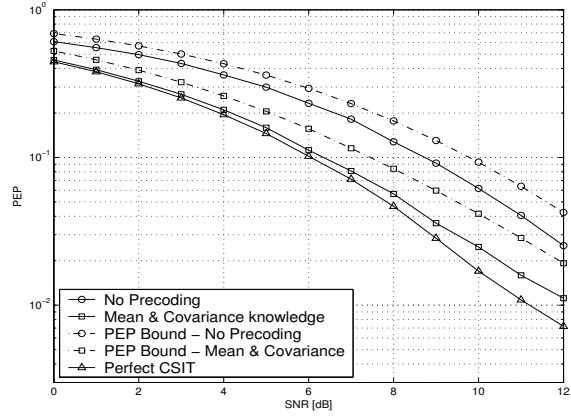

Figure 1: PEP vs. SNR for different levels of CSIT, $\rho = 0.9$ and $\gamma = 40\%$
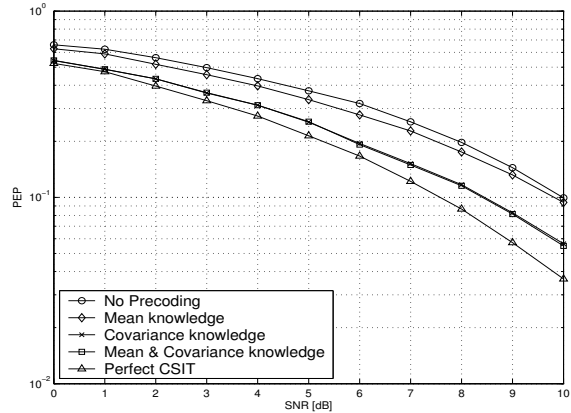

Figure 2: PEP vs. SNR for different levels of CSIT, $\rho = 0.9$ and $\gamma = 10\%$

6. CONCLUSIONS

In this paper, techniques for combining mean and covariance information have been presented. Both sources of information can be either prior (e.g. correlated channel with

LOS) or posterior (given by a Bayesian approach). We provide a general precoding solution for PEP minimization when combining both sources of partial CSIT at the transmitter, and analyze some cases of special interest. Simulation results illustrate the performance benefits that can be reached in MIMO systems. The results show how mean and covariance information should be combined in order to exploit the available sources of CSIT.

Appendix A.

The optimization problem described in (13) can be expressed as

$$
\begin{cases}
\min_{\Psi} tr \left[m_{\mathbf{h}} C_{\mathbf{hh}}^{-1} \left(\eta \Psi + C_{\mathbf{hh}}^{-1} \right)^{-1} C_{\mathbf{hh}}^{-1} m_{\mathbf{h}}^H \right] \\
\qquad - N_R \log \det(\eta \Psi + C_{\mathbf{hh}}^{-1}) \\
s.t. \; tr(\Psi) = 1
\end{cases}
\tag{24}
$$

The solution is obtained by means of the Karush-Kuhn-Tucker (KKT) conditions. Defining the Lagrangian as

$$
L(\Psi, \mu) = tr \left[m_{\mathbf{h}} C_{\mathbf{hh}}^{-1} \left(\eta \Psi + C_{\mathbf{hh}}^{-1} \right)^{-1} C_{\mathbf{hh}}^{-1} m_{\mathbf{h}}^H \right] \tag{25}
$$
$$
- N_R \log \det(\eta \Psi + C_{\mathbf{hh}}^{-1}) + \mu \; [tr(\Psi) - 1]
$$

where μ is the Lagrange multiplier associated with the equality constraint. Differentiating $L(\Psi, \mu)$ w.r.t. Ψ we get

$$
\mu \Phi \Phi - \eta N_R \Phi - \eta C_{\mathbf{hh}}^{-1} m_{\mathbf{h}}^H m_{\mathbf{h}} C_{\mathbf{hh}}^{-1} = 0 \tag{26}
$$

where the change of variable $\Phi = \eta \Psi + C_{\mathbf{hh}}^{-1}$ has been used for clarity. The solution for Ψ to the quadratic matrix equation described above is given by

$$
\Psi = \left\{ \frac{1}{2\mu} \left[N_R \mathbf{I}_{N_T} + \left(N_R{}^2 \mathbf{I}_{N_T} + \frac{4\mu}{\eta} C_{\mathbf{hh}}^{-1} m_{\mathbf{h}}^H m_{\mathbf{h}} C_{\mathbf{hh}}^{-1} \right)^{\frac{1}{2}} \right] - \frac{1}{\eta} C_{\mathbf{hh}}^{-1} \right\}_+
\tag{27}
$$

where $\{ \cdot \}_+$ takes the positive semidefinite part.

Appendix B.

Introducing the eigenvalue decompositions $C_{\mathbf{hh}}^{-1} m_{\mathbf{h}}^H m_{\mathbf{h}} C_{\mathbf{hh}}^{-1} = \mathbf{U} \Sigma \mathbf{U}^H$ and $\Psi + \frac{1}{\eta} C_{\mathbf{hh}}^{-1} = \mathbf{V} \Lambda \mathbf{V}^H$ in equation (24) and eliminating constant terms, the minimization problem becomes

$$
\begin{cases}
\min_{\Psi} tr \left(\frac{1}{\eta} \Lambda^{-1} \mathbf{V}^H \mathbf{U} \Sigma \mathbf{U}^H \mathbf{V} \right) - N_R \log \det(\Lambda) \\
s.t. \; tr(\Lambda) = \beta
\end{cases}
\tag{28}
$$

where $\beta = 1 + \frac{1}{\eta} tr \left[C_{\mathbf{hh}}^{-1} \right]$ and the properties $tr(AB) = tr(BA)$ and $\mathbf{V}^H \mathbf{V} = \mathbf{I}_{N_T}$ have been used. Since the solution we seek assumes Ψ positive semidefinite, also $\Lambda - \frac{1}{\eta} \mathbf{V}^H C_{\mathbf{hh}}^{-1} \mathbf{V}$ is assumed PSD. The optimum $\mathbf{V}$ that minimizes the first term can be chosen as $\mathbf{V} = \mathbf{U}$ [4].
Let $\Sigma = diag(\sigma_1, , \dots, \sigma_{N_T})$ and $\Lambda = diag(\lambda_1, \dots, \lambda_{N_T})$. The Lagrangian in this case is given by

$$
L(\lambda_i, \mu) = \sum_{i=1}^{N_T} \left(\frac{1}{\eta} \frac{\sigma_i}{\lambda_i} - N_R \log \lambda_i \right) + \mu \left(\sum_{i=0}^{N_T} \lambda_i - \beta \right) \tag{29}
$$

where μ is the Lagrange multiplier corresponding to the power constraint. Differentiating $L(\lambda_i, \mu)$ w.r.t. λ_i we get

$$
\lambda_i = \frac{N_R + \sqrt{N_R{}^2 + 4 \frac{\mu \sigma_i}{\eta}}}{2\mu} \tag{30}
$$

Thus, the solution for Ψ is given by

$$
\Psi = \mathbf{U} \Lambda \mathbf{U}^H - \frac{1}{\eta} C_{\mathbf{hh}}^{-1} \tag{31}
$$

Appendix C.

In order to minimize the objective function in (22) subject to the power constraint, introduce the following eigenvalue decompositions: $m_{\mathbf{h}_\perp}^H m_{\mathbf{h}_\perp} = \mathbf{V}_{m_\perp} \Lambda_{m_\perp} \mathbf{V}_{m_\perp}^H$ and $\Psi_\perp = \mathbf{V}_{\Psi_\perp} \Lambda_{\Psi_\perp} \mathbf{V}_{\Psi_\perp}^H$. By applying the following inequality $tr(AB) \leq \sum_i \lambda_i(A) \lambda_i(B)$, it can be seen that (22) is minimized (the trace is maximized) by setting $\mathbf{V}_{\Psi_\perp} = \mathbf{V}_{m_\perp}$. Let $\Lambda_{m_\perp} = diag(\lambda_{m_\perp,1}, \dots, \lambda_{m_\perp, N_T})$ and $\Lambda_{\Psi_\perp} = diag(\lambda_{\Psi_\perp,1}, \dots, \lambda_{\Psi_\perp, N_T})$ ordered decreasingly. The optimization problem becomes

$$
\begin{cases}
J_\perp = \min_{\lambda_{\Psi_\perp,i}} -\eta \sum_{i=1}^{N_T} \lambda_{\Psi_\perp,i} \lambda_{m_\perp,i} \\
s.t. \; \sum_{i=1}^{N_T} \lambda_{\Psi_\perp,i} = P_\perp
\end{cases}
\tag{32}
$$

Clearly, the function described above is minimized (the summation is maximized) if all the power is transmitted along the strongest eigenvalue, $\lambda_{m_\perp,1}$. Hence, the solution is given by choosing $\lambda_{\Psi_\perp,1} = P_\perp$ and $\lambda_{m_\perp,i} = 0, i = 2, 3, \dots, N_T$. With this choice, the value of the objective function becomes $J_\perp = -\eta P_\perp \lambda_{m_\perp,1}$.

7. REFERENCES

[1] S. Ali Jafar, A. Goldsmith, "Transmitter Optimization and Optimality of Beamforming for Multiple Antenna Systems," *IEEE Trans. on Wireless Comm.*, vol. 3, no. 4, pp. 1165-1175, July 2004.

[2] H. Sampath and A.Paulraj, "Linear Precoding for Space-Time Coded Systems With Known Fading Correlations," *IEEE Communications Letters*, Vol. 6, no. 6, June 2002.

[3] R. de Francisco and D.T.M. Slock, "Linear Precoding for MIMO Transmission with Partial CSIT," *Proc. 6th IEEE Workshop on Signal Processing Advances in Wireless Communications*, New York, USA, June 2005.

[4] G. Jongren, M. Skoglund, B. Ottersten, "Combining Beamforming and Orthogonal Space-Time Block coding," *IEEE Trans. on Inf. Theory*, vol. 48, no. 3, pp. 611-627, March 2002.

[5] A. Medles and D.T.M. Slock, "Linear Convolutive Space-Time Precoding for Spatial Multiplexing MIMO Systems," *Proc. 39th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, USA, Oct. 2001.

[6] A.J. Paulraj, R. Nabar, and D. Gore, "Introduction to space-time wireless communications," *Cambridge University Press*, Cambridge, 2003.

[7] V. Tarokh, H. Jafarkhami, and A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.