Multiple-Input Multiple-Output Gaussian Channels: Optimal Covariance for Non-Gaussian Inputs

Miguel R. D. Rodrigues*, Fernando Pérez-Cruz†, and Sergio Verdú†

*Instituto de Telecomunicações, Department of Computer Science, University of Porto, Portugal
†Department of Electrical Engineering, Princeton University, New Jersey, U.S.A.

***Abstract*—We investigate the input covariance that maximizes the mutual information of deterministic multiple-input multiple-output (MIMO) Gaussian channels with arbitrary (not necessarily Gaussian) input distributions, by capitalizing on the relationship between the gradient of the mutual information and the minimum mean-squared error (MMSE) matrix. We show that the optimal input covariance satisfies a simple fixed-point equation involving key system quantities, including the MMSE matrix. We also specialize the form of the optimal input covariance to the asymptotic regimes of low and high snr. We demonstrate that in the low-snr regime the optimal covariance fully correlates the inputs to better combat noise. In contrast, in the high-snr regime the optimal covariance is diagonal with diagonal elements obeying the generalized mercury/waterfilling power allocation policy. Numerical results illustrate that covariance optimization may lead to significant gains with respect to conventional strategies based on channel diagonalization followed by mercury/waterfilling or waterfilling power allocation, particularly in the regimes of medium and high snr.**

Index Terms: Optimal Input Covariance, Multiple-Input Multiple-Output Systems, Mutual Information, MMSE, Gaussian Noise

I. Introduction

Gaussian inputs maximize the mutual information of linear channels with Gaussian noise. The optimal input covariance eigenvectors diagonalize the channel, whereas the optimal input covariance eigenvalues obey the waterfilling policy [1]. For a variety of reasons ranging from complexity to the physics of the medium (as in magnetic recording), in practice non-Gaussian constellations, such as BPSK and QAM, are used in lieu of Gaussian inputs. In fact, the intersymbol interference channel with constrained input distributions received considerable attention. For example, Hirt [2] studied the binary-input linear intersymbol interference channel. Shamai *et al.* [3],[4] derived various bounds on the capacity and the information rate of intersymbol interference channels. Others have resorted to the development of practical algorithms to determine the information rates of general finite-state source/channel models [5]. Zhang *et al.* [6] consider applications to magnetic recording.

The use of arbitrary input distributions rather than the conventional (and optimal) Gaussian ones, considerably complicates the optimization problem due to the absence of explicit and tractable mutual information expressions. The first step towards the resolution of this class of optimization problems was taken in [7], by exploiting the relationship between the mutual information and the minimum mean-squared error (MMSE) [8],[9]. In particular, [7] considers power allocation for parallel non-interfering Gaussian channels with arbitrary inputs, showing that the optimal policy follows a generalization of classic waterfilling, known as mercury/waterfilling. Reference [10] considers optimal power allocation for (interfering) multiple-input multiple-output (MIMO) Gaussian channels with arbitrary inputs. The design of optimal precoders (from a mutual information perspective) for MIMO Gaussian channels with arbitrary inputs is investigated in [11],[12].

In this paper, we consider the input covariance that maximizes mutual information as a function of the non-Gaussian input constellation, the channel matrix and the signal-to-noise ratio. To that end, we capitalize on the relationship between the gradient of the mutual information with respect to system parameters of interest [9]. In Section II we introduce the channel model. For maximum conceptual simplicity, we choose a real-valued model, which as the dimension grows can be used to model linear intersymbol interference. Section III determines the form of the optimal input covariance as a function of various system quantities. Sections IV and V specialize the form of the optimal input covariance to the asymptotic regimes of low and high snr, respectively. Finally, Section VI provides numerical results for a simple 2×2 example.

II. Channel Model

We consider the following real-valued vector channel model:

$$\mathbf{y} = \sqrt{\mathsf{snr}}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{y}$ is the n-dimensional vector of receive symbols, $\mathbf{x}$ is the m-dimensional vector of transmit symbols, $\mathbf{w}$ is a n-dimensional Gaussian noise random vector with mean zero and covariance $\mathbf{\Sigma_w} = \mathcal{E}\left[\mathbf{w}\mathbf{w^T}\right] = \mathbf{I}$, and the deterministic $n \times m$ matrix $\mathbf{H}$ models the deterministic channel gains from each input to every output [1].

We constrain the input symbols to belong to conventional zero-mean real constellations (e.g. BPSK). The objective is to determine the input covariance $\mathbf{\Sigma_x} = \mathcal{E}\left[\mathbf{x}\mathbf{x^T}\right]$ that maximizes $I(\mathbf{x};\mathbf{y})$, subject to the total power constraint $\mathsf{tr}\left(\mathbf{\Sigma_x}\right) = 1$. Note that the covariance matrix embodies two properties of operational significance: (i) the power of the various inputs; and (ii)

[1] It is also straightforward to consider more general noise covariances, by applying a pre-whitening filter that only affects the definition of the channel matrix.

the correlations between the various inputs. In this sense, the covariance optimization problem represents a generalization of the power allocation problems considered in [7],[10].

III. Optimal Input Covariance

We pose the optimization problem:

$$\max_{\mathbf{\Sigma_x}} I(\mathbf{x};\mathbf{y}) \tag{2}$$

subject to the total power constraint $\mathsf{tr}\big(\mathbf{\Sigma_x}\big) = 1$ and to the positive semidefinite constraint $\mathbf{\Sigma_x} \succeq 0$.

Theorem 1: The input covariance $\mathbf{\Sigma_x^*} \succeq 0$ that solves (2) satisfies:

$$\mathbf{\Sigma_x^*} = \alpha \mathbf{H}^T\mathbf{H}\mathbf{E} \tag{3}$$

where the MMSE matrix $\mathbf{E} = \mathcal{E}\big[(\mathbf{x}-\mathcal{E}[\mathbf{x}|\mathbf{y}])(\mathbf{x}-\mathcal{E}[\mathbf{x}|\mathbf{y}])^T\big]$ is a function of the optimal input covariance $\mathbf{\Sigma_x^*}$ and α is chosen to satisfy the unit-trace constraint.

Proof: See Appendix A. ∎

Theorem 1 follows from the Karush-Kuhn-Tucker Theorem [13]. The KKT conditions are satisfied by any of the critical points (minimum, maximum or saddle point) in an optimization problem. In concave optimization problems, the KKT conditions are satisfied by one solution, the global maximum. Unfortunately, the covariance optimization problem is nonconcave except in specific cases, e.g., Gaussian inputs or diagonal channels. The covariance optimization problem is also concave for low snr. Consequently, Theorem 1 represents only a necessary condition to the optimal input covariance, which does not uniquely identify it (note that the MMSE matrix is a function of input covariance in Theorem 1). However, it is possible to compute the global optimal for general snr using an iterative procedure: Initially, we determine the unique globally optimal input covariance for a low enough snr, ensuring the optimization problem concavity. Subsequently, increasing the snr gradually, we determine the optimal covariance for a new higher snr value using the optimal covariance for the lower snr value as the starting point of the optimization algorithm.

It is straightforward to show that the known special cases can also be recovered from Theorem 1:

- *Interfering Channels with Gaussian Inputs*: Let the singular-value decomposition of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal $n \times n$ and $m \times m$ matrices, respectively, and $\mathbf{\Lambda} = \mathsf{diag}(\lambda_i)$. The optimal input covariance is:

$$\bar{\mathbf{\Sigma}}_\mathbf{x}^* = \mathbf{V}\mathsf{diag}(\bar{\sigma}_i^*)\mathbf{V}^T \tag{4}$$

  with $\bar{\sigma}_i^*$ obeying the waterfilling solution [1]:

$$\bar{\sigma}_i^* = \frac{1}{\lambda} - \frac{1}{\mathsf{snr}\cdot\lambda_i^2}, \qquad \lambda \le \mathsf{snr}\cdot\lambda_i^2 \tag{5}$$

$$\bar{\sigma}_i^* = 0, \qquad \lambda > \mathsf{snr}\cdot\lambda_i^2 \tag{6}$$

  with λ such that $\sum_i \bar{\sigma}_i^* = 1$.

- *Noninterfering Channels with Arbitrary Inputs*: Let the channel matrix $\mathbf{H} = \mathsf{diag}(h_i)$. The optimal input covariance is:

$$\mathbf{\Sigma_x^*} = \mathsf{diag}(\sigma_i^*) \tag{7}$$

  with σ_i^* obeying the mercury/waterfilling solution [7]:

$$\sigma_i^* = \frac{1}{\mathsf{snr}\cdot h_i^2}\mathsf{mmse}^{-1}\Big(\frac{\lambda}{\mathsf{snr}\cdot h_i^2}\Big), \qquad \lambda \le \mathsf{snr}\cdot h_i^2 \tag{8}$$

$$\sigma_i^* = 0, \qquad \lambda > \mathsf{snr}\cdot h_i^2 \tag{9}$$

  with λ such that $\sum_i \sigma_i^* = 1$.

The optimal input covariance is given by the fixed-point equation in Theorem 1. However, we can also exploit the gradient of the mutual information with respect to the input covariance [9], which is the basis of Theorem 1, to determine the optimal input covariance iteratively as follows:

$$\mathbf{\Sigma_x^{(k+1)}} = \Big[\mathbf{\Sigma_x^{(k)}} + \mu\mathsf{snr}\mathbf{H}^T\mathbf{H}\mathbf{E}{\mathbf{\Sigma_x^{(k)}}}^{-1}\Big]^+ \tag{10}$$

where μ is a small constant. The operation $\big[\,\cdot\,\big]^+$ denotes the projection onto the feasible set $\mathbf{\Sigma_x^{(k+1)}} \succeq 0$ and $\mathsf{tr}\big(\mathbf{\Sigma_x^{(k+1)}}\big) = 1$.

The projection of a matrix $\mathbf{\Sigma_0}$ onto the feasible set $\mathbf{\Sigma} \succeq 0$ and $\mathsf{tr}\big(\mathbf{\Sigma}\big) = 1$ is the solution to the optimization problem:

$$\min_{\mathbf{\Sigma}} \|\mathbf{\Sigma} - \mathbf{\Sigma_0}\|^2 \tag{11}$$

subject to the constraints $\mathbf{\Sigma} \succeq 0$ and $\mathsf{tr}\big(\mathbf{\Sigma}\big) = 1$. The solution is:

$$\mathbf{\Sigma}^* = \sum_i \max(0, \lambda + \lambda_i)\mathbf{u}_i\mathbf{u}_i^T \tag{12}$$

where $\mathbf{u}_i$ and λ_i are the eigenvectors and the eigenvalues, respectively, of $\mathbf{\Sigma_0} + {\mathbf{\Sigma_0}}^T$, and λ is such that $\mathsf{tr}\big(\mathbf{\Sigma}^*\big) = 1$.

IV. Low-snr Regime

We now consider the structure of the optimal input covariance for MIMO Gaussian channels with arbitrary input distributions in the regime of low snr. We show that in this regime the optimal input covariance for arbitrary inputs is identical to the optimal covariance for Gaussian inputs.

The optimal input covariance can be inferred from the low-snr expansion of the mutual information $I(\mathbf{x};\mathbf{y})$ given by [14]:

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= \frac{\partial I(\mathbf{x};\mathbf{y})}{\partial\mathsf{snr}}\bigg|_{\mathsf{snr}=0}\mathsf{snr} + \frac{\partial^2 I(\mathbf{x};\mathbf{y})}{\partial\mathsf{snr}^2}\bigg|_{\mathsf{snr}=0}\frac{\mathsf{snr}^2}{2} + \mathcal{O}(\mathsf{snr}^3) \\ &= \mathsf{tr}\big(\mathbf{H}\mathbf{\Sigma_x}\mathbf{H}^T\big)\mathsf{snr} + \ddot{I}(0)\frac{\mathsf{snr}^2}{2} + \mathcal{O}(\mathsf{snr}^3) \end{aligned} \tag{13}$$

or from the low-snr expansion of the minimum mean-squared error

$$\mathsf{mmse}(\mathbf{x};\mathbf{y}) = \mathcal{E}\big[\|\mathbf{H}\mathbf{x} - \mathbf{H}\mathcal{E}[\mathbf{x}|\mathbf{y}]\|^2\big] = \mathsf{tr}\big(\mathbf{H}\mathbf{E}\mathbf{H}^T\big) \tag{14}$$

given by:

$$\begin{aligned} \mathsf{mmse}(\mathbf{x};\mathbf{y}) &= \frac{\partial I(\mathbf{x};\mathbf{y})}{\partial\mathsf{snr}}\bigg|_{\mathsf{snr}=0} + \frac{\partial^2 I(\mathbf{x};\mathbf{y})}{\partial\mathsf{snr}^2}\bigg|_{\mathsf{snr}=0}\mathsf{snr} + \mathcal{O}(\mathsf{snr}^2) \\ &= \mathsf{tr}\big(\mathbf{H}\mathbf{\Sigma_x}\mathbf{H}^T\big) + \ddot{I}(0)\mathsf{snr} + \mathcal{O}(\mathsf{snr}^2) \end{aligned} \tag{15}$$

Note that the quantity $\ddot{I}(0)$, which is a function of the input covariance $\mathbf{\Sigma_x}$, is a key low-power performance measure since the bandwidth required to sustain a given rate with a given (low) power is proportional to $-\ddot{I}(0)$ [15]. Equivalently, in view of (14), the optimal input covariance can also be inferred from the low-snr expansion of the MMSE matrix given by:

$$\mathbf{E} = \mathcal{E}\left[\left(\mathbf{x}-\mathcal{E}[\mathbf{x}|\mathbf{y}]\right)\left(\mathbf{x}-\mathcal{E}[\mathbf{x}|\mathbf{y}]\right)^T\right] = \\ = \mathbf{\Sigma_x} - \mathsf{snr}\mathbf{\Sigma_x}\mathbf{H}^T\mathbf{H}\mathbf{\Sigma_x} + \mathcal{O}(\mathsf{snr}^2) \quad (16)$$

Theorem 2: In the low-snr regime, the optimal input covariance for MIMO Gaussian channels with arbitrary inputs is:

$$\mathbf{\Sigma}_\mathbf{x}^*(\mathsf{snr}) = \bar{\mathbf{\Sigma}}_\mathbf{x}^*(\mathsf{snr}) + \mathcal{O}(\mathsf{snr}^2) \quad (17)$$

where $\bar{\mathbf{\Sigma}}_\mathbf{x}^*$ is the optimal input covariance for Gaussian inputs.

Proof: The Theorem follows immediately from Theorem 1 by noting that as $\mathsf{snr} \to 0$ the MMSE matrix for Gaussian inputs:

$$\mathbf{E} = \left(\mathbf{\Sigma_x}^{-1} + \mathsf{snr}\mathbf{H}^T\mathbf{H}\right)^{-1} = \mathbf{\Sigma_x} - \mathsf{snr}\mathbf{\Sigma_x}\mathbf{H}^T\mathbf{H}\mathbf{\Sigma_x} + \mathcal{O}(\mathsf{snr}^2) \quad (18)$$

is identical to the MMSE matrix for arbitrary inputs:

$$\mathbf{E} = \mathbf{\Sigma_x} - \mathsf{snr}\mathbf{\Sigma_x}\mathbf{H}^T\mathbf{H}\mathbf{\Sigma_x} + \mathcal{O}(\mathsf{snr}^2) \quad (19)$$

■

The structure of the optimal input covariance follows directly from the Gaussian result (see (4), (5) and (6)) as $\mathsf{snr} \to 0$. In particular, the optimal input covariance is given by:

$$\mathbf{\Sigma}_\mathbf{x}^* = \mathbf{V}\mathsf{diag}(\sigma_i^*)\mathbf{V}^T \quad (20)$$

with σ_i^* given by:

- When $\mathsf{argmax}_i \lambda_i^2 = k$ is unique or the strongest channel eigenmode is unique, then $\sigma_i^* = 1, i = k$, and $\sigma_i^* = 0, i \neq k$.
- When $\mathsf{argmax}_i \lambda_i^2 = \{k_1, \ldots, k_n\} = \mathcal{K}$ is plural, then $\sigma_i^* = \frac{1}{|\mathcal{K}|}, i \in \mathcal{K}$, and $\sigma_i^* = 0, i \notin \mathcal{K}$

It is interesting to note that in the low-snr regime the input covariance is such that the inputs are fully correlated to better combat the system noise (as long as the strongest channel eigenmode is unique).

V. HIGH-snr REGIME

We now consider the structure of the optimal input covariance for MIMO Gaussian channels with arbitrary discrete input distributions in the regime of high snr. We conclude that it is optimal to have independent and equiprobable inputs with an appropriate fraction of the available power.

We have recently determined upper and lower bounds to the MMSE and the mutual information [12].

Theorem 3: The MMSE for the channel model in (1) is bounded as follows:

$$\frac{1}{M}\frac{e^{-d_{min}^2\mathsf{snr}/4}}{d_{min}\sqrt{snr}}\left(\sqrt{\pi} - \frac{4.37}{d_{min}^2\mathsf{snr}}\right) \leq \mathsf{mmse}(\mathbf{x};\mathbf{y}) \leq \\ \leq (M-1)\frac{e^{-d_{min}^2\mathsf{snr}/4}}{d_{min}\sqrt{snr}}\sqrt{\pi} \quad (21)$$

where M is the number of transmit vectors and d_{min} is the minimum distance between (noiseless) receive vectors[2]:

$$d_{min} = \min_{\substack{\overline{\mathbf{x}},\mathbf{x} \\ \mathbf{x}\neq\overline{\mathbf{x}}}} ||\mathbf{H}\mathbf{x} - \mathbf{H}\overline{\mathbf{x}}|| \quad (22)$$

We use the fact that there is at least a pair of points at minimum distance to prove the lower bound and that at most every pair of points is at minimum distance to prove the upper bound [12]. Upper and lower bounds to the mutual information follow from the MMSE bounds by using the relation between mutual information and MMSE.

Theorem 4: Suppose that the input entropy is finite. Then, the mutual information for the channel model in (1) is bounded as follows:

$$H(\mathbf{x}) - \frac{2(M-1)\sqrt{\pi}}{d_{min}^3\sqrt{snr}}e^{-d_{min}^2\mathsf{snr}/4} \leq I(\mathbf{x};\mathbf{y}) \leq \\ \leq H(\mathbf{x}) - \frac{2e^{-d_{min}^2\mathsf{snr}/4}}{Md_{min}^3\sqrt{snr}}\left(\sqrt{\pi} - \frac{4.37+2\sqrt{\pi}}{d_{min}^2\mathsf{snr}}\right) \quad (23)$$

The proof of Theorem 4 is identical to the proof of Theorem 7 in [12][3]. The optimal input covariance can now be inferred from the upper and lower bounds to the mutual information and the MMSE.

Theorem 5: In the high-snr regime, the optimal input covariance for MIMO Gaussian channels with arbitrary discrete inputs is diagonal with elements obeying the generalized mercury/waterfilling policy.

Proof: We have to maximize both the minimum distance between (noiseless) receive vectors and the input entropy to maximize the mutual information bounds (which are tight for high snr [12]). Maximization of the minimum distance between (noiseless) receive vectors depends only on the input powers, and is independent of the correlation between inputs. On the other hand, maximization of the entropy depends only on the correlation between the inputs, and is independent of the input powers. Consequently, the maximization problems are decoupled:

- Maximization of the entropy is achieved with independent and equiprobable inputs, hence the correlation between the inputs is zero (that is, $\mathbf{\Sigma_x}$ is diagonal).

[2]Note that we take M to be the number of transmit vectors rather than the number of points in a constellation. Note also that we take d_{min} to be the minimum distance between (noiseless) receive vectors rather than the more conventional minimum distance between points in a constellation.

[3]In [12] we use $\log_2 M$ instead of $H(\mathbf{x})$, as the inputs are uncorrelated.

- Maximization of the minimum distance between (noiseless) receive vectors for a diagonal covariance matrix is identical to maximization of the minimum distance between (noiseless) receive vectors for a diagonal real-valued power allocation matrix. Consequently, the diagonal elements of the optimal diagonal covariance matrix obey the generalized mercury/waterfilling power allocation policy for MIMO Gaussian channels with arbitrary discrete inputs [10],[11],[12].

■

VI. Numerical Results

We cast further insight into the structure of the optimal input covariance for MIMO Gaussian channels, by considering a simple 2×2 non-diagonal channel with BPSK inputs. The channel matrix is:

$$\mathbf{H} = \begin{bmatrix} 1 & 0.3 \\ 0.5 & 1 \end{bmatrix} \tag{24}$$

Figure 1 depicts the optimal input covariance whereas Figures 2 and 3 depict the (channel input) covariances resulting from conventional strategies based on channel diagonalization (via singular value decomposition) followed by mercury/waterfilling or by waterfilling power allocation, respectively. The covariances in Figures 2 and 3 are given by $\boldsymbol{\Sigma} = \mathbf{V}\mathsf{diag}(\sigma_i)\mathbf{V}^T$, where σ_i follows the mercury/waterfilling or the waterfilling solution, assuming independent BPSK inputs[4]. It is interesting to note that in the regime of low snr the various strategies are identical. The optimal input covariance assigns all the available power to the strongest channel eigenmode so that the inputs are fully correlated to better combat noise in this noise-limited regime. In the regime of medium to high snr, the optimal input covariance differs from the input covariances based on singular value decomposition followed by mercury/waterfilling or by waterfilling power allocation. In particular, in the regime of high snr the optimal input covariance assigns an appropriate fraction of the available power to the inputs in order to maximize the minimum distance between the (noiseless) receive vectors; additionally, in the regime of high snr the optimal input covariance also fully uncorrelates the inputs. Finally, Figure 4 shows that in the low-snr regime the different strategies yield equal mutual information; in contrast, in the medium- to high-snr regime the optimal input covariance yields higher mutual information than the conventional strategies.

VII. Conclusions

The optimization of the input covariance is a very relevant problem for linear channels with constrained input distributions, e.g. magnetic recording channels. By capitalizing on the relationship between the gradient of the mutual information with respect to system parameters of interest and the MMSE matrix, we have shown that the optimal input covariance obeys a simple fixed-point equation. We have also specialized the structure of the optimal input covariance to the low- and high-snr regimes. For low snr, the optimal input covariance fully correlates the inputs to better combat the noise in this noise-limited regime, just like for Gaussian inputs. For high snr, the optimal input covariance is diagonal with elements obeying the generalized mercury/waterfilling policy. We have also illustrated that covariance optimization may lead to significant gains in relation to conventional strategies based on channel diagonalization followed by mercury/waterfilling or by waterfilling power allocation.

[4]Note that the input covariance resulting from a strategy based on channel diagonalization followed by the waterfilling power allocation is optimal for Gaussian inputs.

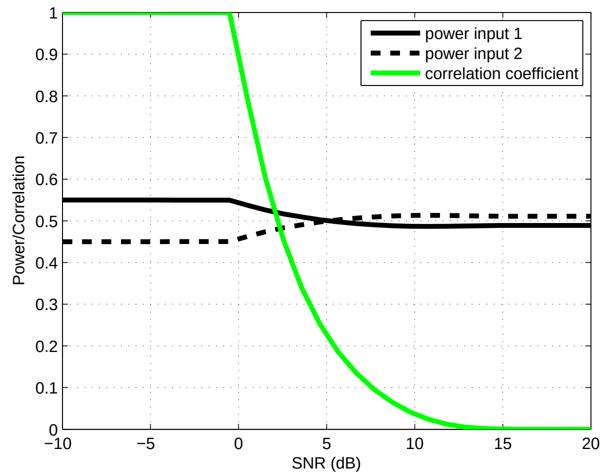

Fig. 1. Optimal input covariance for the non-diagonal channel with BPSK inputs.

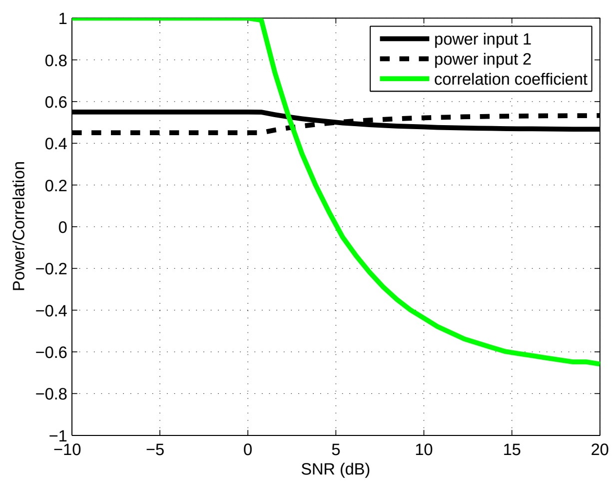

Fig. 2. Channel input covariance resulting from singular value decomposition followed by mercury/waterfilling power allocation.

Appendix A
Proof of Theorem 1

The Karush-Kuhn-Tucker Theorem [13] yields the set of necessary conditions (the KKT conditions or first-order conditions) for the input covariance to be a critical point (maximum, minimum or saddle-point) to the optimization problem. Let us

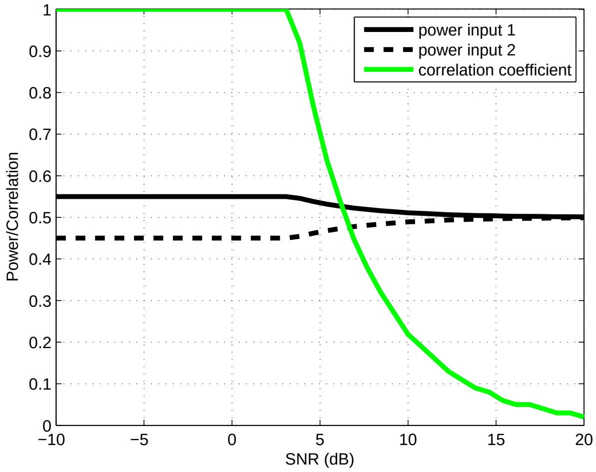

Fig. 3. Channel input covariance resulting from singular value decomposition followed by waterfilling power allocation.

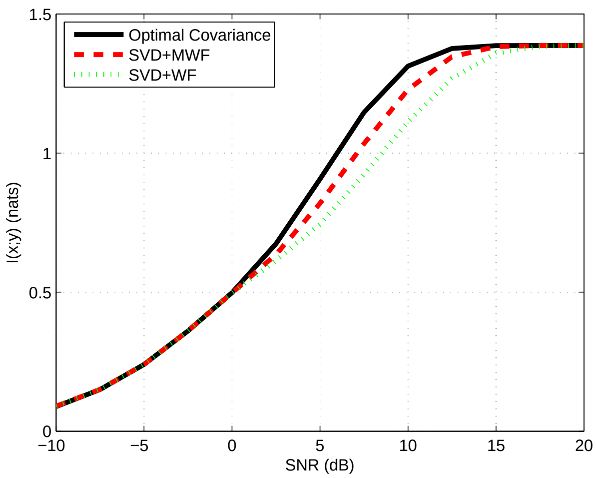

Fig. 4. Mutual information vs. snr for the non-diagonal channel with BPSK inputs.

define the Lagrangian of the optimization problem as follows:

$$\mathfrak{L}\left(\boldsymbol{\Sigma}_{\mathbf{x}}, \boldsymbol{\Psi}, \lambda\right) = -I(\mathbf{x};\mathbf{y}) - \mathsf{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}\right) - \lambda\left(1 - \mathsf{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}}\right)\right) \quad (25)$$

where λ and $\boldsymbol{\Psi}$ are the Lagrange multipliers associated with the problem constraints. The Karush-Kuhn-Tucker conditions state that:

$$\begin{aligned} &\nabla_{\boldsymbol{\Sigma}_{\mathbf{x}}}\mathfrak{L}\left(\boldsymbol{\Sigma}_{\mathbf{x}}, \boldsymbol{\Psi}, \lambda\right)\Big|_{\boldsymbol{\Sigma}_{\mathbf{x}}^*} = -\nabla_{\boldsymbol{\Sigma}_{\mathbf{x}}} I(\mathbf{x};\mathbf{y})\Big|_{\boldsymbol{\Sigma}_{\mathbf{x}}^*} - \boldsymbol{\Psi} + \lambda\mathbf{I} = 0 \\ &\mathsf{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}^*\right) = 0, \boldsymbol{\Psi} \succeq 0, \boldsymbol{\Sigma}_{\mathbf{x}}^* \succeq 0 \\ &\lambda\left(1 - \mathsf{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}}^*\right)\right) = 0, \lambda \geq 0 \end{aligned} \quad (26)$$

or, exploiting the relation between the gradient of the mutual information with respect to the input covariance and the MMSE matrix [9], $\nabla_{\boldsymbol{\Sigma}_{\mathbf{x}}} I(\mathbf{x};\mathbf{y})\boldsymbol{\Sigma}_{\mathbf{x}} = \mathsf{snr}\mathbf{H}^T\mathbf{H}\mathbf{E}$,

$$\begin{aligned} &-\mathsf{snr}\mathbf{H}^T\mathbf{H}\mathbf{E} - \boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}^* + \lambda\boldsymbol{\Sigma}_{\mathbf{x}}^* = 0 \\ &\mathsf{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}^*\right) = 0, \boldsymbol{\Psi} \succeq 0, \boldsymbol{\Sigma}_{\mathbf{x}}^* \succeq 0 \\ &\lambda\left(1 - \mathsf{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}}^*\right)\right) = 0, \lambda \geq 0 \end{aligned} \quad (27)$$

Now, the condition

$$\lambda\boldsymbol{\Sigma}_{\mathbf{x}}^* - \boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}^* = \mathsf{snr}\mathbf{H}^T\mathbf{H}\mathbf{E} \quad (28)$$

is equivalent to the condition

$$\lambda{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Sigma}_{\mathbf{x}}^* - {\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Psi}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2} = \mathsf{snr}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\mathbf{H}^T\mathbf{H}\mathbf{E} \quad (29)$$

where ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}$ is the unique positive semidefinite matrix such that ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2} = \boldsymbol{\Sigma}_{\mathbf{x}}^*$. The matrix ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Psi}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}$ is positive semidefinite because the matrices $\boldsymbol{\Psi}$ and ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}$ are also positive semidefinite, so the condition $\mathsf{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}_{\mathbf{x}}^*\right) = \mathsf{tr}\left({\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Psi}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\right) = 0$ forces the matrix ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Psi}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}$ to have zero diagonal elements and, in turn, zero nondiagonal elements too (that is, the matrix ${\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}\boldsymbol{\Psi}{\boldsymbol{\Sigma}_{\mathbf{x}}^*}^{1/2}$ corresponds to the null matrix). Consequently, the input covariance $\boldsymbol{\Sigma}_{\mathbf{x}}^* \succeq 0$ that solves the optimization problem satisfies:

$$\lambda\boldsymbol{\Sigma}_{\mathbf{x}}^* = \mathsf{snr}\mathbf{H}^T\mathbf{H}\mathbf{E} \quad (30)$$

or

$$\boldsymbol{\Sigma}_{\mathbf{x}}^* = \alpha\mathbf{H}^T\mathbf{H}\mathbf{E} \quad (31)$$

The value of α is chosen to satisfy the unit-trace constraint.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, USA: Wiley, 1991.

[2] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 380–388, 5 1988.

[3] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1527–1539, 11 1991.

[4] S. Shamai and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1388–1404, 9 1996.

[5] D. Arnold, H.-A. Loeliger, P. O.Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, 8 2006.

[6] Z. Zhang, T. M. Duman, and E. M. Kurtas, "Information rates of binary-input intersymbol interference channels with signal-dependent media noise," *IEEE Transactions on Magnetics*, vol. 39, no. 1, pp. 599–607, 1 2003.

[7] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 7 2006.

[8] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 4 2005.

[9] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, 1 2006.

[10] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "Generalized mercury/waterfilling for multiple-input multiple-output channels," in *45th Allerton Conference on Communication, Control, and Computing*, September 2007.

[11] ——, "Optimal precoding for digital subscriber lines," in *IEEE International Conference on Communications*, May 2008.

[12] ——, "Optimal linear precoding for multiple-input multiple-output Gaussian channels with arbitrary inputs," *preprint*, 2008.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[14] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1567–1580, 8 2004.

[15] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory, Special Issue on Shannon Theory: Perspective, Trends and Applications*, vol. 48, no. 6, pp. 1319–1343, 6 2002.