

Generalized Mercury/Waterfilling for Multiple-Input Multiple-Output Channels

Fernando Pérez-Cruz
Department of Electrical Engineering
Engineering Quadrangle
Princeton University
Princeton, New Jersey 08544
Email: fp@princeton.edu

Miguel R. D. Rodrigues
Instituto de Telecomunicações
Computer Science Department
University of Porto
4169 - 007 Porto (Portugal)
Email: mrodrigues@dcc.fc.up.pt

Sergio Verdú
Department of Electrical Engineering
Engineering Quadrangle
Princeton University
Princeton, New Jersey 08544
Email: verdu@princeton.edu

Abstract—We determine the power-allocation policy that maximizes the mutual information for general multiple-input multiple-output Gaussian channels with arbitrary input distributions, by capitalizing on the recent relationship between mutual information and minimum mean squared error (MMSE). In this context, we put forth a novel interpretation of the optimal power-allocation procedure that generalizes the mercury/waterfilling algorithm, previously proposed for parallel non-interfering channels. In this generalization the mercury level accounts for the sub-optimal (non-Gaussian) input distribution and the interferences between inputs.

Index Terms: Optimum Power Allocation, MIMO, Mutual Information, MMSE, Waterfilling

I. INTRODUCTION

Power allocation for mutually interfering channels to maximize mutual information yields the maximal achievable rates by reliable error control codes [1]. Minimizing either the probability of error or the mean squared error are other criteria readily used for addressing this relevant problem [2], [3], [4]. For parallel non-interacting channels corrupted with independent Gaussian noise, the mutual information is maximized, under a total power constraint, by independent Gaussian inputs with power selected according to the waterfilling allocation [1].

Maximization of mutual information has also been considered for general multiple-input multiple-output (MIMO) channels, where the channels are mutually interfering and the noise covariance is arbitrary. The maximum mutual information of this channel is achieved by imposing a covariance structure to the Gaussian input vector that satisfies two conditions. First, we need to transmit using the right eigenvectors decomposition of the channel, which separates the vector channel into a set of parallel non-interfering subchannels [5]. Second, the power allocation follows the waterfilling policy based on the signal-to-noise ratio in each subchannel [1].

Practical structural constraints dictate the use of discrete constellations, such as PSK and QAM, which depart significantly from the optimal Gaussian distributions. Thus, it is of interest to revisit the power allocation problem under the constraint that each input follows a specified discrete constellation. Moreover, even if the inputs are allowed to be Gaussian, it is of interest to optimize the power allocation

when no linear processing of the independent inputs is allowed in order to combat the interference introduced by the channel.

The lack of explicit mutual information expressions, when the inputs are non-Gaussian, can be circumvented using the relation between the mutual information and the minimum mean squared error (MMSE), found in [6] (and generalized in [7]) for Gaussian channels. Reference [8] exploits this relation to obtain the optimal power allocation for parallel non-interacting channels with arbitrary input distributions. The mercury/waterfilling solution of [8] can be interpreted as a generalization of the waterfilling power-allocation policy, in which the base level is modified by the addition of mercury to account for the input constellations. The optimal power-allocation policy for discrete inputs may significantly deviate from the power allocation obtained with Gaussian inputs and with heuristic approaches that take into account the discreteness of the input distribution introducing a power gap [9]–[12].

Yet, no solution exists to date for the optimal power allocation for general MIMO channels with arbitrary inputs. In this contribution, we show the mercury/waterfilling interpretation for optimal power allocation can be extended for interfering channels. In this case, the mercury-level accounts for two factors: the non-Gaussianness of the input distribution and the interference (caused by the other inputs).

The rest of the paper is organized as follows. In Section II we define the problem set up and notation. We derive the optimal power-allocation policy for any channel matrix in Section III, leaving the generalization of the mercury/waterfilling interpretation for Section IV. In Section V we illustrate the mercury/waterfilling reinterpretation for MIMO channels with some representative examples. We conclude the paper with some final remarks and proposed further work in Section VI.

II. SYSTEM SET UP

We consider throughout the paper the following deterministic MIMO channel model:

$$\mathbf{y} = \sqrt{\text{snr}}\mathbf{H}\mathbf{P}\mathbf{x} + \mathbf{w} \quad (1)$$

where the n -dimensional vector \mathbf{y} and the m -dimensional complex random variable \mathbf{x} represent, respectively, the re-

ceived vector and the independent zero-mean unit-variance transmitted information vector. The $n \times m$ complex-valued matrix \mathbf{H} corresponds to the deterministic channel gains (known to both encoder and decoder) and \mathbf{w} is the n -dimensional complex Gaussian noise with independent zero-mean unit-variance components¹. The snr is a scaling factor that accounts for the total transmitted power. The $m \times m$ diagonal real-valued matrix \mathbf{P} , i.e. $(\mathbf{P})_{ij} = \sqrt{p_j} \delta_{ij}$, plays the leading role in the power-allocation strategy, since it has to be optimized to maximize the information transmission rate, subject to a power constraint. We assume that the input distributions are fixed, representing some arbitrary signalling scheme. Thus, in our problem setup in addition to coding of the independent information streams that lead to \mathbf{x} , the only way to combat the channel interference available to the designer is the power allocation.

This optimization problem is directly instilled with operational significance. The maximum information transmission rate or *capacity* of a MIMO channel is determined by the input distribution that maximizes the mutual information, given a power constraint. A multidimensional Gaussian distribution with an appropriate covariance matrix achieves the capacity for these channels. In contrast, we propose to maximize the information transmission rate with a constrained input distribution, as ~~we are limiting the input distribution to be~~ $\mathbf{P}\mathbf{x}$, where the random variable \mathbf{x} is fixed a priori. Clearly, this maximum of the mutual information is less than the channel capacity (equal only if the input distributions are allowed to be Gaussian). However, this situation naturally captures the behavior of many communications systems, since the inputs must be usually drawn from discrete constellations.

Notation: Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and italics denote scalars. The superscript $(\cdot)^\top$ and $(\cdot)^\dagger$ denote transpose and hermitian operations. $(\cdot)^*$ denotes optimum, $\text{Tr}\{\cdot\}$ denotes the trace of a matrix and $\Re\{\cdot\}$ denotes real part of a complex number. $(\mathbf{X})_{ij}$ and x_{ij} denotes the i^{th} element of the j^{th} column of the matrix \mathbf{X} , \mathbf{x}_j denotes the j^{th} column of \mathbf{X} , and x_i denotes the i^{th} element of \mathbf{x} . $\|x\|$, $\|\mathbf{x}\|$ and $\|\mathbf{X}\|$ denote, respectively, the absolute value, L^2 -norm and the Frobenius norm, which is given by $\|\mathbf{X}\| = \sqrt{\text{Tr}\{\mathbf{X}\mathbf{X}^\dagger\}}$. δ_{ij} is the Kronecker's delta function and it is one if $i = j$ and zero otherwise, and $\mathbf{1}$ is a column-vector of ones with the appropriate dimensions.

III. OPTIMAL POWER-ALLOCATION POLICY

The mercury/waterfilling algorithm in [8] sets the optimal power for parallel non-interacting channels (diagonal \mathbf{H} and \mathbf{P} in (1)) with arbitrary inputs. It involves two phases: a mercury-pouring phase followed by a water-pouring one. The mercury-pouring phase amends the base level to account for the mismatch between the actual input and the optimal Gaussian distributions, while the water-pouring phase represents

¹It is also possible to consider non-identity noise covariance matrices, since we could add a whitening filter in the receiver that only affects the definition of the channel matrix.

the power allocation. In [8] the authors capitalized on the relationship between the mutual information and the MMSE [6] to obtain this power-allocation policy, because there are not explicit mutual information expressions for arbitrary input distributions. In this section, we determine the form of the optimal power-allocation policy for MIMO channels with arbitrary input distributions. We later demonstrate this policy for interfering channels represents an extension of the mercury/waterfilling algorithm. In this extension the mercury level accounts for both the mismatch between the actual input and an optimal Gaussian distribution, as well as, the interference between inputs.

The power-allocation problem can be cast as a linearly-constrained nonlinear optimization problem:

$$\max_{p_j} \{I(\mathbf{x}; \mathbf{y})\} \quad (2)$$

subject to:

$$\sum_{j=1}^m p_j \leq 1 \quad (3)$$

$$p_j \geq 0 \quad \forall j = 1, \dots, m \quad (4)$$

Theorem 1: The optimal power allocation $\{p_j^*\}_{j=1}^m$ that solves (2) subject to (3) and (4) satisfies:

$$p_j^* = \lambda^{-1} (\text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} \quad \forall j = 1, \dots, m \quad (5)$$

and

$$\lambda = \text{Tr} \{ \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \} \quad (6)$$

to ensure the p_j^* add up to 1, where

$$(\mathbf{E})_{ij} = E [(x_i - E[x_i|\mathbf{y}])^\dagger (x_j - E[x_j|\mathbf{y}])], \quad (7)$$

which is known as the MMSE matrix [13].

Proof: See Appendix A ■

If \mathbf{H} is diagonal, so it is \mathbf{E} , and (5) can be simplified to:

$$p_j^* = \lambda^{-1} p_j^* \text{snr} \|h_{jj}\|^2 E [\|x_j - E[x_j|\mathbf{y}]\|^2] \quad (8)$$

rearranging (8), we can express the optimal power allocation as follows:

$$p_j^* = \begin{cases} \frac{1}{\text{snr} \|h_{jj}\|^2} \text{mmse}_j^{-1} \left(\frac{\lambda}{\text{snr} \|h_{jj}\|^2} \right), & \lambda \leq \text{snr} \|h_{jj}\|^2 \\ 0, & \lambda > \text{snr} \|h_{jj}\|^2 \end{cases} \quad (9)$$

where

$$E [\|x_j - E[x_j|\mathbf{y}]\|^2] = \text{mmse}_j(\text{snr} \|h_{jj}\|^2 p_j^*) \quad (10)$$

The optimal power allocation in (9) is identical to Theorem 2 in [8] for parallel non-interfering channels.

The proof of Theorem 1 relies on the KKT conditions and the relation between the mutual information and the MMSE [6], [7]. The KKT conditions are satisfied by any critical point (minimum, maximum or saddle point). Only when the problem is strictly concave a unique $\{p_j^*\}_{j=1}^m$ satisfies the KKT conditions and they provide the global maximum. In general, the power allocation as stated in (2)-(4) is nonconcave. It

becomes concave in some specific cases, e.g. for Gaussian input distributions with arbitrary channel matrices [14]-[16] or diagonal channel matrices with arbitrary input distributions [8]. The problem is also concave for low snr, but is in general nonconcave for high snr. Thus, it is important to realize that Theorem 1 gives a necessary condition for the optimum power allocation, but in general, it does not identify uniquely the optimum power allocation.

Consequently, to compute a global optimum for any snr we can first obtain the unique solution for a low enough snr to ensure the problem is concave. Then, we increase the snr by small amounts until the desired snr is achieved. For each increased snr value, a new power allocation is obtained, using as starting point the power allocation for the previous snr. This optimization method is typically used for solving highly nonconcave optimization problems. We could also choose several starting points, find a local maxima, and keep the power allocation matrix with largest mutual information.

Solution (5) does not provide much insight about how the power is allocated and, for arbitrary distributions, we must allocate the power using numerical procedures. In the following section, we provide an interpretation for this optimal power-allocation policy, extending the mercury/waterfilling algorithm put forward in [8] for non-interfering channels. The proposed interpretation ~~is based on the first order conditions (see Appendix A), thereby it holds for any local minima that fulfils (5).~~ The interpretation would give us the optimal power allocation when we have obtained the global optimum.

IV. MERCURY/WATERFILLING REINTERPRETATION

The power allocation for general channel matrices can be understood as an extension of the mercury/waterfilling for non-interfering channels in [8]. In this case, the mercury level accounts for two factors: the non-optimal input distributions and the interferences between inputs. The water and base levels are defined as they were for the non-interfering case and the mercury fills the gap between the base level and the water level to ensure each input's power is correctly allocated.

A. Definitions

Definition 1: The *water level* is $1/\lambda$, the inverse of the Lagrange multiplier associated with the inequality constraint in (3).

Definition 2: The *base level* associated with the j^{th} input is given by

$$b_j = \frac{1}{\left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\substack{p_j = 1 \\ x_j \sim \mathcal{N}(0, 1)}}} - 1 = \frac{1}{\text{snr} \|\mathbf{h}_j\|^2} \quad (11)$$

These definitions correspond to the water and base levels for waterfilling [1] and mercury/waterfilling [8] for non-interfering channels. The base level definition ensures the difference between the water and the base levels is 1, if the j^{th} input uses a Gaussian distribution and is assigned all the available power.

Definition 3: The height of the *mercury level* associated with the j^{th} input is given by

$$m_j = \frac{1}{\left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\mathbf{P}=\mathbf{P}^*}} - p_j^* \quad (12)$$

For $p_j^* > 0$

$$\left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\mathbf{P}=\mathbf{P}^*} = \frac{\text{snr}}{\sqrt{p_j^*}} (\mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} = \lambda^{-1} \quad (13)$$

and $p_j^* = \lambda^{-1} - m_j$.

For $p_j^* = 0$, the derivative of the mutual information is not given by (13), because it is an indeterminate form. The value of the mercury level for $p_j^* = 0$ is computed in Appendix B and it is shown to be above the water level. The mercury level for $p_j^* = 0$ is defined to ensure it is a continuous function of the snr. If $p_j^* = 0$, we would not need to defined the mercury level as the j^{th} input is not assigned any power. However to reach a meaningful reinterpretation for the mercury/waterfilling algorithm, we define the mercury level as in (12) and Appendix B, thereby it accounts for the non-optimal inputs distribution as well as the interferences caused by the other inputs.

In the special case of Gaussian inputs we can add the following results:

Theorem 2: The partial derivative of $m_j - b_j$ with respect to p_i is nonnegative and with respect to snr is nonpositive for Gaussian inputs with any channel matrix \mathbf{H} .

Proof: See Appendix C ■

Theorem 3: The available power is equally divided between all inputs, as the snr tends to infinity, if the inputs are independent Gaussian distributions and $\mathbf{H}\mathbf{H}^\dagger$ is invertible.

Proof: See Appendix D ■

Theorems 2 and 3 allow for a new interpretation of the mercury/waterfilling for Gaussian inputs in which the mercury and water are simultaneously poured, as we detail in Section IV-B.

B. Mercury/Waterfilling Illustration

In the mercury/waterfilling interpretation for non-interfering channels in [8], the mercury level for the j^{th} channel directly depends on $\text{snr} \|h_{jj}\|^2$, through the (tabulated) inverse MMSE, and the water level². The water level is numerically calculated ensuring that $\{p_j^*\}_{j=1}^m$ add up to 1. For non-diagonal \mathbf{H} matrices, we solve the optimization problem in (2)-(4), jointly obtaining the power allocation (primal variables) and the water level (Lagrange multiplier). Thus the generalized mercury/waterfilling interpretation facilitates understanding how the power is divided between the different inputs, but as in [8], it is not a procedure for setting the optimal power allocation.

Figure 1 depicts the generalized mercury/waterfilling reinterpretation for MIMO channels. If $m_j \geq \lambda^{-1}$ the j^{th} input

²The water level depends on every $\text{snr} \|h_{ii}\|^2$, hence indirectly the mercury level depends on ~~every inputs~~ channel response. u

is not assigned power. If $m_j < \lambda^{-1}$, then $p_j^* = \lambda^{-1} - m_j$. The mercury level for $p_j^* = 0$ is the main difference between the mercury/waterfilling for mutually interfering channels and the one in [8] for non-interfering channels. For non-interfering channels the mercury level is zero, if $p_j^* = 0$. For interfering channels the mercury level can be nonzero, if $p_j^* = 0$. This added mercury accounts for the interference the others cause on the j^{th} input. This also means that the mercury level can be above the water level, as depicted in Figure 1 for channels 3 and 4, which would not be possible in the mercury/waterfilling interpretation for non-interfering channels in [8].

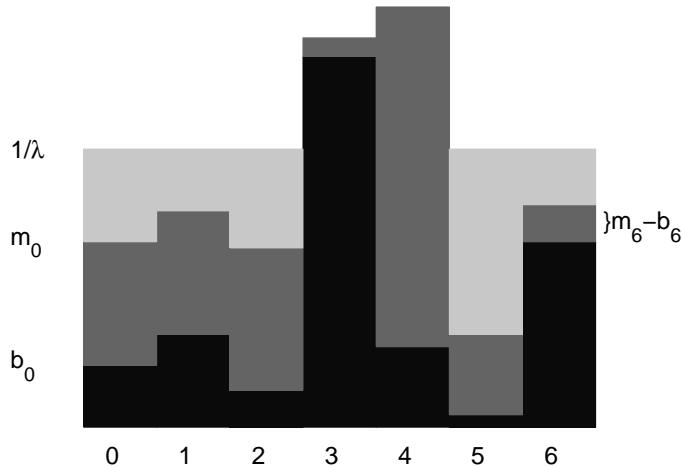


Fig. 1. Illustration of the mercury/waterfilling power allocation for a seven-input MIMO channel. The power for each input is the difference between the water and mercury level. The water is the light grey shaded area, the mercury is the dark grey shaded area and the base level is the black shaded area. The mercury in channels 3 and 4 are only due to the interferences caused by the other inputs, because their allocated power is zero.

The base level can be understood as a barrier for an input to be assigned power. The mercury/waterfilling interpretation introduces another barrier, namely the mercury level. In non-interfering channels, the mercury level accounts for the mismatch between the arbitrary input and the optimal Gaussian distributions. This barrier only kicks in when an input is assigned power. Loosely speaking, the mercury level measures what we are losing for using a given input instead of an optimal Gaussian distribution. This barrier depends on the used distribution, through the inverse of the MMSE, and all the other inputs channel responses and distributions, through the water level. Hence power gap approaches [9]-[12] typically used to amend the waterfilling from Gaussian inputs clearly fail to assign the correct power to each input, because they do not take into account the effect of the other inputs to define this power gap.

For interfering channels, the mercury level additionally accounts for the interference. If the inputs are Gaussian, the barrier ~~an inputs faces~~ is given by the base level (channel's strength versus noise) and the mercury level (channel's strength versus interferences). In fact, the nature of the mercury level is the same as the base level, as the interferences are Gaussian. The base and mercury levels are the barriers the

noise and the other inputs impose on each specific input to be allocated power. In this case, the mercury accounts for the interference the various inputs cause on one another. There is no added mercury for an input that gets all the available power, since this input cannot be interfered by others.

We can put forth the following mercury/waterfilling intuition for Gaussian inputs over interfering channels. Instead of pouring mercury to amend the base level and then water to obtain the power allocation, we pour both mercury and water little by little. We first pour a bit of water, which settles in the input with the strongest channel (largest $\|\mathbf{h}_j\|$). This water corresponds to a bit of mercury in the other inputs, due to the interference cause by the strongest channel. We keep on adding water and the corresponding mercury progressively, until there is no more water to add. The water ~~column~~ is the power allocated to each channel. This intuition is based on the mercury level for each input being an increasing function of the power of the remaining inputs, as proven in Theorem 2.

For non-Gaussian inputs, the mercury level accounts for both the interference and the input distribution. However, it is not possible to decompose the mercury in two different levels, one for the interference and another for the input distribution, since how the interference affects each input depends on all inputs' distributions (interferer and interferee).

For a power allocation where input j is assigned all available power ($p_j^* = 1$) and the other inputs do not transmit ($p_i^* = 0 \quad \forall i \neq j$), the nonzero mercury level for input j ($m_j > b_j$) only accounts for the mismatch between the input and Gaussian distributions, because there is no interference from the other inputs. In contrast, the nonzero mercury level for all the other inputs ($m_i > b_i \quad \forall i \neq j$) only accounts for the interference caused by the j^{th} input.

For any other power allocation, the mercury level accounts for both the interference and the input distribution mismatch. For discrete input distributions, the interference vanishes as the snr increases, so the mercury level only depends on the input distributions not being Gaussian. The mercury level is not in general an increasing function of the other inputs power, so it is not possible to put forward the operational description of the mercury/waterfilling interpretation for Gaussian inputs.

V. EXAMPLES

We now cast further insight into the mercury/waterfilling interpretation for MIMO channels by analyzing four cases of interest: diagonal channel matrix with Gaussian inputs in V-A; diagonal channel matrix with BPSK inputs in V-B; non-diagonal channel matrix with Gaussian inputs in V-C; and non-diagonal channel matrix with BPSK inputs in V-D. We consider real-valued communications systems with the following channel matrices:

$$\mathbf{H} = \begin{bmatrix} \sqrt{1.09} & 0 \\ 0 & \sqrt{0.34} \end{bmatrix} \quad \text{and} \quad \mathbf{H} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 0.5 \end{bmatrix} \quad (14)$$

The first channel model allows recovering the standard waterfilling or the mercury/waterfilling interpretations and is used for comparison purposes. The second model allows

drawing conclusions into the operation of the general mercury/waterfilling interpretation for interfering channels.

A. Gaussian inputs and diagonal channel matrix

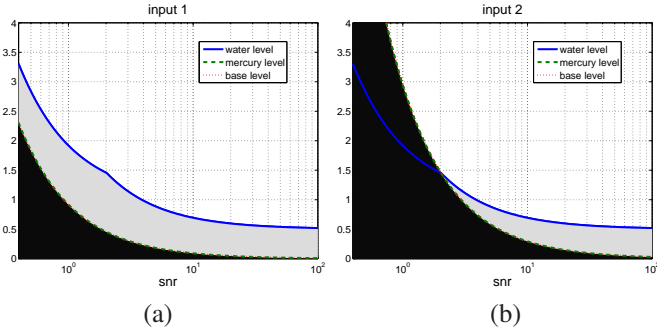


Fig. 2. Power allocation for input 1 in (a) and input 2 in (b) for Gaussian inputs and diagonal channel matrix. The power for each input is the difference between the water and mercury/base level. The water is the light grey shaded area.

In Figure 2, we show the base (=mercury) and water levels for the diagonal channel model with Gaussian inputs as a function of the snr. In these pictures, we can see the standard waterfilling interpretation, in which all the power is assigned to the strongest input (largest $\|h_{jj}\|$) for low snr, and it is evenly split between the inputs for high snr.

B. Non-Gaussian inputs and diagonal channel matrix

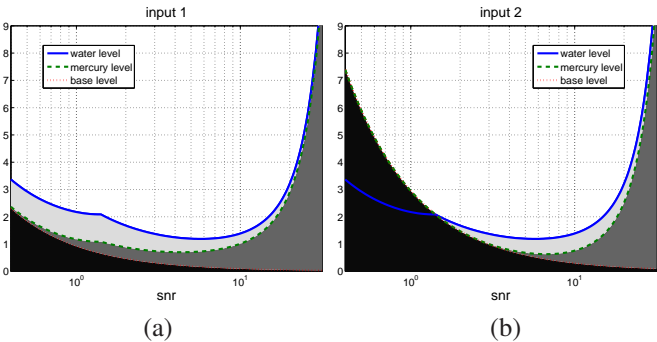


Fig. 3. Power allocation for input 1 in (a) and input 2 in (b) for BPSK inputs and diagonal channel matrix. The power for each input is the difference between the water and mercury level. The water is the light grey shaded area and the mercury is the dark grey shaded area.

In Figure 3, we show the base, mercury and water levels for the diagonal channel model with BPSK inputs. For the input with highest channel value (first input) we always add mercury to the base level. This mercury compensates for the mismatch between the input distribution and the optimal Gaussian distribution. It is interesting to note that the mercury increases with the snr and it tends to the base level as the snr tends to zero. For the input with the lowest channel value (second input), we only add mercury to the base level once this input is allocated some power. In fact, for zero power the input distribution could be Gaussian, and we should not add mercury to the base level. For discrete input distributions, the mercury

grows exponentially with snr as proven in [8], and so does the water level. For high snr the mercury level (mismatched between Gaussian and the input distribution) determines the power allocation, because the base level is (almost) zero for every input.

C. Gaussian inputs and non-diagonal channel matrix

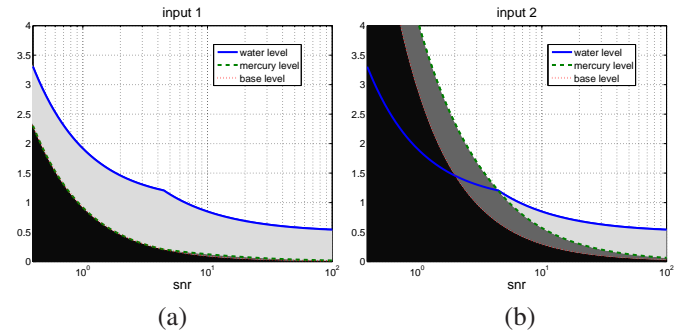


Fig. 4. Power allocation for input 1 in (a) and input 2 in (b) for Gaussian inputs and non-diagonal channel matrix. The power for each input is the difference between the water and mercury level. The water is the light grey shaded area and the mercury is the dark grey shaded area.

In Figure 4, we show the base, mercury and water levels for the interfering channel model with Gaussian inputs. For low snr values, the strongest input (input 1) gets all the power and therefore it is not interfered by the second input, consequently, there is no added mercury. The weakest input (input 2) is not assigned any power until the snr increase is able to compensate not only the noise but also the interference by the first input.

For high snr values, the second input is assigned some power, as it is able to overcome both the noise and the interference. The mercury level for the first input start rising once the second input is assigned some power, due to interference it causes. Although it is barely noticeable in the figure, because the mercury tends to zero when the snr tends to infinity for a full-rank $\mathbf{H}\mathbf{H}^\dagger$, as predicted by Theorem 3. This also guarantees that the power is equally divided between inputs as $\text{snr} \rightarrow \infty$.

D. Non-Gaussian inputs and non-diagonal channel matrix

In Figure 5, we show the base, mercury and water levels for the interfering channel model with BPSK inputs. For low snr, the first input is assigned all the power. The mercury level for input 1 depends only on the distribution mismatch, since there is not any interference from the second input. For low snr, the base, mercury and water levels of the first input are identical to the levels of input 1 in Example V-B. The mercury level for input 2 is only due to interferences, since input 2 is not assigned any power it cannot be penalized for its distribution. This mercury level is not identical to the mercury level in the previous case (Gaussian input and non-diagonal channel matrix) because the interference comes from a different input distribution. We can understand in this simple example why the mercury cannot be divided in two different types.

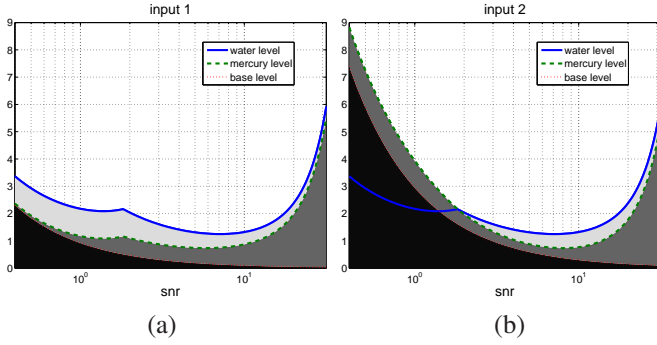


Fig. 5. Power allocation for input 1 in (a) and input 2 in (b) for BPSK inputs and non-diagonal channel matrix. The power for each input is the difference between the water and mercury level. The water is the light grey shaded area and the mercury is the dark grey shaded area.

For high snr, the mercury level is mainly due to the input distribution mismatch, because for discrete input distributions the interferences have (almost) vanished. But it is not equal to example V-B because the mercury level depends on \mathbf{H} , not on the norm of its columns.

E. Discussion

In Figures 2-5, we can see that the base levels for input 1 and 2 are identical, as $\|\mathbf{h}_1\| = \sqrt{1.09}$ and $\|\mathbf{h}_2\| = \sqrt{0.34}$ for both channel matrices in (14). For low snr, input 1 is assigned all the power. It can be seen that in this case there is no mercury in Examples V-A and V-C, because we are working with Gaussian inputs and in V-C there are no interferences. In Examples V-B and V-D the mercury added is the same one, because it is only due to distribution mismatch. For input 2 in Examples V-C and V-D, we add some mercury to account for the interferences and there is no mercury in Examples V-A and V-B because there are not any interferences or distribution mismatched. The mercury in V-C and V-D are not identical because the interferee uses a different input distribution.

For high snr, the examples with Gaussian inputs share equally the available power as the base and mercury levels vanishes, because $\mathbf{H}\mathbf{H}^\dagger$ is full rank. For Examples V-B and V-D the mercury level only depends on the inputs distributions, as there are no interferences, but the mercury levels are different because the channel matrices are different and the mercury level depends on the channel response to every input.

VI. CONCLUSIONS AND FURTHER WORK

A natural extension of this work reported in [17] is to drop the structural constraint that no linear processing of the independent input vector is allowed. In such a case, the problem of optimum power allocation generalizes to one of optimal precoding where the matrix \mathbf{P} is no longer constrained to be diagonal. Interestingly, for non-Gaussian inputs the maximum information transmission rate is not achieved by diagonal \mathbf{P} , even for diagonal channel matrices. The optimization of the precoding matrix \mathbf{P} can significantly increase the rates achieve for practical structures using discrete inputs, thereby yielding **higher** information transmission rates in

many practical scenarios, ranging from DSL to multi-antenna communications.

ACKNOWLEDGEMENT

Fernando Pérez-Cruz is supported by AI-COM Marie Curie Fellowship from the 6th Research Framework Programme of the European Union. Fernando Pérez-Cruz is also a faculty member at the Signal Theory and Communications Department in ~~Carlos III University~~ (Spain).

APPENDIX A

PROOF OF THEOREM 1

The possible solutions to (2) subject to (3) and (4) are characterized by the Karush-Kuhn-Tucker theorem [18], which states the necessary conditions for the set $\{p_j^*\}_{j=1}^m$ to be a critical point, known as the KKT or first order conditions. To compute the KKT conditions we first need to build the Lagrangian for our constrained optimization problem:

$$\mathcal{L}(\mathbf{p}, \lambda, \boldsymbol{\eta}) = -I(\mathbf{x}; \mathbf{y}) + \lambda \left(\sum_{j=1}^m p_j - 1 \right) - \sum_{j=1}^m \eta_j p_j \quad (15)$$

in which the Lagrange multipliers λ and η_j for the inequality constraints have to be nonnegative.

The KKT conditions are given by:

$$\frac{\partial \mathcal{L}(\mathbf{p}, \lambda, \boldsymbol{\eta})}{\partial p_j} = -\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} + \lambda - \eta_j = 0 \quad (16)$$

$$\lambda \left(\sum_{j=1}^m p_j - 1 \right) = 0 \quad (17)$$

$$p_j \eta_j = 0 \quad (18)$$

$$\lambda, \eta_j \geq 0 \quad (19)$$

The derivative of $I(\mathbf{x}; \mathbf{y})$ with respect to p_j is given by:

$$\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} = \frac{\text{snr}}{\sqrt{p_j}} (\mathbf{H}^\dagger \mathbf{H} \mathbf{P} \mathbf{E})_{jj}, \quad (20)$$

which can be computed from Equation 21 in [7] and the chain rule.

For $p_j^* > 0$:

$$\lambda = \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \Big|_{\mathbf{P}=\mathbf{P}^*} = \frac{\text{snr}}{\sqrt{p_j^*}} (\mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} \geq 0, \quad (21)$$

because $\eta_j = 0$ for nonzero p_j^* . λ is nonzero because $\mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E}$ is positive semi-definite. For $p_j^* = 0$:

$$(\mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} = 0, \quad (22)$$

The optimal solution can be expressed as:

$$p_j^* = \frac{\text{snr}}{\lambda} \sqrt{p_j^*} (\mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} = \lambda^{-1} (\text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E})_{jj} \quad (23)$$

where we have combined (21) and (22) to obtain a single equation for all p_j^* . $\lambda = \text{Tr} \{ \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \}$ to ensure that the $\{p_j^*\}_{j=1}^m$ add up to 1.

APPENDIX B
MERCURY LEVEL FOR $p_j^* = 0$

For $p_j^* = 0$, (13) is an indeterminate form and we cannot compute the mercury level from it. We can rewrite the mercury level as follows:

$$m_j^{-1} = \left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\mathbf{P}=\mathbf{P}^*} = \text{snr} \|\mathbf{h}_j\|^2 - \text{snr} \sum_{i \neq j} \mathbf{h}_j^\dagger \mathbf{h}_i \sqrt{p_i^*} E \left[\left(\frac{E[x_j | \mathbf{y}]}{\sqrt{p_j^*}} \right)^\dagger E[x_i | \mathbf{y}] \right] \quad (24)$$

and

$$\begin{aligned} \frac{E[x_j | \mathbf{y}]}{\sqrt{p_j^*}} &= \int \frac{x_j}{\sqrt{p_j^*}} \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} d\mathbf{x} \\ &= \frac{\sqrt{\text{snr}}}{\sqrt{(2\pi)^n}} \int \exp\left(-\frac{\|\mathbf{a}\|^2}{2}\right) \Re\{\mathbf{a}^\dagger \mathbf{h}_j x_j\} x_j \frac{p(\mathbf{x})}{p(\mathbf{y})} d\mathbf{x} \end{aligned} \quad (25)$$

where

$$\begin{aligned} p(\mathbf{y} | \mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2} \|\mathbf{y} - \sqrt{\text{snr}} \mathbf{H} \mathbf{P}^* \mathbf{x}\|^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2} \|\mathbf{a} - \sqrt{\text{snr}} \mathbf{h}_j \sqrt{p_j^*} x_j\|^2\right) \\ &= \frac{\exp\left(-\frac{\|\mathbf{a}\|^2}{2}\right)}{\sqrt{(2\pi)^n}} \left(1 + \sqrt{\text{snr}} p_j^* \Re\{\mathbf{a}^\dagger \mathbf{h}_j x_j\}\right) + \mathcal{O}(p_j^*) \end{aligned} \quad (26)$$

and

$$\mathbf{a} = \mathbf{y} - \sqrt{\text{snr}} \mathbf{H} \mathbf{P}^* \mathbf{x} + \sqrt{\text{snr}} \mathbf{h}_j \sqrt{p_j^*} x_j, \quad (27)$$

To obtain the last equality in (25), we can see that $\mathcal{O}(p_j^*)/\sqrt{p_j^*} = 0$ and that for any matrix \mathbf{P}

$$\frac{1}{\sqrt{(2\pi)^n}} \int \exp\left(-\frac{\|\mathbf{a}\|^2}{2}\right) x_j \frac{p(\mathbf{x})}{p(\mathbf{y})} d\mathbf{x} = 0, \quad (28)$$

because x_j is zero-mean and \mathbf{a} does not depend on x_j .

We can show that the mercury level is above the water level if $p_j^* = 0$, using the first order conditions in Appendix A. For $p_j^* = 0$, (16) becomes

$$\eta_j = \lambda - \left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\mathbf{P}=\mathbf{P}^*} \geq 0, \quad (29)$$

hence

$$m_j \geq \lambda^{-1} \quad (30)$$

APPENDIX C
PROOF OF THEOREM 2

We first prove that the derivative of $m_j - b_j$ with respect to snr is ~~less than zero~~ for Gaussian inputs. First, we rearrange

$m_j - b_j$ as follows:

$$\begin{aligned} m_j - b_j &= \frac{1}{\text{snr} \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j} - p_j - \frac{1}{\text{snr} \|\mathbf{h}_j\|^2} \\ &= \frac{\|\mathbf{h}_j\|^2 - \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j}{\text{snr} \|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j} - p_j \\ &= \frac{\mathbf{h}_j^\dagger \mathbf{H} \mathbf{P}^* (\mathbf{I} + \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^*)^{-1} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{h}_j}{\|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j} - p_j \\ &= \frac{\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j}{\|\mathbf{h}_j\|^2 \left(\|\mathbf{h}_j\|^2 - \text{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)} - p_j \end{aligned} \quad (31)$$

where we have used the matrix inversion lemma to obtain the last two equalities and have defined:

$$\mathbf{M}_1 = \mathbf{H} \mathbf{P}^* (\mathbf{I} + \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^*)^{-1} \mathbf{P}^* \mathbf{H}^\dagger \quad (32)$$

The derivative of $m_j - b_j$ is given by:

$$\begin{aligned} \frac{\partial(m_j - b_j)}{\partial \text{snr}} &= \frac{\mathbf{h}_j^\dagger \frac{\partial \mathbf{M}_1}{\partial \text{snr}} \mathbf{h}_j \left(\|\mathbf{h}_j\|^2 - \text{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)}{\|\mathbf{h}_j\|^2 \left(\|\mathbf{h}_j\|^2 - \text{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)^2} \\ &\quad + \frac{\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \left[\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j + \text{snr} \mathbf{h}_j^\dagger \frac{\partial \mathbf{M}_1}{\partial \text{snr}} \mathbf{h}_j\right]}{\|\mathbf{h}_j\|^2 \left(\|\mathbf{h}_j\|^2 - \text{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)^2} \\ &= -\frac{\|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger \mathbf{M}_1^2 \mathbf{h}_j - \left(\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)^2}{\|\mathbf{h}_j\|^2 \left(\|\mathbf{h}_j\|^2 - \text{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)^2} \end{aligned} \quad (33)$$

where the derivative of \mathbf{M}_1 with respect to snr is given by:

$$\begin{aligned} \frac{\partial \mathbf{M}_1}{\partial \text{snr}} &= -\mathbf{H} \mathbf{P}^* (\mathbf{I} + \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^*)^{-1} \mathbf{P}^* \mathbf{H}^\dagger \times \\ &\quad \mathbf{H} \mathbf{P}^* (\mathbf{I} + \text{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^*)^{-1} \mathbf{P}^* \mathbf{H}^\dagger = -\mathbf{M}_1^2 \end{aligned} \quad (34)$$

We can see that

$$\begin{aligned} \left(\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j\right)^2 &= \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \\ &= \text{Tr} \left\{ \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right\} \\ &= \text{Tr} \left\{ \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \right\} \\ &\leq \text{Tr} \left\{ \mathbf{h}_j \mathbf{h}_j^\dagger \right\} \text{Tr} \left\{ \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \right\} \\ &= \|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger \mathbf{M}_1^2 \mathbf{h}_j \end{aligned} \quad (35)$$

where we have used the following property for positive semi-definite matrices: $\text{Tr}\{\mathbf{A}\mathbf{B}\} \leq \text{Tr}\{\mathbf{A}\}\text{Tr}\{\mathbf{B}\}$ [19], which concludes the proof that $m_j - b_j$ decreases with the snr.

We use the first equality in (31) to compute its derivative

with respect to p_i :

$$\begin{aligned} \frac{\partial(m_j - b_j)}{\partial p_i} &= -\frac{\text{snr} \mathbf{h}_j^\dagger \frac{\partial(\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1}}{\partial p_i} \mathbf{h}_j}{\left(\text{snr} \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j\right)^2} - \delta_{ij} \\ &= \frac{\mathbf{h}_j^\dagger \left[(\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_i \mathbf{h}_i^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \right] \mathbf{h}_j}{\left(\mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j\right)^2} - \delta_{ij} \\ &= \frac{\left\| \mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_i \right\|^2}{\left(\mathbf{h}_j^\dagger (\mathbf{I} + \text{snr} \mathbf{H} \mathbf{P}^* \mathbf{H}^\dagger)^{-1} \mathbf{h}_j\right)^2} - \delta_{ij} \quad (36) \end{aligned}$$

This derivative is nonnegative for $i \neq j$ and zero for $i = j$.

APPENDIX D PROOF OF THEOREM 3

The derivative of the mutual information with respect to p_j for Gaussian inputs is given by:

$$\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} = \mathbf{h}_j^\dagger (\mathbf{H} \mathbf{P}^2 \mathbf{H}^\dagger + \mathbf{I} / \text{snr})^{-1} \mathbf{h}_j \quad (37)$$

Let's assume that all the samples are assigned the same power ($p_j = 1/m$, $\forall p_j$), then if $\mathbf{H} \mathbf{H}^\dagger$ is full-rank:

$$\lim_{\text{snr} \rightarrow \infty} \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} = m \mathbf{h}_j^\dagger (\mathbf{H} \mathbf{H}^\dagger)^{-1} \mathbf{h}_j = m \quad \forall j \quad (38)$$

This means that $1/\lambda$ tends to $1/m$ and the mercury level tends to zero. Therefore, equal power allocation for ~~infinity~~ ∞ snr fulfills the conditions of Theorem 1.

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, USA: Wiley, 1991.
- [2] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2004.
- [3] S. Serbetli and A. Yener, "MMSE transmitter design for correlated MIMO systems with imperfect channel estimates: power allocation trade-offs," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2295–2304, 2006.
- [4] K. H. Bizaki and A. Falahati, "Power loading by minimizing the average symbol error rate on MIMO THP systems," in *9th International Conference on Advanced Communication Technology*, 2007.
- [5] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, 1998.
- [6] 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, 2005.
- [7] 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, 2006.
- [8] 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, 2006.
- [9] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, p. 2384–2415, 1998.
- [10] J. A. C. Bingham, P. S. Chow, and J. M. Cioffi, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, p. 773–775, 1995.

- [11] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Transactions on Communications*, vol. 48, no. 1, p. 23–27, 2000.
- [12] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Prentice-Hall, 1999.
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [14] R. G. Gallager, *Information Theory and Reliable Communication*. New York, USA: Wiley, 1968.
- [15] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *European Transactions on Communications*, vol. 10, no. 6, pp. 585–595, 1999.
- [16] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.
- [17] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "Optimal precoding for multiple-input multiple-output Gaussian channels with arbitrary inputs," *In preparation*, 2007.
- [18] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.
- [19] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, USA: Academic Press, 1979.