PILOT DESIGN FOR MIMO CHANNEL ESTIMATION: AN ALTERNATIVE TO THE KRONECKER STRUCTURE ASSUMPTION

John Flåm[⋆]*, Emil Björnson*[†‡] *and Saikat Chatterjee*[‡]

[⋆]Department of Electronics and Telecommunications, NTNU,Trondheim, Norway
[†]Alcatel-Lucent Chair on Flexible Radio, SUPELEC, Gif-sur-Yvette, France
[‡]School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

This work seeks to design a pilot signal, under a power constraint, such that the channel can be estimated with minimum mean square error. The procedure we derive does not assume Kronecker structure on the underlying covariance matrices, and the pilot signal is obtained in three main steps. Firstly, we solve a relaxed convex version of the original minimization problem. Secondly, its solution is projected onto the feasible set. Thirdly we use the projected solution as starting point for an augmented Lagrangian method. Numerical experiments indicate that this procedure may produce pilot signals that are far better than those obtained under the Kronecker structure assumption.

1. PROBLEM STATEMENT

Consider the following multiple-input-multiple-output (MIMO) communication model

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{w}. \tag{1}$$

Here $\mathbf{z}$ is the observed output, $\mathbf{w}$ is random noise, $\mathbf{H}$ is a random channel matrix that we wish to estimate, and $\mathbf{s}$ is a pilot vector to be designed for that purpose. In order to estimate $\mathbf{H}$ with some confidence, we should typically send at least as many pilot vectors at there are columns in $\mathbf{H}$, although this is not strictly necessary when the columns are strongly correlated [1]. In order to utilize the channel estimate for subsequent data transmission, we also assume that the time needed to transmit the pilots is only a fraction of the coherence time. This assumption typically holds in flat, block-fading MIMO systems [2, 3, 1]. With p transmitted pilots, model (1) can be written in matrix form as

$$\mathbf{Z} = \mathbf{H}\mathbf{S} + \mathbf{W}. \tag{2}$$

We assume that $\mathbf{H} \in \mathbb{C}^{m\times n}$ and that the *pilot matrix* $\mathbf{S} \in \mathbb{C}^{n\times p}$. Vectorizing equation (2) then gives [4, Lemma 2.11]

$$\underbrace{\mathrm{vec}(\mathbf{Z})}_{\mathbf{y}} = \underbrace{\left(\mathbf{S}^T \otimes \mathbf{I}_m\right)}_{\mathbf{G}} \underbrace{\mathrm{vec}(\mathbf{H})}_{\mathbf{x}} + \underbrace{\mathrm{vec}(\mathbf{W})}_{\mathbf{n}}, \tag{3}$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, the $\mathrm{vec}(\cdot)$ operator stacks the columns of a matrix into a column vector, $\otimes$ denotes the Kronecker product, and $(\cdot)^T$ denotes transposition. In this work, we assume a Bayesian setting where a prior knowledge is available. Specifically, we assume that the vectorized channel $\mathbf{x}$ and vectorized noise $\mathbf{n}$ are independent and circular symmetric complex Gaussian distributed as

$$\mathbf{x} \sim \mathcal{CN}\left(\mathbf{u}_\mathbf{x}, \mathbf{C}_{\mathbf{xx}}\right) \tag{4}$$

$$\mathbf{n} \sim \mathcal{CN}\left(\mathbf{u}_\mathbf{n}, \mathbf{C}_{\mathbf{nn}}\right). \tag{5}$$

The estimator for $\mathbf{x}$ which has minimum mean square error (MMSE) is the mean of the posterior distribution, which is given by [1]

$$\mathbf{u}_\mathbf{x} + \left(\mathbf{C}_{\mathbf{xx}}^{-1} + \mathbf{G}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^H\mathbf{C}_{\mathbf{nn}}^{-1}\left(\mathbf{y} - \mathbf{G}\mathbf{u}_\mathbf{x} - \mathbf{u}_\mathbf{n}\right).$$

Here, $(\cdot)^H$ denotes the complex conjugate transpose. The MMSE associated with this estimator is given by the trace of the posterior covariance matrix,

$$\mathrm{Tr}\left(\left(\mathbf{C}_{\mathbf{xx}}^{-1} + \mathbf{G}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{G}\right)^{-1}\right), \tag{6}$$

where $\mathrm{Tr}\left(\cdot\right)$ denotes the trace operator. When designing $\mathbf{S}$ in (3), our objective is to estimate $\mathbf{x}$ from the observation $\mathbf{y}$ with as small MSE as possible. As constraint, we will impose a total power limitation on the transmitted pilots. Utilizing (6), and $\mathbf{G} = \mathbf{S}^T \otimes \mathbf{I}_m$, this optimization problem can be formulated as

$$\min_{\mathbf{S}} \quad \mathrm{Tr}\left(\left(\mathbf{C}_{\mathbf{xx}}^{-1} + \left(\mathbf{S}^T \otimes \mathbf{I}_m\right)^H \mathbf{C}_{\mathbf{nn}}^{-1}\left(\mathbf{S}^T \otimes \mathbf{I}_m\right)\right)^{-1}\right) \tag{7}$$

$$\text{s.t.} \quad \|\mathbf{S}\|_2^2 \triangleq \mathrm{Tr}\left(\mathbf{S}^H\mathbf{S}\right) \leq \sigma. \tag{8}$$

The objective function in (7) is MMSE, for a *given* $\mathbf{S}$. The constraint in (8) represents an upper bound on the squared Frobenius norm of $\mathbf{S}$.

2. BACKGROUND AND MOTIVATION

The literature on pilot design for MIMO channel estimation is rich, because (7)-(8) is a non convex problem and therefore difficult to optimize without making limiting assumptions. This work offers an alternative approach to those works that assume *Kronecker structure* on $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$. The Kronecker structure assumption is the assumption that the covariance matrices in (4) and (5) factorize as Kronecker products [1]:

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{X}_T^T \otimes \mathbf{X}_R \text{ and } \mathbf{C}_{\mathbf{nn}} = \mathbf{N}_T^T \otimes \mathbf{N}_R. \tag{9}$$

Here, $\mathbf{X}_R$ is the spatial covariance matrix at the receiver, and $\mathbf{X}_T$ is at the transmitter. Similarly, $\mathbf{N}_T$ is the temporal noise covariance matrix, and $\mathbf{N}_R$ is the spatial noise covariance matrix.

Such Kronecker factorizations allow for tractable analysis. Moreover, exploiting the Weichselberger channel model [5], it has been demonstrated in [1] that this assumption may provide good pilots even when the Kronecker structure does not hold. In general, however, assuming Kronecker structure imposes quite severe restrictions on the spatial correlation of the MIMO channel [6]. The main reason

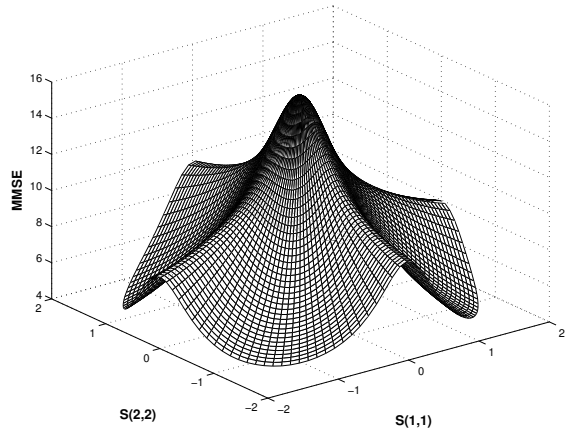

Fig. 1. Example of MMSE when $\mathbf{S} \in \mathbb{R}^{2\times 2}$, $\|\mathbf{S}\|_2^2 \leq 4$ and $\mathbf{S}$ is restricted to be a diagonal matrix.

is that arbitrary covariance matrices do generally not factorize like this. Therefore, the present works avoids this assumption, and offers an alternative approach.

For smooth optimization problems, as described by (7), (8), we can arrive at a solution that is at least first order optimal (zero gradient), from an arbitrary initial starting point [7]. The challenge is that our problem is generally not convex in $\mathbf{S}$, and the number of local minima may be large. Therefore, we propose a procedure that most often provides a better starting point than a completely random one. From this starting point, we proceed iteratively towards a local minimum. Briefly the idea goes as follows. First, we solve a relaxed convex version of the original optimization problem. Next, we project that solution onto the nearest candidate in the feasible set. Finally we move iteratively from the projected solution towards a local minimizer by employing an augmented Lagrangian method. These three steps are described in the next three sections, respectively.

Before proceeding, we mention that our approach does not always produce the best pilot matrix. In some scenarios, the pilot matrix resulting from the Kronecker structure assumption, e.g. [1], may prove better. This should not be considered a problem; we merely provide the designer with an alternative pilot matrix. Equipped with alternatives, the designer can compute the MMSE associated with each alternative, and simply choose the best one. This is valuable, especially when the channel and noise processes are stationary.

3. A RELAXED CONVEX PROBLEM

It is not difficult to generate examples showing that the problem defined by (7) and (8) is generally not convex in $\mathbf{S}$. Fig. 1 illustrates one case, when $\mathbf{S} \in \mathbb{R}^{2\times 2}$, $\|\mathbf{S}\|_2^2 \leq 4$ and $\mathbf{S}$ is restricted to be diagonal. The implication is that we must generally contend with a *local* minimizer. Such a minimizer tends to depend critically on the starting point. This section therefore derives a starting point which in many cases is better than an arbitrary one.

Note from (3) that a power constraint on the pilots $\|\mathbf{S}\|_2^2 \leq \sigma$, transfers into a corresponding power constraint $\|\mathbf{G}\|_2^2 \leq \gamma = m\sigma$. If we now consider the latter constraint only, and disregard the fact that any feasible $\mathbf{G} \in \mathbb{C}^{pm\times nm}$ must factorize as $\mathbf{S}^T \otimes \mathbf{I}_m$, we may formulate the following relaxed optimization problem

$$\min_{\mathbf{S}} \quad \mathrm{Tr}\left(\left(\mathbf{C}_{\mathbf{xx}}^{-1} + \mathbf{G}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{G}\right)^{-1}\right) \tag{10}$$

$$\text{s.t.} \quad \mathrm{Tr}\left(\mathbf{G}^H\mathbf{G}\right) \leq \gamma. \tag{11}$$

This problem has a convex structure, which will become clear shortly. Its solution, which can be efficiently obtained, must then be projected on to the set of feasible $\mathbf{G}$:s, defined by

$$\mathbb{G} := \left\{\mathbf{G} = \mathbf{S}^T \otimes \mathbf{I}_m, \text{ where } \|\mathbf{S}\|_2^2 \leq \sigma\right\}. \tag{12}$$

Finally the result of the projection is treated as a starting point, and updated iteratively towards a local minimum. Note that this approach, in contrast to [1], allows for completely arbitrary covariance matrices.

The remainder of this section presents the solution for the problem defined by (10) and (11), where $\mathbf{G}$ can have arbitrary structure. We assume that $\mathbf{G} \in \mathbb{C}^{r\times c} = \mathbb{C}^{pm\times nm}$. Observe that $\mathbf{C}_{\mathbf{xx}}$ is $nm \times nm$ and $\mathbf{C}_{\mathbf{nn}}$ is $pm \times pm$. We introduce the following singular value decompositions (SVD)

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{U}_{\mathbf{x}}\Sigma_{\mathbf{x}}\mathbf{U}_{\mathbf{x}}^H, \quad \mathbf{C}_{\mathbf{nn}}^{-1} = \mathbf{U}_{\mathbf{n}}\Sigma_{\mathbf{n}}^{-1}\mathbf{U}_{\mathbf{n}}^H, \tag{13}$$

with

$$\Sigma_{\mathbf{x}}(1,1) \geq \Sigma_{\mathbf{x}}(2,2) \geq \cdots \geq \Sigma_{\mathbf{x}}(nm,nm) > 0, \tag{14}$$

$$\Sigma_{\mathbf{n}}^{-1}(1,1) \geq \Sigma_{\mathbf{n}}^{-1}(2,2) \geq \cdots \geq \Sigma_{\mathbf{n}}^{-1}(pm,pm) > 0\,, \tag{15}$$

Throughout, $\mathbf{B}(i,j)$ will denote the element on the i-th row and j-th column of matrix $\mathbf{B}$. In order to rewrite the optimization problem (10)-(11) in a more convenient form, we now assume that

$$\mathbf{G} = \mathbf{U}_{\mathbf{n}}\mathbf{F}\mathbf{U}_{\mathbf{x}}^H. \tag{16}$$

Observe that this introduces no restrictions on $\mathbf{G}$: If $\mathbf{F}$ can be any $pm \times nm$ matrix, then so can $\mathbf{G}$, because both $\mathbf{U}_{\mathbf{n}}$ and $\mathbf{U}_{\mathbf{x}}$ are unitary. Exploiting (16), (13) and (7), it is straightforward to verify that the optimization simplifies to

$$\min_{\mathbf{F}} \mathrm{Tr}\left(\left(\Sigma_{\mathbf{x}}^{-1} + \mathbf{F}^H\Sigma_{\mathbf{n}}^{-1}\mathbf{F}\right)^{-1}\right) \tag{17}$$

$$\text{s.t.} \qquad \mathrm{Tr}\left(\mathbf{F}^H\mathbf{F}\right) \leq \gamma. \tag{18}$$

Applying [1, Lemma 1], it can be shown that the optimal $\mathbf{F}$ is diagonal[1]. If we define the compact notation $\mathbf{F}^H(i,i)\mathbf{F}(i,i) = f_i^2$, $\Sigma_{\mathbf{x}}^{-1}(i,i) = \sigma_x^{-1}(i)$ and $\Sigma_{\mathbf{n}}^{-1}(i,i) = \sigma_n^{-1}(i)$ the optimization problem can therefore be written as

$$\min_{\mathbf{F}} \sum_{i=1}^{nm} \frac{1}{\sigma_x^{-1}(i) + f_i^2\sigma_n^{-1}(i)} \tag{19}$$

$$\text{s.t.} \qquad \sum_{i=1}^{nm} f_i^2 \leq \gamma. \tag{20}$$

This is clearly a convex problem in f_i^2, for which we know the KKT conditions define the optimal solution. The solution can be derived as

$$f_i^2 = \left(0, \sqrt{\frac{\sigma_n(i)}{\alpha}} - \frac{\sigma_n(i)}{\sigma_x(i)}\right)^+, \tag{21}$$

[1]In fact, it can also be shown that the optimal $\mathbf{F}$ is such that $\mathbf{F}^H\Sigma_{\mathbf{n}}^{-1}\mathbf{F}$ has decreasingly ordered diagonal elements. We do not have to rely on that property at this point, because it follows naturally.

where $(0,q)^+$ denotes the maximum of 0 and q, and $\alpha > 0$ is a Lagrange multiplier chosen such that

$$\gamma = \sum_{i=1}^{nm} \left(0, \sqrt{\frac{\sigma_n(i)}{\alpha}} - \frac{\sigma_n(i)}{\sigma_x(i)}\right)^+. \quad (22)$$

Observe that both the objective function and the constraint depend on $\mathbf{F}$ only via the squared elements $f_i^2 = \mathbf{F}^H(i,i)\mathbf{F}(i,i)$. This implies that the optimal solution for $\mathbf{F}$ is not unique: any $\mathbf{F}$ satisfying (21) and (22) is optimal, and we may for instance choose $\mathbf{F}$ to be purely real. Inserting such an $\mathbf{F}$ into (16), produces an optimal $\mathbf{G}$ matrix.

Note finally that the solution (21) satisfies the constraint (20) with equality. The explanation for this is straightforward: The objective function $\text{Tr}\left(\mathbf{W}^{-1}\right)$, as in (17), is strictly convex in the eigenvalues of any positive definite matrix $\mathbf{W}$. Hence, for a matrix $\mathbf{F}$ that does not fulfill (18) with equality, we can always reduce (17) by updating $\mathbf{F}$ to $\eta\mathbf{F}$, where $\eta > 1$ without violating the constraint. The implication is that we need not consider the interior of the constraint region, only its boundary. An entirely similar argument goes of course for the original problem (7), and therefore we can conclude that a solution should satisfy (8) with equality.

4. PROJECTING ONTO THE FEASIBLE SET

We cannot expect that an optimal matrix $\mathbf{G}$, as given in the previous section, factorizes as required by (12). Moreover, we are actually interested in the underlying $\mathbf{S}$. To that end, and since we know that a solution will spend all the available power, a natural approach is to select the $\mathbf{S}$ which solves

$$\min_{\mathbf{S}} \left\| \mathbf{G} - \left(\mathbf{S}^T \otimes \mathbf{I}_m\right) \right\|_2^2 \quad (23)$$

$$\text{s.t.} \qquad \text{Tr}\left(\mathbf{S}^H\mathbf{S}\right) = \sigma. \quad (24)$$

From the definition of the Kronecker product we then have

$$\mathbf{S}^T \otimes \mathbf{I}_m = \begin{bmatrix} \mathbf{S}(1,1)\mathbf{I}_m & \cdots & \mathbf{S}(n,1)\mathbf{I}_m \\ \vdots & \ddots & \vdots \\ \mathbf{S}(1,p)\mathbf{I}_m & \cdots & \mathbf{S}(n,p)\mathbf{I}_m \end{bmatrix}.$$

If we partition $\mathbf{G}$ into a similar block structure, such that

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \cdots & \mathbf{G}_{n,1} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{1,p} & \cdots & \mathbf{G}_{n,p} \end{bmatrix},$$

where each block $\mathbf{G}_{i,j}$ is $m \times m$, it can be verified that

$$\left\| \mathbf{G} - \left(\mathbf{S}^T \otimes \mathbf{I}_m\right) \right\|_2^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \|\mathbf{G}_{i,j} - \mathbf{S}(i,j)\mathbf{I}_m\|_2^2. \quad (25)$$

As for the constraint, note that

$$\text{Tr}\left(\mathbf{S}^H\mathbf{S}\right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbf{S}(i,j)\mathbf{S}^*(i,j) = \sigma, \quad (26)$$

where $(\cdot)^*$ denotes complex conjugation. The projection is therefore the solution to

$$\min_{\mathbf{S}} \sum_{i=1}^{n}\sum_{j=1}^{p} \|\mathbf{G}_{i,j} - \mathbf{S}(i,j)\mathbf{I}_m\|_2^2$$
$$\text{s.t.} \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbf{S}(i,j)\mathbf{S}^*(i,j) = \sigma.$$

This is a convex problem with convex constraints. The solution can be derived as

$$\mathbf{S}(i,j) = \frac{\text{Tr}\left(\mathbf{G}_{i,j}\right)}{m+\beta}, \quad (27)$$

where β is a Lagrange multiplier chosen such that

$$\sum_{i=1}^{n}\sum_{j=1}^{p} \frac{\text{Tr}\left(\mathbf{G}_{i,j}\right)^* \text{Tr}\left(\mathbf{G}_{i,j}\right)}{(m+\beta)^2} = \sigma. \quad (28)$$

Observe that if $\beta = 0$ satisfies (28), we see from (27) that $\mathbf{S}(i,j)$ becomes the mean of the diagonal elements of block $\mathbf{G}_{i,j}$.

5. UPDATING TO A LOCAL MINIMUM

The result of the projection of (27)-(28) produces a feasible pilot matrix, but that pilot matrix is in general not even a first order optimal solution to the original problem (7)-(8). Therefore it should be treated as a starting point for subsequent optimization. We will move from this starting point towards a local optimum using an augmented Lagrangian method. The latter is also known as the *method of multipliers*. The core idea is to replace a constrained problem by a sequence of unconstrained problems. A good introduction to this method, along with algorithms for implementation, can be found in [7]. Therefore we do not present the full details of the method here, but rather focus on some key ingredients.

Let the objective function in (7) be denoted by $g(\mathbf{S})$. Because we know that a solution will spend all the available power, we substitute the inequality contraint (8) by the equality constraint $c(\mathbf{S}) = \sigma - \text{Tr}\left(\mathbf{S}^H\mathbf{S}\right) = 0$. The augmented Lagrangian function is then given by

$$\mathcal{L}\left(\mathbf{S}, \lambda, \mu\right) = g(\mathbf{S}) - \lambda c(\mathbf{S}) + \frac{1}{2\mu}c^2(\mathbf{S}), \quad (29)$$

where λ is a Lagrange multiplier and μ is a penalty parameter. The derivative of this function with respect to $\mathbf{S}$ is

$$\nabla_{\mathbf{s}}\mathcal{L}\left(\mathbf{S}, \lambda, \mu\right) = \nabla_{\mathbf{s}} g(\mathbf{S}) - \left(\lambda - \frac{c(\mathbf{S})}{\mu}\right)\nabla_{\mathbf{s}} c(\mathbf{S}). \quad (30)$$

Because $\nabla_{\mathbf{s}} g(\mathbf{S})$ and $\nabla_{\mathbf{s}} c(\mathbf{S})$ are key elements in the augmented Lagrangian method, we derive them next.

5.1. The gradient of the objective and the constraint

The objective function can be expressed as $\text{Tr}\left(\mathbf{W}^{-1}\right)$, where

$$\mathbf{W} = \mathbf{C}_{\mathbf{xx}}^{-1} + \left(\mathbf{S}^T \otimes \mathbf{I}_m\right)^H \mathbf{C}_{\mathbf{nn}}^{-1}\left(\mathbf{S}^T \otimes \mathbf{I}_m\right). \quad (31)$$

In order to find the derivative of the objective function w.r.t. $\mathbf{S}$, it is convenient to take the approach suggested in [4]. That is, to first identify the differential, and then use this to obtain the derivative.

	$\mathbf{S}$	$\mathbf{S}_{rand}$	$\mathbf{S}_{kron}$
Winner rate	0.5080	0.3160	0.1760
Average normalized MMSE	0.0657	0.0674	0.0748

Table 1. Performance for a specific class of covariance matrices.

Without displaying the preceding steps here, the gradient of the objective function w.r.t. $\mathbf{S}$, expressed as a $1 \times np$ row vector is:

$$-\text{vec}^T \left(\mathbf{S}^* \otimes \mathbf{I}_m\right) \left(\mathbf{C}_{\mathbf{nn}}^{-1} \otimes \mathbf{W}^{-1}\mathbf{W}^{-1}\right) \left(\mathbf{I}_p \otimes \mathbf{R}\right), \quad (32)$$

where

$$\mathbf{R} = \left(\mathbf{K}_{m,n} \otimes \mathbf{I}_m\right) \left(\mathbf{I}_n \otimes \text{vec}\left(\mathbf{I}_m\right)\right), \quad (33)$$

and $\mathbf{K}_{m,n}$ is the *commutation* matrix [4, Definition 2.9]. In order to obtain $\nabla_{\mathbf{s}} g(\mathbf{S})$, we split the row vector (32) into p equally long sub vectors and take these as the columns of $\nabla_{\mathbf{s}} g(\mathbf{S})$. The derivative of the constraint w.r.t. $\mathbf{S}$ is simply

$$\nabla c_{\mathbf{s}}(\mathbf{S}) = \mathbf{S}^*. \quad (34)$$

For space reasons, we do not present the full algorithmic framework of the augmented Lagrangian method here. Instead we refer the reader to [7, Framework 17.3]. With the gradients given in (32) and (34), the algorithm is straightforward to implement.

6. NUMERICAL RESULTS

This section compares experimentally the performance of our method with that of [1, Heuristic 1], for a particular class of noise and channel covariance matrices, which we describe shortly. The augmented Lagrangian method is implemented as described in [7, Framework 17.3], using the following parameters:

$$\mu_k = \tau_k = \frac{1}{k}.$$

As initial values, we select $\lambda_0 = \mu_0 = \tau_0 = 1$. We assume a case where all matrices in (2) are 2×2. Consequently, $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$ are 4×4. We study the average MMSE over 500 different scenarios where $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$ are generated randomly. For each scenario, the covariance matrices are generated as follows. Let $\mathbf{A}$ be a realization of a 4×4 matrix with i.i.d. elements $\mathcal{N}(0,1)$. Compute $\mathbf{C}_{\mathbf{xx}} = \text{abs}(\mathbf{A}^T)\text{abs}(\mathbf{A})$, where the abs$(\cdot)$ operator turns the sign of the negative elements. $\mathbf{C}_{\mathbf{nn}}$ is generated independently in the same manner. Observe that these matrices are symmetric, and positive definite with probability one. We assume that $\sigma = 4$ in (8). Table 1 summarizes the results. In Table 1, $\mathbf{S}$, $\mathbf{S}_{kron}$ and $\mathbf{S}_{rand}$ denote the pilot matrices that result from our method, from [1, Heuristic 1], and from an augmented Lagrangian method with random starting point, respectively. The 'winner rate' represents the share of scenarios where a method outperforms the two other methods. The normalized MMSE is defined as

$$\frac{\text{Tr}\left(\left(\mathbf{C}_{\mathbf{xx}}^{-1} + \left(\mathbf{S}^T \otimes \mathbf{I}_m\right)^H \mathbf{C}_{\mathbf{nn}}^{-1} \left(\mathbf{S}^T \otimes \mathbf{I}_m\right)\right)^{-1}\right)}{\text{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right)}.$$

This is just the standard MMSE normalized with the power of the channel that we wish to estimate.

Table 1 indicates that, for covariance matrices generated as described, the proposed method is better than that of [1, Heuristic 1] on average. In fact, for this setup, an augmented Lagrangian method with random starting point is also better than [1, Heuristic 1] on average. These observations underline that pilot matrices based on the Kronecker structure assumption should be used with some care, and that other alternatives could be worth exploring. Also, Table 1 indicates that our method is better than using a random starting point on average. This examples focuses on the average performance over multiple scenarios. In a single scenario, with stationary settings, one can evaluate several alternatives and select the best one. We have observed, in such cases, that the difference between the different methods may be substantial.

The Kronecker structure assumption allows for a closed form solution. It may not always turn out to be the best, but it can be derived with very limited complexity. In, contrast the augmented Lagrangian method is based on an iterative algorithm. The speed of convergence depends on the initial values and how one updates the parameters, but it will invariably introduce much higher computational load. Under stationary or slowly varying statistics, that effort may still pay off.

7. CONCLUSION

We have described a procedure which obtains a pilot matrix for MIMO channel estimation when the structure on the underlying covariance matrices is arbitrary. In particular, we do not rely on Kronecker structure. The procedure is based on a convex relaxation of the original problem. Its solution is projected onto the feasible set, and used as starting point for an augmented Lagrangian method. Numerical experiments indicate that this procedure may produce pilot signals that are better than those obtained under the Kronecker structure assumption.

8. REFERENCES

[1] E. Björnson and B. Ottersten, "A Framework for Training-Based Estimation in Arbitrarily Correlated Rician MIMO Channels With Rician Disturbance," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1807 –1820, march 2010.

[2] D. Katselis, E. Kofidis, and S. Theodoridis, "On Training Optimization for Estimation of Correlated MIMO Channels in the Presence of Multiuser Interference," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4892 –4904, oct. 2008.

[3] M. Biguesh and A.B. Gershman, "Training-based MIMO Channel Estimation: a Study of Estimator Tradeoffs and Optimal Training Signals," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 884 – 893, march 2006.

[4] A. Hjørungnes, *Complex-Valued Matrix Derivatives : with Applications in Signal Processing and Communications*, Cambridge University Press, Cambridge, 2011.

[5] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 1, pp. 90 – 100, jan. 2006.

[6] C. Oestges, "Validity of the Kronecker Model for MIMO Correlated Channels," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, may 2006, vol. 6, pp. 2818 – 2822.

[7] Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, 2006.