OPTIMAL TRAINING FOR ML AND LMMSE CHANNEL ESTIMATION IN MIMO SYSTEMS

Geert Leus and Alle-Jan van der Veen

Delft University of Technology - Fac. EEMCS
Mekelweg 4, 2628 CD Delft, The Netherlands
{leus,allejan}@cas.et.tudelft.nl

ABSTRACT

In this paper, we present some optimal training designs for maximum likelihood (ML) and linear minimum mean square error (LMMSE) channel estimation for multiple-input multiple-output (MIMO) systems. As optimization criterion, the channel mean square error (MSE) is chosen. The key idea is not to restrict the channel estimation to a single transmitted symbol block, but to possibly exploit multiple symbol blocks, assuming the channel remains constant over these blocks. This leads to some new optimal training designs.

1. INTRODUCTION

Optimal training for MIMO channel estimation has received a lot of attention lately. Both multi-carrier [1, 2, 3] and single-carrier [4, 5, 6] MIMO systems have been investigated, using a wide range of criteria to optimize the training. In this paper, we focus on single-carrier MIMO systems and investigate optimal training, in the sense of the channel MSE, for ML as well as LMMSE channel estimation. Note that LMMSE optimal training requires feedback of a few optimal training parameters that depend on the channel statistics and the noise variance. However, since these system characteristics do not change fast, the amount of feedback information is relatively low.

In addition, we do not restrict the channel estimation to a single transmitted symbol block, but possibly exploit multiple symbol blocks, assuming the channel remains constant over these blocks. As a result, we obtain a more general design approach. Simulation results are carried out to investigate the performance of ML and LMMSE channel estimation under ML and LMMSE optimal training.

Notation: Matrices and column vectors are written in boldface uppercase and lowercase, respectively. For a matrix or column vector, superscript ^{T} is the transpose and ^{H} is the complex conjugate transpose. $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{0}_{M\times N}$ is the $M \times N$ all-zero matrix. vec($\mathbf{A}$) is a stacking of the columns of a matrix $\mathbf{A}$ into a column vector. $[\mathbf{A}]_{i,j}$ denotes the (i, j)th element of the matrix $\mathbf{A}$, whereas $[\mathbf{A}]_{i,:}$ and $[\mathbf{A}]_{:,i}$ respectively denote the ith row and column of the matrix $\mathbf{A}$. $\|\cdot\|$ represents the Frobenius norm and $\otimes$ is the Kronecker product. Finally, E$(\cdot)$ and tr$(\cdot)$ respectively denote the expectation and trace operator.

This reasearch work is supported by NWO-STW under the VIDI program (DTC.6577) and the VICI program (DTC.5893).

2. DATA MODEL

Let us consider a MIMO system with A_t transmit antennas and A_r receive antennas. Suppose $\mathbf{x}(n)$ represents the $A_t \times 1$ symbol vector sequence transmitted at the A_t transmit antennas. Assuming symbol rate sampling at each receive antenna, the $A_r \times 1$ sample vector sequence received at the A_r receive antennas is then given by

$$\mathbf{y}(n) = \sum_{l=0}^{L} \mathbf{H}(l)\mathbf{x}(n-l) + \mathbf{e}(n), \tag{1}$$

where $\mathbf{e}(n)$ is the $A_r \times 1$ additive noise vector sequence on the A_r receive antennas, which we assume to be zero-mean white (spatially and temporally) Gaussian with variance σ_e^2, and $\mathbf{H}(l)$ is the $A_r \times A_t$ MIMO channel of order L. Note that we will often make use of the vectorized form of $\mathbf{H}(l)$, which is obtained by stacking its columns: $\mathbf{h}(l) = \text{vec}\{\mathbf{H}(l)\}$.

In this paper, we focus on estimating $\mathbf{H}(l)$ (or $\mathbf{h}(l)$) without assuming any structure on it. Hence, no calibration of the different transmit/receive antennas is required. We assume a burst of N symbol vectors is transmitted, in the form of K symbol blocks, where each symbol block consists of N_t training symbol vectors, surrounded at each side by $N_d/2$ unknown data symbol vectors, i.e., $N = K(N_t + N_d)$. The N_t training symbol vectors in the kth symbol block can be collected into

$$\mathbf{x}_k = [\mathbf{x}^T(n_k), \ldots, \mathbf{x}^T(n_k + N_t - 1)]^T,$$

where $n_k = k(N_t + N_d) + N_d/2$ indicates the start of the training symbol vectors in the kth symbol block. Since we only focus on conventional training-based channel estimation, we will only focus on the received sample vectors that solely depend on the training symbol vectors. The $N_t - L$ received sample vectors that solely depend on the N_t training symbol vectors in the kth symbol block yield the vector

$$\mathbf{y}_k = [\mathbf{y}^T(n_k + L), \ldots, \mathbf{y}^T(n_k + N_t - 1)]^T,$$

which can be expressed as

$$\mathbf{y}_k = \boldsymbol{\mathcal{H}}\mathbf{x}_k + \mathbf{e}_k, \tag{2}$$

where $\mathbf{e}_k$ is similarly defined as $\mathbf{y}_k$ and $\boldsymbol{\mathcal{H}}$ is the $A_r(N_t - L) \times A_t N_t$ block Toeplitz matrix representing the convolution by the channel:

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{H}(L) & \cdots & \mathbf{H}(0) & & \\ & \ddots & & \ddots & \\ & & \mathbf{H}(L) & \cdots & \mathbf{H}(0) \end{bmatrix}.$$

Alternatively, we can write the convolution operation (2) as a linear operation on the channel coefficient vector $\mathbf{h} = [\mathbf{h}^T(0), \ldots, \mathbf{h}^T(L)]^T$, which gives

$$\mathbf{y}_k = (\boldsymbol{\mathcal{X}}_k \otimes \mathbf{I}_{A_r})\mathbf{h} + \mathbf{e}_k, \tag{3}$$

where $\boldsymbol{\mathcal{X}}_k$ is the $(N_t - L) \times A_t(L+1)$ block Toeplitz symbol matrix given by

$$\boldsymbol{\mathcal{X}}_k = \begin{bmatrix} \mathbf{x}^T(n_k + L) & \cdots & \mathbf{x}^T(n_k) \\ \vdots & & \vdots \\ \mathbf{x}^T(n_k + N_t - 1) & \cdots & \mathbf{x}^T(n_k + N_t - L - 1) \end{bmatrix}. \tag{4}$$

Stacking $\mathbf{y}_k$ for $k = 0, 1, \ldots, K-1$, we obtain $\mathbf{y} = [\mathbf{y}_0^T, \ldots, \mathbf{y}_{K-1}^T]^T$, which can be expressed as

$$\mathbf{y} = (\boldsymbol{\mathcal{X}} \otimes \mathbf{I}_{A_r})\mathbf{h} + \mathbf{e}, \tag{5}$$

where $\mathbf{e}$ is similarly defined as $\mathbf{y}$, and $\boldsymbol{\mathcal{X}} = [\boldsymbol{\mathcal{X}}_0^T, \ldots, \boldsymbol{\mathcal{X}}_{K-1}^T]^T$.

Many different channel estimation procedures can be applied to (5). In this paper, we restrict ourselves to ML and LMMSE channel estimation. For each of them, we investigate optimal training designs.

3. ML CHANNEL ESTIMATION

Assuming $\boldsymbol{\mathcal{X}}$ has full column rank, which can be guaranteed by design, it is easy to derive from (5) that the ML channel estimate is given by

$$\mathbf{h}_{ML} = [(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}})^{-1} \otimes \mathbf{I}_{A_r}](\boldsymbol{\mathcal{X}}^H \otimes \mathbf{I}_{A_r})\mathbf{y}. \tag{6}$$

This ML channel estimate is unbiased, and the channel MSE can be expressed as

$$\begin{aligned} J_{ML} &= \mathrm{E}(\|\mathbf{h}_{ML} - \mathbf{h}\|^2) \\ &= \sigma_e^2 A_r \,\mathrm{tr}[(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}})^{-1}]. \end{aligned} \tag{7}$$

Note that the ML channel estimation problem can actually be decoupled into the different receive antennas, and is often presented as such. However, for LMMSE channel estimation, which will be discussed in the next section, the correlation between the different receive antennas will come into the picture, and the problem cannot be decoupled anymore.

3.1. Optimal Training

We now design $\boldsymbol{\mathcal{X}}$ such that J_{ML} is minimized under a total training power constraint. In other words, we consider the following problem

$$\min_{\{\mathbf{x}_k\}} \mathrm{tr}[(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}})^{-1}] \quad \text{s.t.} \quad \sum_{k=0}^{K-1} \|\mathbf{x}_k\|^2 = E. \tag{8}$$

To solve this, observe that

$$\mathrm{tr}[(\boldsymbol{\mathcal{X}}^{(H}\boldsymbol{\mathcal{X}})^{-1}] \geq \sum_{l=0}^{L}\sum_{i=1}^{A_t} \frac{1}{\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2}, \tag{9}$$

where equality is obtained if $\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}$ is diagonal. We proceed by looking for the minimum of the right hand side of (9) under the total training power constraint, and we subsequently try to realize this minimum by a training design for which $\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}$ is diagonal, in order to obtain equality in (9).

We will consider the following two cases: the number of training symbols $N_t \geq 2L+1$ and $N_t = L+1$. For the remaining case where $L+1 < N_t < 2L+1$, the optimization problem is hard to solve in analytical form.

3.2. Case $N_t \geq 2L+1$.

It is clear that in order to minimize the right hand side of (9), we should make sure that no training power is wasted. Hence, we should take $\mathbf{x}(n_k + l) = \mathbf{x}(n_k + N_t - 1 - l) = \mathbf{0}_{A_t \times 1}$, for $l = 0, \ldots, L-1$ and $k = 0, \ldots, K-1$. This means that for a fixed value of i the columns $[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}$, $l = 0, \ldots, L$, are shifted versions of each other, whereas for a fixed value of l the columns $[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}$, $i = 1, \ldots, A_t$, have no entries in common. Hence, defining $\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2 = \lambda_i$, we need to solve

$$\min_{\{\lambda_i\}} \sum_{l=0}^{L}\sum_{i=1}^{A_t} \frac{1}{\lambda_i} \quad \text{s.t.} \quad \sum_{i=1}^{A_t} \lambda_i = E \text{ and } \lambda_i \geq 0.$$

It is clear that the solution is given by $\lambda_i = E/A_t$. In order to force $\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}$ to be diagonal, we should thus design $\boldsymbol{\mathcal{X}}$ such that

$$\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} = \frac{E}{A_t}\mathbf{I}_{A_t(L+1)}.$$

As an example, consider $N_t = M(L+1)+L$ with $M \geq 1$. An optimal solution is then given by using dispersed training symbol vectors separated by L zero vectors (see Figure 1):

$$\begin{aligned} \mathbf{x}_k = [&\mathbf{0}_{A_tL\times1}^T, \mathbf{t}_{kM}^T, \mathbf{0}_{A_tL\times1}^T, \mathbf{t}_{kM+1}^T, \mathbf{0}_{A_tL\times1}^T, \ldots \\ &\ldots, \mathbf{0}_{A_tL\times1}^T, \mathbf{t}_{(k+1)M-1}^T, \mathbf{0}_{A_tL\times1}^T]^T \end{aligned} \tag{10}$$

where $\mathbf{T} = [\mathbf{t}_0, \ldots, \mathbf{t}_{KM-1}]$ satisfies $\mathbf{T}\mathbf{T}^H = E/A_t\mathbf{I}_{A_t}$, which requires $K \geq A_t/M$.

3.3. Case $N_t = L+1$.

We observe from (4) that in this case $\boldsymbol{\mathcal{X}}_k = \mathbf{x}_k^T$, and consequently $\boldsymbol{\mathcal{X}} = [\mathbf{x}_0, \ldots, \mathbf{x}_{K-1}]^T$. This means that all columns $[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}$, $l = 0, \ldots, L$ and $i = 1, \ldots, A_t$, have no entries in common. This means that $\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2$ is independent for different values of l and i, and can thus be denoted as $\lambda_{l,i}$. Hence, defining $\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2 = \lambda_{l,i}$, we need to solve

$$\min_{\{\lambda_{l,i}\}} \sum_{l=0}^{L}\sum_{i=1}^{A_t} \frac{1}{\lambda_{l,i}} \quad \text{s.t.} \quad \sum_{l=0}^{L}\sum_{i=1}^{A_t} \lambda_{l,i} = E \text{ and } \lambda_{l,i} \geq 0.$$

It is clear that the solution is given by $\lambda_{l,i} = E/(A_t(L+1))$. In order to force $\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}$ to be diagonal, we should thus design $\boldsymbol{\mathcal{X}}$ such that

$$\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} = \frac{E}{A_t(L+1)}\mathbf{I}_{A_t(L+1)}.$$

This is easy to obtain as long as $K \geq A_t(L+1)$, since there is no structure in $\boldsymbol{\mathcal{X}}$.

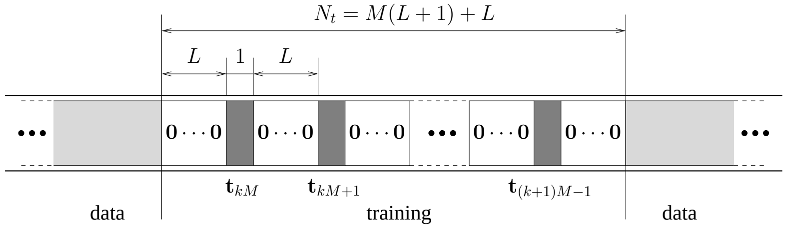

Fig. 1. Optimal training structure if $N_t \geq 2L+1$.

4. LMMSE CHANNEL ESTIMATION

From (5) and the fact that the additive noise $\mathbf{e}(n)$ is zero-mean white (spatially and temporally) Gaussian with variance σ_e^2, it is easy to derive that the LMMSE channel estimate is given by

$$\mathbf{h}_{LMMSE} = [(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2 \mathbf{R}_h^{-1}]^{-1}(\boldsymbol{\mathcal{X}}^H \otimes \mathbf{I}_{A_r})\mathbf{y}, \quad (11)$$

where $\mathbf{R}_h = \mathrm{E}(\mathbf{h}\mathbf{h}^H)$. This LMMSE channel estimate is biased, and the channel MSE can be expressed as

$$\begin{aligned} J_{LMMSE} &= \mathrm{E}(\|\mathbf{h}_{LMMSE} - \mathbf{h}\|^2) \\ &= \sigma_e^2 \ \mathrm{tr}\{[(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2 \mathbf{R}_h^{-1}]^{-1}\}. \qquad (12) \end{aligned}$$

Note that in contrast to the ML channel estimate, the LMMSE estimate requires the knowledge of the channel statistics and the noise variance.

4.1. Optimal Training

We now design $\boldsymbol{\mathcal{X}}$ such that J_{LMMSE} is minimized under a total training power constraint. In other words, we consider the following problem

$$\min_{\{\mathbf{x}_k\}} \mathrm{tr}\{[(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2 \mathbf{R}_h^{-1}]^{-1}\} \quad \text{s.t.} \quad \sum_{k=0}^{K-1} \|\mathbf{x}_k\|^2 = E. \qquad (13)$$

To solve this, observe that

$$\mathrm{tr}\{[(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2 \mathbf{R}_h^{-1}]^{-1}\} \geq \sum_{l=0}^{L}\sum_{i=1}^{A_t}\sum_{j=1}^{A_r} \frac{1}{\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2 + \sigma_e^2[\mathbf{R}_h^{-1}]_{n_{l,i,j},n_{l,i,j}}}, \quad (14)$$

where $n_{l,i,j} = lA_tA_r + (i-1)A_r + j$, and equality is obtained if $(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2\mathbf{R}_h^{-1}$ is diagonal. As before, we proceed by looking for the minimum of the right hand side of (14) under the total training power constraint, and we subsequently try to realize this minimum by a training design for which $(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2\mathbf{R}_h^{-1}$ is diagonal, in order to obtain equality in (14). Since this last diagonalization step is not always possible, we assume for simplicity that $\mathbf{R}_h$ is diagonal. As before, we only consider the $N_t \geq 2L+1$ case and the $N_t = L+1$ case.

4.2. Case $N_t \geq 2L+1$.

As before, it is clear that in order to minimize the right hand side of (14), we should make sure that no training power is wasted. Hence, we should take $\mathbf{x}(n_k + l) = \mathbf{x}(n_k + N_t - 1 - l) = \mathbf{0}_{A_t \times 1}$, for $l = 0, \ldots, L-1$ and $k = 0, \ldots, K-1$. Hence, as in the ML case, we can define $\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2 = \lambda_i$, and solve

$$\min_{\{\lambda_i\}} \sum_{l=0}^{L}\sum_{i=1}^{A_t}\sum_{j=1}^{A_r} \frac{1}{\lambda_i + \sigma_e^2[\mathbf{R}_h^{-1}]_{n_{l,j,i},n_{l,j,i}}}$$
$$\text{s.t.} \quad \sum_{i=1}^{A_t} \lambda_i = E \quad \text{and} \quad \lambda_i \geq 0.$$

The solution can be obtained numerically and will be denoted by $\lambda_{i,opt}$. Given this solution, we can finally force $(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2\mathbf{R}_h^{-1}$ to be diagonal by designing $\boldsymbol{\mathcal{X}}$ such that

$$\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} = \mathbf{I}_{L+1} \otimes \boldsymbol{\Lambda},$$

where

$$[\boldsymbol{\Lambda}]_{i,i'} = \begin{cases} \lambda_{i,opt} & \text{if } i' = i \\ 0 & \text{otherwise} \end{cases}.$$

If $N_t = M(L+1) + L$ with $M \geq 1$, we can obtain this by using the training strategy of (10) (see Figure 1) and designing $\mathbf{T} = [\mathbf{t}_0, \ldots, \mathbf{t}_{KM-1}]$ such that $\mathbf{T}\mathbf{T}^H = \boldsymbol{\Lambda}$, which again requires $K \geq A_t/M$.

4.3. Case $N_t = L+1$.

As before, we observe from (4) that in this case $\boldsymbol{\mathcal{X}}_k = \mathbf{x}_k^T$, and consequently $\boldsymbol{\mathcal{X}} = [\mathbf{x}_0, \ldots, \mathbf{x}_{K-1}]^T$. Hence, as in the ML case, we can define $\|[\boldsymbol{\mathcal{X}}]_{:,lA_t+i}\|^2 = \lambda_{l,i}$, and solve

$$\min_{\{\lambda_{l,i}\}} \sum_{l=0}^{L}\sum_{i=1}^{A_t}\sum_{j=1}^{A_r} \frac{1}{\lambda_{l,i} + \sigma_e^2[\mathbf{R}_h^{-1}]_{n_{l,j,i},n_{l,j,i}}}$$
$$\text{s.t.} \quad \sum_{l=0}^{L}\sum_{i=1}^{A_t} \lambda_{l,i} = E \quad \text{and} \quad \lambda_{l,i} \geq 0.$$

The solution can be obtained numerically and will be denoted by $\lambda_{l,i,opt}$. Given this solution, we can finally force $(\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}}) \otimes \mathbf{I}_{A_r} + \sigma_e^2\mathbf{R}_h^{-1}$ to be diagonal by designing $\boldsymbol{\mathcal{X}}$ such that

$$\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} = \boldsymbol{\Lambda}, \qquad (15)$$

where

$$[\boldsymbol{\Lambda}]_{(l-1)A_t+i,(l'-1)A_t+i'} = \begin{cases} \lambda_{l,i,opt} & \text{if } l' = l \text{ and } i' = i \\ 0 & \text{otherwise} \end{cases}.$$

This is again easy to obtain as long as $K \geq A_t(L+1)$, since there is no structure in $\boldsymbol{\mathcal{X}}$.

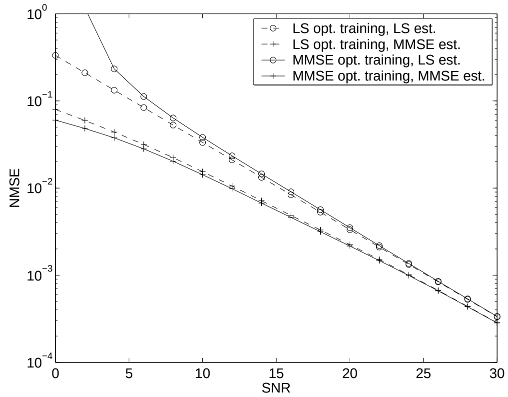

Fig. 2. Performance results for $N_t = 2L + 1$ and $K = 2$.

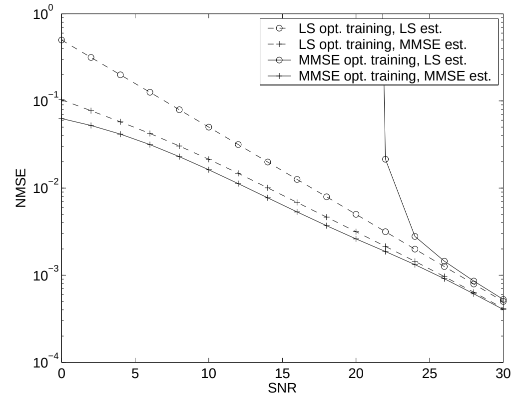

Fig. 3. Performance results for $N_t = L + 1$ and $K = 4$.

4.4. Remarks

The assumption that $\mathbf{R}_h$ is diagonal might not be very realistic. However, it turns out that this training design procedure also works well for MIMO channels with correlated channel taps, although it is suboptimal in that case.

It can be observed that the optimal training parameters $\lambda_{i,opt}$ and $\lambda_{l,i,opt}$ depend on the channel statistics and the noise variance. Hence, these optimal training parameters need to be computed at the receiver and be fed back to the transmitter. Fortunately, the channel statistics and the noise variance vary slowly and the optimal training parameters do not have to be fed back very often, thereby reducing the relative amount of feedback information.

5. SIMULATIONS

We consider a MIMO system with $A_t = 2$ transmit antennas and $A_r = 2$ receive antennas. The channel order we simulate is $L = 1$. We assume the different channel taps are Rayleigh fading and that the channel covariance matrix is given by

$$\mathbf{R}_h = \begin{bmatrix} 50 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 50 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 50 & 0 \\ 0 & 1 \end{bmatrix}.$$

For each setup, we study four different possibilities: ML channel estimation with ML optimal training, ML channel estimation with LMMSE optimal training, LMMSE channel estimation with ML optimal training, and LMMSE channel estimation with LMMSE optimal training. As a performance measure, we consider the channel MSE (see (7) and (12) for the channel MSE of ML and LMMSE channel estimation, respectively), normalized with the average channel energy:

$$NMSE = \frac{\mathrm{E}(\|\mathbf{h}_{LMMSE} - \mathbf{h}\|^2)}{\mathrm{E}(\|\mathbf{h}\|^2)}.$$

The signal-to-noise ratio (SNR) per transmit and receive antenna is defined as

$$SNR = \frac{\mathrm{E}(\|\mathbf{h}\|^2)}{\sigma_e^2 A_t A_r},$$

where it is assumed that the average energy of a training symbol is equal to one.

First, we take $N_t = 2L + 1 = 7$ and $K = 2$, and implement the optimal training strategy of Sections 3.2 and 4.2. The simulation results are shown in Figure 2. Next, we take $N_t = L + 1 = 4$ and $K = 4$, and implement the optimal training strategy of Sections 3.3 and 4.3. The simulation results are shown in Figure 3. As expected, LMMSE optimal training outperforms LS optimal training if LMMSE channel estimation is considered. On its turn, LMMSE channel estimation with LS optimal training outperforms LS channel estimation with LS optimal training. The worst performance is obtained by LS channel estimation with LMMSE optimal training. The reason for this is that at low SNR some $\lambda_{i,opt}$ or $\lambda_{l,i,opt}$ values get close to zero, which means that $\boldsymbol{\mathcal{X}}$ becomes almost singular and J_{ML} gets very large. Hence, when using LMMSE optimal training, one should always incorporate some kind of regularization in the channel estimation procedure.

6. REFERENCES

[1] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Communications*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[2] T.-L. Tung and K. Yao, "Channel estimation and optimal power allocation for a multiple-antenna OFDM system," *Eurasip Journal on Applied Signal Proc.*, pp. 330–339, March 2002.

[3] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Proc.*, vol. 51, no. 6, pp. 1615–1624, June 2003.

[4] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," in *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002, pp. 1107–1111.

[5] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Communications*, vol. 2, no. 2, pp. 384–391, Mar. 2003.

[6] B. Hassibi and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Info. Th.*, vol. 48, no. 4, pp. 951–963, Apr. 2003.