

# TRAINING-BASED BAYESIAN MIMO CHANNEL AND CHANNEL NORM ESTIMATION

*Emil Björnson and Björn Ottersten*

ACCESS Linnaeus Center, Signal Processing Lab, Royal Institute of Technology (KTH),  
SE-100 44 Stockholm, Sweden (*{emil.bjornson,bjorn.ottersten}@ee.kth.se*)

## ABSTRACT

Training-based estimation of channel state information in multi-antenna systems is analyzed herein. Closed-form expressions for the general Bayesian minimum mean square error (MMSE) estimators of the channel matrix and the squared channel norm are derived in a Rayleigh fading environment with known statistics at the receiver side. When the second-order channel statistics are available also at the transmitter, this information can be exploited in the training sequence design to improve the performance. Herein, mean square error (MSE) minimizing training sequences are considered. The structure of the general solution is developed, with explicit expressions at high and low SNRs and in the special case of uncorrelated receive antennas. The optimal length of the training sequence is equal or smaller than the number of transmit antennas.

**Index Terms**— Channel matrix, Squared Frobenius norm, MMSE estimation, Rayleigh fading, Training optimization.

## 1. INTRODUCTION

The performance of wireless communication systems can be drastically improved by using multiple antennas, but the potential gains come with the requirement of accurate channel state information (CSI) at both the transmitter and the receiver [1, 2]. In many practical multiple-input multiple-output (MIMO) systems, the long-term channel statistics can be regarded as known, through reverse-link estimation or a negligible signalling overhead. Instantaneous CSI needs however to be estimated at the receiver and then fed back with finite precision. Channel estimation techniques are commonly divided into three categories: training-based, semi-blind, and blind. In the former, the estimation is entirely based on transmission of training sequences (known a priori to the receiver) [3, 4]. The other extreme is blind estimation, which only exploits some known structure of the received data. The combination of these techniques is called semi-blind estimation. None of the approaches are clearly superior; training improves estimation, but the time and power spent on training is taken away from the actual data transmission. It is indicated in [5] that the semi-blind approach is preferable, in particular in comparison to blind estimation. Herein, we consider training-based estimation, without claiming the approach to be optimal.

By nature, the channel is stochastic, which motivates Bayesian estimation—that is, modeling of the current state as a realization from a multi-variate probability density function (PDF). Hence, the channel statistics need to be known a priori, but this is usually not a limitation (as mentioned above). Yet, there is a large amount of literature on estimation of deterministic MIMO channels, which is more

analytically tractable. The survey in [6] compares the deterministic and Bayesian approach and, as expected, the latter is superior.

Herein, we consider training-based Bayesian minimum mean square error (MMSE) estimation of a Rayleigh fading MIMO channel. The problem of finding the *linear* MMSE estimator was considered in [3, 4], but although they both claim to use a general linear form, we show herein that the structure in [3] is in fact restrictive. The *general* MMSE estimator of the channel is derived herein, and it coincides with the linear estimator in [4] and achieves a lower mean square error (MSE) than [3]. Herein, we also derive the MMSE estimator of the squared Frobenius norm of the channel. This is an extension of [7] and is of great interest since the squared norm describes the signal-to-noise ratio (SNR) of many systems and since indirect calculation from an estimated channel matrix usually is sub-optimal. For both the channel matrix and the squared norm estimators, the structure of the optimal training sequences will be analyzed and explicit expressions will be given at high and low SNRs, as well as in the case of uncorrelated receive antennas. Outlines of some of the proofs are provided herein, while all details are given in [8].

## 2. SYSTEM MODEL

We consider a flat and block-fading MIMO system with a single base station equipped with an array of  $n_T$  transmit antennas. There are one or several mobile users, each with an array of  $n_R$  receive antennas. The symbol-sampled complex baseband equivalent of the narrowband channel to user  $k$  at symbol slot  $t$  is modeled as

$$\mathbf{y}_k(t) = \mathbf{H}_k \mathbf{x}(t) + \mathbf{n}_k(t), \quad (1)$$

where  $\mathbf{x}(t) \in \mathbb{C}^{n_T}$  and  $\mathbf{y}_k(t) \in \mathbb{C}^{n_R}$  are the transmitted and received signals, respectively, and  $\mathbf{n}_k(t) \in \mathbb{C}^{n_R}$  is white complex Gaussian noise with zero-mean and variance  $\mu$ . The channel is represented by  $\mathbf{H}_k \in \mathbb{C}^{n_R \times n_T}$  and modeled as Rayleigh fading with the covariance matrix  $\mathbf{R}_k$ . Thus,  $\text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ , where the vectorization operator  $\text{vec}(\cdot)$  gives the column stacking of a matrix. Most of the results herein are derived for Kronecker-structured covariance matrices, which means that they can be separated as the Kronecker product  $\mathbf{R}_k = \mathbf{R}_{T,k}^T \otimes \mathbf{R}_{R,k}$ . Here,  $\mathbf{R}_{T,k} \in \mathbb{C}^{n_T \times n_T}$  and  $\mathbf{R}_{R,k} \in \mathbb{C}^{n_R \times n_R}$  represent the antenna correlation at the transmitter and the receiver, respectively. The channel and noise statistics are known at both the transmitter and the receiver.

### 2.1. Training-Based Estimation

In this paper, we consider estimation of the channel matrix  $\mathbf{H}_k$  and of the squared channel norm  $\|\mathbf{H}_k\|^2$ , where  $\|\cdot\|^2$  denotes the squared Frobenius norm (i.e., the sum of the squared absolute values of all the matrix elements). The receiver knows the statistics, but in order to estimate a function of the unknown realization of  $\mathbf{H}_k$ , the transmitter typically needs to send a collection of known training vec-

This work is supported in part by the FP6 project Cooperative and Opportunistic Communications in Wireless Networks (COOPCOM), Project Number: FP6-033533. Björn Ottersten is also with securityandtrust.lu, University of Luxembourg.

tors that spans  $\mathbb{C}^{n_T}$ . These vectors can for example be the columns of a unitary  $n_T \times n_T$  matrix. When the total training power  $\mathcal{P}$  is allowed to vary, it was in [9] that optimal number of training vectors is exactly  $n_T$  in spatially uncorrelated systems. In systems with non-identical eigenvalue distribution (i.e., spatial correlation) of the channel covariance  $\mathbf{R}_k$ , even fewer vectors might be optimal [7, 8].

Let the training matrix  $\mathbf{P} \in \mathbb{C}^{n_T \times n_T}$  represent the training sequence and fulfill the power constraint  $\text{tr}(\mathbf{P}^H \mathbf{P}) = \mathcal{P}$ , where the total training power  $\mathcal{P} > 0$  is a design parameter. The columns of this matrix are typically transmitted as  $\mathbf{x}(t)$  at  $n_T$  consecutive symbols slots (e.g.,  $t = 1, \dots, n_T$ ) in (1). The received matrix  $\mathbf{Y}_k = [\mathbf{y}_k(1), \dots, \mathbf{y}_k(n_T)]$  of the training transmission is

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{P} + \mathbf{N}_k, \quad (2)$$

where  $\mathbf{N}_k = [\mathbf{n}_k(1), \dots, \mathbf{n}_k(n_T)]$  is the combined noise matrix.

### 3. MMSE CHANNEL AND NORM ESTIMATION

Next, we consider MMSE estimation of  $\mathbf{H}_k$  and  $\|\mathbf{H}_k\|^2$  at a receiver  $k$  that knows the training sequence  $\mathbf{P}$ , the received signal  $\mathbf{Y}_k$  from the system model in (2), and the channel and noise statistics.

From Bayesian theory [10], the MMSE estimator of a vector  $\mathbf{h}$  from an observation  $\mathbf{y}$  can be expressed as

$$\hat{\mathbf{h}}_{\text{MMSE}} = E\{\mathbf{h}|\mathbf{y}\} = \int \mathbf{h} f(\mathbf{h}|\mathbf{y}) d\mathbf{h}, \quad (3)$$

where  $E\{\cdot\}$  denotes the expected value and  $f(\mathbf{h}|\mathbf{y})$  is the conditional PDF of  $\mathbf{h}$  given  $\mathbf{y}$ . The MMSE estimator minimizes the MSE  $E\{\|\mathbf{h} - \hat{\mathbf{h}}_{\text{MMSE}}\|^2\}$  and the optimum can be calculated as the trace of the covariance of  $f(\mathbf{h}|\mathbf{y})$  averaged over  $\mathbf{y}$ .

#### 3.1. MMSE Channel Estimation

There are many reasons for estimating the channel matrix  $\mathbf{H}_k$  at the receiver. Instantaneous CSI can, for example, be used for receive processing (improved interference suppression and simplified detection) and feedback (to employ beamforming and rate adaptation). The following theorem gives the MMSE estimator of the channel matrix and does not require a Kronecker structured  $\mathbf{R}_k$ .

**Theorem 1.** *The MMSE estimator  $\hat{\mathbf{H}}_{\text{MMSE}}$  of  $\mathbf{H}_k$  with the observation  $\mathbf{Y}_k$ , training sequence  $\mathbf{P}$ , and  $\text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$  is*

$$\text{vec}(\hat{\mathbf{H}}_{\text{MMSE}}) = \mathbf{R}_k \tilde{\mathbf{P}}^H (\tilde{\mathbf{P}} \mathbf{R}_k \tilde{\mathbf{P}}^H + \mu \mathbf{I})^{-1} \text{vec}(\mathbf{Y}_k), \quad (4)$$

where  $\tilde{\mathbf{P}} \triangleq (\mathbf{P}^T \otimes \mathbf{I})$ . The corresponding MSE is

$$E\{\|\text{vec}(\mathbf{H}_k) - \text{vec}(\hat{\mathbf{H}}_{\text{MMSE}})\|^2\} = \text{tr}\left\{\left(\mathbf{R}_k^{-1} + \frac{\tilde{\mathbf{P}}^H \tilde{\mathbf{P}}}{\mu}\right)^{-1}\right\}, \quad (5)$$

where  $\text{tr}\{\cdot\}$  denotes the matrix trace.

*Proof.* The theorem follows from that the conditional distribution of  $\text{vec}(\mathbf{H}_k)$  is  $\mathcal{CN}\left(\left(\frac{\tilde{\mathbf{P}}^H \tilde{\mathbf{P}}}{\mu} + \mathbf{R}_k^{-1}\right)^{-1} \frac{\tilde{\mathbf{P}}^H}{\mu} \text{vec}(\mathbf{Y}_k), \left(\mathbf{R}_k^{-1} + \frac{\tilde{\mathbf{P}}^H \tilde{\mathbf{P}}}{\mu}\right)^{-1}\right)$ , which is shown using Bayes' formula and some identification.  $\square$

**Remark 1.** Observe that the *general* MMSE channel estimator in Theorem 1 is linear; it has the form  $\text{vec}(\hat{\mathbf{H}}_{\text{MMSE}}) = \mathbf{A} \text{vec}(\mathbf{Y})$  and coincides with the *linear* MMSE estimator derived in [4]. Thus, the estimator in (4) is also the linear MMSE estimator for any PDF of  $\text{vec}(\mathbf{H}_k)$  with the same mean and covariance, with (5) as its MSE.

**Remark 2.** The linear estimator proposed in [3] was claimed to be the linear MMSE estimator, but this statement is only valid in the special Kronecker-structured case with uncorrelated receive antennas ( $\mathbf{R}_{R,k} = \lambda^{(R)} \mathbf{I}$ ). In general, the estimator presented in [3] corresponds to a subset of linear estimators that fulfills  $\mathbf{A} = (\mathbf{A}_o^T \otimes \mathbf{I})$ .

#### 3.1.1. MSE Minimizing Training Sequences

The MMSE estimator in (4) and its MSE in (5) depend on the training matrix  $\mathbf{P}$ . Next, we will analyze the impact of the training matrix design on the estimation performance. In multi-user systems where several spatially separated users want to estimate their channels simultaneously, any unitary matrix (scaled to fit the power constraint) is optimal; the explanation is that the training sequence needs to be based on information available at all receivers. When only a single user is active, the training sequence can however be tailored for the spatial properties of the covariance matrix of this user. The intuition is that more power should be allocated to estimate the channel along strong eigenmodes than along weak eigenmodes.

The next theorem characterizes the MSE minimizing training matrix  $\mathbf{P}$ , and the user indices have been dropped for brevity.

**Theorem 2.** *Let the covariance matrix  $\mathbf{R}$  fulfill the Kronecker model  $\mathbf{R} = \mathbf{R}_T^T \otimes \mathbf{R}_R$ , and let  $\mathbf{R}_T = \mathbf{U}_T \mathbf{\Lambda}_T \mathbf{U}_T^H$  be the eigenvalue decomposition of  $\mathbf{R}_T$ . The training matrix  $\mathbf{P}$  that minimizes the MSE in (5) has the structure  $\mathbf{P} = \mathbf{U}_T \mathbf{\Sigma} \mathbf{V}^H$ , where  $\mathbf{V}$  is an arbitrary unitary matrix and  $\mathbf{\Sigma} = \text{diag}(\sqrt{\sigma_1}, \dots, \sqrt{\sigma_{n_T}})$ . When  $\mathbf{P}$  satisfies this structure, the MSE is convex in the training powers  $\sigma_j$ .*

*Let the eigenvalues of  $\mathbf{R}_T$  and  $\mathbf{R}_R$  be denoted  $\lambda_1^{(T)}, \dots, \lambda_{n_T}^{(T)}$  and  $\lambda_1^{(R)}, \dots, \lambda_{n_R}^{(R)}$ , respectively. Then, the MSE minimizing training power allocation  $\sigma_1, \dots, \sigma_{n_T}$  is ordered in the same way as  $\lambda_1^{(T)}, \dots, \lambda_{n_T}^{(T)}$  and is given by the following system of equations:*

$$\sum_{i=1}^{n_R} \frac{\mu (\lambda_j^{(T)} \lambda_i^{(R)})^2}{(\mu + \sigma_j \lambda_j^{(T)} \lambda_i^{(R)})^2} = \alpha, \quad (6)$$

for all  $j$  such that  $\alpha < \sum_i (\lambda_j^{(T)} \lambda_i^{(R)})^2 / \mu$  and  $\sigma_j = 0$  otherwise. The parameter  $\alpha \geq 0$  is chosen to fulfill the constraint  $\sum_j \sigma_j = \mathcal{P}$ .

Let  $\tilde{n} = \text{rank}(\mathbf{R}_T)$  be the number of non-zero eigenvalues of  $\mathbf{R}_T$  and let  $\tilde{m}$  be the multiplicity of its largest eigenvalue. Then, the asymptotic solution at high SNR (defined as  $\mathcal{P}/\mu$ ) is  $\sigma_j = \mathcal{P}/\tilde{n}$  for all  $j$  such that  $\lambda_j^{(T)} > 0$  and  $\sigma_j = 0$  for all  $j$  such that  $\lambda_j^{(T)} = 0$  (i.e., equal power allocation among all non-zero eigenmodes). The asymptotic solution at low SNR is  $\sigma_j = \mathcal{P}/\tilde{m}$  for all  $j$  such that  $\lambda_j^{(T)} = \max_i \lambda_i^{(T)}$  and  $\sigma_j = 0$  otherwise (i.e., selective allocation to the strongest eigenmode, with multiplicity).

In the case of uncorrelated receive antennas,  $\mathbf{R}_R = \lambda^{(R)} \mathbf{I}$ , the waterfilling solution of (6) transforms into an explicit expression:

$$\sigma_j = \sqrt{\frac{\mu}{\alpha} - \frac{\mu}{\lambda_j^{(T)} \lambda^{(R)}}} \quad \text{for } \alpha < (\lambda_j^{(T)} \lambda^{(R)})^2 / \mu, \quad (7)$$

and  $\sigma_j = 0$  otherwise, where  $\alpha$  should fulfill the power constraint.

**Remark 3.** The structure of the MSE minimizing training matrix was proved in [4], along with somewhat similar asymptotic results. The optimal power allocation for the special case of uncorrelated receive antennas coincides with that proposed in [3]. With arbitrary correlation, the restrictions made in [3] leads to simpler training power expressions, but also to suboptimal performance (see Section 4).

**Remark 4.** Observe that the MSE minimizing training power allocation in Theorem 2 has the statistical waterfilling structure, and that it depends on the available training power  $\mathcal{P}$  and the spatial correlation. When the correlation is pronounced, we can expect some of the powers  $\sigma_j$  to be zero. Say that  $m < n_T$  of them are non-zero, then there is no reason for the training sequence to contain more than  $m$  vectors. This stands in contrast to [9] that states that the optimal number *always* is equal to  $n_T$ , without taking the spatial correlation into account. In the special case of uncorrelated receive antennas, the optimal training length  $m$  can be derived explicitly [8].

### 3.2. MMSE Channel Norm Estimation

Next, we consider MMSE estimation of the squared channel norm  $\|\mathbf{H}_k\|^2$ . This norm corresponds directly to the SNR in space-time coded systems and has a large impact on the SNR in many other types of systems [11]. Based on the structure of the solution in Theorem 2, we limit the analysis to training matrices whose left singular vectors coincide with the eigenvectors of the transmit-side covariance  $\mathbf{R}_{T,k}$ :  $\mathbf{P} = \mathbf{U}_{T,k} \boldsymbol{\Sigma} \mathbf{V}^H$ . It is our conjecture that the MSE minimizing training matrix for norm estimation has this form, exactly as in the channel estimation case. For this type of eigen-training matrices, the following theorem gives the MMSE estimator.

**Theorem 3.** Let  $\mathbf{H}_k$  have the distribution  $\text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ , where the eigenvalue decomposition of the Kronecker-structured covariance matrix is  $\mathbf{R}_k = (\mathbf{U}_{T,k}^T \otimes \mathbf{U}_{R,k}) \boldsymbol{\Lambda}_k (\mathbf{U}_{T,k}^T \otimes \mathbf{U}_{R,k})^H$ . The MMSE estimator  $\widehat{\rho}_k$  of  $\rho_k \triangleq \|\mathbf{H}_k\|^2$ , with the observation  $\mathbf{Y}_k$  and the training sequence  $\mathbf{P} = \mathbf{U}_{T,k} \boldsymbol{\Sigma} \mathbf{V}^H$ , is

$$\widehat{\rho}_k = \mu \mathbf{1}^T \mathbf{B}_k \mathbf{1} + \widetilde{\mathbf{y}}_k^H \widetilde{\boldsymbol{\Sigma}} \widetilde{\mathbf{B}}_k^2 \widetilde{\boldsymbol{\Sigma}}^H \widetilde{\mathbf{y}}_k, \quad (8)$$

where  $\widetilde{\mathbf{y}}_k \triangleq \text{vec}(\mathbf{U}_{R,k}^H \mathbf{Y}_k \mathbf{V})$ ,  $\mathbf{B}_k \triangleq \boldsymbol{\Lambda}_k (\widetilde{\boldsymbol{\Sigma}} \boldsymbol{\Lambda}_k \widetilde{\boldsymbol{\Sigma}}^H + \mu \mathbf{I})^{-1}$ ,  $\widetilde{\boldsymbol{\Sigma}} \triangleq (\boldsymbol{\Sigma}^T \otimes \mathbf{I})$ , and  $\mathbf{1} \triangleq [1, \dots, 1]^T$ . The corresponding MSE is

$$E\{|\rho_k - \widehat{\rho}_k| \}^2 = \mathbf{1}^T \mathbf{B}_k (2\mu \widetilde{\boldsymbol{\Sigma}} \boldsymbol{\Lambda}_k \widetilde{\boldsymbol{\Sigma}}^H + \mu^2 \mathbf{I}) \mathbf{B}_k \mathbf{1}. \quad (9)$$

*Proof.* In the single-antenna case, we achieve the conditional PDF

$$f(\rho_k | \mathbf{Y}_k) = \frac{\mathcal{P} \lambda + \mu}{\lambda \mu} e^{-\rho_k \frac{\mathcal{P} \lambda + \mu}{\lambda \mu}} e^{-\varrho_y \frac{\mathcal{P} \lambda}{\mu(\mathcal{P} \lambda + \mu)}} I_0\left(\frac{2}{\mu} \sqrt{\mathcal{P} \rho_k \varrho_y}\right),$$

using Bayes' formula and some calculations. Here,  $\varrho_y = \|\mathbf{Y}_k\|^2$  and  $I_\nu(\cdot)$  is the modified Bessel function of the first kind. The estimator in the MIMO case follows by integration and by separating the problem into  $n_T n_R$  independent single-antenna problems.  $\square$

#### 3.2.1. MSE Minimizing Training Sequences

Next, we consider MSE minimization based on the training sequence  $\mathbf{P} = \mathbf{U}_{T,k} \boldsymbol{\Sigma} \mathbf{V}^H$  used in Theorem 3. There are two types of sequences that can be represented by a matrix like this. In a system with multiple active users,  $\mathbf{P}$  cannot depend on any of the users. Hence, we need  $\boldsymbol{\Sigma} = \sqrt{\mathcal{P}/n_T} \mathbf{I}$ , which means that  $\mathbf{P}$  becomes a (scaled) arbitrary unitary matrix. If only a single user is active, we are allowed to choose any  $\boldsymbol{\Sigma} = \text{diag}(\sqrt{\sigma_1}, \dots, \sqrt{\sigma_{n_T}})$  that fulfills the power constraint  $\text{tr}(\mathbf{P} \mathbf{P}^H) = \text{tr}(\boldsymbol{\Sigma} \boldsymbol{\Sigma}^H) = \mathcal{P}$ . As for the channel estimator, we therefore use the training sequence to minimize the MSE. Contrary to the optimization in Section 3.1, the MSE of the norm estimator is not convex in the training powers, which makes it difficult to derive a closed-form solution. The following theorem will however give the asymptotic solution at high and low SNRs, as well as explicit expressions for the case of uncorrelated receive antennas. The user indices have been dropped for brevity.

**Theorem 4.** Let the eigenvalues of  $\mathbf{R}_T$  and  $\mathbf{R}_R$  be denoted  $\lambda_1^{(T)}, \dots, \lambda_{n_T}^{(T)}$  and  $\lambda_1^{(R)}, \dots, \lambda_{n_R}^{(R)}$ , respectively. Then, the MSE minimizing training power allocation  $\sigma_1, \dots, \sigma_{n_T}$  is given as one of the solutions to the following system of equations:

$$\sum_{l=1}^{n_R} \frac{2\sigma_j \mu (\lambda_j^{(T)} \lambda_l^{(R)})^4}{(\mu + \sigma_j \lambda_j^{(T)} \lambda_l^{(R)})^3} = \alpha, \quad (10)$$

for all  $\sigma_j > 0$  (among  $j = 1, \dots, n_T$ ) and  $\sigma_j = 0$  otherwise. The parameter  $\alpha \geq 0$  should fulfill the power constraint  $\sum_j \sigma_j = \mathcal{P}$ .

The asymptotic solution at high SNR is  $\sigma_j = \mathcal{P} \sqrt{\lambda_j^{(T)}} / \sum_i \sqrt{\lambda_i^{(T)}}$  for all  $j$  (i.e., proportional allocation with  $\lambda_j^{(T)/2}$  as the proportionality coefficient). The asymptotic solution at low SNR is  $\sigma_j = \mathcal{P}$  for some  $j$  such that  $\lambda_j^{(T)} = \max_i \lambda_i^{(T)}$  and  $\sigma_j = 0$  otherwise (i.e., selective allocation to one of the strongest eigenmodes).

In the case of uncorrelated receive antennas,  $\mathbf{R}_R = \lambda^{(R)} \mathbf{I}$ , the MSE minimizing solution is given by either  $\sigma_j = 0$  or

$$\sigma_j(m) = \sqrt{\frac{8\mu \lambda_j^{(T)} \lambda^{(R)}}{3\alpha}} \cos\left(\frac{\pi(-1)^m - \phi}{3}\right) - \frac{\mu}{\lambda_j^{(T)} \lambda^{(R)}}, \quad (11)$$

for  $m = 0, 1$ , where  $\phi = \arctan \sqrt{\frac{8(\lambda_j^{(T)} \lambda^{(R)})^3}{27\mu\alpha} - 1}$  and  $\alpha \geq 0$  is chosen to fulfill the power constraint. The latter two potential solutions only exist when  $\alpha \leq 8(\lambda_j^{(T)} \lambda^{(R)})^3 / (27\mu)$  and represents solutions in different intervals:  $0 \leq \sigma_j(1) \leq \frac{\mu}{2\lambda_j^{(T)} \lambda^{(R)}} \leq \sigma_j(0) < \infty$ .

Assume, without loss of generality, that the eigenvalues  $\lambda_j^{(T)}$  are ordered non-increasingly. Then,  $\sigma_j \geq \sigma_i$  for all  $j < i$ . Thus, if  $\sigma_i$  is given by  $m=0$  in (11), then  $\sigma_j$  is also given by  $m=0$  for  $1 \leq j < i$ .

*Remark 5.* Although the MSE is non-convex, the theorem shows that the asymptotic solutions (of the special type  $\mathbf{P} = \mathbf{U}_{T,k} \boldsymbol{\Sigma} \mathbf{V}^H$ ) at high and low SNRs can be derived explicitly. Observe that the power allocation is similar with that for channel matrix estimation at low SNR, while the asymptotic behaviors are different at high SNR. In the case of uncorrelated receive antennas, the theorem shows that the explicit solution lies in a closed set with a cardinality that scales with  $n_T$  as  $2^{n_T}$ , which becomes the worst case complexity.

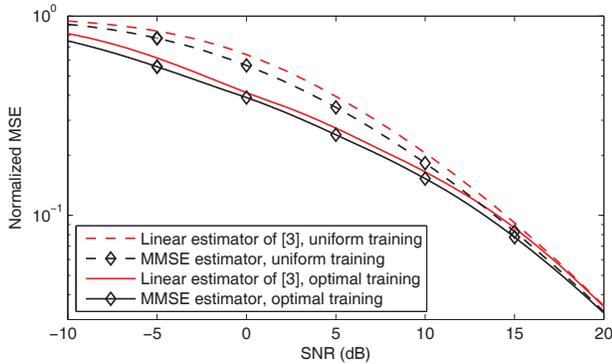
*Remark 6.* Under the additional constraint that  $\sigma_j \geq \frac{\mu}{2\lambda_j^{(T)} \lambda_l^{(R)}}$  for all  $j, l$ , the MSE becomes convex. Thus, the system of equations in (10) has a unique solution, which in the case of  $\mathbf{R}_R = \lambda^{(R)} \mathbf{I}$  is given by  $m = 0$  in (11) for all  $\sigma_j$  larger than the new lower bound.

## 4. NUMERICAL EXAMPLES

In this section, the performance of the proposed MMSE estimators of the channel matrix and the squared channel norm will be illustrated numerically. We consider a Kronecker-structured system where the transmitter and the receiver are equipped with four antennas each. The antenna correlation follows the exponential model [12], which in principle models a uniform linear array (ULA) with the correlation between adjacent antennas as a parameter. The antenna correlation at the transmitter and receiver side is fixed at 0.8 and 0.6, respectively.

The normalized MSEs, defined as  $E\{\|\mathbf{H} - \widehat{\mathbf{H}}_{\text{MMSE}}\|^2\} / \text{tr}(\mathbf{R})$ , of different channel estimators are given in Fig. 1 as a function of the SNR, defined as  $\text{SNR} = \mathcal{P}/\mu$ . The simulation compares the performance of the MMSE channel estimator derived in Theorem 1 with the linear estimator proposed in [3], for the two cases of a uniform training sequence ( $\mathbf{P} = \sqrt{\mathcal{P}/n_T} \mathbf{I}$ ) and MSE minimizing training. The latter estimator has previously been claimed to minimize the MSE, but it is clear from Fig. 1 that the correct MMSE estimator derived herein gives a better MSE performance; the difference is non-negligible, both with and without training optimization. It is also clear that training optimization can improve the performance considerably at low SNR, while the advantage disappears asymptotically at high SNR. This confirms the result in Theorem 2 that uniform training is asymptotically optimal at high SNR.

In Fig. 2, the normalized MSEs, defined as  $E\{\|\mathbf{H}\|^2 - \widehat{\|\mathbf{H}\|_{\text{MMSE}}}\|^2\} / \text{tr}(\mathbf{R} \mathbf{R}^H)$ , for different squared norm estimators



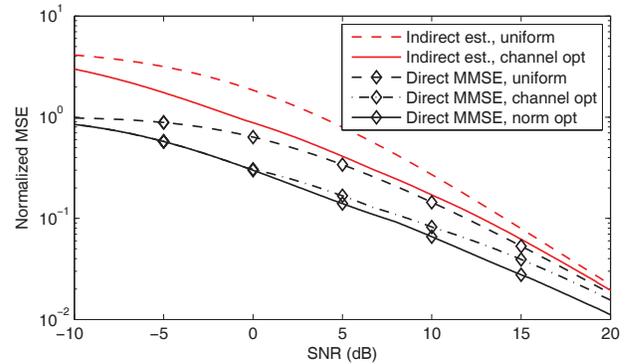
**Fig. 1.** The normalized MSEs of channel matrix estimation as a function of the SNR for two estimators: the linear estimator of [3] and the optimal MMSE estimator derived herein. The two cases of uniform training and MSE minimizing training are considered.

are given as a function of the SNR. The simulation compares the performance of the MMSE estimator derived in Theorem 3 with the performance achieved by first estimating the channel matrix using Theorem 1 and then calculating the squared norm. In the latter case, uniform and channel matrix MSE minimizing training are considered. For the MMSE estimator, three different training sequences are considered: uniform, channel matrix MSE minimizing, and channel norm MSE minimizing.

The first observation from Fig. 2 is that the indirect approach gives poor performance at low SNR (even worse than the statistical estimator  $\hat{\rho}_{\text{stat}} = \text{tr}\{\mathbf{R}\}$  that would give unit normalized MSE), while it approaches the performance of the MMSE estimator at high SNR. It is clear that the performance can be considerably improved by training optimization. Obviously, the best performance is achieved by optimizing over the MSE, but the MMSE estimator achieved with a training sequence optimized for channel matrix estimation gives a considerable performance gain in comparison with uniform training (especially at low SNR, while both have suboptimal asymptotic behaviors at high SNR). This is probably the most important case in practice; the training sequence will be used to maximize the knowledge of the channel matrix at the receiver, but the received training signal can simultaneously be used to calculate an MMSE estimate of the squared channel norm (e.g., for the purpose of feedback).

## 5. CONCLUSIONS

In this paper, training-based estimation of the channel matrix and the squared Frobenius norm of the channel have been studied for Rayleigh fading MIMO systems. By assuming that the channel statistics are known at the receiver, the MMSE estimators and their corresponding MSEs were derived in closed-form for general training sequences (a condition was required for the squared norm). When the statistics also are known at the transmitter, the training sequences may be optimized to minimize the MSE and the optimal sequence length will be smaller or equal to the number of transmit antennas. The MSE minimizing training sequence for channel matrix estimation was derived and can be expressed explicitly at high and low SNRs and for uncorrelated receive antennas. For channel norm estimation, the MSE is non-convex, but the asymptotic solutions were derived and expressions were given for low-complexity computation of the MSE minimizing training sequence for uncorrelated receive antennas and in a special case when an additional power constraint was imposed. Finally, the performance of the general MMSE channel estimators were illustrated numerically.



**Fig. 2.** The normalized MSEs of channel norm estimation as a function of the SNR for two estimators: the direct MMSE estimator derived herein and indirect estimation from an MMSE estimated channel matrix. The performance of the direct estimator is shown with uniform, channel matrix MSE minimizing, and channel norm MSE minimizing training. The performance of the indirect estimator is shown with uniform and channel matrix MSE minimizing training.

## 6. REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecom.*, vol. 10, pp. 585–595, 1999.
- [3] M. Biguesh and A.B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884–893, 2006.
- [4] J.H. Kotecha and A.M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 546–557, 2004.
- [5] E. De Carvalho and D.T.M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE SPAWC'97*, 1997.
- [6] F.A. Dietrich and W. Utschick, "Pilot-assisted channel estimation based on second-order statistics," *IEEE Trans. Signal Process.*, vol. 53, pp. 1178–1193, 2005.
- [7] E. Björnson and B. Ottersten, "Pilot-based Bayesian channel norm estimation in Rayleigh fading multi-antenna systems," in *Proc. Nordic Radio Science and Commun. (RVK'08)*, 2008.
- [8] E. Björnson and B. Ottersten, "A unified framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, Submitted for publication.
- [9] B. Hassibi and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, 2003.
- [10] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.
- [11] E. Björnson and B. Ottersten, "Post-user-selection quantization and estimation of correlated Frobenius and spectral channel norms," in *Proc. IEEE PIMRC'08*, 2008.
- [12] S.L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, pp. 369–371, 2001.