Training-Based MIMO Channel Estimation: A Study of Estimator Tradeoffs and Optimal Training Signals

Mehrzad Biguesh, *Member, IEEE,* and Alex B. Gershman, *Fellow, IEEE*

***Abstract*—In this paper, we study the performance of multiple-input multiple-output channel estimation methods using training sequences. We consider the popular linear least squares (LS) and minimum mean-square-error (MMSE) approaches and propose new scaled LS (SLS) and relaxed MMSE techniques which require less knowledge of the channel second-order statistics and/or have better performance than the conventional LS and MMSE channel estimators. The optimal choice of training signals is investigated for the aforementioned techniques. In the case of multiple LS channel estimates, the best linear unbiased estimation (BLUE) scheme for their linear combining is developed and studied.**

***Index Terms*—Multiple-input multiple-output (MIMO) channel estimation, optimal training signals.**

I. Introduction

Because of the growing demand for high data rates in wireless communication systems, array-based transceivers and space diversity methods have recently become an intensive area of research [1]–[7]. It has been shown both analytically and using field tests that in rich scattering environments, multiple-input multiple-output (MIMO) techniques can greatly increase the capacity of wireless systems [2], [3], [6].

However, to use the advantages that MIMO systems can offer, an accurate channel state information (CSI) is required at the transmitter and/or receiver. For example, the performance of transmit beamforming is entirely determined by the accuracy of the CSI at the transmitter. If space-time coding is used, then the availability of an accurate CSI at the receiver is crucial for the performance of space-time decoders. Therefore, an accurate channel estimation plays a key role in MIMO communications [8]–[10].

One of the most popular and widely used approaches to the MIMO channel estimation is to employ pilot signals (also referred to as training sequences) and then to estimate the channel based on the received data and the knowledge of training symbols.

Manuscript received August 7, 2004; revised April 28, 2005. This work was supported in part by the Wolfgang Paul Award Program of the Alexander von Humboldt Foundation, Germany, and the German Ministry of Education and Research; in part by the Premier's Research Excellence Award Program of the Ministry of Energy, Science, and Technology of Ontario; in part by the Natural Sciences and Engineering Research Council of Canada; and in part by Communication and Information Technology Ontario. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Behrouz Farhang-Boroujeny.

M. Biguesh is with the Department of Electrical and Electronic Engineering, Shiraz University, Shiraz, Iran.

A. B. Gershman was with McMaster University, Hamilton, ON, Canada. He is now with the Communication Systems Group, Darmstadt University of Technology, Darmstadt D-64283, Germany.

Digital Object Identifier 10.1109/TSP.2005.863008

Recently, there has been a growing interest in training-based MIMO channel estimation. In [11], the problem of training sequence design for MIMO channels has been linked with the channel capacity. There are several works where different training schemes are considered for both the flat-fading and frequency-selective MIMO cases [12]–[17]. For example, in [12], the maximum likelihood (ML) method has been considered for BLAST training, and orthogonal pilot signals have been investigated. Further study of this estimator is reported in [13]. In [14], a least squares (LS) training-based channel estimation technique for orthogonal frequency-division multiplexing systems with multiple transmit antennas is developed. In [15], the conventional LS channel estimate is improved using the minimum mean-square-error (MMSE) symbol estimate, and training design issues are discussed. In [16], optimal choices of training signals are investigated in the case of multiple transmit antennas and single receive antenna in application to several training-based channel estimation schemes, including the linear LS and MMSE estimators. In [17], a pilot symbol aided modulation [18] is used to estimate doubly selective fading channels, and an MMSE-based training scheme with orthogonal training is considered. In [19], a general discussion on optimal MIMO training schemes is given based on the LS criterion. Blind and semiblind MIMO channel estimation techniques are discussed in [20] as alternatives to the training-based channel estimation.

In this paper (see also [21] and [22]), we extend the results of [16] to the MIMO case and study the performance of training-based flat block-fading MIMO channel estimation. Four training-based channel estimators are considered, which offer different tradeoffs in terms of performance and a priori required knowledge of the channel second-order statistics.

First, the traditional LS method is considered, which does not require any knowledge about the channel parameters. Then, a refined version of the LS estimator is developed, which is referred to as the scaled LS (SLS) estimator. It is shown that the proposed SLS estimator offers a substantially improved performance relative to the LS method but requires that the ratio of the trace of a specifically defined matrix of channel correlations and the receiver noise power be known a priori.

Then, the linear MMSE channel estimator is studied. The latter technique is shown to be able to outperform both the LS and SLS estimators, but it requires the full a priori knowledge of the aforementioned matrix of channel correlations and the receiver noise power.

Finally, the relaxed MMSE (RMMSE) method is introduced, which represents a simplified and approximate version of the linear MMSE method that requires only the knowledge of the trace of the matrix of channel correlations and the receiver noise power.

For the LS, SLS, and MMSE techniques, the optimal choices of training matrices are studied and the channel estimation errors are analyzed.

Moreover, in the case of multiple LS channel estimates, an optimal scheme for their linear combining is developed using the so-called best linear unbiased estimation (BLUE) approach. The effect of such a combining on the performance of channel matrix estimation is studied.

II. Background

Let us consider a flat block-fading MIMO system with t transmit and r receive antennas. The $r \times 1$ complex received signal vector can be expressed as [19]

$$\mathbf{s}_i = \mathbf{H}\mathbf{p}_i + \mathbf{v}_i \tag{1}$$

where $\mathbf{H}$ is the $r \times t$ complex random channel matrix, $\mathbf{p}_i$ is the $t \times 1$ complex vector of the transmitted signals, and $\mathbf{v}_i$ is the $r \times 1$ complex zero-mean white noise vector.

It should be stressed that in any statistical expectation below, the matrix $\mathbf{H}$ is treated as random. At the same time, any estimator of $\mathbf{H}$ is supposed to obtain an estimate of a particular *realization* of this random matrix that corresponds to the current block of the received data.

In order to estimate the channel matrix $\mathbf{H}$, let $N \geq t$ training signal vectors $\mathbf{p}_1, \ldots, \mathbf{p}_N$ be transmitted. The corresponding $r \times N$ matrix $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$ of the received signals can be expressed as

$$\mathbf{S} = \mathbf{H}\mathbf{P} + \mathbf{V} \tag{2}$$

where

$$\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_N] \tag{3}$$

is the $t \times N$ *training matrix* and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ is the $r \times N$ matrix of sensor noise.

The task of a channel estimation algorithm is to recover the channel matrix $\mathbf{H}$ based on the knowledge of $\mathbf{S}$ and $\mathbf{P}$.

III. LS Channel Estimator

Knowing $\mathbf{P}$ and received data, the realization of the channel matrix can be estimated using the LS approach as [25]

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{S}\mathbf{P}^{\dagger} \tag{4}$$

where $\mathbf{P}^{\dagger} = \mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}$ is the pseudoinverse of $\mathbf{P}$ and $(\cdot)^H$ denotes the Hermitian transpose. We will use the following transmitted training power constraint:

$$\|\mathbf{P}\|_F^2 = \mathcal{P} \tag{5}$$

where $\mathcal{P}$ is a given constant value and $\|\cdot\|_F$ is the Frobenius matrix norm. Let us find $\mathbf{P}$ which minimizes the channel estimation error subject to the transmitted power constraint (5). This is equivalent to the following optimization problem:

$$\min_{\mathbf{P}} \mathrm{E}\left\{\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LS}}\|_F^2\right\} \quad \text{subject to} \quad \|\mathbf{P}\|_F^2 = \mathcal{P}. \tag{6}$$

Using (2) and (4), we have that $\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{V}\mathbf{P}^{\dagger}$ and, hence, the objective function in (6) can be rewritten as

$$\begin{aligned} J_{\mathrm{LS}} &= \mathrm{E}\left\{\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LS}}\|_F^2\right\} \\ &= \mathrm{E}\left\{\|\mathbf{V}\mathbf{P}^{\dagger}\|_F^2\right\} \\ &= \sigma_n^2 r \,\mathrm{tr}\{\mathbf{P}^{\dagger H}\mathbf{P}^{\dagger}\} \\ &= \sigma_n^2 r \,\mathrm{tr}\{(\mathbf{P}\mathbf{P}^H)^{-1}\} \end{aligned} \tag{7}$$

where we have used the fact that $\mathrm{E}\{\mathbf{V}^H\mathbf{V}\} = \sigma_n^2 r\mathbf{I}$. Here, σ_n^2 is the receiver noise power, $\mathbf{I}$ is the identity matrix, and $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix.

Using (7), the optimization problem (6) can be equivalently written in the following form:

$$\min_{\mathbf{P}} \mathrm{tr}\{(\mathbf{P}\mathbf{P}^H)^{-1}\} \quad \text{subject to} \quad \mathrm{tr}\{\mathbf{P}\mathbf{P}^H\} = \mathcal{P}. \tag{8}$$

It can be straightforwardly shown[1] that any training matrix is optimal for (8) if it satisfies the equation

$$\mathbf{P}\mathbf{P}^H = \frac{\mathcal{P}}{t}\mathbf{I}. \tag{9}$$

Therefore, any training matrix with *orthogonal rows* of the same norm $\sqrt{\mathcal{P}/t}$ is optimal. Similar facts have been earlier noted for various cases and from different points of view in [12], [16], [17], [19], and [27].

From (9) it follows that there is an infinite number of choices of the optimal training matrix and that each such choice is *receiver-independent*. Hence, any training matrix that satisfies (9) is optimal for *all receivers*.

Additional constraints on $\mathbf{P}$ may be dictated by particular implementation issues. For example, if the peak transmitted power per antenna is limited, all the elements of the optimal training matrix should have the same magnitude. To satisfy this constraint, a properly normalized submatrix of the discrete Fourier transform (DFT) matrix can be used [16]

$$\mathbf{P} = \sqrt{\frac{\mathcal{P}}{Nt}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_N & \cdots & W_N^{N-1} \\ \vdots & \vdots & & \vdots \\ 1 & W_N^{t-1} & \cdots & W_N^{(t-1)(N-1)} \end{bmatrix} \tag{10}$$

where $W_N = e^{j2\pi/N}$.

For optimal training which satisfies (9), the LS channel estimate (4) yields

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \frac{t}{\mathcal{P}}\mathbf{S}\mathbf{P}^H = \mathbf{H} + \frac{t}{\mathcal{P}}\mathbf{V}\mathbf{P}^H \tag{11}$$

i.e., the estimation error is $(t/\mathcal{P})\mathbf{V}\mathbf{P}^H$.

Using (9) along with (7), we obtain that the channel estimation error under optimal training is given by

$$\min_{\mathbf{P}} J_{\mathrm{LS}} = \frac{\sigma_n^2 t^2 r}{\mathcal{P}}. \tag{12}$$

It is noteworthy that the error in (12) is proportional to the square of t. This may cause a certain restriction of the number of transmit antennas as compared to the number of receive antennas used.

[1] See Section V where the solution to a more general problem than (8) is derived.

IV. Scaled LS Channel Estimation

The LS estimate (4) does not necessarily lead to the estimate of $\mathbf{H}$ with the lowest MSE [23]. Therefore, it is meaningful to optimally scale the LS channel estimate to further reduce the channel estimation error. Scaled LS estimators that further reduce the MSE by allowing for a bias is a common approach that has emerged in statistics [23], [24].

Using this idea and (4), we express the channel estimation error in the following form:

$$\begin{aligned} &\mathrm{E}\left\{\|\mathbf{H}-\gamma\hat{\mathbf{H}}_{\mathrm{LS}}\|_F^2\right\} \\ &\quad = \mathrm{tr}\{\mathrm{E}\{(\mathbf{H}-\gamma\hat{\mathbf{H}}_{\mathrm{LS}})^H(\mathbf{H}-\gamma\hat{\mathbf{H}}_{\mathrm{LS}})\}\} \\ &\quad = (1-\gamma)^2\mathrm{tr}\{\mathbf{R_H}\} \\ &\qquad + \gamma^2\sigma_n^2 r\,\mathrm{tr}\{(\mathbf{PP}^H)^{-1}\} \\ &\quad = (J_{\mathrm{LS}}+\mathrm{tr}\{\mathbf{R_H}\}) \\ &\qquad \cdot\left(\gamma-\frac{\mathrm{tr}\{\mathbf{R_H}\}}{J_{\mathrm{LS}}+\mathrm{tr}\{\mathbf{R_H}\}}\right)^2 \\ &\qquad + \frac{J_{\mathrm{LS}}\mathrm{tr}\{\mathbf{R_H}\}}{J_{\mathrm{LS}}+\mathrm{tr}\{\mathbf{R_H}\}} \end{aligned} \tag{13}$$

where $\hat{\mathbf{H}}_{\mathrm{LS}}$ is the LS channel estimate (4), γ is the scaling factor, $\mathbf{R_H} = \mathrm{E}\{\mathbf{H}^H\mathbf{H}\}$ is the matrix of channel correlations,[2] and J_{LS} is given by (7). Clearly, (13) is minimized with

$$\gamma_{\circ} = \frac{\mathrm{tr}\{\mathbf{R_H}\}}{J_{\mathrm{LS}}+\mathrm{tr}\{\mathbf{R_H}\}} \tag{14}$$

and the minimum of (13) with respect to γ is given by

$$\begin{aligned} J_{\mathrm{SLS}} &= \min_{\gamma}\mathrm{E}\left\{\|\mathbf{H}-\gamma\hat{\mathbf{H}}\|_F^2\right\} \\ &= \mathrm{E}\left\{\|\mathbf{H}-\gamma_{\circ}\hat{\mathbf{H}}\|_F^2\right\} \\ &= \frac{J_{\mathrm{LS}}\mathrm{tr}\{\mathbf{R_H}\}}{J_{\mathrm{LS}}+\mathrm{tr}\{\mathbf{R_H}\}} < J_{\mathrm{LS}} \end{aligned} \tag{15}$$

which means that the SLS estimation error is always lower than the LS estimation error. Note that the difference between these errors becomes especially pronounced when the channel is weak or the transmitted power is small, i.e., when $\mathrm{tr}\{\mathbf{R_H}\} \ll \sigma_n^2 t^2 r/\mathcal{P}$.

Assuming that the values of $\mathrm{tr}\{\mathbf{R_H}\}$ and σ_n^2 are given in advance and using (4), (7), and (14), we obtain that the SLS channel estimate can be written as

$$\begin{aligned} \hat{\mathbf{H}}_{\mathrm{SLS}} &= \gamma_{\circ}\hat{\mathbf{H}}_{\mathrm{LS}} \\ &= \frac{\mathrm{tr}\{\mathbf{R_H}\}}{\sigma_n^2 r\,\mathrm{tr}\{(\mathbf{PP}^H)^{-1}\}+\mathrm{tr}\{\mathbf{R_H}\}}\mathbf{SP}^{\dagger}. \end{aligned} \tag{16}$$

Note that the SLS estimator (16) is a function of the ratio $\mathrm{tr}\{\mathbf{R_H}\}/\sigma_n^2$. Therefore, this ratio has to be known (or preliminary estimated) when using the SLS approach.

Obviously, the requirement of knowing $\mathrm{tr}\{\mathbf{R_H}\}$ is less restrictive than that of knowing $\mathbf{R_H}$, which, in turn, is less restrictive than that of knowing $\mathbf{H}$ itself. In practice, the need of knowing $\mathrm{tr}\{\mathbf{R_H}\}$ can be avoided in the SLS estimator by means of using the following LS-based *consistent* sample estimate:

$$\mathrm{tr}\{\widehat{\mathbf{R_H}}\} = \mathrm{tr}\left\{\hat{\mathbf{H}}_{\mathrm{LS}}^H\hat{\mathbf{H}}_{\mathrm{LS}}\right\} \tag{17}$$

in (14) instead of $\mathrm{tr}\{\mathbf{R_H}\}$. The resulting estimator will be referred to as the LS-SLS estimator.

In the case of orthogonal training (9), we have

$$\mathrm{tr}\{\widehat{\mathbf{R_H}}\} = \frac{t^2}{\mathcal{P}^2}\mathrm{tr}\{\mathbf{P}^H\mathbf{SS}^H\mathbf{P}\} = \frac{t}{\mathcal{P}}\mathrm{tr}\{\mathbf{SS}^H\}. \tag{18}$$

The optimal training matrix for the SLS channel estimation method can be found by solving the following constrained optimization problem:

$$\min_{\mathbf{P}} J_{\mathrm{SLS}} \quad \text{subject to} \quad \mathrm{tr}\{\mathbf{PP}^H\} = \mathcal{P}. \tag{19}$$

Since $\mathrm{tr}\{\mathbf{R_H}\} > 0$, from (15) we see that J_{SLS} is a monotonically increasing function of J_{LS}. Note that $\mathrm{tr}\{\mathbf{R_H}\}$ is not a function of $\mathbf{P}$, therefore, J_{LS} is the only term in (15) which depends on $\mathbf{P}$. This means that the optimization problems (19) and (6) are *equivalent*. Therefore, the optimal choice of training matrix for the SLS channel estimator is the same as for the LS approach.

Using (12) and (15), we obtain that the MSE of the SLS estimator (16) under the optimal probing is given by

$$\min_{\mathbf{P}} J_{\mathrm{SLS}} = \frac{\sigma_n^2 t^2 r\,\mathrm{tr}\{\mathbf{R_H}\}}{\sigma_n^2 t^2 r + \mathcal{P}\mathrm{tr}\{\mathbf{R_H}\}}. \tag{20}$$

According to (20), increasing the number of transmit and/or receive antennas, we have

$$\lim_{r\to\infty}\min_{\mathbf{P}} J_{\mathrm{SLS}} = \lim_{t\to\infty}\min_{\mathbf{P}} J_{\mathrm{SLS}} = \mathrm{tr}\{\mathbf{R_H}\} \tag{21}$$

which means that asymptotically and under optimal training, the SLS channel estimation error is determined only by the strength of the channel itself.

V. MMSE Channel Estimation

Let us obtain a linear estimator that minimizes the estimation MSE of $\mathbf{H}$ [25]. It can be expressed in the following general form:

$$\mathbf{H}_{\mathrm{MMSE}} = \mathbf{SA}_o \tag{22}$$

where $\mathbf{A}_o$ has to be obtained so that the MSE is minimized

$$\begin{aligned} \mathbf{A}_o &= \arg\min_{\mathbf{A}}\mathrm{E}\left\{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2\right\} \\ &= \arg\min_{\mathbf{A}}\mathrm{E}\left\{\|\mathbf{H}-\mathbf{SA}\|_F^2\right\}. \end{aligned} \tag{23}$$

Using (2), the estimation error can be expressed as

$$\begin{aligned} \mathcal{E} &= \mathrm{E}\left\{\|\mathbf{H}-\mathbf{SA}\|_F^2\right\} \\ &= \mathrm{tr}\{\mathbf{R_H}\} - \mathrm{tr}\{\mathbf{R_H PA}\} - \mathrm{tr}\{\mathbf{A}^H\mathbf{P}^H\mathbf{R_H}\} \\ &\quad + \mathrm{tr}\left\{\mathbf{A}^H\left(\mathbf{P}^H\mathbf{R_H P}+\sigma_n^2 r\mathbf{I}\right)\mathbf{A}\right\}. \end{aligned} \tag{24}$$

The optimal $\mathbf{A}$ can be found from $\partial\mathcal{E}/\partial\mathbf{A} = 0$ and is given by

$$\mathbf{A}_o = \left(\mathbf{P}^H\mathbf{R_H P}+\sigma_n^2 r\mathbf{I}\right)^{-1}\mathbf{P}^H\mathbf{R_H}. \tag{25}$$

Hence, the linear MMSE estimator of $\mathbf{H}$ can be written as

$$\hat{\mathbf{H}}_{\mathrm{MMSE}} = \mathbf{S}\left(\mathbf{P}^H\mathbf{R_H P}+\sigma_n^2 r\mathbf{I}\right)^{-1}\mathbf{P}^H\mathbf{R_H}. \tag{26}$$

[2]Note that this matrix does not correspond to the conventional definition of the channel correlation matrix $\mathrm{E}\{\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\}$, where $\mathrm{vec}(\cdot)$ is the vectorization operator stacking the columns of a matrix on top of each other.

The performance of this estimator is characterized by the error matrix $\mathbf{E} = \mathbf{H} - \hat{\mathbf{H}}_{\rm MMSE}$ with zero mean and

$$\mathbf{R_E} = \mathrm{E}\{\mathbf{E}\mathbf{E}^H\} = \left(\mathbf{R}_\mathbf{H}^{-1} + \frac{1}{\sigma_n^2 r}\mathbf{P}\mathbf{P}^H\right)^{-1}. \quad (27)$$

Therefore, the MMSE estimation error can be computed as

$$\begin{aligned} J_{\rm MMSE} &= \mathrm{tr}\{\mathbf{R_E}\} \\ &= \mathrm{tr}\left\{\left(\mathbf{R}_\mathbf{H}^{-1} + \frac{1}{\sigma_n^2 r}\mathbf{P}\mathbf{P}^H\right)^{-1}\right\}. \end{aligned} \quad (28)$$

To minimize (28) subject to the transmit power constraint $\mathrm{tr}\{\mathbf{P}\mathbf{P}^H\} = \mathcal{P}$, we can use the Lagrange multiplier method. The problem can be written as

$$L(\mathbf{P},\mu) = \mathrm{tr}\left\{\left(\mathbf{R}_\mathbf{H}^{-1} + \frac{1}{\sigma_n^2 r}\mathbf{P}\mathbf{P}^H\right)^{-1}\right\} + \mu(\mathrm{tr}\{\mathbf{P}\mathbf{P}^H\} - \mathcal{P}) \quad (29)$$

where μ is the Lagrange multiplier.

Using the chain differentiation rule for matrices [16], we obtain that the optimal training matrix should satisfy the equation

$$\mathbf{P}\mathbf{P}^H = \frac{\sigma_n^2 r}{\sqrt{\mu}}\mathbf{I} - \sigma_n^2 r \mathbf{R}_\mathbf{H}^{-1}. \quad (30)$$

Using the constraint $\mathrm{tr}\{\mathbf{P}\mathbf{P}^H\} = \mathcal{P}$, (30) can be expressed as

$$\mathbf{P}\mathbf{P}^H = \frac{1}{t}\left(\mathcal{P} + \sigma_n^2 r\,\mathrm{tr}\left\{\mathbf{R}_\mathbf{H}^{-1}\right\}\right)\mathbf{I} - \sigma_n^2 r \mathbf{R}_\mathbf{H}^{-1}. \quad (31)$$

Note that the constraint that the matrix $\mathbf{P}\mathbf{P}^H$ should be positive semidefinite is ignored in (29). Therefore, (31) provides a sensible solution to the optimal probing matrix design only if the matrix $(1/t)(\mathcal{P} + \sigma_n^2 r\,\mathrm{tr}\{\mathbf{R}_\mathbf{H}^{-1}\})\mathbf{I} - \sigma_n^2 r \mathbf{R}_\mathbf{H}^{-1}$ in the right-hand side of (31) is positive semidefinite [i.e., if the signal-to-noise ratio (SNR) $\mathcal{P}/\sigma_n^2$ is sufficiently high].

Interestingly, in the uncorrelated channel ($\mathbf{R_H} \propto \mathbf{I}$) or high SNR ($\sigma_n^2/\mathcal{P} \to 0$) cases, (31) reduces to (9). Therefore, in these cases the LS, SLS, and MMSE approaches all have the same condition on optimal training matrices. Otherwise, the condition for the optimal training matrix of the MMSE estimator can be different from that of the LS and SLS estimators.

Using (31), we obtain that, if the matrix $(1/t)(\mathcal{P} + \sigma_n^2 r\,\mathrm{tr}\{\mathbf{R}_\mathbf{H}^{-1}\})\mathbf{I} - \sigma_n^2 r \mathbf{R}_\mathbf{H}^{-1}$ remains positive semidefinite, then

$$\min_{\mathbf{P}} J_{\rm MMSE} = \frac{t^2}{\mathrm{tr}\left\{\mathbf{R}_\mathbf{H}^{-1}\right\} + \mathcal{P}/\left(\sigma_n^2 r\right)}. \quad (32)$$

To find the solution to (29) which is valid for any SNR, let us use the eigenvalue decomposition of $\mathbf{R_H}$ in the form

$$\mathbf{R_H} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H \quad (33)$$

where $\mathbf{Q}$ is the unitary eigenvector matrix and $\boldsymbol{\Lambda}$ is the diagonal matrix with nonnegative eigenvalues. Using this notation, (28) can be rewritten as

$$\begin{aligned} J_{\rm MMSE} &= \mathrm{tr}\left\{\left(\mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^H + \frac{1}{\sigma_n^2 r}\mathbf{P}\mathbf{P}^H\right)^{-1}\right\} \\ &= \mathrm{tr}\left\{\left(\boldsymbol{\Lambda}^{-1} + \frac{1}{\sigma_n^2 r}\mathbf{Q}^H\mathbf{P}\mathbf{P}^H\mathbf{Q}\right)^{-1}\right\} \\ &= \mathrm{tr}\{(\boldsymbol{\Lambda}^{-1} + \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H)^{-1}\} \end{aligned} \quad (34)$$

where

$$\tilde{\mathbf{P}} = \frac{1}{\sqrt{\sigma_n^2 r}}\mathbf{Q}^H\mathbf{P}. \quad (35)$$

Using $\tilde{\mathbf{P}}$, the total transmit power (5) can be replaced with the following equivalent form:

$$\mathrm{tr}\{\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H\} = \frac{1}{\sigma_n^2 r}\mathrm{tr}\{\mathbf{Q}^H\mathbf{P}\mathbf{P}^H\mathbf{Q}\} = \frac{\mathcal{P}}{\sigma_n^2 r}. \quad (36)$$

To minimize (34), we use the following lemma.

Lemma 1: For positive definite $m \times m$ matrix $\mathbf{A}$, the following inequality holds:

$$\mathrm{tr}\{\mathbf{A}^{-1}\} \geq \sum_{i=1}^{m}(a_{i,i})^{-1} \quad (37)$$

where $a_{i,i}$ is the ith diagonal element of $\mathbf{A}$ and the equality holds if $\mathbf{A}$ is diagonal.

Proof: See [25, p. 65]. □

Based on this lemma, the minimum of (34) is achieved if $\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$ has the following diagonal structure:

$$\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H = \mathrm{diag}(|\tilde{p}_1|^2, \ldots, |\tilde{p}_t|^2). \quad (38)$$

Using Lagrange multiplier method and taking into account (34), (36), and (38), the optimal training matrix of the MMSE method can be found by minimizing the function

$$\begin{aligned} L(\tilde{\mathbf{P}},\mu) &= \mathrm{tr}\{(\boldsymbol{\Lambda}^{-1} + \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H)^{-1}\} \\ &\quad + \mu\left[\mathrm{tr}\{\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H\} - \mathcal{P}/\left(\sigma_n^2 r\right)\right] \\ &= \sum_{i=1}^{t}\left(\lambda_i^{-1} + |\tilde{p}_i|^2\right)^{-1} \\ &\quad + \mu\sum_{i=1}^{t}\left(|\tilde{p}_i|^2 - \mathcal{P}/\left(\sigma_n^2 r t\right)\right). \end{aligned} \quad (39)$$

Note that the matrix $\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$ in (38) is positive semidefinite. Therefore, in contrast to (29), the positive semidefiniteness of $\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$ is explicitly taken into account in (39). Setting $\partial L(\tilde{\mathbf{P}},\mu)/\partial\tilde{p}_i = 0$ for $i = 1,\ldots,t$ yields

$$\tilde{p}_i\left[\left(\lambda_i^{-1} + |\tilde{p}_i|^2\right)^{-2} - \mu\right] = 0, \quad i = 1,\ldots,t. \quad (40)$$

From (40), we obtain the following water-filling-type solution to (39):

$$\tilde{p}_i = \begin{cases} \sqrt{\mu_\circ - \lambda_i^{-1}}, & \text{if } \lambda_i^{-1} < \mu_\circ \\ 0, & \text{otherwise} \end{cases} \quad (41)$$

where the positive constant $\mu_\circ = 1/\sqrt{\mu}$ should be tuned so that the transmitted power constraint (36) is satisfied. Related water-filling results in the context of channel capacity are reported in [26].

If $N = t$, then the optimal $\tilde{\mathbf{P}}$ can be written in the following matrix form:

$$\tilde{\mathbf{P}} = ([\mu_\circ \mathbf{I} - \mathbf{\Lambda}^{-1}]^+)^{1/2} \quad (42)$$

where the operator $[\cdot]^+$ replaces all negative entries of a real matrix by zeros and leaves all nonnegative entries unchanged. Inserting (42) into (35), we have that the optimal training matrix can be written as

$$\mathbf{P} = \sqrt{\sigma_n^2 r}\mathbf{Q}([\mu_\circ \mathbf{I} - \mathbf{\Lambda}^{-1}]^+)^{1/2} \quad (43)$$

where the constant $\mu_\circ$ has to be adjusted to satisfy the transmitted power constraint (5).

To find the optimal training matrix $\mathbf{P}$ for an arbitrary $N \geq t$, we will use the following two lemmas.

Lemma 2: If $\mathbf{P}$ is an optimal $t \times t$ training matrix for the MMSE estimator, then the $t \times N$ $(N > t)$ matrix $\breve{\mathbf{P}} = [\mathbf{P}, \mathbf{0}]$ is also an optimal training matrix, where $\mathbf{0}$ is $t \times (N - t)$ zero matrix.

Lemma 3: If $\mathbf{P}$ is an optimal $t \times N$ $(N \geq t)$ training matrix for MMSE estimator, then $\breve{\mathbf{P}} = \mathbf{P}\mathbf{U}$ is also an optimal training matrix, where $\mathbf{U}$ is an arbitrary $N \times N$ unitary matrix.

The proofs of these lemmas are straightforward if we note that $\mathbf{P}\mathbf{P}^H = \breve{\mathbf{P}}\breve{\mathbf{P}}^H$. Therefore, if $\mathbf{P}$ is an optimal training matrix, then, according to (27) and (28), the matrix $\breve{\mathbf{P}}$ is also optimal.

Using Lemmas 2 and 3 and taking into account (43), we obtain that if $N \geq t$, then the optimal $t \times N$ training matrix is given by

$$\mathbf{P} = \sqrt{\sigma_n^2 r}\mathbf{Q}\left[([\mu_\circ \mathbf{I} - \mathbf{\Lambda}^{-1}]^+)^{1/2}, \mathbf{0}_{t\times(N-t)}\right]\mathbf{U} \quad (44)$$

where, as before, the constant factor $\mu_\circ$ must be tuned to satisfy the transmitted power constraint.

Equation (44) shows that, from the MMSE viewpoint, it is equivalent either to concentrate the whole transmitted power in t training samples (which corresponds to the case $\mathbf{U} = \mathbf{I}$) or, alternatively, to spread it over $N > t$ samples by means of a proper choice of the matrix $\mathbf{U}$. The second way may be more suitable if there is an additional peak power per antenna constraint.

VI. RMMSE Channel Estimation

The MMSE channel estimator (26) assumes the perfect knowledge of the matrix $\mathbf{R_H}$. This assumption may be unrealistic in practical applications. Therefore, we relax it and simplify the MMSE estimator by using the matrix $\alpha\mathbf{I}$ in lieu of $\mathbf{R_H}$ in (26) where the parameter α is adjusted to minimize the MSE.

Replacing $\mathbf{R_H}$ with $\alpha\mathbf{I}$ in (26) and applying the matrix inversion lemma, we can rewrite this equation as

$$\begin{aligned}\hat{\mathbf{H}} &= \alpha\mathbf{S}\left(\alpha\mathbf{P}^H\mathbf{P} + \sigma_n^2 r\mathbf{I}\right)^{-1}\mathbf{P}^H \\ &= \frac{\alpha}{\sigma_n^2 r}\mathbf{S}\left(\mathbf{I} - \frac{\alpha t}{\alpha\mathcal{P} + \sigma_n^2 rt}\mathbf{P}^H\mathbf{P}\right)\mathbf{P}^H. \end{aligned} \quad (45)$$

Using (45) and assuming the orthogonal training (9), the channel MSE can be computed as

$$\begin{aligned} J &= \mathrm{E}\left\{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2\right\} \\ &= \left(\frac{r\xi^2\mathcal{P}^2}{t} + \sigma_n^2 r\mathcal{P}\right) \\ &\quad \cdot \left(\frac{\alpha t}{\alpha\mathcal{P} + \sigma_n^2 rt} - \frac{\xi^2 t}{\xi^2\mathcal{P} + \sigma_n^2 t}\right)^2 \\ &\quad + \frac{\xi^2\sigma_n^2 rt^2}{\xi^2\mathcal{P} + \sigma_n^2 t} \end{aligned} \quad (46)$$

where $\xi^2 \triangleq \mathrm{tr}\{\mathbf{R_H}\}/rt$. Obviously, (46) is minimized with

$$\alpha_{\mathrm{opt}} = \xi^2 r = \mathrm{tr}\{\mathbf{R_H}\}/t. \quad (47)$$

Interestingly, (47) is also the value which minimizes $\|\mathbf{R_H} - \alpha\mathbf{I}\|_F$ (i.e., the value which gives the best approximation of $\mathbf{R_H}$ in terms of a weighted identity matrix).

Using (47), we have that for any training matrix $\mathbf{P}$, the RMMSE channel estimator is given by

$$\hat{\mathbf{H}}_{\mathrm{RMMSE}} = \mathbf{S}\left[\mathbf{P}^H\mathbf{P} + \frac{\sigma_n^2 rt}{\mathrm{tr}\{\mathbf{R_H}\}}\mathbf{I}\right]^{-1}\mathbf{P}^H \quad (48)$$

where $\mathrm{tr}\{\mathbf{R_H}\}$ is assumed to be known or estimated. Interestingly, (48) can be viewed as a diagonally loaded version of the LS estimator (4) with the diagonal loading factor of $\sigma_n^2 rt/\mathrm{tr}\{\mathbf{R_H}\}$. Also, it can be readily seen that if the orthogonal training is used, then the RMMSE estimator of (48) coincides with the SLS estimator of (16).

Using (46), we obtain that the RMMSE estimation error for any orthogonal training matrix that satisfies (9) is given by

$$J_{\mathrm{RMMSE}} = \frac{\mathrm{tr}\{\mathbf{R_H}\}\sigma_n^2 rt^2}{\mathrm{tr}\{\mathbf{R_H}\}\mathcal{P} + \sigma_n^2 rt^2}. \quad (49)$$

Using (49) along with (12), we have

$$\frac{J_{\mathrm{RMMSE}}}{J_{\mathrm{LS}}} = \frac{\mathrm{tr}\{\mathbf{R_H}\}\mathcal{P}}{\mathrm{tr}\{\mathbf{R_H}\}\mathcal{P} + \sigma_n^2 rt^2}. \quad (50)$$

From (50) we see that, if $\sigma_n^2 > 0$, then $J_{\mathrm{RMMSE}} < J_{\mathrm{LS}}$ and, therefore, the proposed RMMSE channel estimation technique performs always better than the LS channel estimator. The improvement of the RMMSE estimator over the LS estimator is especially pronounced if the SNR is low (i.e., if $\mathrm{tr}\{\mathbf{R_H}\}\mathcal{P} \ll \sigma_n^2 rt^2$).

The RMMSE estimator (48) is a function of $\mathrm{tr}\{\mathbf{R_H}\}$ and σ_n^2 and, therefore, it requires these parameters to be known. In practice, the estimate (17) can be used in lieu of the exact value of $\mathrm{tr}\{\mathbf{R_H}\}$.

In the case of orthogonal training matrix of (9), inserting (18) into (48), we obtain the following LS-RMMSE estimator:

$$\hat{\mathbf{H}}_{\text{LS}-\text{RMMSE}} = \frac{t\mathrm{tr}\{\mathbf{SS}^H\}}{\mathcal{P}\left(\mathrm{tr}\{\mathbf{SS}^H\} + \sigma_n^2 rt\right)}\mathbf{SP}^H. \tag{51}$$

This estimator corresponds to the RMMSE technique which uses the LS estimate (18) of $\mathrm{tr}\{\mathbf{R_H}\}$. In contrast to the original RMMSE estimator (48), the LS-RMMSE estimator (51) does not require any knowledge of $\mathrm{tr}\{\mathbf{R_H}\}$.

VII. Combining Multiple LS Channel Estimates

In all the previous sections, the case of a single channel estimate has been addressed. In this section, we extend the optimal training approach to the case of *multiple* LS channel estimates. If there are multiple training periods available within the channel coherency time, it may be inconvenient to store and process long amounts of data that are formed by accumulation of multiple received data blocks that correspond to different probing periods [16]. A good alternative here would be to obtain a particular channel estimate for each training period and then to store these estimates dynamically rather than storing the data itself. Using this approach, the final channel estimate can be obtained based on a proper combination of such previously computed particular estimates.

Let us have K estimates $\hat{\mathbf{H}}_1, \ldots, \hat{\mathbf{H}}_K$ of the MIMO channel which are computed using the LS estimator (4) based on the training matrices $\mathbf{P}_1, \ldots, \mathbf{P}_K$, respectively. The channel is assumed to be quasi-static (fixed) at the interval involving K training periods and $\mathcal{P}_k = \|\mathbf{P}_k\|_F^2$ is the transmitted power during the kth training period.

We aim to improve the performance of MIMO channel estimation by combining the values of $\hat{\mathbf{H}}_k (k = 1, \ldots, K)$ in a linear way as

$$\hat{\mathbf{H}} = \sum_{k=1}^{K} \alpha_k \hat{\mathbf{H}}_k \tag{52}$$

where α_k are unknown weight coefficients.

Let us obtain the optimal weight coefficients by means of minimizing the error in (52). Then, these coefficients can be found by solving the following optimization problem:

$$\min_{\alpha_1,\ldots,\alpha_K} \mathrm{E}\left\{\left\|\mathbf{H} - \sum_{k=1}^{K}\alpha_k\hat{\mathbf{H}}_k\right\|_F^2\right\} \text{ subject to } \sum_{k=1}^{K}\alpha_k = 1 \tag{53}$$

where the constraint in (53) guarantees that the final channel estimate is unbiased. This problem corresponds to the so-called BLUE estimator; see [25].

The solution to (53) is given by the following lemma which is an extension of one of the results in [16] to the MIMO case.

Lemma 4: The optimal weights $\{\alpha_k\}_{k=1}^K$ are given by

$$\alpha_k = \frac{1}{\mathrm{tr}\left\{\left(\mathbf{P}_k\mathbf{P}_k^H\right)^{-1}\right\}\sum_{l=1}^{K} 1/\mathrm{tr}\left\{\left(\mathbf{P}_l\mathbf{P}_l^H\right)^{-1}\right\}}. \tag{54}$$

Proof: See the Appendix. □

It is important to stress that the optimal weight coefficients α_k in (54) are user-independent, i.e., they are the same for any user.

Choosing optimal orthogonal weighting matrices in each training period, we have

$$\mathrm{tr}\left\{\left(\mathbf{P}_k\mathbf{P}_k^H\right)^{-1}\right\} = \frac{t^2}{\mathcal{P}_k} \tag{55}$$

$$\sum_{l=1}^{K} 1/\mathrm{tr}\left\{\left(\mathbf{P}_l\mathbf{P}_l^H\right)^{-1}\right\} = \frac{\mathcal{P}_{\text{tot}}}{t^2} \tag{56}$$

where

$$\mathcal{P}_{\text{tot}} = \sum_{k=1}^{K}\mathcal{P}_k \tag{57}$$

is the total transmitted power during the K training periods.

Inserting (55) and (56) into (54), we obtain that, if orthogonal training matrices are used, the expression for optimal weight coefficients can be simplified to

$$\alpha_k = \frac{\mathcal{P}_k}{\mathcal{P}_{\text{tot}}}. \tag{58}$$

In this case, the channel estimation error is given by

$$\begin{aligned}
\mathrm{E}\left\{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2\right\} &= \mathrm{E}\left\{\left\|\mathbf{H} - \sum_{k=1}^{K}\frac{\mathcal{P}_k}{\mathcal{P}_{\text{tot}}}\hat{\mathbf{H}}_k\right\|_F^2\right\} \\
&= \mathrm{E}\left\{\left\|\sum_{k=1}^{K}\frac{\mathcal{P}_k}{\mathcal{P}_{\text{tot}}}(\mathbf{H} - \hat{\mathbf{H}}_k)\right\|_F^2\right\} \\
&= \frac{t^2}{\mathcal{P}_{\text{tot}}^2}\mathrm{E}\left\{\left\|\sum_{k=1}^{K}\mathbf{V}_k\mathbf{P}_k^H\right\|_F^2\right\} \\
&= \frac{\sigma_n^2 t^2 r}{\mathcal{P}_{\text{tot}}^2}\mathrm{tr}\left\{\sum_{k=1}^{K}\mathbf{P}_k\mathbf{P}_k^H\right\} \\
&= \frac{\sigma_n^2 t^2 r}{\mathcal{P}_{\text{tot}}}
\end{aligned} \tag{59}$$

where $\mathbf{V}_k$ is the receiver zero-mean white noise during the kth training period. Here, we have used the property $\mathrm{E}\{\mathbf{V}_k^H\mathbf{V}_l\} = \sigma_n^2 t\delta_{k,l}\mathbf{I}$, where $\delta_{k,l}$ is the Kronecker delta.

We observe that, similar to (12), the error in (59) is independent of the channel realizations used. Comparing (59) with (12), we see that the optimal linear combining of multiple estimates reduces the estimation error by a factor of $\mathcal{P}_{\text{tot}}/\mathcal{P}$. For example, if each training has the same power ($\mathcal{P}_k = \mathcal{P}; k = 1, 2, \ldots, K$), then $\mathcal{P}_{\text{tot}} = K\mathcal{P}$ and the estimation error is reduced by a factor of K.

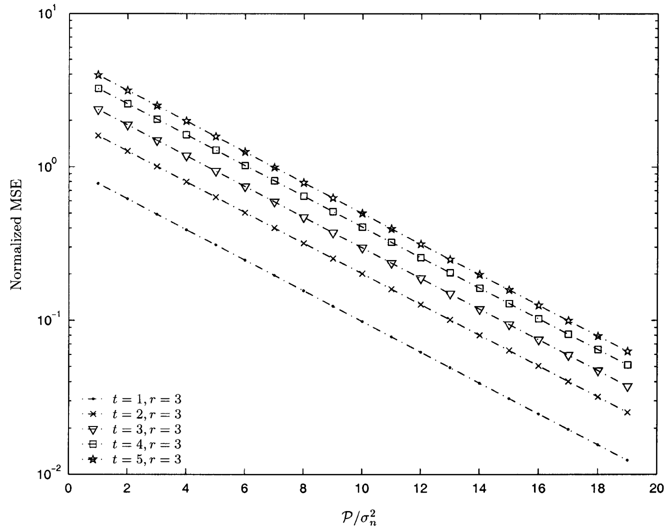


Fig. 1. Channel estimation MSEs of the LS estimator versus $\mathcal{P}/\sigma_n^2$.

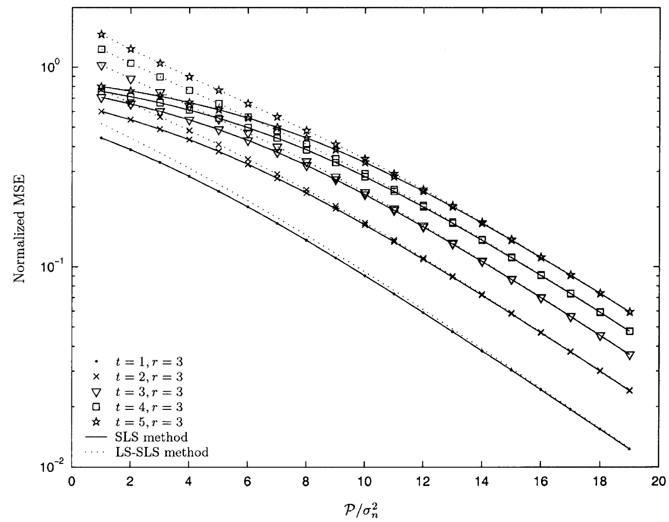


Fig. 2. Channel estimation MSEs of SLS and LS-SLS estimators versus $\mathcal{P}/\sigma_n^2$. As there is no difference between the SLS and RMMSE estimators and between the LS-SLS and LS-RMMSE estimators in the case of orthogonal training, this figure is also valid for the RMMSE-based techniques.

VIII. Numerical Examples

In this section, we compare the performance of the LS, SLS, MMSE, and RMMSE channel estimators numerically.

Throughout all our examples, we assume that $N = t$. The channel coefficients and the receiver noise are assumed to be circular complex Gaussian random variables with the unit variance. It is assumed that the matrix $\mathbf{R_H}$ has the following structure:

$$[\mathbf{R_H}]_{n,m} = r\varepsilon^{|n-m|}, \qquad 0 \leq \varepsilon < 1 \tag{60}$$

where n and m are the indexes of the array sensors. This covariance matrix model is frequently used in the literature; see [28]–[30] and references therein. Each point in our figures is obtained by averaging over 5000 independent simulation runs.

In Fig. 1, we display the normalized MSE $J_{\rm LS}/(tr)$ of the LS channel estimator with optimal training versus SNR$= \mathcal{P}/\sigma_n^2$. The parameter $r = 3$ is fixed, while the parameter t is varied in this figure. It can be seen from Fig. 1 that, as can be expected from Section III, the performance of the LS estimator decreases with the number of transmit antennas.

In Fig. 2, the normalized MSEs of the SLS estimator (16) and the LS-SLS estimator are displayed versus SNR, where the SLS estimator assumes that $\mathrm{tr}\{\mathbf{R_H}\}$ is perfectly known, whereas the LS-SLS estimator corresponds to (16), where the LS-based estimate (17) is used instead of $\mathrm{tr}\{\mathbf{R_H}\}$. Both estimators tested use the optimal training. The parameter $r = 3$ is fixed, while the parameter t is varied.

Comparing Figs. 1 and 2, we observe that at low SNRs, both the SLS and LS-SLS estimators have substantially lower MSEs than the LS estimator. From Fig. 2, it is also clear that the SLS estimator with the perfect knowledge of $\mathrm{tr}\{\mathbf{R_H}\}$ slightly outperforms the LS-SLS estimator at low SNRs, while at high SNRs the performances of both these estimators are nearly identical.

Fig. 3(a) and (b) displays the normalized MSEs of the MMSE estimator versus SNR in the cases of $\varepsilon = 0.4$ and $\varepsilon = 0.8$, respectively. Both orthogonal probing and optimal probing (which is not orthogonal because of nonzero channel correlation) are tested. Similar to the previous figures, the parameter $r = 3$ is fixed, while the parameter t is varied.

From Fig. 3, it follows that the MMSE estimator performs better than the LS and SLS techniques. From Fig. 3(a), it also follows that at low values of ε (weakly correlated channels), orthogonal probing is nearly optimal for the MMSE channel estimator. However, if ε is large (i.e., if the channel is highly correlated) as in Fig. 3(b), then the orthogonal probing used instead of the optimal one can substantially reduce the performance of the MMSE estimator. This effect is especially pronounced when the number of transmit antennas is large and the SNR is low.

Note that, as mentioned in Section VI, the SLS and RMMSE estimators coincide in the case of orthogonal training. Accordingly, the same is true for the LS-SLS and LS-RMMSE estimators. Therefore, Fig. 2 is also valid for the RMMSE-based estimators. From Fig. 2, we see that at high SNRs, the performances of the RMMSE and LS-RMMSE estimators are nearly identical. However, at low SNRs the RMMSE estimator substantially outperforms the LS-RMMSE estimator. This is especially true when t is high.

Fig. 4 compares the normalized MSEs of the LS, SLS, MMSE, and RMMSE channel estimators versus SNR. The cases $r = t = 2$ and $r = t = 8$ are considered, $\varepsilon = 0.8$ is assumed, and orthogonal training is used in this figure. We can observe that the LS estimator has the worst performance among all methods tested. As can be expected from the theoretical part of this paper, the performances of the SLS and RMMSE estimators are identical. The MMSE estimator has the best performance among the methods tested, but it requires more a priori knowledge about the channel than any of the other techniques tested. Therefore, the proposed SLS and RMMSE

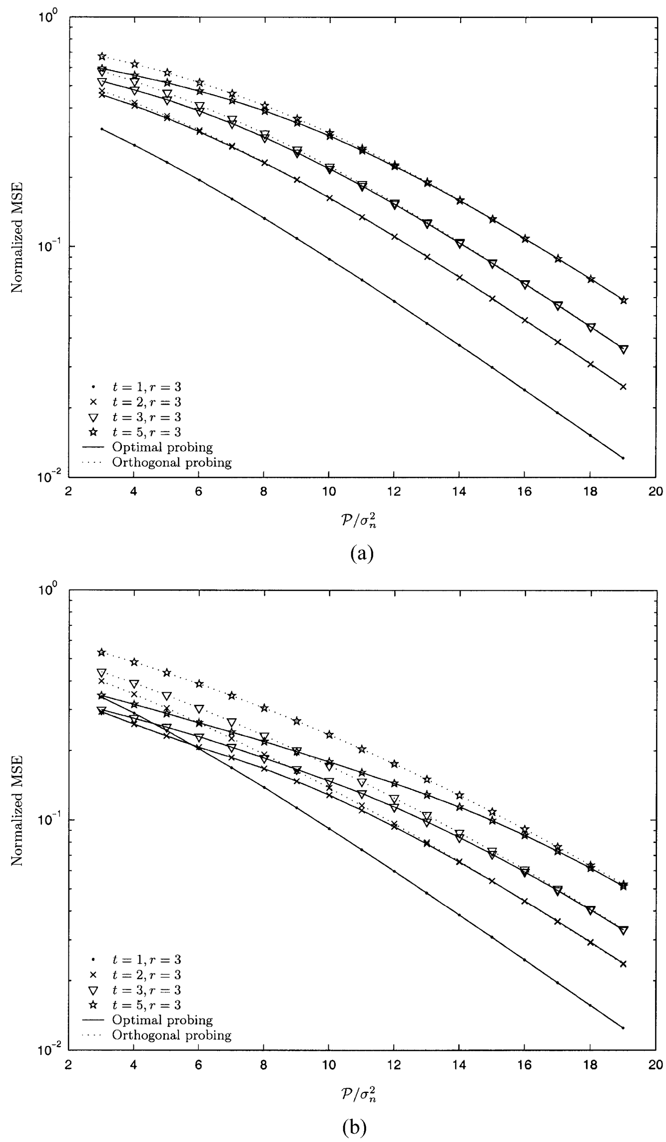


Fig. 3. Channel estimation MSEs of the MMSE estimator versus $\mathcal{P}/\sigma_n^2$: (a) $\varepsilon = 0.4$ and (b) $\varepsilon = 0.8$.

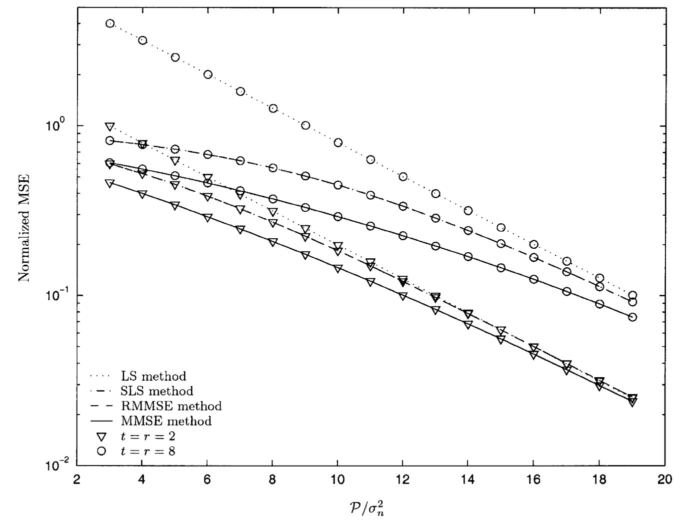


Fig. 4. Channel estimation MSEs of the LS, SLS, MMSE and RMMSE estimators versus $\mathcal{P}/\sigma_n^2$ in the case of orthogonal training.

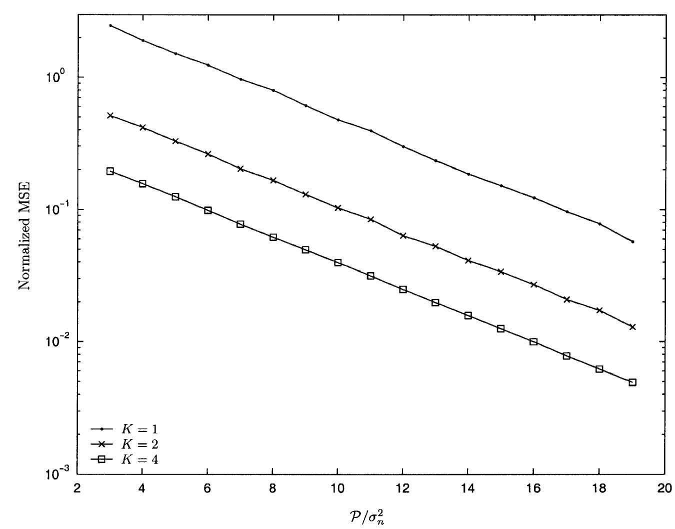


Fig. 5. Channel estimation MSEs of the BLUE estimator versus $\mathcal{P}/\sigma_n^2$.

estimators (which are required to know about the channel much less than the MMSE estimator) provide a good tradeoff between the achieved performance and the required channel knowledge.

In our last example, the case of multiple LS channel estimates is assumed. In Fig. 5, the normalized MSEs of the LS-based BLUE estimator are shown versus SNR in the case of optimal training. In this figure, $r = 2$ and $t = 4$ are assumed and the value of K is varied.

Fig. 5 demonstrates substantial improvements that, as expected from Section VII, can be achieved when the BLUE estimator is used in the case of multiple channel estimates.

IX. Conclusion

The performance of several training-based MIMO channel estimation methods has been studied. The popular LS and MMSE approaches have been considered, and new scaled LS and relaxed MMSE techniques have been proposed that require less knowledge of the channel second-order statistics and/or have better performance than the conventional LS and MMSE channel estimators, therefore offering attractive tradeoffs in terms of performance and a priori required knowledge of the channel parameters. For each of the considered techniques, channel estimation performances and the aforementioned tradeoffs have been investigated and the optimal choice of training matrices has been studied.

In the case of multiple LS channel estimates, the best linear unbiased estimation scheme for their linear combining has been proposed and studied.

Numerical examples have further illustrated the aforementioned tradeoffs between different channel estimators and validated the advantages of optimal training.

APPENDIX
PROOF OF LEMMA 4

To solve (53), let us insert (4) into the objective function of this problem and rewrite it as

$$\begin{aligned}\min_{\alpha_1,\ldots,\alpha_K} \mathrm{E}\left\{\left\|\mathbf{H}-\sum_{k=1}^{K}\alpha_k\hat{\mathbf{H}}_k\right\|_F^2\right\} &= \mathrm{E}\left\{\mathrm{tr}\left\{\left(\sum_{m=1}^{K}\alpha_m\mathbf{V}_m\mathbf{P}_m^{\dagger H}\right)^H\left(\sum_{l=1}^{K}\alpha_l\mathbf{V}_l\mathbf{P}_l^{\dagger H}\right)\right\}\right\}\\ &= \mathrm{tr}\left\{\left(\sum_{m=1}^{K}\sum_{l=1}^{K}\alpha_m\alpha_l^*\mathbf{P}_m^{\dagger}\mathbf{P}_l^{\dagger H}\mathrm{E}\left\{\mathbf{V}_m\mathbf{V}_l^H\right\}\right)\right\}\\ &= \sigma_n^2 r\,\mathrm{tr}\left\{\sum_{l=1}^{K}|\alpha_l|^2\left(\mathbf{P}_l\mathbf{P}_l^H\right)^{-1}\right\}\end{aligned} \tag{61}$$

where we have taken into account that $\mathrm{E}\{\mathbf{V}_m\mathbf{V}_l^H\}=\delta_{ml}\sigma_n^2 r\mathbf{I}$. Hence, the problem can be written in the form

$$\min_{\alpha_1,\ldots,\alpha_K}\sum_{k=1}^{K}|\alpha_k|^2\kappa_k \quad \text{subject to} \sum_{k=1}^{K}\alpha_k=1 \tag{62}$$

where $\kappa_k=\mathrm{tr}\{(\mathbf{P}_k\mathbf{P}_k^H)^{-1}\}$. Note that problems similar to (62) frequently arise in beamforming, multiuser detection, and other fields; see, for example, [31] and references therein. Solving (62), we immediately obtain (54) and the lemma is proved.

ACKNOWLEDGMENT

The authors are grateful to Prof. E. G. Larsson and Prof. Z.-Q. Luo for their useful comments on earlier versions of this paper, and to anonymous reviewers whose comments helped to improve the manuscript significantly.

REFERENCES

[1] D. Gerlach and A. Paulraj, "Spectrum reuse using transmitting antenna arrays with feedback," in *Proc. ICASSP*, vol. IV, Adelaide, Australia, Apr. 1994, pp. 97–100.

[2] I. Telatar, "Capacity of multiple antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[3] 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, Mar. 1998.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high-data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[6] C. C. Martin, J. H. Winters, and N. R. Sollenberger, "Miltiple-input multiple-output (MIMO) radio channel measurements," *Proc. VTC'00—Fall*, vol. 2, pp. 774–779, Sep. 2000.

[7] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998.

[8] L. Tong and S. Perreau, "Blind channel estimation: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[9] A. Grant, "Joint decoding and channel estimation for linear MIMO channels," in *Proc. IEEE Wireless Communications Networking Conf.*, vol. 3, Chicago, IL, Sep. 2000, pp. 1009–1012.

[10] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, pp. 2515–2528, Oct. 2002.

[11] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2515–2528, Apr. 2003.

[12] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Sep. 1999.

[13] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 549–553, Oct. 2002.

[14] Y. Li, "Optimal training sequences for OFDM systems with multiple transmit antennas," *Proc. GLOBECOM'00*, vol. 3, pp. 1478–1482, 2000.

[15] A. Scaglione and A. Vosoughi, "Turbo estimation of channel and symbols in precoded MIMO systems," in *Proc. ICASSP*, vol. 4, Montreal, PQ, Canada, May 2004, pp. 413–416.

[16] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP J. Appl. Signal Process. (Special Issue on Advances in Smart Antennas)*, pp. 1330–1339, Sep. 2004.

[17] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmission over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1351–1366, May 2003.

[18] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[19] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[20] E. de Carvalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (Global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, pp. 1053–1064, Apr. 2004.

[21] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: Optimal training and estimator tradeoffs," in *Proc. ICC'04*, vol. 5, Paris, France, Jun. 2004, pp. 2658–2662.

[22] ——, "On channel estimation and optimum training for MIMO systems," in *Proc. IEEE Sensor Array Multichannel Signal Processing Workshop*, Sitges, Spain, July 2004.

[23] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, pp. 2177–2188, Aug. 2004.

[24] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, Aug. 1973.

[25] S. M. Kay, *Fundumentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[26] N. Jindal and A. J. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2895–2909, Nov. 2003.

[27] S. D. Silverstein, "Application of orthogonal codes to the calibration of active phased array antennas for communication satellites," *IEEE Trans. Signal Process.*, vol. 45, pp. 206–218, Jan. 1997.

[28] A. B. Gershman, C. F. Mecklenbräuker, and J. F. Böhme, "Matrix fitting approach to direction of arrival estimation with imperfect spatial coherence of wavefronts," *IEEE Trans. Signal Process.*, vol. 45, pp. 1894–1899, Jul. 1997.

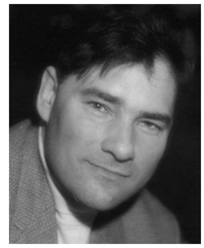
[29] A. B. Gershman, P. Stoica, M. Pesavento, and E. Larsson, "The stochastic Cramer–Rao bound for direction estimation in unknown noise fields," *Proc. Inst. Elect. Eng. Radar, Sonar Navig.*, vol. 149, no. 1, pp. 2–8, Feb. 2002.

[30] J. Ringelstein, A. B. Gershman, and J. F. Böhme, "Direction finding in random inhomogeneous media in the presence of multiplicative noise," *IEEE Signal Process. Lett.*, vol. 7, pp. 269–272, Oct. 2000.

[31] H. L. Van Trees, *Optimum Array Processing: Part IV, Detection, Estimation and Modulation Theory*. New York: Wiley Interscience, 2002.

Mehrzad Biguesh (S'89–M'02) was born in Shiraz, Iran. He received the B.S. degree in electronics engineering from Shiraz University, Shiraz, in 1991 and the M.S. and Ph.D. degrees in telecommunications (with honors) from Sharif University of Technology (SUT), Tehran, Iran, in 1994 and 2000, respectively.

From November 1998 to August 1999, he was with INRS-Telecommunications, University of Quebec, Quebec, PQ, Canada, as a Doctoral Trainee. From 1999 to 2001, he was with the Iran Telecom Research Center. During 2000–2001, he was also associated with SUT. From March 2002 to December 2004, he was a Postdoctoral Fellow at the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. In January 2005, he joined the Department of Electrical and Electronics Engineering, Shiraz University, as an Assistant Professor. His research interests include sensor array and MIMO signal processing, wireless communications, and radar signal processing.



Alex B. Gershman (M'97–SM'98–F'06) received the diploma and the Ph.D. degree in radiophysics from Nizhny Novgorod State University, Russia, in 1984 and 1990, respectively.

From 1984 to 1999, he held several full-time and visiting research appointments in Russia, Switzerland, and Germany. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a full Professor in 2002. From 2003 to 2005, he was a Visiting Professor at the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. Since April 2005, he has been with Darmstadt University of Technology, Darmstadt, Germany, as a full Professor and Head of the Communication Systems Group. His research interests are in the area of signal processing and communications with primary emphasis on statistical signal and array processing, adaptive beamforming, spatial diversity in MIMO communications, multiuser communications, and parameter estimation and detection. He has (co)authored several book chapters and more than 90 journal and 150 conference papers on these subjects. He has also coedited (with Y. Hua and Q. Cheng) *High-Resolution and Robust Signal Processing* (Berlin, Germany: Marcel Dekker, 2003) and (with N. D. Sidiropoulos) *Space-Time Processing for MIMO Communications* (New York: Wiley, 2005).

Dr. Gershman was a corecipient of the 2004 IEEE Signal Processing Society Best Paper Award for his paper on robust minimum variance beamforming. He was a recipient of the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research; the 2001 Wolfgang Paul Award from the Alexander von Humboldt Foundation, Germany; the 2000 Premier's Research Excellence Award, Ontario, Canada; the 1995–1996 Alexander von Humboldt Fellowship (Germany); the 1994 Swiss Academy of Engineering Science and Branco Weiss Fellowship (Switzerland); the 1994 Outstanding Young Scientist Presidential Fellowship (Russia); and the 1993 International Union of Radio Science Young Scientist Award. He is currently serving as Editor-in-Chief of IEEE SIGNAL PROCESSING LETTERS. He has been on the Editorial Boards of the *EURASIP Journal on Wireless Communications and Networking* and *EURASIP Signal Processing* from 2003 and 2005, respectively. He was Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1999–2005) and Guest Editor of the *EURASIP Journal of Applied Signal Processing* (2004). He is Vice-Chair of the Sensor Array and Multichannel (SAM) Technical Committee (TC) and a member of the Signal Processing Theory and Methods (SPTM) TC of the IEEE Signal Processing Society. He was Technical Cochair of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Darmstadt, Germany, December 2003, and Cochair of the International ITG/IEEE Workshop on Smart Antennas, Duisburg, Germany, April 2005. He is General Cochair of the First IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Puerto Vallarta, Mexico, December 2005; and Technical Cochair of the Fourth IEEE Sensor Array and Multichannel Signal Processing Workshop, Waltham, MA, June 2006. He is Tutorial Chair of EUSIPCO 2006, Florence, Italy, September 2006.