A LOW-COST APPROACH TO TRAINING-BASED MIMO CHANNEL ESTIMATION IN INTERFERENCE-LIMITED ENVIRONMENTS*

Dimitrios Katselis, a Eleftherios Kofidis, b and Sergios Theodoridis a

a Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis, Ilissia, 157 84 Athens, Greece.
b Department of Statistics and Insurance Science, University of Piraeus, 80, Karoali & Dimitriou str., 185 34 Piraeus, Greece.

ABSTRACT
The problem of optimizing the training signal for the estimation of multiple-input multiple-output (MIMO) fading channels has been of a great interest in the last few years, due to its central role in combining accuracy in estimation with bandwidth efficiency. The general case of correlated channels and colored interference was recently addressed [3]. It was shown that the optimal training for the Linear Minimum Mean Squared Error (LMMSE) estimator, in the sense of minimizing the channel estimation MSE subject to a constraint on the total transmit power, consists of a joint water-filling along the eigenmodes of the desired channel and interference covariance matrices. Thus, the resulting scheme relies on the assumption of the availability of these matrices at the transmitter, which is, in practice, realized via a feedback path. In this paper, that scheme is revisited, with the aim of reducing its requirements for side information and transmit beamforming. Inspired by an interpretation of the LMMSE estimator as a two-step procedure, we investigate possible gains (and tradeoffs) from an alternative scheme, in which the processing of the received data is performed on both of its dimensions, temporal and spatial. The simulation results demonstrate that this approach provides a good trade-off between estimation performance and low feedback communication and beamforming overheads.

1. INTRODUCTION
The construction of optimal training signals for MIMO channel estimation has received a lot of attention recently, mainly for the single-user case. The adopted optimization criteria include channel estimation accuracy and/or capacity or error rate bounds (see [3] and references therein). The MMSE-optimum training in a multiuser context was reported in [7], for the special case of an uncorrelated fading channel. The design was based on the (common) assumption of the a-priori knowledge of the interference covariance at the transmitter (via, e.g., covariance feedback [1]). Recently, the more general case of a multiuser MIMO network with correlated fading channels was addressed in [3]. In that work, the LMMSE channel estimator was considered and the optimal training scheme, in the sense of minimizing the channel MSE subject to a constraint on the total transmit power, was derived and shown to consist of a joint water-filling over the eigenmodes of the channel and interference correlation matrices. Consequently, the resulting so-called Bayesian Minimum-Disturbance Estimator (B-MDE) also relies on the availability of the channel and interference second-order statistics at the transmitter site and requires eigen-beamforming for its training. Thus, the improvement in the estimation performance from the optimal training comes at the price of an increased communication overhead for the system1 as well as higher complexity of the transmitter during the training period.

In this paper, we revisit the B-MDE scheme with the goal of reducing the required side-information and the complexity of the optimal training. Inspired by an alternative formulation of the LMMSE estimator as a two-step procedure, we propose a novel linear estimation scheme consisting of two filters, each operating on a different side of the received data matrix. The first filter is applied on the right side (temporal dimension) and is devoted to the minimization of the interference power, subject to a zero-forcing constraint that ensures the preservation of the desired signal component. The resulting channel estimate is further refined through a Wiener filter that exploits the a priori knowledge about the channel statistics. In this step, the data matrix is filtered on its left side (spatial dimension). Under the (quite realistic) assumption of

---

*This work was supported by a grant of the General Secretariat for Research and Technology (GSRT) of Greece, under the programme ΠΕΝΕΔ-2003.

1Efficient ways of feeding back the required information from the receiver to the transmitter can be conceived but they can also introduce inaccuracies (see, e.g., [7], where a circulant approximation of a Toeplitz correlation matrix is feedback instead).
an interference-limited environment with one dominant interferer [6], we show that, with this “two-sided” processing, the transmission modes at training need only be aligned with the interference and not the channel eigen-directions, unlike the B-MDE of [3]. Moreover, the optimal power allocation across the transmit spatial dimensions is derived. We also show that the channel estimator can be greatly simplified when optimal training is employed. In view of this fact that this alternative approach is based on processing the data from both sides, we will call this new estimator the two-sided B-MDE (2S B-MDE). The simulation results demonstrate that this provides a good trade-off between estimation performance and reduced feedback communication and beamforming overheads. In particular, despite its relative simplicity, the novel scheme is seen to always outperform the B-MDE variant of analogous training requirements.

Notation. We employ bold lowercase and uppercase letters to represent vectors and matrices, respectively. Superscripts $^T$ and $^H$ stand for transposition and conjugate transposition, respectively. $\| \cdot \|_F$ is the Frobenius norm. The expectation and matrix trace operators are denoted by $E(\cdot)$ and $\text{tr}(\cdot)$, respectively. $I_{m}$ is the $m$-th order identity matrix.

2. SIGNAL AND SYSTEM MODEL

We consider a MIMO communications system with $M_T$ transmit and $M_R$ receive antennas, and $L$ interfering sources. All the channels, desired and interfering, are assumed narrowband (flat fading) and quasi-static, i.e., constant over a frame. The received signal vector can thus be expressed as

$$y(n) = H_0 x_0(n) + \sum_{i=1}^{L} H_i x_i(n) + w(n)$$

(1)

In the above model, $y(n)$ is the received $M_R \times 1$ vector at time $n$. $H_0$ denotes the $M_R \times M_T$ channel matrix for the desired transmitter-receiver pair, and $H_i$ is the channel matrix for the $i$th interferer, $i = 1, 2, \ldots, L$. The desired and interfering signals transmitted at time $n$ are in the $M_T \times 1$ vectors $x_0(n)$ and $x_i(n)$, respectively. Let $P_i$, $i = 0, 1, \ldots, L$ denote the average power of the (scalar) symbols transmitted by the $i$th source. The desired source is assumed i.i.d. whereas the interfering sources may be temporally correlated. We assume that all inputs are zero-mean and uncorrelated with each other and with their channels. The $M_R \times 1$ vector $w(n)$ represents the background noise at the receiver front end and is assumed zero mean and temporally and spatially white with variance $\sigma_w^2$.

Furthermore, we assume that the channels are accurately described by the well-known model [6]:

$$H_i = R_{t_i}^{1/2} H_{w_i} R_{r_i}^{1/2}$$

(2)

where $H_{w_i}$, $i = 0, 1, \ldots, L$ is an uncorrelated channel, represented by an $M_R \times M_T$ matrix of i.i.d. zero mean, unit variance, circularly symmetric complex Gaussian entries, and $R_{t_i}$ and $R_{r_i}$ are the $i$th transmit and receive fade correlation matrices, respectively, normalized so that $\text{tr}(R_{t_i}) = M_T$ and $\text{tr}(R_{r_i}) = M_R$.

3. A TWO-STEP CHANNEL ESTIMATOR

3.1. The Bayesian Minimum-Disturbance Estimator (B-MDE)

Consider that $M$ transmitted vectors per frame are devoted to training. Grouping together the corresponding $M$ received vectors in an $M_R \times M$ matrix, we can write (1) as:

$$Y \triangleq \begin{bmatrix} y(n) & y(n-1) & \cdots & y(n-M+1) \end{bmatrix} = H_0 \begin{bmatrix} x_0(n) & x_0(n-1) & \cdots & x_0(n-M+1) \end{bmatrix} X_0 + \sum_{i=1}^{L} H_i X_i \mathcal{T}_i + W$$

(3)

where $X_i$, $i = 1, 2, \ldots, L$ are constructed in a manner analogous with $X_0$ with i.i.d. entries, and the $M \times M$ matrices $\mathcal{T}_i$ represent linear temporal filters, or

$$Y = H_0 X_0 + \sum_{i=1}^{L} H_i X_i \mathcal{T}_i + W = H_0 X_0 + \mathcal{E}$$

(4)

where $\mathcal{E} \triangleq \sum_{i=1}^{L} H_i X_i \mathcal{T}_i + W$ is the total interference (interfering signals plus noise) in the training interval of $M$ time slots. Note that the $M_T \times M$ matrix $X_0$ is known to the receiver and will be referred to hereafter as the training matrix.

Let $R_E = E(\mathcal{E}^H \mathcal{E})$ denote the (temporal) correlation matrix of the interference term and $R_{H_0} = E(H_0^H H_0)$ be the (transmit) correlation of the desired channel. The LMMSE channel estimate is then computed as

$$\hat{H}_0 = Y \hat{C}$$

(5)

and corresponding MSE [2, 3]:

$$\text{MSE} = \text{tr} \left( R_{H_0}^{-1} + X_0 R_E^{-1} X_0^H \right) \left( R_{H_0}^{-1} \right)^{-1}$$

(6)

The problem of choosing $X_0$ so that the channel estimation MSE is minimized was addressed in [3]. To exclude the trivial solution, the optimization was performed subject to a constraint on the total training data energy, namely.

$$\min_{X_0} \text{MSE}$$

subject to

$$\text{tr}(X_0 X_0^H) \leq E_T$$
where $E_T$ is the available energy for training. The solution was shown to be based upon the eigenvalue decompositions of the above correlation matrices. Let $R_{H_0} = Q \hat{K} Q^H$ and $R_\epsilon = GAG^H$ be the eigenvalue decompositions (EVDs) of the channel and interference correlation matrices, respectively, with $\hat{K} = \text{diag}(\kappa_1, \kappa_2, \ldots, \kappa_M)$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$, and let their eigenvalues, $\kappa_i$ and $\lambda_i$, be arranged in a descending and an ascending order, respectively. The eigenvalues of $X_0 R_\epsilon^{-1} X_0^H$ are denoted by $\mu_1, \mu_2, \ldots, \mu_{M_T}$, and assumed to be arranged in a descending order. The optimal training matrix is then constructed as follows [3]:

**Proposition 1** The training matrix optimizing the criterion (7), (8) is given by

$$X_0^* = Q \left[ \begin{array}{cccc} \sqrt{\mu_1 \lambda_1} & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\mu_{M_T} \lambda_{M_T}} \end{array} \right] G^H$$

where

$$\mu_i^* = \left\{ \begin{array}{ll} \frac{E_T + \sum_{i=1}^{m_*} \lambda_i / \kappa_i}{E_T + \sum_{i=1}^{m} \lambda_i / \kappa_i}, & i = 1, 2, \ldots, m_* \\ 0, & i = m_* + 1, \ldots, M_T \end{array} \right.$$are the optimal values for the $\mu_i$'s, and $m_*$ is so chosen as to ensure their positivity:

$$m_* = \max \left\{ m \in \{1, 2, \ldots, M_T\} : \frac{\sqrt{\lambda_m}}{\kappa_m} \sum_{i=1}^{m} \sqrt{\lambda_i} = \sum_{i=1}^{m} \lambda_i < E_T \right\}$$

### 3.2. LMMSE Estimation in Two Steps

It will be instructive for the rest of the paper to interpret the LMMSE estimator as a two-step procedure, as follows. In the first step, the power of the interference is minimized under a constraint that preserves the desired component in (4). Let the $M \times M_T$ matrix $C_1$ represent the corresponding filter. Then its output will be written as:

$$Y_1 = Y C_1 = H_0 X_0 C_1 + \epsilon C_1 = H_0 X_0 C_1 + \epsilon_1$$

with $\epsilon_1$ denoting the output interference. The optimization criterion for the selection of $C_1$ is given by $\text{min}_C \ E[\|\epsilon_1\|^2_F]$ subject to (s.t.) the “zero-forcing” constraint that $X_0 C_1 = I_{M_T}$. The solution to this problem is known to be given by the Gauss-Markov estimator [2]

$$C_1 = R_\epsilon^{-1} X_0^H (X_0 R_\epsilon^{-1} X_0^H)^{-1}$$

if and only if $X_0$ is full row rank. The resulting (unbiased) channel estimate,

$$Y_1 = H_0 + \epsilon_1,$$ is then improved, by minimizing $E[\|Y_1 C_2 - H_0\|^2_F]$, through an $M_T \times M_T$ filter $C_2$. Thus, in this second step the channel statistics are also taken into account and the solution is given by the Wiener filter

$$C_2 = R_{Y_1}^{-1} R_{Y_1, H_0}$$

where $R_{Y_1} = E(Y_1 Y_1^H)$ and $R_{Y_1, H_0} = E(Y_1 H_0^H)$. Making use of the assumptions in Section 2 and utilizing the matrix inversion lemma [2], one can verify that the cascade of these two steps is equivalent to the LMMSE estimator in (5), that is, $C = C_1 C_2$.

### 3.3. The Two-Sided Bayesian Minimum-Disturbance Estimator (2S B-MDE)

An overbar notation, $(\cdot)$, will be used to distinguish the quantities involved in the scheme developed here from their counterparts in the B-MDE as described above. Channel estimation will be now formulated as a two-step procedure, where processing in each step is performed on a different side of the data matrix. The first filter is as in the LMMSE above, namely $C_1$, which processes the data on the right side (temporal dimension). Call the second-step filter $\tilde{C}_2$. This is to be applied on the left side (spatial dimension):

$$\tilde{Y}_2 = \tilde{C}_2 Y_1 = \tilde{C}_2 H_0 + \tilde{C}_2 \epsilon_1$$

Once again, the MSE criterion, $\min_{\tilde{C}_2} E[\|\tilde{C}_2 Y_1 - H_0\|^2_F]$, is adopted and hence the solution is given by the Wiener filter:

$$\tilde{C}_2 = \tilde{R}_{Y_1, H_0}^{-1} \tilde{R}_{Y_1}^{-1}$$

where $\tilde{R}_{Y_1} = E(Y_1 Y_1^H)$ is the (spatial) autocorrelation of $Y_1$, and $\tilde{R}_{Y_1, H_0} = E(Y_1 H_0^H)$ is its cross-correlation with $H_0$. From (11) and the model assumptions in Section 2 the following formula for $\tilde{C}_2$ results:

$$\tilde{C}_2 = \tilde{R}_{H_0}^{-1} (\tilde{R}_{H_0} + \tilde{R}_{\epsilon_1})^{-1}$$

where $\tilde{R}_{H_0} = E(H_0 H_0^H)$ is the channel (receive) correlation matrix and $\tilde{R}_{\epsilon_1} = E(\epsilon_1 \epsilon_1^H)$ the residual interference (spatial) correlation matrix. We can also easily obtain the following alternative expression for $\tilde{C}_2$:

$$\tilde{C}_2 = \left( \tilde{R}_{H_0}^{-1} + \tilde{R}_{\epsilon_1}^{-1} \right)^{-1} \tilde{R}_{\epsilon_1}^{-1}$$

The minimum MSE turns out to be given by:

$$\text{MSE} = tr \left[ \left( \tilde{R}_{H_0}^{-1} + \tilde{R}_{\epsilon_1}^{-1} \right)^{-1} \right]$$
4. OPTIMAL TRAINING DESIGN

We will use the same criterion as before for the optimization of the training input, namely the minimization of the MSE (given by (16)) under a limited energy budget:

$$
\min_{X_0} \text{tr} \left( R_{H_0}^{-1} + R_{\tilde{e}_i}^{-1} \right) 
$$

s.t. \( \text{tr}(X_0X_0^H) \leq E_T \) (17)

The cost function in (17) has to be analyzed first to clarify its dependence on \( X_0 \). To this end, we may appeal to our assumptions of uncorrelatedness and the model (2). Moreover, we will make the reasonable assumption that the noise is negligible compared to the colored interference (interference-limited scenario) and there is only one \((L = 1)\) dominant interferer, whose channel fading can be considered as spatially uncorrelated at the receiver side, i.e., \( R_{e_i} = I_{M_R} \) (due, e.g., to large angle spread) [6]. One can then verify that the term \( R_{\tilde{e}_i} \) can be written as:

$$
R_{\tilde{e}_i} = \text{tr}(T_1C_1C_1^HT_1^H)M_TP_1I_{M_R} \quad (19)
$$

We can also arrive at the following expression for the desired channel receive covariance:

$$
\tilde{R}_{H_0} = M_T R_{v_0} \quad (20)
$$

which suggests that this matrix can be considered as being known at the receiver [4]. Substituting the above expressions in (16) and representing the matrix \( R_{v_0} \) in terms of its EVD, we obtain the following expression for the MSE:

$$
\text{MSE} = \sum_{k=1}^{M_R} \left[ 1 - \frac{1}{M_T \lambda_{v_k}} - \frac{1}{\text{tr}(T_1C_1C_1^HT_1^H)M_TP_1} \right]^{-1} \quad (21)
$$

where \( \lambda_{v_k}, k = 1, 2, \ldots, M_R \) are the eigenvalues of \( R_{v_0} \). In view of (10) and the expression \( R_{e_i} = M_TM_R P_1 T_1^H \) for the temporal interference correlation matrix, one can write:

$$
\text{tr}(T_1C_1C_1^HT_1^H) = \frac{1}{M_TM_R P_1} \text{tr} \left( X_0 R_{e_i}^{-1} X_0^H \right)^{-1} \quad (22)
$$

It is clear from (21) that it is only through the last relation that the MSE depends on \( X_0 \). Moreover, this kind of dependence is crucial for the relaxation of the requirement for beamforming along the channel eigen-directions in the new training scheme. Indeed, let \( X_0 = US \Sigma V^H \) be the singular value decomposition (SVD) of \( X_0 \) where \( U \) and \( V \) are unitary matrices of dimensions \( M_T \times M_T \) and \( M \times M \), respectively, and \( \Sigma \) is the \( M_T \times M \) matrix containing the singular values of \( X_0 \) (say, in nondecreasing order): \( \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_{M_T} \). It is readily verified that the MSE in (21) is an increasing function of \( \text{tr}(T_1C_1C_1^HT_1^H) \). Furthermore, this quantity is independent of the matrix \( U \), while the constraint (18) only depends on \( \Sigma \). Thus, the training optimization problem reduces to the unconstrained minimization of (22) with respect to \( \Sigma \) and \( V \).

The solution to this problem is detailed in the following (the proof is omitted due to lack of space):

**Proposition 2** The training matrix optimizing the criterion (17), (18) is given by

$$
\bar{X}_0 = U \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{M_T} & 0 & \cdots & 0 \end{bmatrix} G^H \quad (23)
$$

where \( \sigma_i = \sqrt{\frac{E_T}{\Sigma_{j=1}^{M_R} \lambda_j}} \) for \( i = 1, 2, \ldots, M_T \), and \( U \) can be any unitary \( M_T \times M_T \) matrix.

**Remarks.**

1. Note that the above scheme only involves interference correlation (its \( M_T \) smallest eigenvalues and corresponding eigenvectors) and is completely transparent to any channel-related information. This results in a considerable reduction in the training cost, with respect to the B-MDE scheme of [3], both in terms of feedback overhead and transmitter complexity. Note, also, that if no account is taken of the interference temporal color, that is, if the matrix \( G \) is also allowed to be arbitrarily chosen, equal power sharing results and the scheme is completely relieved of the need to feedback any information and perform any kind of beamforming. Interestingly, as shown in the simulations section, using this simple training with the 2S B-MDE can lead to a noticeable performance gain over the corresponding B-MDE scheme.

2. Employing a training matrix as in Proposition 2 can also be seen to result in considerable simplifications for the receive filters. If \( X_0 \) is as in (23), then eq. (10) becomes

$$
C_1 = G_{M_T} \begin{bmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{M_T} \end{bmatrix} U^H \quad (24)
$$

with \( G_{M_T} \) denoting the \( M \times M_T \) matrix consisting of the first \( M_T \) columns of \( G \). Since \( U \) can be any unitary matrix, the identity is a legitimate choice, i.e., \( U = I_{M_T} \). If, moreover, \( R_{e_i} \) (assumed to be) a scaled identity matrix, then we can also choose \( G = I_M \). It is easy to see that, in this case, eq. (24) is equivalent to simply taking the first \( M_T \) columns of \( Y \) and multiplying them with \( \sqrt{\frac{E_T}{M_T}} \). With \( X_0 \) as in (23), \( \bar{C}_2 \) is also simplified as

$$
\bar{C}_2 = R_{v_0} \begin{bmatrix} \frac{(\sum_{i=1}^{M_T} \lambda_i)^2}{E_T M_T M_R} I_{M_R} \\
\end{bmatrix}^{-1} \quad (25)
$$

and is reduced to a simple scaling for \( R_{v_0} = I_{M_R} \). Observe that, in contrast to the receive filter in (5) which requires the channel transmit correlation, it is the receive correlation that is employed in the above receiver.
5. SIMULATION RESULTS

We assume $L = 1$ interferer with its source resulting from a linear temporal filtering of an i.i.d. QPSK sequence. Both channels, desired and interfering, are generated according to the model (2). The transmit and receive correlation matrices follow the exponential model [5], namely, they are built as Hermitian with entries $(R)_{i,j} = r^{j-i}, j \geq i$, where $r$ is the (complex) normalized correlation coefficient with magnitude $\rho = |r| < 1$. For the interferer, the receive channel fading is spatially uncorrelated ($\rho = 0$). The novel scheme, in both the optimal and sub-optimal (assuming white interference) versions, is compared with the B-MDE of [3] as well as its variants that are based upon the assumption of uncorrelated channel and (i) temporally white and (ii) colored interference (same with [7]). We consider three different antenna setups: (a) the symmetric case, with $M_T = M_R$, and two asymmetric cases, (b) $M_R > M_T$, and (c) $M_R < M_T$. In Fig. 1, the channel estimation MSE, normalized with the channel size $M_RM_T$, is plotted versus the signal-to-interference ratio (SIR), for an interference-to-noise ratio (INR) of 20 dB. Due to lack of space, only results for strongly ($\rho = 0.9$) correlated channels are shown here. The 2S B-MDE is seen to always lie between the B-MDE variant (ii) of analogous training requirements and the optimal B-MDE scheme. Moreover, the sub-optimal 2S B-MDE, with no need for feedback or eigenbeamforming, compares quite favorably to the corresponding B-MDE variant (i). In fact, as seen in Fig. 1(c), despite its simplicity, it can be comparable to or even better than the B-MDE variant (ii) for sufficiently low SIR values. As a conclusion, the proposed scheme, in view of its quite good performance and low requirements for feedback and beamforming, could provide a highly efficient, low cost solution for channel estimation in interference-limited environments.

6. REFERENCES


**Fig. 1.** MSE performance of the estimators for strong channel correlation ($\rho = 0.9$); INR=20 dB. (a) $M_T = 4$ transmit and $M_R = 4$ receive antennas, with $M = 8$ training vectors. (b) $M_T = 4$, $M_R = 8$, $M = 10$. (c) $M_T = 8$, $M_R = 4$, $M = 10$. 