Training-based estimation of correlated MIMO fading channels in the presence of colored interference

Dimitrios Katselis\textsuperscript{a}, Eleftherios Kofidis\textsuperscript{b,}\textsuperscript{*}, Sergios Theodoridis\textsuperscript{a}

\textsuperscript{a}Department of Informatics and Telecommunications, University of Athens, Panepistimioupolis, 157 84 Athens, Greece

\textsuperscript{b}Department of Statistics and Insurance Science, University of Piraeus, 185 34 Piraeus, Greece

Received 19 April 2006; received in revised form 24 February 2007; accepted 27 February 2007

Available online 12 March 2007

Abstract

In this paper, training-based estimation of correlated block fading channels in a multiple-antenna, multi-user environment is considered. The linear minimum mean squared error (LMMSE) estimator is presented first. The problem of optimally designing the training data set, so as to minimize the mean squared channel estimation error subject to a total transmit power constraint, is then addressed. The design is based on the assumption of the availability of the channel and interference second-order statistics at the transmitter. It is shown that the optimal transmission directions are dictated jointly by the eigen-decompositions of the channel and interference covariance matrices. Their roles, in the channel estimation and interference suppression tasks, respectively, are revealed in the optimal transmit beamformer structure. The simulation results demonstrate that the gain in the estimation performance from using the optimal training sequence increases considerably with increasing spatial fading correlation, especially in strong interference environments.

Copyright \textcopyright{} 2007 Elsevier B.V. All rights reserved.

Keywords: Beamforming; Covariance feedback; Flat fading; Interference suppression; Channel estimation; Multiple-input multiple-output (MIMO); Minimum mean square error (MSE); Optimal training; Power allocation; Water filling

1. Introduction

Wireless communication systems employing multiple antennas for transmission and reception have attracted a lot of interest in the last decade or so [1–4]. This is due to the considerable increase in spectral efficiency and reliability that multiple-input multiple-output (MIMO) systems can offer, in fading environments [5], compared to single-antenna systems. Moreover, these gains are possible, via spatial multiplexing and/or space-time coding, at no extra cost in bandwidth. More recently, the potential of MIMO transceivers for increased data rates has started to be investigated in multiuser environments as well, where spatial diversity is combined with multiuser diversity [6].

In order to fully exploit the advantages of MIMO systems, perfect channel state information (CSI) is commonly required at the transmitter and/or the receiver. Of course, such a requirement cannot be met even in the simplest cases, due to the presence of background noise and the power and bandwidth limitations of wireless systems. CSI statistical...
information may be employed at the transmitter, to implement transmit beamforming. At the receiver, channel estimates, as accurate as possible, are desirable. These are commonly obtained with the aid of training (pilot) input data.

Channel estimation, a well-studied problem in the context of single-antenna communications, gets more difficult in MIMO systems, due to the increased number of channel coefficients that need to be estimated. Hence, the problem of choosing the training sequence in an optimal way becomes of a particular importance here. In the more daunting scenario of a MIMO multiuser network, the problems of channel estimation and optimal training design are even more challenging and crucial, since the colored multiple-access interference must be suppressed as well [7,8].

The problem of constructing optimal training signals for MIMO channel estimation has been studied in several works, for the single-user case, using the channel estimation accuracy (e.g., [9–16]) and/or capacity or error rate bounds (e.g., [11,17,18]) as optimization criteria. The MMSE-optimum training in a multiuser context was reported in [19], for spatially uncorrelated channels. The scenario of both colored interference and correlated channel fading was only recently considered in [20]; however, the results are confined to multiple-input single-output (MISO) systems. In all of these works, the design is based on the assumption of the knowledge of the interference and/or channel covariances at the transmitter (via, e.g., covariance feedback [21]).

In this paper, the more general case of a multiuser MIMO network with correlated fading channels is addressed. First, the linear minimum mean squared error (LMMSE) channel estimator for this scenario is considered. The training data matrix that minimizes the MSE of the corresponding channel estimate is then derived, under a constraint on the total transmit power. It is built upon the eigen-decomposition of the channel and interference correlation matrices, which are, as usual, assumed to be known at the transmitter. The structure of the optimal transmit beamformer reveals the role played by the eigen-modes of the channel and interference correlation matrices in the channel estimation and interference reduction tasks, respectively. The optimal power allocation scheme is also derived. The simulation results demonstrate the effect of the optimal training on the channel estimation accuracy, over a wide range of signal-to-interference-and-noise ratio (SINR) values, for several system sizes and correlation strengths. Gains in the performance of a channel-based detector are also illustrated through a bit error rate (BER) simulation example.

The rest of the paper is organized as follows. The system model is described in Section 2. In Section 3, the LMMSE channel estimator is presented. The optimal training data matrix is determined in Section 4, and some practical issues are discussed. Simulation results are presented in Section 5. Section 6 concludes the paper.

Notation. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. Superscripts T and H stand for transposition and conjugate transposition. \( \| \cdot \|_F \) is the Frobenius norm. The expectation and matrix trace operators are denoted by \( E(\cdot) \) and \( \text{tr}(\cdot) \), respectively. \( I_m \) denotes the \( m \times m \) identity matrix.

2. System model

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

\[
y(n) = H_0 x_0(n) + \sum_{i=1}^{L} H_i x_i(n) + w(n). \tag{1}
\]

In the above model, \( y(n) \) denotes the received \( M_R \times 1 \) vector at time \( n \). \( H_0 \) is the \( M_R \times M_T \) matrix of the channel gains for the desired transmitter–receiver pair, and \( H_i, i = 1, 2, \ldots, L \), is the channel matrix for the \( i \)th interferer. The desired and interfering signals transmitted at time \( n \) are in the \( M_T \times 1 \) vectors \( x_0(n) \) and \( x_i(n), i = 1, 2, \ldots, L \), respectively. All signals are assumed zero mean, possibly with temporal and/or spatial correlation. The interfering signals are assumed uncorrelated with the desired one. The \( M_R \times 1 \) vector \( w(n) \) represents the background noise at the receiver front end, which is assumed to have zero mean and be uncorrelated with the rest of the signals.

3. LMMSE channel estimation

Consider that \( M \) time slots per frame are devoted to training. Grouping together the corresponding \( M \)
received vectors in an $M_R \times M$ matrix, (1) can be written as
\[ Y \triangleq [y(n) \ y(n-1) \ \ldots \ y(n-M+1)] = H_0 \begin{bmatrix} x_0(n) & x_0(n-1) & \ldots & x_0(n-M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_L(n) & x_L(n-1) & \ldots & x_L(n-M+1) \end{bmatrix} x_0 + [H_1 \ H_2 \ \ldots \ H_L] \begin{bmatrix} x_1(n) \\ x_1(n-1) \\ \vdots \\ x_1(n-M+1) \\ x_2(n) \\ x_2(n-1) \\ \vdots \\ x_2(n-M+1) \\ \vdots \\ x_L(n) \\ x_L(n-1) \\ \vdots \\ x_L(n-M+1) \end{bmatrix} \]

\[ + [w(n) \ w(n-1) \ \ldots \ w(n-M+1)] w \]

or
\[ Y = H_0 X_0 + H_{\text{int}} X_{\text{int}} + W = H_0 X_0 + \varepsilon, \quad (2) \]

where $\varepsilon \triangleq H_{\text{int}} X_{\text{int}} + W$ denotes 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.

The goal of this section is to derive the LMMSE channel estimator for a given training matrix $X_0$. For a proper estimation of $H_0$, at least as many observations as unknowns are required, hence the assumption that $M \geq M_T$ [9,17] is adopted in the sequel. Let $C$ denote the corresponding $M \times M_T$ filter, operating on the data $Y$ from the right. Then the optimization criterion can be written as
\[ \min_C E(\|YC - H_0\|_F^2) \quad (3) \]

and the solution for $C$ is given by the Wiener filter [22],
\[ C = R^{-1}_Y R_{YH_0}, \quad (4) \]

where $R_Y = E(YH)Y$ is the (temporal) autocorrelation matrix of $Y$ and $R_{YH_0} = E(YH_0)$ its cross-correlation with $H_0$. Using (2) and the assumption of uncorrelatedness between desired and interfering signals, the above expression for the optimum $C$ becomes
\[ C = (X_0^H R_{H_0} X_0 + R_\varepsilon)^{-1} X_0^H R_{H_0}, \quad (5) \]

where $R_\varepsilon = E(\varepsilon\varepsilon^H)$ is the (temporal) correlation of the interference term and $R_{H_0} = E(H_0^H H_0)$ is the (transmit) channel correlation. Mobilizing the matrix inversion lemma [22] one can easily arrive at the following alternative expression for $C$:
\[ C = R_\varepsilon^{-1} X_0^H (X_0 R_\varepsilon^{-1} X_0^H + R_{H_0}^{-1})^{-1}. \quad (6) \]

The above is summarized in the following:

**Proposition 1.** The LMMSE estimate of $H_0$ in (2) is given by
\[ \hat{H}_0 = YC = YR_\varepsilon^{-1} X_0^H (X_0 R_\varepsilon^{-1} X_0^H + R_{H_0}^{-1})^{-1}. \quad (7) \]

One can easily see that this result generalizes to the present general setup the LMMSE estimators derived earlier for the cases of white interference [16] and uncorrelated channel [19]. Using (5) in (3), the minimum value of MSE can now be computed as
\[ \min_C E(\|YC - H_0\|_F^2) = \text{tr}[(R_{H_0} - R_{H_0} X_0 R_\varepsilon^{-1} X_0^H)^{-1} X_0^H R_{H_0}] \]
\[ = \text{tr}[(R_{H_0}^{-1} + X_0 R_\varepsilon^{-1} X_0^H)^{-1}], \quad (8) \]

where the last equality follows directly from the matrix inversion lemma [22].

4. **Optimal training design**

The MSE corresponding to the LMMSE estimator (7) is given by Eq. (8), for any $M_T \times M$ training matrix $X_0$. The question that now arises is how one can choose $X_0$ so that the MSE of the channel estimate be minimized. This problem is addressed in this section. In order to exclude trivial solutions for $X_0$, a constraint on its total energy, corresponding to a limited transmit power budget, is included in the optimization. The problem can then be formulated as
\[ \min_{X_0} \text{tr}[(R_{H_0}^{-1} + X_0 R_\varepsilon^{-1} X_0^H)^{-1}] \quad \text{s.t.} \ \text{tr}(X_0 X_0^H) \leq E_T, \quad (9) \]

where $E_T$ denotes the available energy for training. Note that the trace function in (9) equals the sum of the inverses of the eigenvalues of the matrix $R_{H_0}^{-1} + X_0 R_\varepsilon^{-1} X_0^H$. Let $R_{H_0} = QKQ^H$ and
\( \mathbf{R}_d = \mathbf{G} \Lambda \mathbf{G}^H \) be the eigenvalue decompositions (EVDs) of the channel and interference correlation matrices, respectively, with \( \mathbf{K} = \text{diag}(\kappa_1, \kappa_2, \ldots, \kappa_M) \) and \( \mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M) \), and let their eigenvalues, \( \kappa_i \) and \( \lambda_i \), be arranged in a descending and an ascending order, respectively. The eigenvalues of \( \mathbf{X}_0 \mathbf{R}_d^{-1} \mathbf{X}_0^H \) are denoted by \( \mu_1, \mu_2, \ldots, \mu_M \), and assumed to be arranged in a descending order. The solution to the above optimal training design problem is provided in the following proposition:

**Proposition 2.** The training matrix optimizing the criterion (9), (10) is given by

\[
\mathbf{X}_0 = \mathbf{Q} \sqrt{\mu_1 \lambda_1} \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\mu_2 \lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\mu_M \lambda_M} & 0 & \cdots & 0 \end{bmatrix} \mathbf{G}^H,
\]

(11)

where

\[
\mu_i^* = \begin{cases} \sqrt{\frac{E_T + \sum_{j=1}^m \frac{\lambda_j}{\kappa_j}}{\kappa_i}} - \frac{1}{\kappa_i}, & i = 1, 2, \ldots, m^*, \\ 0, & i = m^* + 1, \ldots, M \end{cases}
\]

(12)

are the optimal values for the \( \mu_i^* \)s, and \( m^* \) is so chosen as to ensure their positivity:

\[
m^* = \max \left\{ m \in \{1, 2, \ldots, M\} : \sqrt{\frac{\mu_i^*}{\kappa_i}} \alpha_i \sum_{i=1}^m \sqrt{\lambda_i} < E_T \right\}.
\]

(13)

**Proof.** See Appendix A.

A physical interpretation of the optimal training matrix \( \mathbf{X}_0 \) in terms of transmit beamforming may readily result as follows. Write (11) in the more compact form

\[
\mathbf{X}_0 = \mathbf{Q}_{mv} \text{diag}(\sqrt{\mu_1 \lambda_1}, \sqrt{\mu_2 \lambda_2}, \ldots, \sqrt{\mu_M \lambda_M}) \mathbf{G}_{mv}^H,
\]

(14)

where the matrices \( \mathbf{Q}_{mv} \) and \( \mathbf{G}_{mv} \) consist of the first \( m^* \) columns of \( \mathbf{Q} \) and \( \mathbf{G} \), respectively, and recall the orderings of the corresponding eigenvalues, namely, \( \kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_M \) and \( \lambda_1 \geq \lambda_2 \leq \cdots \leq \lambda_M \). It is thus seen that utilizing \( \mathbf{X}_0 \) for training amounts to transmitting in directions which result from an optimal combination of the strong channel eigen-directions \( \mathbf{Q}_{mv} \) and those eigen-directions along which the interference-plus-noise is weak \( \mathbf{G}_{mv} \). The development of this scheme was based on both the channel and the interference statistics, with the goal of minimizing the channel estimation MSE. Therefore, the corresponding channel estimator, Eq. (7), will be referred to hereafter as the **Bayesian minimum-disturbance estimator (B-MDE)**.

Using Proposition 2 in (8), the following expression for the MSE of the B-MDE results:

\[
\text{MMSE} = \frac{\left(\sum_{i=1}^{m^*} \sqrt{\lambda_i}\right)^2}{E_T + \sum_{i=1}^{m^*} \frac{(\lambda_i/\kappa_i)}{\kappa_i}} + \sum_{i=m^*+1}^{M_T} \kappa_i.
\]

(15)

**Remarks.** (1) It is easily verified that, with the optimal training matrix \( \mathbf{X}_0 \), the formula for the LMMSE channel estimate in (7) takes the following simpler and more insightful form:

\[
\hat{\mathbf{H}}_0 = \mathbf{Y} \mathbf{Q}_{mv} \text{diag}(\zeta_1, \zeta_2, \ldots, \zeta_{m^*}) \mathbf{Q}_{mv}^H,
\]

(16)

where

\[
\zeta_i = \sqrt{\frac{\kappa_i}{\mu_i^*}}, \quad i = 1, 2, \ldots, m^*.
\]

(17)

Note how the structure of the B-MDE, as exhibited in Eq. (16), reveals its mixed role, as a two-step procedure: the interference is reduced in the first step, and this facilitates the channel estimation task performed in the second step.

(2) One can easily see that the above results specialize to those reported in the literature for the special cases of white interference (i.e., \( \mathbf{R}_d \propto \mathbf{I}_M \)), e.g., [11,16] and/or white channel [19]. Moreover, the above training scheme was also arrived at in [20] for the class of MISO systems.

(3) If both the interference and the desired channel are white, i.e., \( \mathbf{R}_d = \lambda \mathbf{I}_M \) and \( \mathbf{R}_{I} = \kappa \mathbf{I}_{M_T} \), then one can choose the matrices \( \mathbf{G} \) and \( \mathbf{Q} \) in Proposition 2 as identities. Moreover, in that case, \( m^* \) in (12) turns out to be equal to \( M_T \), and all diagonal entries of \( \mathbf{X}_0 \) become equal to \( \sqrt{E_T/M_T} \), corresponding to a uniform power sharing among the antennas. Furthermore, the \( \zeta_i \)s in (17) are now all equal to \( \zeta = \sqrt{E_T/M_T/(E_T/M_T + \lambda/\kappa)} \) and the LMMSE estimation in (16) reduces to keeping only the first \( M_T \) received vectors and multiplying them with \( \zeta \). Thus, in this scenario, the last \( M - M_T \) columns of \( \mathbf{Y} \), which, as seen in (14), are nothing but interference, are not needed unless the interference power \( \lambda \) has to be estimated as well. Finally, the MMSE in (15) will then be given by

\[
\text{MMSE} = M_T \lambda/(E_T/M_T + \lambda/\kappa).
\]
4.1. Practical considerations

In the training scheme developed above, the availability of the eigen-decompositions of the channel and interference correlation matrices at the transmitter is presumed. This is a common assumption in such scenarios, see, e.g., [11,16,19]. For slowly time-varying environments, as it is the case studied in this paper, this requirement can be met by assigning the receiver the task of estimating these statistics and feeding them back to the transmitter, through a low-rate feedback channel [21,23]. This is possible in view of the fact that the variation rate of the channel and interference second-order statistics is much lower than that of the channels themselves. Hence, an accurate estimate of $\hat{R}_H$ can be obtained on the basis of a number of recently computed channel estimates. Using the uncorrelatedness assumption for the desired and the interfering signals, the interference statistics can then be estimated from the relation $\hat{R}_Y = X^H_0 R_{H_0} X_0 + R_\xi$. Furthermore, the feedback overhead can be considerably reduced if the structure of the correlation matrices (i.e., Hermitian and Toeplitz) is adequately exploited. For example, the interference correlation matrix is viewed as circulant in [19] (which is close to be true for large enough $M$) and hence the associated eigenvectors and eigenvalues can be approximated by the DFT basis vectors and the DFT of the corresponding Hermitian sequence, respectively. As a result, only the estimates of its $M$ smallest eigenvalues need to be computed and fed back to the transmitter, along with their indices.

On the other hand, in cases where the channel and signal statistics and the network parameters (number of interferers, antenna array sizes, etc.) are time invariant (or remain constant over a relatively large number of frames), one can resort to an offline estimation of the correlation matrices and their eigen-decompositions. Moreover, commonly used channel models [24] can be exploited to provide rough estimates of the channel correlation matrix. Under the above assumptions, such a “batch” scheme can offer a good performance with a significant reduction in the computational and communication overheads. This approach was adopted in conducting the simulation experiments in this paper.

5. Simulation results

The effect of the use of the B-MDE on the channel estimation performance is demonstrated in this section via simulation results for a wide range of input SINRs and for various degrees of channel correlation and antenna setups. Three (i.e., $L = 3$) interferers are considered, with transmit antenna arrays of the same size ($M_T$) with the desired transmitter. All channels are assumed frequency flat and quasi-stationary, with their coefficients changing independently from one block to another. The desired channel and two of the interfering channels are spatially correlated. Their correlation follows the well-known separable (Kronecker) model, $H = R_{H_{w}}^{1/2}H_{w} R_{H_{w}}^{1/2}$, where the matrix $H_{w}$ represents a white channel, with i.i.d. zero-mean, unit variance, circularly symmetric, complex Gaussian entries, and $R_{H_{w}}^{1/2}$ and $R_{H_{w}}^{1/2}$ are square roots of the receive and transmit fading correlation matrices, respectively [24]. Both correlation matrices follow the exponential model [25], i.e., they are built as Hermitian with entries

$$\langle R \rangle_{ij} = r^{j-i}, \quad j > i,$$

where $r$ is the (complex) normalized correlation coefficient with magnitude $\rho \triangleq |r| < 1$. The interfering sources are generated through a linear temporal filtering of i.i.d. QPSK sequences.

Three LMMSE estimators are compared in terms of their MSE performance. The optimal estimator, developed in Section 4, and two sub-optimal variants, based upon the assumption of uncorrelated channel ($Q = I_{M_T}$ in (16)), and (i) temporally white and (ii) colored interference are studied. Three different antenna setups are considered, corresponding to the symmetric case of equal numbers of transmit and receive antennas, and two asymmetric cases of more antennas at the receiver/transmitter. The last two cases can be viewed as resembling the extreme models of single-input multi-output (SIMO) and MISO systems, respectively.

The channel estimation MSE, normalized with the channel size $M_R M_T$, is plotted in Figs. 1–3 for weakly ($\rho = 0.1$), moderately ($\rho = 0.5$), and strongly ($\rho = 0.9$) correlated channels, respectively, versus input SINR. The latter is defined as the ratio of the channel input signal power to the interference plus noise power. In Fig. 1, the curves for the optimal scheme and its variant assuming white channel and colored interference are almost indistinguishable. This should be expected in view of the fact that the assumption of a diagonal $R_{H_{w}}$ underlying the sub-optimal scheme is close to hold true in this case. However, as seen in Figs. 2 and 3, as the channel correlation becomes more significant, the B-MDE estimator greatly outperforms its simplified variants, with the difference in performance being more
apparent in the low SINR region. Regarding the effect of the different relative antenna array sizes, one can see that, as expected, the SIMO-like case (b) is the

Fig. 1. MSE performance of the estimators for weak channel correlation ($\rho = 0.1$). (a) $M_T = 4$ transmit and $M_R = 4$ receive antennas, with $M = 8$ training vectors. (b) $M_T = 4$, $M_R = 10$, $M = 12$. (c) $M_T = 10$, $M_R = 4$, $M = 12$.

Fig. 2. As in Fig. 1, with medium channel correlation ($\rho = 0.5$).
easy (lowest MSE) and the MISO-like (c) is the difficult one (highest MSE), with the symmetric case (a) lying in between. A significant gap between the two sub-optimal schemes is observed in all cases, demonstrating the crucial role played by the interference eigen-directions in the estimation performance. The fact that this difference in performance accentuates as the number of antennas at the receiver increases over that at the transmitter could be explained by noting that such a system configuration allows more of the undesired power to interfere with the desired signals. In contrast, as one can observe especially in Fig. 3, it is in systems with less receive than transmit antennas that the channel correlation has its most prominent effect.

Computing accurate channel estimates may be important on its own. Nevertheless, it adds to the practical value of the above results to also examine the gains from the B-MDE scheme in the error rate performance of a detector which is based on the available channel estimate. Fig. 4 provides such an example for a system with $M_T = 4$ transmit and $M_R = 4$ receive antennas, with a strongly correlated desired channel. One dominant interferer, spatially white at both the transmit and receive sides [2], is assumed to be present. Both the desired and the interfering users employ, for data transmission, the full-rate space-time codes proposed in [26] for MISO systems, properly extended here to a MIMO system.

Fig. 4. Performance of the MMSE detector of [26] with strongly correlated channel ($\rho = 0.9$). The channel is perfectly known to the clairvoyant detector. Channel estimated at $\text{SINR} = 0\, \text{dB}$. Interference-limited environment, with $\text{INR} = 20\, \text{dB}$. $M_T = 4$ transmit and $M_R = 4$ receive antennas, with $M = 5$ training vectors.
The MMSE receiver is adopted [26]. Estimates of the channel were computed, with only \( M = 5 \) training samples and at an input SINR of 0 dB, using all three of the LMMSE estimators tested above. The corresponding BERs are plotted in Fig. 4 versus the received signal-to-interference ratio (SIR). Both channel estimation and data detection were performed with the assumption of an interference-limited environment, at an interference-to-noise ratio (INR) of 20 dB. As a performance benchmark, the BER for the detector with ideal channel knowledge (clairvoyant) is included in Fig. 4. These results show the gain that may also result from employing the optimal B-MDE scheme in a channel-based detector, as compared to its simplified variants.

### 6. Conclusions

Channel estimation for frequency flat MIMO channels with block correlated fading was addressed in the context of a colored interference environment. Adopting the MSE as the optimization criterion, the linear optimal (LMMSE) channel estimator was derived for this scenario. The problem of optimally designing the training input, in the sense of minimizing the channel MSE subject to a total transmit power constraint, was also considered. The structure of the optimal training matrix was shown to imply transmission directions that result through a combination of the dominant and the minimal eigen-vectors of the channel and interference correlation matrices, respectively. The optimal combination is suggested by the power allocation scheme. The simulation results demonstrated that the gain in the estimation (and detection) performance from using the optimal training sequence increases considerably with increasing fading correlation, especially in strong interference environments.

### Appendix A. Proof of proposition 2

The problem stated in Eqs. (9) and (10) was also studied in [19], for the case of a white channel, i.e., \( R_{H_0} \propto I_{M_t} \). The approach followed was to formulate a relaxed version of the problem and ensure that its solution satisfies the constraint of the original problem. An analogous approach is to be taken here to address the more general case of a correlated channel matrix. Define the matrix \( \tilde{X} = X_0 R_{\tilde{F}}^{-1/2} \), where \( R_{\tilde{F}}^{-1/2} \) denotes a Hermitian positive definite square root of \( R_{\tilde{F}}^{-1} \). Then one can write \( X_0 R_{\tilde{F}}^{-1} X_0^H = \tilde{X} \tilde{X}^H \). Furthermore, \( \text{tr}(X_0 X_0^H) \) in (10) can be written as \( \text{tr}(X_0 X_0^H) = \text{tr}(\tilde{X} R_{\tilde{F}} \tilde{X}^H) = \text{tr}(\tilde{X}^H \tilde{X} R_{\tilde{F}}) \). The constraint (10) can now be relaxed by using the following trace inequality [19, Lemma 1]:

**Lemma 1.** Let \( A \) and \( B \) be two Hermitian \( M \times M \) matrices. Arrange the eigenvalues \( a_1, a_2, \ldots, a_M \) of \( A \) in a descending order and the eigenvalues \( b_1, b_2, \ldots, b_M \) of \( B \) in an ascending order. Then \( \text{tr}(AB) \geq \sum_{i=1}^{M} a_i b_i \).

Applying this result with \( A = \tilde{X}^H \tilde{X} \) and \( B = R_{\tilde{F}} \), and using the fact that the matrices \( \tilde{X} \tilde{X}^H \) and \( \tilde{X}^H \tilde{X} \) have the same nonzero eigenvalues, \(^2\) yields

\[
\text{tr}(X_0 X_0^H) = \text{tr}(\tilde{X}^H \tilde{X} R_{\tilde{F}}) \geq \sum_{i=1}^{M_T} \mu_i \lambda_i.
\]

The above suggests the following relaxed optimization problem:

\[
\min_{X_0} \quad \text{tr}[R_{H_0}^{-1} + X_0 R_{\tilde{F}}^{-1} X_0^H]^{-1} (19)
\]

s.t. \( \mu_i \lambda_i \leq E_T \) and \( \mu_1 \geq \mu_2 \geq \cdots \geq \mu_{M_T} \geq 0. \)

(20)

Consider the singular value decomposition (SVD) of \( X_0 \):

\[
X_0 = U \Sigma V^H,
\]

where \( U \) and \( V \) are \( M_T \times M_T \) and \( M \times M \) unitary matrices containing its left and right singular vectors, respectively, and

\[
\Sigma = \begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_{M_T} & 0 & \cdots & 0
\end{bmatrix}
\]

is the matrix of its singular values, \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{M_T} > 0 \). The solution for the optimal \( X_0 \) will be determined in terms of the optimal factors \( U, \Sigma, \) and \( V \). Let \( T N T^H \) with \( N = \text{diag}(v_1, v_2, \ldots, v_{M_T}) \) be the EVD of the matrix \( R_{H_0}^{-1} + X_0 R_{\tilde{F}}^{-1} X_0^H \). Then the objective function in (19) can be written as

\[
\text{tr}[R_{H_0}^{-1} + X_0 R_{\tilde{F}}^{-1} X_0^H]^{-1} = \text{tr}(N^{-1}) = \text{tr}[T^H R_{H_0}^{-1} T + T^H X_0 R_{\tilde{F}}^{-1} X_0^H T]^{-1}.
\]

(22)

\(^2\)Note that these are, in fact, the squared singular values of \( \tilde{X} \).
Let the EVD of $X_0R_0^{-1}X_0^H$ for the optimal choice of $U$ be given by $ZMZ^H$, where $M = \text{diag}(\mu_1, \mu_2, \ldots, \mu_{M_T})$. Using the above in (22) and substituting for the EVD of $R_{H_0}^{-1}$ yields

$$\text{tr}(R_{H_0}^{-1} + X_0R_0^{-1}X_0^H) = \text{tr}(T^HQK^{-1}Q^HT + T^HZMZ^HT)^{-1}).$$

Note that the above function is, in fact, independent of $T$. It will be convenient for the rest of the proof to let $T = Z$. With this choice of $T$, the problem in (19), (20) takes the following simpler form:

$$\begin{align*}
\min_{\mu_i, z_i, i = 1, \ldots, M_T} & \quad \sum_{i=1}^{M_T} \left( \mu_i + \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2 \right)^{-1} \\
\text{s.t.} & \quad \sum_{i=1}^{M_T} \mu_i z_i \leq E_T \\
& \quad \mu_1 \geq \mu_2 \geq \cdots \geq \mu_{M_T} \geq 0,
\end{align*}$$

(23)

where $z_i, i = 1, 2, \ldots, M_T$, the columns of $Z$, are constrained to be orthonormal, and $q_i, i = 1, 2, \ldots, M_T$ denote the columns of $Q$. Now consider the corresponding Lagrangian function

$$\Gamma(\mu_1, \mu_2, \ldots, \mu_{M_T}) = \sum_{i=1}^{M_T} \left( \mu_i + \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2 \right)^{-1} + \gamma \left( \sum_{i=1}^{M_T} \mu_i z_i - E_T \right),$$

(25)

where the dependence on the vectors $z_i$ is not shown. As it will be seen later on, these in fact do not appear in the formulae for the optimal $\mu_i$'s. Furthermore, the ordering of the $\mu_i$'s is not incorporated in the above function and will be seen to naturally emerge in the optimal solution. Setting the derivatives of (25) w.r.t. $\mu_i, i = 1, 2, \ldots, M_T$, equal to zero and taking the power constraint into account yields

$$\mu_i^* = \frac{E_T + \sum_{j=1}^{M_T} \lambda_j \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2}{\sqrt{K_j \sum_{j=1}^{M_T} \sqrt{\lambda_j}}} - \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2.$$

(26)

The cost function in (23) then becomes

$$\sum_{i=1}^{M_T} \left( \mu_i^* + \frac{M_T}{K_j} |q_j^H z_i|^2 \right)^{-1}$$

$$= \sum_{i=1}^{M_T} \frac{\sqrt{\lambda_i} \sum_{k=1}^{M_T} \sqrt{\lambda_k}}{E_T + \sum_{k=1}^{M_T} \lambda_k \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2}$$

$$= \frac{\sum_{k=1}^{M_T} \sqrt{\lambda_k}}{E_T + \sum_{k=1}^{M_T} \lambda_k \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2}.$$

(27)

It is thus clear that $Z$ should now be chosen so as to optimize the following criterion:

$$\max \sum_{k=1}^{M_T} \frac{\lambda_k}{K_j} \sum_{j=1}^{M_T} \frac{1}{K_j} |q_j^H z_i|^2,$$

(28)

where $z_i, k = 1, 2, \ldots, M_T$, are orthonormal. (29)

One can readily verify that the above cost function can be equivalently written in the form

$$\text{tr}(Z^HQK^{-1}Q^HZ\Lambda_{M_T}),$$

(30)

where $\Lambda_{M_T} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{M_T})$ contains the first $M_T$ eigenvalues in $\Lambda$. It will now be shown that the optimal choice for $Z$ is $Z = Q$. To this end, the following lemma [27, p. 183] will prove useful:

**Lemma 2.** Consider any two matrices $A, B \in \mathbb{C}^{m \times n}$ and their singular values, $\sigma(A), \sigma(B), i = 1, 2, \ldots, q$, with $q = \min\{m, n\}$, arranged in a descending order. Then:

(a) The following inequality holds: $|\text{tr}(A^HB)| \leq \sum_{i=1}^{q} \sigma(A)\sigma(B)$. (b) There exist unitary matrices $P_1$ and $P_2$ such that $\max\{|\text{tr}(P_1A^HP_2B)| : P_1 \in \mathbb{C}^{n \times k}, P_2 \in \mathbb{C}^{m \times k} \text{ are unitary}\} = \sum_{i=1}^{q} \sigma(A)\sigma(B)$. (c) \[P = Z^HQ.\] Then (30) becomes

$$\text{tr}(PK^{-1}P^HA_{M_T})$$

(31)

and it suffices to prove that $P = I_{M_T}$. Note first that $P$ is unitary and (31) is nonnegative. However, since the diagonal entries of the matrices $K^{-1}$ and $\Lambda_{M_T}$ are in an ascending and not a descending order, Lemma 2 cannot be directly applied to this trace function. Nevertheless, with the aid of the $M_T$th-order exchange matrix $J$, (31) can be equivalently written as

$$\text{tr}(PK^{-1}P^HA_{M_T}),$$

where $P = JPJ$, and the diagonal entries of the matrices $K^{-1} \triangleq JK^{-1}J$ and $\Lambda_{M_T} \triangleq JA_{M_T}J$ are arranged in a descending order. Then, with $A = K^{-1}$, $B = \Lambda_{M_T}$ and $P_1 = P_2 = P$, it follows from Lemma 2 that the maximum value of (31) is $\sum_{j=1}^{M_T} (\lambda_j/k_j)$. This value is obviously attained with the choice $P = P = I_{M_T}$. It remains to prove that this is the only maximizing value for $P$. 


This is easy to see by considering the existence of a second choice for \( \mathbf{P} \), say \( \mathbf{P} \neq \mathbf{I}_{M_T} \), and observing that this would imply the equality
\[
\text{tr}(\mathbf{P}^H \mathbf{K}^{-1}(\mathbf{P}^H - \mathbf{K}^{-1}) \mathbf{A}_{M_T}^1) = 0.
\]
Since no other constraint has been imposed on \( \mathbf{A}_{M_T} \) apart from the fact that it is diagonal with ascending positive diagonal entries, for the last equality to hold true for any such matrix \( \mathbf{A}_{M_T} \), the matrix \( \mathbf{P}^H \mathbf{K}^{-1}(\mathbf{P}^H - \mathbf{K}^{-1}) \) must have zeros on its diagonal. In view of the fact that the diagonal entries of \( \mathbf{K}^{-1} \) are not all equal to each other\(^3\) and ordered, this can only happen if \( \mathbf{P} = \mathbf{I}_{M_T} \).

With \( \mathbf{Z} = \mathbf{Q} \), i.e., \( z_j = q_{ij}, j = 1, 2, \ldots, M_T \), in (26), the following expression for the optimal values of the \( \mu_i \)'s results:
\[
\mu_i = \frac{E_T + \sum_{j=1}^{M_T} \lambda_j / K_j}{\sqrt{\lambda_i} / \sum_{j=1}^{M_T} \sqrt{\lambda_j}} - \frac{1}{K_i}, \quad i = 1, 2, \ldots, M_T.
\]
(32)

Note that this solution ensures the ordering of the \( \mu_i \)'s as required by the constraints of the problem. This can be easily established by considering two successive \( \mu_i \)'s, say \( \mu_i \) and \( \mu_{i+1} \). Then:
\[
\mu_i - \mu_{i+1} = \frac{E_T + \sum_{j=1}^{M_T} \lambda_j / K_j}{\sum_{j=1}^{M_T} \sqrt{\lambda_j}} \left( \frac{1}{\sqrt{\lambda_i}} - \frac{1}{\sqrt{\lambda_{i+1}}} \right) + \left( \frac{1}{K_{i+1}} - \frac{1}{K_i} \right).
\]
From the ordering assumed for the \( \lambda_i \)'s and \( K_i \)'s, it follows that neither of the two terms on the right-hand side of the above equation can be negative. Thus, \( \mu_i - \mu_{i+1} > 0, i = 1, 2, \ldots, M_T - 1 \).

The positivity constraint for the \( \mu_i \)'s has to be satisfied as well. Define the parameter \( m^* \) as follows:
\[
m^* = \max\{m \in \{1, 2, \ldots, M_T\} : \mu_{m} > 0\}
\]
\[
= \max\left\{ m \in \{1, 2, \ldots, M_T\} : \sqrt{\frac{\mu_m}{\sum_{j=1}^{m} \sqrt{\lambda_j}}} \right\}
\]
\[
- \sum_{j=1}^{m} \left( \frac{\lambda_j}{\sqrt{\lambda_j}} \right) < E_T \}. \quad (33)
\]
One can easily check that \( m^* \) is well defined and represents the index of the smallest nonzero \( \mu_i \). In view of (32), this leads to the final formula for the optimal \( \mu_i \)'s, Eq. (12).

---

\(^3\)The contrary would imply lack of fading correlation for the desired channel.

What about the optimal choices for \( \mathbf{U}, \mathbf{\Sigma}, \) and \( \mathbf{V} \) in (21)? Observe that with \( \mathbf{Z} \) being fixed to \( \mathbf{Q} \), the minimum value of the cost function in (19) becomes \( \text{tr}[\mathbf{K}^{-1} + \mathbf{M} \mathbf{K}^{-1}] \), whereby the exact pairing of the \( \kappa_i \)'s and \( \mu_i \)'s is deduced. One can directly conclude from the relation
\[
\mathbf{M} = \mathbf{Q}^H \mathbf{X}_0 \mathbf{R}^{-1} \mathbf{X}_0^H \mathbf{Q}
\]
\[
= \mathbf{Q}^H \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \mathbf{G}^{-1} \mathbf{G} \mathbf{\Sigma} \mathbf{V}^H \mathbf{U}^H \mathbf{Q}
\]
(34)
that the choice \( \mathbf{V} = \mathbf{G} \) implies \( \mu_i = \sigma_i^2 / \lambda_i, \ i = 1, 2, \ldots, M_T \), and the resulting solution satisfies the constraint (10):
\[
\text{tr}(\mathbf{X}_0 \mathbf{Q}^H) = \sum_{i=1}^{M_T} \sigma_i^2 = \sum_{i=1}^{M_T} \mu_i^* \lambda_i.
\]

It then follows from Eq. (34) that \( \mathbf{U} = \mathbf{Q} \). Putting together all the above results in (22), the formula in Eq. (11) results for the optimal training matrix.

References