Error-Entropy based Channel State Estimation of Spatially Correlated MIMO-OFDM

H.D. Tuan¹, H.H. Kha¹ and H.H. Nguyen²
¹University of New South Wales, Sydney, Australia. Email: {h.d.tuan,h.k.ha}@unsw.edu.au
²University of Saskatchewan, Saskatoon, Canada. Email: ha.nguyen@usask.ca.

Abstract—This paper deals with optimized training sequences to estimate multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) channel states in the presence of spatial fading correlations. The optimization criterion is the entropy minimization of the error between the high dimensional and correlated channel state and its estimator. The globally optimized training sequences are exactly solved by a semi-definite programming (SDP) of tractable computational complexity \( O((M_t(M_t+1)/2)^2) \), where \( M_t \) is the transmit antenna number. With new tight two-sided bounds for the objective function, the optimal value of the generic SDP can be approximately solved by the standard water-filling algorithm. Intensive simulation results are provided to illustrate the performance of our methods.

Index terms. Error-entropy, training sequences, spatial correlation, MIMO, OFDM, convex programming.

I. INTRODUCTION

Accurate knowledge of channel state estimation (CSE) at both receiver and transmitter sides of mobile wireless multi-input multi-output orthogonal frequency division multiplexing (MIMO-OFDM) systems is the base to greatly increase communication capacity by space-time exploratory tools such as pre-coding and beamforming. In fact, the channel capacity can be gained with CSE available not only at the receiver but also at transmitter, which obviously depends on the estimation quality [1]. For broadband wireless systems, which experience frequency-selective fading, OFDM can turn the channel into parallel flat fading sub-channels. However, it is not efficient to estimate these flat fading sub-channels separately because (i) they are not independent but rather correlated in time and space due to inadequate antenna spacing or scattering, and (ii) their total dimension is much larger than that of channel dimension. This is why CSE in MIMO-OFDM systems cannot be readily extended from that for flat fading channels [2]. In fact, CSE in MIMO-OFDM is a highly-structured matrix problem, which cannot be easily factorized for tractably computational solutions. The first estimator for spatially correlated MIMO-OFDM channels was obtained in [2] but its proposed training solution is locally optimized at extreme SNR regimes only. More recently, the globally optimized training sequences and estimator have been more thoroughly and correctly addressed in [3].

All modern wireless communication systems are digital, which means that the channel estimation is in fact further distorted by digitalization such as variable-rate quantization and compression under limited bit budget. Thus, traditional mean square error (MSE) is not a true distortion measure for quality of digitalized estimators. On the other hand, it is well known [4], [5] that the Shannon distortion-rate function [6] defined by the error-entropy instead of MSE is a much more appropriate measure, especially for jointly estimating and quantizing random correlated vectors. Only recently, there is an increasing interest in error-entropy minimization for CSE of flat-fading MIMO systems [7], [8] but only locally optimized solutions at extreme (low or high) SNR regimes have been appropriately addressed.

In the present paper, we also adopt error-entropy minimization (EEM) in optimizing the training sequences for CSE of spatially correlated frequency-selective MIMO-OFDM channels, which in fact needs a quite different convexity analysis and supporting mathematical framework than the global MSE minimization (MSEM) based CSE treated in [3]. Our contributions are two-fold: 1) To show in Section III that the problem can be lossless recast as tractable semi-definite programming (SDP) and thus is computationally solved by available SDP solvers. The dimension of the only matrix variable in this SDP is fixed at \( M_t \times M_t \), where \( M_t \) is the number of transmit antennas. This means that the computational complexity of this SDP-based solution is \( O((M_t(M_t+1)/2)^2) \), which is not only low but also does not depend on the size of the training sequences. Moreover, the correlation between channels can be exploited to reduce the sub-carrier number used for channel estimation while maintaining computational tractability. The CSE quality is maintained even only a small portion of OFDM sub-carriers is used for training (within one OFDM block, the rate occupancy of training sequences can be reduced to 1/256); 2) To provide in Section IV both-sided tight upper and lower bounds for the matrix objective function in general cases such that the corresponding two-sided bound optimization problems can be solved by the standard water-filling algorithms of low numerical complexity. Our intensive simulation V shows that these water-filling based solutions effectively yield nearly global minimal value of the original exact EEM.

Notations: Bold capital and lower case letters denote matrices and column vectors, respectively. \( (\cdot)^T \) and \( (\cdot)^H \) denote transpose and Hermitian transpose operations, respectively. The symbol \( \otimes \) is used for the Kronecker product and \( \text{vec}(\mathbf{X}) \) denotes the vectorization operation of matrix \( \mathbf{X} \). \( \text{tr}(\cdot) \) and \( |\cdot| \) stand for the trace and determinant of the matrix, respectively. \( \mathbf{X} \succeq 0 \) means \( \mathbf{X} \) is Hermitian symmetric and positive semi-definite. \( \mathbf{I}_n \) is the identity matrix of dimension \( n \times n \). The expectation operation is \( \mathbf{E}\{\cdot\} \), while \( \mathcal{CN}(0,\sigma^2) \) denotes a circularly symmetric complex Gaussian random variable. Furthermore, \( [A_{ij}]_{i,j=0,1,...,N} \) with matrices \( A_{ij} \) means the matrix.
with block entries $A_{ij}$. Analogously, $\text{diag}[A_{i}]_{i=0,\ldots,N}$ means the matrix with diagonal blocks $A_i$ and zero off-diagonal blocks. All logarithms in the paper are base-2.

II. MIMO-OFDM SYSTEMS AND TRAINING DESIGN

Consider a broadband MIMO-OFDM system with $M_t$ transmit antennas, $M_r$ receive antennas, whose frequency-selective MIMO fading channel is described by the transfer matrix

$$H(z) = \sum_{\ell=0}^{L-1} H_{\ell} z^{-\ell},$$

where each time-varying stationary process $H_{\ell} \in \mathbb{C}^{M_r \times M_t}$ represents the gains of the $\ell$th MIMO path. The elements $(H_{\ell})_{m,n}$ of $H_{\ell}$ are (possibly correlated) circularly symmetric complex Gaussian random variables that remain unchanged over the period of channel estimation.

The spatial correlations of MIMO channels can be modeled by the Kronecker structure of their channel matrix gain [9]

$$H_{\ell} = R_{tt}^{1/2} H_{\text{uf}} (R_{tt}^{1/2})^H, \quad \ell = 0, 1, \ldots, L - 1,$$

where the deterministic Hermitian symmetric matrix $R_{tt} = R_{tt}^{1/2} R_{tt}^{1/2} \in \mathbb{C}^{M_t \times M_t}$ models the correlation between the transmit antennas while $R_{tt}^{1/2} = R_{tt}^{1/2} R_{tt}^{1/2} \in \mathbb{C}^{M_r \times M_r}$ captures the correlation between the receive antennas. $H_{\text{uf}} \in \mathbb{C}^{M_r \times M_t}$ is time-varying stationary process, whose elements are independent and identical distributed zero-mean circularly symmetric complex Gaussian random variables with unit variance.

For $h := (\{c^T(H_0), c^T(H_1), \ldots, c^T(H_{t-1})\}) \in \mathbb{C}^{L M_t M_r}$, its correlation matrix is $R_h := E[hh^H] = \text{diag}[R_{ij} \otimes R_{jk}]_{i,j=0,1,\ldots,L-1}$ Suppose the MIMO-OFDM system uses $M = 2^{\log_2 M} \mathcal{M}$ sub-carriers to turn the frequency-selective fading channel into $\mathcal{M}$ parallel flat fading sub-channels. Each block of length $\mathcal{M}$ goes through an OFDM modulator to form an OFDM block and is transmitted via one transmit antenna. The OFDM cyclic prefix length is chosen to be longer than the channel order, $L - 1$, to avoid the inter block interference (IBI). Thus sequences $x(j) \in \mathcal{M}_t, j = 0, 1, \ldots, M - 1$ are transmitted from $M_t$ antennas on the $j$th sub-carrier. Inside just one OFDM block, $N = 2^{\log_2 N} \ll M$ training symbols are inserted on the 0th, $(2^{\log_2 M} - 2^{\log_2 N})$th, $(N - 1)$th, $(N - 2^{\log_2 N})$th sub-carriers for channel estimation. The training sequences are inserted as $s(k) = x(k2^{\log_2 N} - 2^{\log_2 N}) \in \mathcal{M}_r$, $k = 0, 1, \ldots, N - 1$.

By defining $W_N = e^{-j 2\pi/N}$, the channel transfer function corresponding to the $k2^{\log_2 M} - 2^{\log_2 N}$th sub-channel is $H_{\ell}(k) = H(W_k2^{\log_2 M} - 2^{\log_2 N}) = \sum_{\ell=0}^{L-1} H_{\ell} W_{\ell k}2^{\log_2 M} - 2^{\log_2 N}$. Thus, the normalized input-output equation for each pilot sub-carrier is $r(k) = \sqrt{\frac{E}{M_t}} H_{\ell}(k) s(k) + n(k), k = 0, 1, \ldots, N - 1$, where $\rho$ is the average signal-to-noise-ratio (SNR), $s(k) = (s_0(k), s_1(k), \ldots, s_{M_t - 1}(k))^T \in \mathbb{C}^{M_t}$ is the $k$th received signal vector, $n(k) = (n_0(k), n_1(k), \ldots, n_{M_t - 1}(k))^T \in \mathbb{C}^{M_t}$ is the training vector, and $\hat{h}(k) = (\hat{n}_0(k), \hat{n}_1(k), \ldots, \hat{n}_{M_t - 1}(k))^T$ represents additive white Gaussian noise (AGWN), whose elements are i.i.d $\mathcal{CN}(0,1)$ random variables.

For convenience, define the training symbol matrix $S = [s(0) \ s(1) \ldots s(N - 1)]^T \in \mathbb{C}^{N \times M_t}$. Then, the equation for the received signal for training can be compactly represented as $r = \sqrt{\frac{E}{M_t}} M(S) h + n$, where $r = (r^T(0), r^T(1), \ldots, r^T(N - 1))^T \in \mathbb{C}^{N M_t}$, $n = (n^T(0), n^T(1), \ldots, n^T(N - 1))^T \in \mathbb{C}^{N M_t}$, $M(S) = [I_0 S \ I_1 S \ldots \ I_{L-1} S] \times I_{M_t}$, $F_{\ell} = \text{diag}(W_{\ell})^T_{N=0,1,\ldots,N-1}, \ell = 0, 1, \ldots, L - 1$, $M^H(S)M(S) = (H^T F_{\ell} F_{\ell}^H S) \times I_{M_t}, \ell = 0, 1, \ldots, L - 1$.

There are $M_r N$ measurements for the estimation of $L M_t M_r$ unknown parameters, which are the entries of matrices $H_{\ell} \in \mathbb{C}^{M_r \times M_t}$, $\ell = 0, 1, \ldots, L - 1$. When all the entries of $H_{\ell}$ are independent, to make the estimation problem meaningful, it requires that the number of measurements be not less than the number of unknowns [8], i.e. $N \geq L M_t$. However, as the entries of $H_{\ell}$ are correlated, using large $N$ would make optimization problems much less efficient because there is not much freedom in optimizing solutions. In fact, $N \geq 2L$ is sufficient if the channel correlation is well exploited.

The MSE problem is to obtain an estimator $\hat{h}$ for $h$ with a known (deterministic) training signal $S$. As all random variables $r, h, n$ are Gaussian with zero mean, $h$ is Wiener filter (or MMSE estimator) $h = \sqrt{\frac{E}{M_t}} R_h M(S) (\frac{E}{M_t} M(S) R_h M^H(S) + I)^{-1} r$. Obviously, the error $e = h - \hat{h}$ is a Gaussian random variable with zero mean and covariance $R_e := R_h - \hat{h} = R_h^{-1} + \frac{\rho}{M_t} M^H(S) M(S)^{\dagger}$ in general, training design is to find the training matrix $S$ to obtain the conditional mean $\hat{h}$ of the channel state $h$ (which is considered as time-varying stationary process) under some criterion and subject to the normalized training power constraint $\text{tr}(S^H S) = N M_t$. Traditional estimation methods try to minimize the error between the channel state $h$ and the conditional mean $\hat{h}$ in terms of the least square [10], or mean square [2], [3], [11] $\min \text{tr}(R_e)$. Here we aim at designing the training sequences such that the error-entropy $\log |\hat{r} R_e| \min$ is minimized: $\min \log |R_e| \text{ s.t. } \text{tr}(S^H S) = N M_t$, which is equivalent to

$$\max \log |R_e^{-1} + \frac{\rho}{M_t} M^H(S) M(S)|.$$

The error-entropy is nothing but the Shannon-distribution function [6], which is also the distance between probability distributions of the random variable $h$ and its estimator $\hat{h}$. Thus, it provides an accurate measure of mismatch between $h$ and $\hat{h}$ as random variables. Therefore, it is also robust against the uncertainties with the correlation matrix $R_h$.

III. OPTIMIZED TRAINING SEQUENCE BY CONVEX PROGRAMMING

The key result of this section is presented in the following theorem.
Theorem 1: Suppose that there is Q ∈ CN×Mt of Mt columns qi ∈ CN,i =1 ,...,Mt such that
\[ Q = \{ q_{i} \}_{i=1}^{Mt} \]
for 0 ≤ m < ℓ ≤ L − 1, and
\[ Q = \{ q_{i} \}_{i=1}^{Mt} \]
Under singular value decompositions (SVDs) \( R_{\ell t} = U_{\ell} \Sigma_{\ell} V_{\ell}^{H} \), \( R_{\ell t} = V_{\ell} \Sigma_{\ell} U_{\ell}^{H} \) with \( \Sigma_{\ell} = \text{diag}(\gamma_{1,\ell},...,\gamma_{Mt,\ell}) \), \( \Sigma_{\ell} = \text{diag}(\gamma_{1,\ell},...,\gamma_{Mt,\ell}) \), \( \ell = 0,1,...,L-1 \), the optimization problem in (1) in \( S \subset \mathbb{C}^{N \times M_t} \) is equivalent to the following SDP in \( X \in \mathbb{C}^{M_t \times M_t} \):
\[
\max_{0 \leq X \in \mathbb{C}^{M_t \times M_t}, \text{tr}(X) = N M_t} \sum_{\ell=0}^{L-1} \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}|.
\]
The optimal solution \( S_{\text{opt}} \) of (1) is obtained from the optimal solution \( X_{\text{opt}} \) of (4).

A matrix \( Q \) of \( M_t \) columns \( q_{i} \), \( i = 1,2,...,M_t \) can be constructed as follows. Take \( q_{i} = (q_{i}(0),q_{i}(1),...,q_{i}(N-1))' \) with \( |q_{i}(j)|^2 = 1/\sqrt{N} \), \( i = 0,1,...,N-1 \), so that \( |q_{i}(j)| = 1 \). Then \( q_{i}, i = 2,...,M_t \) are defined from \( q_{1} \) by \( q_{i}(k) = q_{i}(k)W_{N}^{i(k-1)N} \), \( k = 0,1,...,N-1 \). For \( N \geq M_t \), \( K \) is chosen by \( K = \lfloor N/M_t \rfloor \).

On the other hand, a smaller value of \( N \) gives more freedom for optimization because the role of \( X \) in (2) becomes more relevant. However, for \( N < M_t \), the convex optimization problem (4) is only an upper bound of the nonconvex optimization problem (1). For \( N \) even such that \( N \geq 2L \), the choice \( K = L \) results in \( q_{i} = \sum_{k=0}^{N-1} W_{N}^{i(k-1)N} = 1 \).

We end this section by complexity analysis for SDP (4) in the matrix variable \( X \) of dimension \( M_t(M_t-1)/2 \). It can be solved by the existing SDP software for max-det (such as YALMIP with SDP solvers [12]) in polynomial time \( O((M_t(M_t+1)/2)^{2.5}) \) dependent on \( M_t \) only. As the number \( M_t \) is typically low, its computational load is low as well. Nevertheless, in the interest of a much faster computation, specialized convex programming algorithms to find the solution of (4) are developed in the next sections.

IV. CLOSED-FORM ITERATIVE ALGORITHMS VIA TIGHT TWO-SIDED BOUNDS

Define \( R_{\ell t} = \sum_{\ell=0}^{L-1} R_{\ell t} / L \) and \( R_{\ell t} = \text{diag}(\gamma_{i,\ell})_{i=1,2,...,M_t} \), \( \Sigma_{\ell} = \text{diag}(\gamma_{i,\ell})_{i=1,2,...,M_t} \), where \( \gamma_{i,\ell} = \max_{\ell=0,1,...,L-1} \gamma_{\ell}(i,i), \gamma_{i,\ell} = \min_{\ell=0,1,...,L-1} \gamma_{\ell}(i,i), i = 1,2,...,M_t \).

Theorem 2: In terms of \( \Sigma_{\ell} \), the objective function in (4) is two-side bounded by \( L \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| \leq \sum_{\ell=0}^{L-1} \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| \leq L \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| + L-1 \sum_{\ell=0}^{L-1} |M_t \log (|R_{\ell t}|/|R_{\ell t}|) - M_t \log (|\Sigma_{\ell}|/|\Sigma_{\ell}|)|.

Furthermore, in terms of \( \Sigma_{\ell} \), the objective function in (4) is two-side bounded by \( L \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| \leq \sum_{\ell=0}^{L-1} \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| \leq L \log |R_{\ell t}^{-1} \otimes \Sigma_{\ell}^{-1} + \frac{\rho}{M_t} X \otimes I_{M_t}| + L-1 \sum_{\ell=0}^{L-1} |M_t \log (|R_{\ell t}|/|R_{\ell t}|) - M_t \log (|\Sigma_{\ell}|/|\Sigma_{\ell}|)|.

In light of Theorem 2 and making SVD \( \sum_{\ell=0}^{L-1} R_{\ell t} = U \Lambda U^{H} \), \( \Lambda = \text{diag}(\lambda_{1,\ell},...\lambda_{Mt,\ell}) \) to facilitate the variable change \( X = U \text{diag}(y_{1},y_{2},...,y_{Mt}) U^{H} \), problem (4) (5) can be approximated by the following optimization problems in \( y_{i} \geq 0, i = 1,2,...,M_t \).

\[
\max_{\sum_{i=1}^{Mt} y_{i} = N M_t} \sum_{i=1}^{Mt} \sum_{j=1}^{Mt} \left( \lambda_{i,j}^{-1} g_{i,j}^{-1} + \frac{\rho}{LM_t} y_{i} \right)^{-1},
\]

where \( \gamma_{ij} = \text{diag}(\gamma_{ij})_{j=1,2,...,M_t} \in \{ \Sigma_{opt}, \Sigma_{min} \} \).

Problem (5) can be solved in exactly the same manner by using the iterative water filling procedure [13]. Moreover, we can find an optimal solution in a closed-form expression to its upper bound \( y_{i} = \max(\{ M_{t}(\mu - a_{i}^{-1})/\rho, 0 \}) \) with \( a_{i} = \frac{1}{LM_t} \lambda_{i} \sum_{j=1}^{Mt} \gamma_{ij} \) and \( \mu > 0 \) is chosen so that \( \sum_{i=1}^{Mt} y_{i} = N M_t \).

V. SIMULATION RESULTS

Considered in the simulation are MIMO systems with uniform linear antenna arrays at both the transmitter and receiver. For the \( \ell \)th path with path gain \( \sigma_{\ell}^{2} \), the spatial correlation matrices at the transmitter and receiver can be presented by [14] \( [R_{\ell t}]_{m,n} = \sigma_{\ell} e^{-2\pi |n-m|/\Delta_{r} \cos(\theta_{\ell})} e^{-\frac{1}{2}(2\pi |n-m|/\Delta_{t} \sin(\theta_{\ell}) \sigma_{\ell})^{2}}, \)
\( [R_{\ell r}]_{m,n} = \sigma_{\ell} e^{-2\pi |n-m|/\Delta_{r} \cos(\theta_{\ell})} e^{-\frac{1}{2}(2\pi |n-m|/\Delta_{r} \sin(\theta_{\ell}) \sigma_{\ell})^{2}} \),
where \( \Delta_{r} = \frac{d_{t}}{\lambda} \) and \( \Delta_{r} = \frac{d_{r}}{\lambda} \) denote the relative transmit and receive antenna spacings, \( d_{t} \) and \( d_{r} \) are the absolute antenna spacings with wavelength \( \lambda = \frac{c}{f} \) and carrier frequency \( f \). \( \theta_{\ell} \) and \( \theta_{\ell} \) are the angle of departure from the transmit array and the mean angle of arrival at the receive array, \( \sigma_{\ell} = \sigma_{\ell} = 8.6^\circ \) are the cluster angle spreads perceived by the transmitter and receiver. In figure legends, we refer to problems (4) and its upper bound based water-filling given at the end of Section IV as “min-error-entropy” and “water-filling”, while “\( \Sigma_{\text{min}} \) bound”, and “\( \Sigma_{\text{max}} \) bound” are referred to the water filling given at the end of Section IV for (5) with \( \Sigma = \Sigma_{\text{max}} \) and \( \Sigma = \Sigma_{\text{min}} \). For comparison, the performance of the best equi-power mode assigning \( S \) in (1) by \( S = \sqrt{N} Q \) is also plotted. It is worthwhile to point out that a typical MIMO-OFDM system employs \( M = 2^{10} \approx 1024 \) sub-carriers. So in all of the below examples, the multiplexing rates \( \log_{2} M = \log_{2} M = 3/256 \), \( 3/128 \), and \( 1/8 \) are quite low and clearly indications for advantage of the proposed frequency multiplexing training schemes.
Example 1: MIMO-OFDM systems equipped with \( M_t = 4 \) and \( M_t = 6 \), \( M_r = 2 \), \( N = 2^5 = 32 \), \( L = 5 \), \( (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (0.3, 0.2, 0.2, 0.15, 0.15) \), \( d_1 = d_r = 0.5 \lambda \), \( \bar{\theta}_t = 13^\circ \), \( \ell = 0, \ldots, 4 \), \( (\bar{\theta}_{t_0}, \bar{\theta}_{t_1}, \bar{\theta}_{t_2}, \bar{\theta}_{t_3}, \bar{\theta}_{t_4}) = (290^\circ, 300^\circ, 315^\circ, 320^\circ, 335^\circ) \).

Note that \( N \geq LM_t \) so the SDP (4) is equivalent to (1). Figure 1 plots the error-entropy of channel state estimation \( \log(\pi R_t) \) versus SNR in training phase. It can be seen from Figure 1 that all of our proposed methods practically result in the same error-entropy. This suggests that our approximate solution is highly accurate and the bounds employed in the approximate optimization problems are very tight. It can be clearly observed that our proposed methods are significantly better than the equi-power method. Moreover, the error-entropy is significantly reduced as the number of transmit antennas increases from \( M_t = 4 \) to \( M_t = 6 \).

Example 2: All parameters are the same as those in Example 1, except the number of subcarriers used to transmit training symbols is either \( N = LM_t = 16 = 2^4 \) or reduced to \( N = 2L = 8 = 2^3 \). As can been seen from Figures 2, for the same total transmitted power, the smaller number of training symbols can provide better performance for low SNR region. The reason is that, the smaller number of training symbols \( N \) will lead to a higher signal to noise ratio for a given total power. In the low SNR region, the performance improvement in terms of the error-entropy or MSE linearly increases with SNR.

VI. CONCLUSIONS

We have considered the optimal design of training sequences for spatially correlated MIMO-OFDM channels, which is formulated in terms of error-entropy minimization between the channel state and channel output. Such a design problem has been shown transformed into a semi-definite program, which can be efficiently solved by standard SDP solvers. Furthermore, we have also proposed highly precise approximation algorithms with a significantly lower complexity. The numerical results confirm that our proposed methods outperform the previously proposed methods in terms of the error-entropy.

REFERENCES


