

Optimization of Training Sequences for Spatially Correlated MIMO-OFDM

H.D. Tuan[†], V.-J. Luong[†], and H.H. Nguyen[‡]

[†] School of Electrical Engineering and Telecommunication, University of New South Wales, UNSW Sydney, NSW 2052, AUSTRALIA; Email: h.d.tuan@unsw.edu.au, d.v.luong@unsw.edu.au

[‡]Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, Canada. Email: ha.nguyen@usask.ca.

Abstract—The optimal training sequence for channel estimation in spatially correlated multiple-input multiple-output (MIMO) orthogonal frequency-division multiplexing (OFDM) systems has not been found for an arbitrary signal-to-noise ratio (SNR). Only one class of training sequences was proposed in the literature in which the power allocation is given only for the extreme conditions of low and high SNR. Provided in this paper are (i) a necessary and sufficient condition for the optimal training sequence together with a convex programming to find the solution, and (ii) efficient procedures to find the optimal training sequence. Simulation results confirm the superiority of the proposed design over the existing one.

Keywords. MIMO-OFDM, MMSE channel estimation, training sequence, spatially correlated fading.

I. INTRODUCTION

In coherent MIMO-OFDM systems, channel state information is required for data detection at the receiver. In general, the accuracy of channel estimation directly affects the overall system performance. When the channels between multiple transmit and receive antennas are statistically uncorrelated, several channel estimation methods have been proposed (see, for example, [1], [5]). However, in certain MIMO systems where the transmit antennas and/or the receive antennas cannot be placed sufficiently far a part due to the physical constraints, the MIMO channels are spatially correlated. In such systems, developing an accurate channel estimation method becomes a challenging task. To our best knowledge, only the work in [10] considers channel estimation for spatially correlated MIMO-OFDM in the following cases

- There is one dominant tap over all other taps. The training sequences are designed according to the correlation of the dominant tap only. In essential, the frequency selective fading MIMO channel is treated as a flat fading one;
- All transmit correlation matrices as scaled by a fixed matrix (see (13) below);

Both these special cases are analyzed at the extreme scenarios of low and high signal-to-noise ratio (SNR). The total power is completely assigned to the best eigen-mode of transmit correlation at the low SNR, while equi-powers are allocated to all eigen-modes at the high SNR.

The contributions of this paper are twofold. First, a necessary and sufficient condition for the optimal training sequence is obtained. It is shown that the optimization of the training sequence design can be formulated as a convex programming, which is solvable by many existing softwares of polynomial complexity. Second, inspired by our recent results in [8], [7] for flat fading MIMO systems, efficient procedures for finding the optimal training sequences are developed. Our analysis and results are supported by computer simulation.

The rest of the paper is organized as follows. Section II presents the mathematical model of channel estimation for spatially correlated MIMO-OFDM systems and discusses the challenging design

issue. Formulation of the optimal training design based on convex programming is developed in Section III. Section IV and sections V describe tractable optimization algorithms to optimize the training sequences in a particular and general cases of correlation matrices. Simulation results are provided in Section VI and conclusions are drawn in Section VII.

Notation: 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 of two matrices and $\text{vec}(\mathbf{X})$ denotes the vectorization operation of matrix \mathbf{X} while $\langle \mathbf{X} \rangle$ is the trace of \mathbf{X} . $\mathbf{X} \geq 0$ ($\mathbf{X} > 0$, resp.) means \mathbf{X} is Hermitian symmetric and positive semi-definite (positive definite, resp.). \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, $[\mathbf{A}_{ij}]_{i,j=0,1,\dots,N}$ with matrices \mathbf{A}_{ij} means the matrix with block entries \mathbf{A}_{ij} . Analogously, $\text{diag}[\mathbf{A}_i]_{i=0,1,\dots,N}$ means the matrix with diagonal blocks \mathbf{A}_i and zero off-diagonal blocks.

II. MIMO-OFDM SYSTEMS AND CHANNEL ESTIMATION

Consider a frequency-selective fading MIMO communication channel described by the following transfer matrix $\mathbf{H}(z) = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell z^{-\ell}$, where each matrix $\mathbf{H}_\ell \in \mathbf{C}^{M_r \times M_t}$ represents the gains of the ℓ th MIMO path. The elements $(\mathbf{H}_\ell)_{m,n}$ of \mathbf{H}_ℓ are (possibly correlated) circularly symmetric complex Gaussian random variables that remain unchanged over the period of channel estimation. The spatial correlations between the M_t transmit antennas and M_r receive antennas have the following Kronecker structure [2]:

$$\mathbf{H}_\ell = \mathbf{R}_{r\ell}^{1/2} \mathbf{H}_{w\ell} \mathbf{R}_{t\ell}^{1/2}, \quad \ell = 0, 1, \dots, L-1 \quad (1)$$

where the elements of $\mathbf{H}_{w\ell} \in \mathbf{C}^{M_r \times M_t}$ are i.i.d $\mathcal{CN}(0, 1)$, $\mathbf{R}_{t\ell} = \mathbf{R}_{t\ell}^{1/2} \mathbf{R}_{t\ell}^{1/2}$ and $\mathbf{R}_{r\ell} = \mathbf{R}_{r\ell}^{1/2} \mathbf{R}_{r\ell}^{1/2}$ are the deterministic symmetric transmit and receive correlation matrices, respectively.

For the (uncoded) MIMO-OFDM system with N sub-carriers, each of M_t data sequences is divided into blocks of N symbols. Each block 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 to avoid the inter block interference (IBI). By defining $W_N = e^{-j2\pi/N}$, the channel transfer function corresponding to the k th sub-channel is

$$\mathbf{H}_f(k) = \mathbf{H}(W_N^k) = \sum_{\ell=0}^{L-1} \mathbf{H}_\ell W_N^{\ell k}, \quad k = 0, 1, \dots, N-1. \quad (2)$$

The normalized input-output equation for each sub-carrier is

$$\mathbf{r}(k) = \sqrt{\frac{\rho}{M_t}} \mathbf{H}_f(k) \mathbf{s}(k) + \mathbf{n}(k), \quad k = 0, 1, \dots, N-1, \quad (3)$$

where $\mathbf{r}(k) = (r_0(k), r_1(k), \dots, r_{M_r-1}(k))^T \in \mathbf{C}^{M_r}$ is the k th received signal vector, $\mathbf{s}(k) = (s_0(k), s_1(k), \dots, s_{M_t-1}(k))^T \in \mathbf{C}^{M_t}$ is the transmitted signal vector, and $\mathbf{n}(k) = (n_0(k), n_1(k), \dots, n_{M_r-1}(k))^T$ represents additive white Gaussian noise (AWGN) and it consists of i.i.d $\mathcal{CN}(0, 1)$.

In channel estimation, the channel tap matrices $\mathbf{H}_{\ell}, \ell = 0, 1, \dots, L-1$, or equivalently, the vector $\mathbf{h} = (\text{vec}(\mathbf{H}_0)^T, \text{vec}(\mathbf{H}_1)^T, \dots, \text{vec}(\mathbf{H}_{L-1})^T)^T \in \mathbf{C}^{LM_t M_r}$, are estimated at the receiver based on the received signals $\mathbf{r}(k), k = 0, 1, \dots, N-1$ and known transmitted signals $\mathbf{s}(k), k = 0, 1, \dots, N-1$, which are called the training sequences. The optimal training design problem is to find $\mathbf{s}(k), k = 0, 1, \dots, N-1$, or equivalently $\mathbf{S} = [\mathbf{s}(0) \ \mathbf{s}(1) \ \dots \ \mathbf{s}(N-1)]^T \in \mathbf{C}^{N \times M_t}$, under the

normalized power constraint $\langle \mathbf{S}^H \mathbf{S} \rangle = \sum_{k=1}^{N-1} \|\mathbf{s}(k)\|^2 = NM_t$, to

optimize some estimation criterion such as the least mean-square error (LMSE) or minimum mean-square error (MMSE), etc.. This paper adopts the MMSE criterion and the optimization problem is formulated in more detail next. First, rewrite (3) as

$$\mathbf{r} = \sqrt{\frac{\rho}{M_t}} \mathbf{M}(\mathbf{S}) \mathbf{h} + \mathbf{n}, \quad (4)$$

where

$$\mathbf{r} = \begin{pmatrix} \mathbf{r}(0) \\ \mathbf{r}(1) \\ \dots \\ \mathbf{r}(N-1) \end{pmatrix} \in \mathbf{C}^{NM_r}, \quad \mathbf{n} = \begin{pmatrix} \mathbf{n}(0) \\ \mathbf{n}(1) \\ \dots \\ \mathbf{n}(N-1) \end{pmatrix} \in \mathbf{C}^{NM_r},$$

$$\mathbf{M}(\mathbf{S}) [\mathbf{F}_0 \mathbf{S} \ \mathbf{F}_1 \mathbf{S} \ \dots \ \mathbf{F}_{L-1} \mathbf{S}] \otimes \mathbf{I}_{M_r} \in \mathbf{C}^{NM_r \times LM_t M_r} \quad (5)$$

$$\begin{aligned} \mathbf{F}_{\ell} &= \text{diag}\{W_N^{k\ell}\}_{k=0,1,\dots,N-1}, \quad \ell = 0, 1, \dots, L-1, \\ \mathbf{M}^H(\mathbf{S})\mathbf{M}(\mathbf{S}) &= [(\mathbf{S}^H \mathbf{F}_{\ell}^H \mathbf{F}_{\ell} \mathbf{S}) \otimes \mathbf{I}_{M_r}]_{\ell, q=0,1,\dots,L-1}. \end{aligned} \quad (6)$$

Based on the singular value decompositions (SVDs)

$$\mathbf{R}_{t\ell} = \mathbf{U}_{\ell} \mathbf{\Lambda}_{\ell} \mathbf{U}_{\ell}^H, \quad \mathbf{R}_{r\ell} = \mathbf{V}_{\ell} \mathbf{\Upsilon}_{\ell} \mathbf{V}_{\ell}^H, \quad (7)$$

the correlation matrix of noise \mathbf{n} and the channel vector \mathbf{h} are $\mathbf{R}_{\mathbf{n}} = \mathbf{I}_{NM_r}$ and $\mathbf{R}_{\mathbf{h}} =$

$$\begin{pmatrix} \mathbf{R}_{t0} \otimes \mathbf{R}_{r0} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{R}_{t(L-1)} \otimes \mathbf{R}_{r(L-1)} \end{pmatrix} = \mathbf{U}_{\mathbf{h}} \mathbf{\Lambda}_{\mathbf{h}} \mathbf{U}_{\mathbf{h}}^H, \quad (8)$$

where $\mathbf{U}_{\mathbf{h}} = \text{diag}[\mathbf{U}_{\ell} \otimes \mathbf{V}_{\ell}]_{\ell=0,1,\dots,L-1}$, and $\mathbf{\Lambda}_{\mathbf{h}} = \text{diag}[\mathbf{\Lambda}_{\ell} \otimes \mathbf{\Upsilon}_{\ell}]_{\ell=0,1,\dots,L-1}$. The MMSE estimation of \mathbf{h} is thus [4] $\hat{\mathbf{h}} = \sqrt{\frac{\rho}{M_t}} (\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\rho}{M_t} \mathbf{M}^H(\mathbf{S})\mathbf{M}(\mathbf{S}))^{-1} \mathbf{M}^H(\mathbf{S})\mathbf{r}$, and the corresponding MMSE is $E(\|\mathbf{h} - \hat{\mathbf{h}}\|^2) = \langle (\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\rho}{M_t} \mathbf{M}^H(\mathbf{S})\mathbf{M}(\mathbf{S}))^{-1} \rangle$. Thus the optimal training design is

$$\min_{\mathbf{S} \in \mathbf{C}^{N \times M_t}} \langle (\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\rho}{M_t} \mathbf{M}^H(\mathbf{S})\mathbf{M}(\mathbf{S}))^{-1} \rangle : \langle \mathbf{S}^H \mathbf{S} \rangle = NM_t. \quad (9)$$

In general, the above is a very difficult optimization problem. The next section shows how convex programming can be used to find the optimal solution for a particular case of correlation matrices.

III. CONVEX PROGRAMMING FOR OPTIMAL TRAINING SEQUENCES IN THE GENERAL CASE

It follows from (5) and (8) that:

$$\langle (\mathbf{R}_{\mathbf{h}}^{-1} + \frac{\rho}{M_t} \mathbf{M}^H(\mathbf{S})\mathbf{M}(\mathbf{S}))^{-1} \rangle = \langle (\mathbf{\Lambda}_{\mathbf{h}}^{-1} + \frac{\rho}{M_t} \mathbf{Q}(\mathbf{S}))^{-1} \rangle, \quad (10)$$

where $\mathbf{Q}(\mathbf{S}) = [(\mathbf{U}_{\ell}^H \mathbf{S}^H \mathbf{F}_{\ell}^H \mathbf{F}_{\ell} \mathbf{S} \mathbf{U}_{\ell}) \otimes (\mathbf{V}_{\ell}^H \mathbf{V}_{\ell})]_{\ell, q=0,1,\dots,L-1}$. The following conditions have been stated in [10] as necessary conditions for the optimality of a solution \mathbf{S} of (9):

$$\mathbf{U}_{\ell}^H \mathbf{S}^H \mathbf{S} \mathbf{U}_{\ell} \text{ is diagonal, } \ell = 0, 1, \dots, L-1, \quad (11)$$

$$\mathbf{S}^H \mathbf{F}_{\ell}^H \mathbf{F}_{m} \mathbf{S} = 0 \text{ when } 0 \leq \ell \neq m \leq L-1. \quad (12)$$

In general, (11)-(12) are a system of nonlinear equations where the number of equations may exceed the number of unknowns and thus there is no guarantee for its feasibility. The interesting question is "how rich is the feasible solution class?". In this respect, [10] addresses the case when all the transmit correlation matrices $\mathbf{R}_{t\ell}$ are scaled by a fixed matrix $\mathbf{R} = \mathbf{U} \mathbf{\Lambda}_t \mathbf{U}^H$ with unitary matrices $\mathbf{U} \in \mathbf{C}^{M_t \times M_t}$, $\mathbf{V}_{\ell} \in \mathbf{C}^{M_r \times M_r}$ and diagonal matrices $\mathbf{\Lambda}_{t\ell}$ and $\mathbf{\Lambda}_{r\ell}$, i.e.,

$$\mathbf{R}_{t\ell} = \sigma_{\ell} \mathbf{R}, \quad \ell = 0, 1, \dots, L-1. \quad (13)$$

In this special case, \mathbf{U}_{ℓ} are the same as in (11).

Furthermore, condition (12) implies that $\mathbf{S} = \mathbf{Q} \bar{\mathbf{X}}$, $\mathbf{Q} \in \mathbf{C}^{N \times M_t}$, $\bar{\mathbf{X}} \in \mathbf{C}^{M_t \times M_t}$, where $\mathbf{Q} \in \mathbf{C}^{N \times M_t}$ is constructed from M_t columns $\mathbf{q}_i \in \mathbf{C}^N$, $i = 1, \dots, M_t$ of an unitary matrix, which is pre-chosen according to [5] to satisfy

$$\mathbf{q}_i^H \mathbf{F}_{m-\ell} \mathbf{q}_j = 0 \text{ for } i, j = 1, 2, \dots, M_t; 0 \leq \ell \neq m \leq L-1. \quad (14)$$

For instance, when $N \geq LM_t$ such \mathbf{q}_i can be easily constructed as in [5], [1].

At the end of this section, it will be seen that (11) and (12) cannot constitute a necessary optimal condition for (9) for the case of having different \mathbf{U}_{ℓ} in (7).

Our result to correct (11) and (12) is stated in the following theorem.

Theorem 1: Under the existence of \mathbf{Q} that satisfies (14), the optimization problem (9) in $\mathbf{S} \in \mathbf{C}^{N \times M_t}$ is equivalent to the following optimization problem in $\mathbf{X} \in \mathbf{C}^{M_t \times M_t}$:

$$\min_{0 \leq \mathbf{X} \in \mathbf{C}^{M_t \times M_t}} \sum_{\ell=0}^{L-1} \langle (\mathbf{R}_{t\ell}^{-1} \otimes \mathbf{R}_{r\ell}^{-1} + \frac{\rho}{M_t} \mathbf{X} \otimes \mathbf{I}_{M_r})^{-1} \rangle : \langle \mathbf{X} \rangle = NM_t. \quad (15)$$

The optimal solution \mathbf{S}_{opt} of (9) is defined from the optimal solution \mathbf{X}_{opt} of (15) as $\mathbf{S}_{\text{opt}} = \mathbf{Q} \mathbf{X}_{\text{opt}}^{1/2}$.

One can see that the objective function in (15) is convex in \mathbf{X} so (15) is in fact a convex programming. However, for the demand of a much faster computation in channel estimation problem, a specialized convex programming algorithm to find the solution of (15) is developed in the next sections.

Before closing this section, we shall show that (11) in general cannot be a necessary optimality condition for (9) by considering the case of $L = 2$, $M_r = 1$, $M_t > 2$ and \mathbf{U}_0 and \mathbf{U}_1 are different in SVD (7) and \mathbf{R}_{t0} and \mathbf{R}_{t1} are not diagonal. This means that \mathbf{U}_0 and \mathbf{U}_1 are not rotation matrices¹. Without loss of generality, set $\mathbf{R}_{r1} = 1$ and problem (15) is

$$\min_{0 \leq \mathbf{X} \in \mathbf{C}^{M_t \times M_t}} \left[\langle (\mathbf{R}_{t0}^{-1} + \frac{\rho}{M_t} \mathbf{X})^{-1} \rangle + \langle (\mathbf{R}_{t1}^{-1} + \frac{\rho}{M_t} \mathbf{X})^{-1} \rangle \right] \quad (16)$$

under the constraint $\langle \mathbf{X} \rangle = NM_t$. The condition (11) requires that both $\mathbf{U}_0^H \mathbf{X}_{\text{opt}} \mathbf{U}_0$ and $\mathbf{U}_1^H \mathbf{X}_{\text{opt}} \mathbf{U}_1$ must be diagonal for the optimal solution \mathbf{X}_{opt} of (16). This can only be fulfilled for $\mathbf{X}_{\text{opt}} = N \mathbf{I}_{M_t}$. Obviously, such \mathbf{X}_{opt} is not the optimal solution of (16). By obtaining a numerical solution of (16), it can be verified that its optimal solution

¹A matrix is called rotation if it is unitary and each of its columns has only one nonzero component.

does not necessarily satisfy (11). A similar convex programming problem with no available simplified diagonal structure has been considered in [11, Theorem 1].

IV. A MORE TRACTABLE OPTIMIZATION FOR A PARTICULAR CASE OF CORRELATION MATRICES

We consider (15) for the following case

$$\mathbf{R}_{\ell\ell} = \mathbf{U}\mathbf{\Lambda}_\ell\mathbf{U}^H, \ell = 0, 1, \dots, L-1, \quad (17)$$

of different diagonal matrices $\mathbf{\Lambda}_{\ell\ell}$ but the same unitary matrix \mathbf{U} . This includes the case stated in (13). Then (15) can be written in the new variable $\tilde{\mathbf{X}} = \mathbf{U}^H\mathbf{X}\mathbf{U}$ as follows:

$$\min_{0 \leq \tilde{\mathbf{X}} \in \mathbf{C}^{M_t \times M_t}} \sum_{\ell=0}^{L-1} \langle (\mathbf{\Lambda}_\ell^{-1} \otimes \mathbf{\Upsilon}_\ell^{-1} + \tilde{\mathbf{X}} \otimes \mathbf{I}_{M_r})^{-1} \rangle : \langle \tilde{\mathbf{X}} \rangle = NM_t. \quad (18)$$

For any positive definite matrix $\tilde{\mathbf{X}}$, one has

$$\begin{aligned} \text{tr}\{(\mathbf{\Lambda}_\ell^{-1} \otimes \mathbf{\Upsilon}_\ell^{-1} + \tilde{\mathbf{X}} \otimes \mathbf{I}_{M_r})^{-1}\} &\geq \\ \text{tr}\{(\mathbf{\Lambda}_\ell^{-1} \otimes \mathbf{\Upsilon}_\ell^{-1} + \text{diag}\{\tilde{\mathbf{X}}(i, i)\}_{i=1,2,\dots,M_t} \otimes \mathbf{I}_{M_r})^{-1}\}. \end{aligned}$$

Hence the optimal solution of (18) with only one constraint $\text{tr}\{\tilde{\mathbf{X}}\} = NM_t$ must be in the diagonal form $\tilde{\mathbf{X}} = \text{diag}\{\mathbf{y}_1, \dots, \mathbf{y}_{M_t}\}$. This means that the matrix optimization problem (18) in the variable \mathbf{X} is equivalent to the following vector optimization problem in $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_{M_t})^T \in \mathbf{R}^{M_t}$:

$$\min_{\mathbf{y}_i \geq 0, i=1,2,\dots,M_t} \sum_{i=1}^{M_t} f_i(\mathbf{y}_i) : \sum_{i=1}^{M_t} \mathbf{y}_i = NM_t, \quad (19)$$

where $f_i(\mathbf{y}_i) = \sum_{\ell=0}^{L-1} \sum_{j=1}^{M_r} \left(\lambda_{\ell i}^{-1} \gamma_{\ell j}^{-1} + \frac{\rho}{M_t} \mathbf{y}_i \right)^{-1}$, $\mathbf{\Lambda}_\ell = \text{diag}\{\lambda_{\ell 1}, \dots, \lambda_{\ell M_t}\}$, $\mathbf{\Upsilon}_\ell = \text{diag}\{\gamma_{\ell 1}, \dots, \gamma_{\ell M_r}\}$, $\ell = 0, 1, \dots, L-1$.

In [10], asymptotic solutions of the above optimization problem were given for low SNR (i.e., ρ/M_t is small) and high SNR (i.e., ρ/M_r is large) with equal-power allocations ($y_1 = y_2 = \dots = y_{M_r}$). On the other hand, an efficient procedure to locate the optimal solution of (19) was developed in [8], [7] by a few iterative ‘‘water filling’’ steps. The efficiency of the procedure is confirmed by simulation results presented in Section VI.

Furthermore, there is the following approximation for the optimal value of the optimization problem (19). Using the fact that the function $(x^{-1} + a)^{-1}$ is concave in scalar variable $x > 0$, a tight upper bound for $f_i(\mathbf{y}_i)$ is

$$f_i(\mathbf{y}_i) \leq LM_r(a_i^{-1} + \frac{\rho}{M_t}\mathbf{y}_i)^{-1}, \quad a_i := \frac{1}{LM_r} \sum_{\ell=0}^{L-1} \sum_{j=0}^{M_r} \lambda_{\ell i} \gamma_{\ell j} \quad (20)$$

As an approximated solution to the optimal solution of (19) we take the optimal solution of the following optimization problem, which is minimization of an upper bound of the objective function :

$$\min_{\mathbf{y}_i \geq 0, i=1,2,\dots,M_t} \sum_{i=1}^{M_t} \frac{1}{a_i^{-1} + \frac{\rho}{M_t}\mathbf{y}_i} : \sum_{i=1}^{M_t} \mathbf{y}_i = NM_t. \quad (21)$$

The optimal solution of (21) is well known:

$$\mathbf{y}_i = \frac{M_t}{\rho} \max\{(\rho/M_t)^{1/2}\mu - a_i^{-1}, 0\}, \quad (22)$$

where $\mu > 0$ is chosen so that $\sum_{i=1}^{M_t} \mathbf{y}_i = 1$.

In essential, the optimal solution (22) is water-filling based on the

following measure of each eigen-mode i ($i = 1, 2, \dots, M_t$): $e_i = \sum_{\ell=0}^{L-1} \sum_{j=0}^{M_r} \lambda_{\ell i} \gamma_{\ell j}$, that is in contrast to the solution of [10] with total power NM_t allocated to the mode with the maximum measure $\tilde{e}_i = \sum_{\ell=0}^{L-1} \sum_{j=0}^{M_r} (\lambda_{\ell i} \gamma_{\ell j})^2$. As confirmed by simulation in Section VI, indeed it is the optimal solution of (19), while the solution of [10] is not.

V. SIMPLIFIED CONVEX OPTIMIZATION WITH CLOSED LOOP SOLUTIONS

As mentioned, in general there is no available closed loop solution for the optimization problem (15), so in this section we provide some tractable approximations for its optimal solution. Firstly, (15) is equivalently simplified to

$$\min_{0 \leq \mathbf{X} \in \mathbf{C}^{M_t \times M_t}} \sum_{\ell=0}^{L-1} \langle (\mathbf{R}_{\ell\ell}^{-1} \otimes \mathbf{\Upsilon}_\ell^{-1} + \frac{\rho}{M_t} \mathbf{X} \otimes \mathbf{I}_{M_r})^{-1} \rangle : \langle \mathbf{X} \rangle = NM_t. \quad (23)$$

We now develop some nontrivial bounds for the objective in (23).

Theorem 2: For the objective function in (23) the following upper bound holds true

$$\sum_{\ell=0}^{L-1} \langle (\mathbf{R}_{\ell\ell}^{-1} \otimes \mathbf{\Upsilon}_\ell^{-1} + \frac{\rho}{M_t} \mathbf{X} \otimes \mathbf{I}_{M_r})^{-1} \rangle \leq L \langle \left(\left(\sum_{\ell=0}^{L-1} \mathbf{R}_{\ell\ell} \right)^{-1} \otimes \left(\sum_{\ell=0}^{L-1} \mathbf{\Upsilon}_\ell / L \right)^{-1} + \frac{\rho}{M_t} \mathbf{X} \otimes \mathbf{I}_{M_r} \right)^{-1} \rangle.$$

Thus, setting $\mathbf{R} = \frac{1}{L} \sum_{\ell=0}^{L-1} \mathbf{R}_{\ell\ell}$, $\mathbf{\Upsilon}_U = \sum_{\ell=0}^{L-1} \mathbf{\Upsilon}_\ell$, an upper bound optimization for (23) is provided by the following optimization problem

$$\min_{0 \leq \mathbf{X} \in \mathbf{C}^{M_t \times M_t}} \langle (\mathbf{R}^{-1} \otimes \mathbf{\Upsilon}^{-1} + \frac{\rho}{M_t} \mathbf{X} \otimes \mathbf{I}_{M_r})^{-1} \rangle : \langle \mathbf{X} \rangle = NM_t. \quad (24)$$

With a SVD $\sum_{\ell=0}^{L-1} \mathbf{R}_{\ell\ell} = \mathbf{L}\mathbf{R}^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{M_t}\}$,

and $\sum_{\ell=0}^{L-1} \mathbf{\Upsilon}_\ell = \mathbf{\Upsilon}^{-1} = \text{diag}\{\gamma_1, \dots, \gamma_{M_r}\}$; $\gamma_j = \sum_{\ell=0}^{L-1} \gamma_{\ell j}$ and the

earlier variable change, it can be similarly shown that the optimal solution of the upper bound minimization (24) must be in the form $\mathbf{X} = \mathbf{U}\text{diag}\{\mathbf{y}_1, \dots, \mathbf{y}_{M_t}\}\mathbf{U}^H$, where $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_{M_t})^T \in \mathbf{R}^{M_t}$ is the optimal solution of the vector optimization problem (19)

with $f_i(\mathbf{y}_i) = \sum_{j=1}^{M_r} (L\lambda_i^{-1}\gamma_j^{-1} + \frac{\rho}{M_t}\mathbf{y}_i)^{-1}$. As mentioned at the end

of the previous section, the optimal solution can be exactly found by using the iterative water filling procedure [8], [7]. Like (21), an approximated solution of the optimal solution of the upper bound

minimization with $a_i = \frac{1}{LM_r} \lambda_i \sum_{j=1}^{M_r} \gamma_j = \frac{1}{LM_r} \lambda_i \sum_{j=1}^{M_r} \sum_{\ell=0}^{L-1} \gamma_{\ell j}$.

VI. SIMULATION RESULTS

In the simulation the matrices $\mathbf{R}_{\ell\ell}$ and $\mathbf{R}_{r\ell}$ in (1) are defined in uniform antenna array environment as [2, eq. (4)] $[\mathbf{R}_{\ell\ell}]_{m,n} = \sigma_\ell e^{-j2\pi|n-m|\Delta_t \cos(\bar{\theta}_{\ell\ell})} e^{-\frac{1}{2}(2\pi|n-m|\Delta_t \sin(\bar{\theta}_{\ell\ell})\sigma_{\theta_{\ell\ell}})^2}$, $[\mathbf{R}_{r\ell}]_{m,n} = \sigma_\ell e^{-j2\pi|n-m|\Delta_r \cos(\bar{\theta}_{r\ell})} e^{-\frac{1}{2}(2\pi|n-m|\Delta_r \sin(\bar{\theta}_{r\ell})\sigma_{\theta_{r\ell}})^2}$, where

- $\Delta_t = d_t/\lambda$ ($\Delta_r = d_r/\lambda$, resp.) is the relative transmit (receive, resp.) antenna spacing. d_t and d_r are the absolute antenna spacings with wavelength $\lambda = \frac{c}{f_c}$ of the carrier f_c . They are set equal 1 in all simulation.

- $\bar{\theta}_{t\ell}$ and $\bar{\theta}_{r\ell}$ are the mean angle of departure from the transmit array and the mean angle of arrival at the receive array.
- $\sigma_{\bar{\theta}_{t\ell}}^2$ and $\sigma_{\bar{\theta}_{r\ell}}^2$ are the cluster angle spread perceived by the transmitter and receiver, respectively, so the actual angle of transmit (arrival, resp.) is $\theta_{t\ell} = \bar{\theta}_{t\ell} + \hat{\theta}_{t\ell}$ ($\theta_{r\ell} = \bar{\theta}_{r\ell} + \hat{\theta}_{r\ell}$, resp.) with $\hat{\theta}_{t\ell} = \mathcal{CN}(0, \sigma_{\bar{\theta}_{t\ell}}^2)$ ($\hat{\theta}_{r\ell} = \mathcal{CN}(0, \sigma_{\bar{\theta}_{r\ell}}^2)$, resp.). They set equal some specific σ_{θ} ;
- σ_{ℓ}^2 reflects the ℓ -th path channel gain.

It can be seen that large antenna spacing and/or large cluster angle spread results in low spatial fading correlation and vice versa.

For the first example, $L = 5$, $\sigma_{\theta} = 8.6^\circ$, $(\bar{\theta}_{t0}, \bar{\theta}_{t1}, \dots, \bar{\theta}_{t(L-1)}) = (13^\circ, 13^\circ, 13^\circ, 13^\circ, 13^\circ)$, $(\bar{\theta}_{r0}, \bar{\theta}_{r1}, \dots, \bar{\theta}_{r(L-1)}) = (290^\circ, 300^\circ, 315^\circ, 320^\circ, 335^\circ)$, $(\sigma_{\theta_0}^2, \sigma_{\theta_1}^2, \dots, \sigma_{\theta_{L-1}}^2) = (0.3, 0.2, 0.2, 0.15, 0.15)$, where mean departure angles $\bar{\theta}_{t\ell}$ are equal so (13) is met and the asymptotic-based result of [10] is still applied. The Figure 1 demonstrate the MMSE channel estimation performances of four results: our 2 results (one by the iterative water filling (IBP) [7] for the solution of the exact minimization problem (19) and another one by the water filling solution (22) of the upper bound minimization problem (21)), the asymptotic based solution [10] and equi-power allocation. It is clear from there that our results outperform the other existing results. Moreover, the solution of the upper bound minimization problem (22) also results in the minimal MMSE channel estimation as well and so it is preferred because of less computational load.

For the second example with $L = 5$, $\sigma_{\theta} = 8.6^\circ$, $(\bar{\theta}_{t0}, \bar{\theta}_{t1}, \dots, \bar{\theta}_{t(L-1)}) = (13^\circ, 16^\circ, 20^\circ, 24^\circ, 27^\circ)$, $(\bar{\theta}_{r0}, \bar{\theta}_{r1}, \dots, \bar{\theta}_{r(L-1)}) = (290^\circ, 300^\circ, 315^\circ, 329^\circ, 335^\circ)$, $(\sigma_{\theta_0}^2, \sigma_{\theta_1}^2, \dots, \sigma_{\theta_{L-1}}^2) = (0.3, 0.2, 0.2, 0.15, 0.15)$ as the departure angles $\theta_{t\ell}$ are not equal and (13) is not met, the asymptotic-based result of [10] is not applied. There is no existing result for this case except equi-power allocation. The figures 2 show that our result by accepting the solution (22), performs much better than the equi-power solution.

VII. CONCLUSIONS

In this paper, we have revisited, raised and resolved some important technical issues of the MMSE channel estimation and optimal training design for MIMO-OFDM systems operating over spatially correlated fading channels. It was shown that Our proposed design of training sequence outperforms the existing design over the whole range of SNR.

REFERENCES

- [1] I. Barhumi, G. Leus, M. Moonen, Optimal Training Design for MIMO OFDM systems in mobile wireless channels, *IEEE Trans. Signal Processing* 51(2003), 1615-1624.
- [2] H. Bölcskei, Principles of MIMO-OFDM wireless systems, in *CRC Handbook on Signal Processing for Communications*, M. Ibnkahla, Ed., 2004
- [3] R.A. Horn, C.R. Johnson, *Matrix analysis*, Cambridge University Press, 1985.
- [4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.
- [5] Y. G. Li, Simplified channel estimation for OFDM systems with multiple transmit antennas, *IEEE Trans. Wireless Comm* 1(2002), 67-75.
- [6] J. Lofberg, YALMIP: a toolbox for modeling and optimization in Matlab, *Proc. of 2004 IEEE Int. International Symposium on Computer Aided Control System Design*, Taipei, Taiwan, Sept. 2004, pp. 284-188.
- [7] V. Nguyen, H.D. Tuan, H.H. Nguyen, N.N. Tran, Optimal superimposed training design for spatially correlated MIMO channels, *IEEE Trans. Wireless Communication* 7(2008), N.8.
- [8] H.D. Tuan, H.H. Nguyen, N.N. Tran, V. Nguyen, An effective global optimization algorithm for wireless MIMO channel, *Proc. IEEE CAMSAP 2007*, Virgin Island (USA), pp. 225-228.

- [9] H. Tuy, *Convex analysis and global optimization*, Kluwer Academic, 2000.
- [10] H. Zhang, Y. Li, A. Reid, J. Terry, Optimum training symbol design for MIMO OFDM in correlated fading channels, *IEEE Trans. Wireless Comm*, 5(2006), 2343-2347.
- [11] W. Yu, W. Rhee, S. Boyd, J.M. Cioffi, Iterative water-filling for Gaussian vector multiple-access channels, *IEEE Trans. Information Theory* 50(2004), 145-152.
- [12] K. Zhou, J. Doyle, *Essentials of robust control*, Prentice-Hall, 1998.

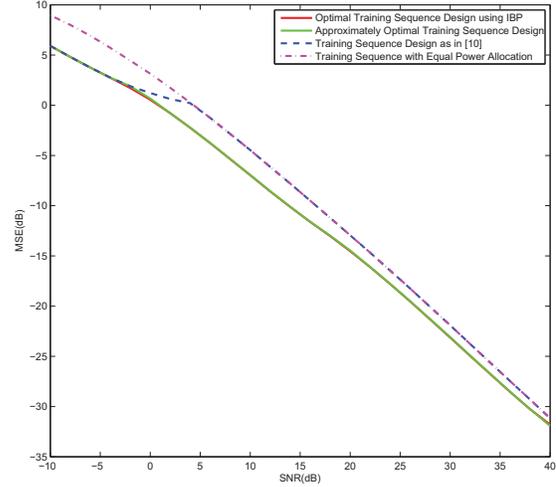


Fig. 1. MMSE channel estimation by different results

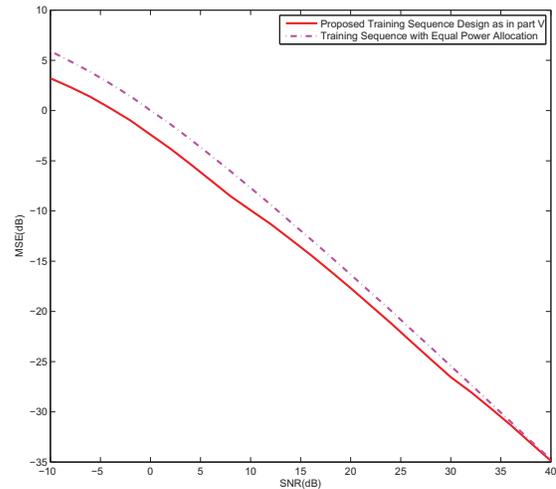


Fig. 2. MMSE channel estimation by different results