TRAINING SIGNAL DESIGN FOR MIMO CHANNEL ESTIMATION WITH CORRELATED DISTURBANCE

Qingjiang Shi[1,2], *Cheng Peng*[1], *Weiqiang Xu*[1], *and Yongchao Wang*[2]

[1]School of Information & Science Technology, Zhejiang Sci-Tech University, Hangzhou, China
[2]The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China
Email:{qing.j.shi, peng18cheng}@gmail.com, wq.xu@126.com, ychwang@mail.xidian.edu.cn

ABSTRACT

This paper studies minimum mean square error (MMSE)–based training signal design for MIMO channel estimation with correlated disturbance (i.e., interference plus noise). First, we consider training signal design for Kronecker–structured MIMO channel estimation where both channel and disturbance are assumed in Kronecker structures. We prove the optimal training sequence structure for arbitrarily Kronecker–structured MIMO channel estimation. Using the optimal training sequence structure, we show that the MSE minimization problem can be globally solved. Second, we consider the training signal design problem in the case of general channel and disturbance model (i.e., without Kronecker structure assumption). We propose a simple iterative algorithm based on block coordinate descent method which can keep the MSE nonincreasing at each iteration. Finally, simulation results indicate good performance of the proposed iterative algorithm by comparing with the optimal training signal design method.

Index Terms— Training signal design, MIMO channel estimation, MMSE, block coordinate descent.

1. INTRODUCTION

Accurate channel state information (CSI) is often desired to improve the performance of MIMO communication. The CSI is used in both transmitter and receiver design and commonly obtained by training–based channel estimation method. In training–based channel estimation, the transmitter first sends a training signal which is known to the receiver and then the receiver does channel estimation based on some criterion, e.g., mean square error (MSE) minimization [1–5] or mutual information (MI) maximization [6,7]. Here we focus on MSE based MIMO channel estimation but our techniques also apply to MI-based MIMO channel estimation.

This research was supported in part by the National Natural Science Foundation of China under grants 61374020, 61302076 and 61372135, Key Project of Chinese Ministry of Education under grant 212066, Zhejiang Provincial Natural Science Foundation of China under grants LY12F02042 and LQ13F010008, and the Science Foundation of Zhejiang Sci-Tech University (ZSTU) under Grant No. 1203805Y.

To improve channel estimation performance, the training signal should be properly designed. Hence, training signal design (i.e., the MSE minimization problem) is a fundamental problem in training-based MIMO channel estimation and has been widely investigated. Most previous works considered channel estimation for either *noise–limited* MIMO channel [1, 2] or *interference–limited* MIMO channel [3, 4], for which, the training signal design problem can be reduced to a convex optimization problem and the optimal structure of the training sequence was explicitly derived. Recently, [5] provided a framework for Kronecker–structured MIMO channel estimation in the presence of both interference and noise (i.e., the so–called *disturbance*). For the disturbance case, [5] showed that the training sequence has the same structure as in the noise or interference case under a *special* Kronecker structure assumption. For the training signal design problem with general Kronecker structure, [5] provided only a *heuristic* solution inspired by the obtained training sequence structure in the special Kronecker–sturctured case.

In this paper, we study MMSE-based training signal design for MIMO channel estimation with correlated disturbance. First, we consider the general Kronecker structured MIMO channel estimation, i.e., both channel and disturbance are in general Kronecker structures. We prove that the optimal training sequence for arbitrarily Kronecker–structured MIMO channel estimation has the same structure as in the special Kronecker–sturctured case. Using the optimal training sequence structure, we show that the MSE minimization problem can be globally solved. Second, we consider the training signal design problem in the case of general channel and disturbance model (i.e., without Kronecker structure assumption). We propose a simple iterative algorithm based on block coordinate descent (BCD) algorithm [11]. The proposed iterative algorithm can monotonically converge to a stationary point of the MSE minimization problem.

Notations: scalars are denoted by lower-case letters; bold-face lower-case letters are used for vectors, while bold-face upper-case letters are for matrices. $\mathbb{R}^{m\times n}$ ($\mathbb{C}^{m\times n}$) denotes the space of $m \times n$ real (complex) matrices. For a matrix $\mathbf{A}$, $\mathrm{Tr}\{\mathbf{A}\}$, $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$ and $\mathbf{A}_{i,j}$ denote its trace, transpose,

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

conjugate transpose, conjugate and the (i, j)-th entry, respectively. $(\mathbf{A})^{ij}_{blk,n}$ denotes the (i, j)-th submatrix obtained by partitioning $\mathbf{A}$ into $n \times n$ blocks. The Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$. $\text{vec}(\mathbf{A})$ denotes a column vector obtained by stacking the columns of $\mathbf{A}$. $\text{diag}(x_1, x_2, \ldots, x_n)$ denotes a diagonal matrix with x_i being its i-th diagonal entry. $\mathbf{I}_n$ denotes n by n identity matrix. $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ means that $\boldsymbol{x}$ is a circularly symmetric complex Gaussian (CSCG) random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. $\mathbb{E}\{\cdot\}$ denotes expectation opertation.

2. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a point–to–point MIMO channel where a transmitter equipped with N_t antennas transmits signal to a receiver with N_r antennas subject to disturbance (i.e., interference plus noise). We assume that the channel is flat and quasi–static block fading with coherence time T_c. Thus the system can be described as

$$\boldsymbol{y}(t) = \mathbf{H}\boldsymbol{x}(t) + \boldsymbol{w}(t), t = 1, 2, \ldots, T_c \quad (1)$$

where $\boldsymbol{x}(t) \in \mathbb{C}^{N_t \times 1}$ and $\boldsymbol{y}(t) \in \mathbb{C}^{N_r \times 1}$ are the transmitted signal and received signal at time slot t, respectively, and $\boldsymbol{w}(t) \in \mathbb{C}^{N_r \times 1}$ represents the disturbance, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix which needs to be estimated.

In training-based channel estimation, the transmitter sends the training signal (also called pilot signal) which the receiver knows beforehand at first T_p $(< T_c)$ time slots and then the receiver performs channel estimation based on the received signal and some system statistical information. The channel model in the training phase can be shortly written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W} \quad (2)$$

which has a vectorized form

$$\text{vec}(\mathbf{Y}) = (\mathbf{X}^T \otimes \mathbf{I}_{N_r})\text{vec}(\mathbf{H}) + \text{vec}(\mathbf{W}) \quad (3)$$

where $\mathbf{X} \triangleq [\boldsymbol{x}(1)\ \boldsymbol{x}(2)\ \ldots\ \boldsymbol{x}(T_p)]$ and similarly for $\mathbf{Y}$ and $\mathbf{W}$.

We assume that the channel matrix $\mathbf{H}$ follows Rayleigh fading $\text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ and the disturbance $\mathbf{W}$ follows the circularly symmetric complex Gaussian distribution $\text{vec}(\mathbf{W}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{S})$. Moreover, we assume that the channel is uncorrelated with disturbance. Under these assumption, performing MMSE channel estimation based on (3) yields the channel estimation error covariance matrix [5]

$$\mathbf{MSE} \triangleq \left(\mathbf{R}^{-1} + (\mathbf{X}^{\mathrm{T}} \otimes \mathbf{I}_{\mathrm{N_r}})^{\mathrm{H}}\mathbf{S}^{-1}(\mathbf{X}^{\mathrm{T}} \otimes \mathbf{I}_{\mathrm{N_r}})\right)^{-1}. \quad (4)$$

The training signal design problem is to find an $\mathbf{X}$ to minimize the trace of the error covariance matrix. Mathematically, it can be formulated as

$$\begin{aligned} &\min_{\mathbf{X}} \text{Tr}\left\{\left(\mathbf{R}^{-1} + (\mathbf{X}^T \otimes \mathbf{I}_{N_r})^H\mathbf{S}^{-1}(\mathbf{X}^T \otimes \mathbf{I}_{N_r})\right)^{-1}\right\} \\ &\text{s.t. } \text{Tr}\{\mathbf{X}\mathbf{X}^H\} \le P_{tr} \end{aligned} \quad (5)$$

where P_{tr} is the total power used for channel training. Problem (5) is nonconvex. Moreover, the Kronecker term $\mathbf{X}^T \otimes \mathbf{I}_{N_r}$ makes solving (5) a cumbersome task. For mathematical convenience, we will reverse the order of $\mathbf{X}$ and $\mathbf{I}_{N_r}$ in the Kronecker term and derive an equivalent problem of problem (5) based on the following Theorem. The proof is simple and omitted to save space.

Theorem 2.1. *Let* $\bar{\mathbf{R}}$ *and* $\bar{\mathbf{S}}$ *be the covariance matrix of* $\text{vec}(\mathbf{H}^H)$ *and* $\text{vec}(\mathbf{W}^H)$ *respectively. The MMSE channel estimation error convariance matrix for* $\text{vec}(\mathbf{H}^H)$ *based on the vectorized conjugate transpose channel model*

$$\text{vec}(\mathbf{Y}^H) = (\mathbf{I}_{N_r} \otimes \mathbf{X}^H)\text{vec}(\mathbf{H}^H) + \text{vec}(\mathbf{W}^H) \quad (6)$$

is

$$\overline{\mathbf{MSE}} = \left(\bar{\mathbf{R}}^{-1} + (\mathbf{I}_{N_r} \otimes \mathbf{X}^H)^H\bar{\mathbf{S}}^{-1}(\mathbf{I}_{N_r} \otimes \mathbf{X}^H)\right)^{-1} \quad (7)$$

and moreover

$$\mathbf{MSE} = \mathbf{Q}\overline{\mathbf{MSE}}\mathbf{Q}^H \quad (8)$$

where $\mathbf{Q}$ *is a* permutation matrix [9] *such that* $\text{vec}(\mathbf{H}) = \mathbf{Q}\left(\text{vec}(\mathbf{H}^H)\right)^*$.

Since we have $\text{Tr}\{\mathbf{MSE}\} = \text{Tr}\{\overline{\mathbf{MSE}}\}$, problem (5) is equivalent to

$$\begin{aligned} &\min_{\mathbf{X}} \text{Tr}\left\{\left(\bar{\mathbf{R}}^{-1} + (\mathbf{I}_{N_r} \otimes \mathbf{X}^H)^H\bar{\mathbf{S}}^{-1}(\mathbf{I}_{N_r} \otimes \mathbf{X}^H)\right)^{-1}\right\} \\ &\text{s.t. } \text{Tr}\{\mathbf{X}\mathbf{X}^H\} \le P_{tr} \end{aligned} \quad (9)$$

In this paper, we attempt to solve problem (9) (i.e., (5)) by considering the special (Kronecker) channel and disturbance model and the general channel and disturbance model.

3. OPTIMAL TRAINING SIGNAL DESIGN UNDER KRONECKER MODEL

We here consider a commonly used channel and disturbance model, i.e., both $\mathbf{S}$ (or $\bar{\mathbf{S}}$) and $\mathbf{R}$ (or $\bar{\mathbf{R}}$) are in Kronecker structures. Specifically, we assume $\bar{\mathbf{R}} = \mathbf{R}_r^T \otimes \mathbf{R}_t$ where $\mathbf{R}_t \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{R}_r \in \mathbb{C}^{N_r \times N_r}$ represent the spatial correlation at transmitter side and receiver side, respectively. Moreover, we assume $\bar{\mathbf{S}} = \mathbf{S}_r^T \otimes \mathbf{S}_q$ where $\mathbf{S}_r \in \mathbb{C}^{N_r \times N_r}$ represents the received spatial correlation and $\mathbf{S}_q \in \mathbb{C}^{T_p \times T_p}$ represents the temporal correlation. Under the Kronecker model assumption and by using the identities $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ and $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ [9], problem (9) reduces to

$$\begin{aligned} &\min_{\mathbf{X}} \text{Tr}\left\{\left(\mathbf{R}_r^{-T} \otimes \mathbf{R}_t^{-1} + \mathbf{S}_r^{-T} \otimes (\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H)\right)^{-1}\right\} \\ &\text{s.t. } \text{Tr}\{\mathbf{X}\mathbf{X}^H\} \le P_{tr}. \end{aligned} \quad (10)$$

Problem (10) is still difficult to solve. A special case of problem (10) was considered in [5] by additionally assuming that $\mathbf{R}_r$ and $\mathbf{S}_r$ are *simultaneously diagonalizable*. For this special case, [5] has shown that the optimal training sequence has the following structure

$$\mathbf{X} = \mathbf{U}_t \mathbf{P} \mathbf{U}_q^H \tag{11}$$

where $\mathbf{P} \in \mathbb{R}^{N_t \times T_p}$ is a *rectangular diagonal matrix*, and $\mathbf{U}_t$ and $\mathbf{U}_q$ are unitary matrices, respectively obtained from eigen-decomposition of $\mathbf{R}_t$ and $\mathbf{S}_q$, i.e., $\mathbf{R}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ and $\mathbf{S}_q = \mathbf{U}_q \mathbf{\Lambda}_q \mathbf{U}_q^H$ ($\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_q$ are diagonal matrices with properly ordered eigenvalues). In this paper, we will show that the optimal structure (11) holds for arbitrary combination of $\mathbf{R}_r$ and $\mathbf{S}_r$, and moreover problem (10) can be globally solved. We have the following proposition.

Proposition 3.1. *Let $\mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$ be the eigen–decomposition of $\mathbf{S}_r^{\frac{T}{2}} \mathbf{R}_r^{-T} \mathbf{S}_r^{\frac{T}{2}}$ with $\mathbf{\Lambda}_r \triangleq \operatorname{diag}\left(\lambda_1^{(r)}, \lambda_2^{(r)}, \ldots, \lambda_{N_r}^{(r)}\right)$. Let $\mathbf{\Lambda}_t \triangleq \operatorname{diag}\left(\lambda_1^{(t)}, \lambda_2^{(t)}, \ldots, \lambda_{N_t}^{(t)}\right)$ with descendingly ordered eigenvalues $\lambda_1^{(t)} \geq \lambda_2^{(t)} \geq \ldots \geq \lambda_{N_t}^{(t)}$ and $\mathbf{\Lambda}_q \triangleq \operatorname{diag}\left(\lambda_1^{(q)}, \lambda_2^{(q)}, \ldots, \lambda_{T_p}^{(q)}\right)$ with ascendingly ordered eigenvalues $\lambda_1^{(q)} \leq \lambda_2^{(q)} \leq \ldots \leq \lambda_{T_p}^{(q)}$. The optimal solution to problem* (10) *could have the following structure*

$$\mathbf{X} = \mathbf{U}_t \mathbf{P} \mathbf{U}_q^H \tag{12}$$

where $\mathbf{P} \in \mathbb{R}^{N_t \times T_p}$ is a rectangular diagonal matrix with $\sqrt{p_i}$ on its main diagonal which can be obtained by solving the following convex optimization problem

$$\begin{aligned} &\min_{\{p_j \geq 0\}} \sum_{i=1}^{N_r} \sum_{j=1}^{M} \frac{\rho_i \lambda_j^{(q)}}{\lambda_i^{(r)} \frac{\lambda_j^{(q)}}{\lambda_j^{(t)}} + p_j} + \sum_{i=1}^{N_t} \frac{\rho_i}{\lambda_i^{(r)}} \sum_{j=M+1}^{N_t} \lambda_j^{(t)} \\ &\text{s.t.} \sum_{j=1}^{M} p_j \leq P_{tr} \end{aligned} \tag{13}$$

where $M \triangleq \min(N_t, T_p)$ and ρ_i is the i-th diagonal entry of the matrix $\mathbf{U}_r^H \mathbf{S}_r^T \mathbf{U}_r$.

Proof. Recall the identities of Kronecker operation [9] $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ and $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. We have (14) where (a) follows from the fact that $\mathbf{S}_r^{\frac{T}{2}} \mathbf{R}_r^{-T} \mathbf{S}_r^{\frac{T}{2}} = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$, $\mathbf{R}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ and $\mathbf{U}_r \otimes \mathbf{U}_t$ is a unitary matrix, (b) follows from the identity $\operatorname{Tr}\{\mathbf{AB}\} = \operatorname{Tr}\{\mathbf{BA}\}$, and (c) is due to the fact that $\left(\mathbf{\Lambda}_r \otimes \mathbf{\Lambda}_t^{-1} + \mathbf{I}_{N_r} \otimes \left(\mathbf{U}_t^H \mathbf{X} \mathbf{S}_q^{-1} \mathbf{X}^H \mathbf{U}_t\right)\right)^{-1}$ is a block diagonal matrix and $\rho_i \mathbf{I}_{N_t}$ is on the main diagonal of the matrix $\mathbf{U}_r^H \mathbf{S}_r^T \mathbf{U}_r \otimes \mathbf{I}_{N_t}$.

By defining $\mathbf{P} = \mathbf{U}_t^H \mathbf{X} \mathbf{U}_q$ and noting that $\operatorname{Tr}\{\mathbf{P}\mathbf{P}^H\} = \operatorname{Tr}\{\mathbf{X}\mathbf{X}^H\}$, problem (10) can be equivalently written as

$$\begin{aligned} &\min_{\mathbf{P}} \sum_{i=1}^{N_r} \rho_i \operatorname{Tr}\left\{\left(\lambda_i^{(r)} \mathbf{\Lambda}_t^{-1} + \left(\mathbf{P} \mathbf{\Lambda}_q^{-1} \mathbf{P}^H\right)\right)^{-1}\right\} \\ &\text{s.t.} \operatorname{Tr}\{\mathbf{P}\mathbf{P}^H\} \leq P_{tr} \end{aligned} \tag{15}$$

Since $\mathbf{\Lambda}_t$ is in descending order and $\mathbf{\Lambda}_q$ is in ascending order, we concludes by invoking Lemma 1 of [5] that $\mathbf{P}$ could be a rectangular diagonal matrix. With some simple manipulations, (15) reduces to (13), which completes the proof. □

Remark 3.1. *For the special case when $\mathbf{R}_r$ and $\mathbf{S}_r$ have the same eigenvectors $\mathbf{U}_r$, it is not difficult to verify that our result is consistent with the result of [5] by noting that problem* (13) *is identical to problem (34) of [5] in this case.*

Remark 3.2. *For general combination of $\mathbf{R}_r$ and $\mathbf{S}_r$, [5] proposed a heuristic solution to* (10) *by assuming the solution structure as* (12) *with*

$$p_j \triangleq \max\left(\sqrt{\frac{\lambda_j^{(q)}}{\alpha}} - \frac{\lambda_j^{(q)}}{\lambda_j^{(t)}}\right)$$

where $\alpha > 0$ is chosen such that the power constraint holds with equality. It is observed from (13) *that this heuristic solution is optimal to problem* (10) *when all λ_i^r are equal to* 1*. This corresponds to the special case when $\mathbf{S}_r = \mathbf{R}_r$.*

By performing KKT analysis on problem (13), we have the following proposition.

Proposition 3.2. *The solution to problem* (13) *must satisfy the system of equations*

$$\sum_{i=1}^{N_r} \frac{\rho_i \lambda_j^{(q)}}{\left(\lambda_i^{(r)} \frac{\lambda_j^{(q)}}{\lambda_j^{(t)}} + p_j\right)^2} = \alpha \tag{16}$$

for all j with $\alpha < \sum_{i=1}^{N_r} \frac{\rho_i \left(\lambda_j^{(t)}\right)^2}{\left(\lambda_i^{(r)}\right)^2 \lambda_j^{(q)}}$ and $p_j = 0$ otherwise. α is the Lagrange multiplier which is chosen such that the power constraint with equality.

Note that, since the left-hand-side of (16) is a decreasing function of p_j, the system of equations (16) given α can be solved by using Bisection method [10]. Moreover, since problem (13) is convex, the optimal α can be also found by using Bisection method . Hence, problem (13) can be easily solved by two-tier Bisection method.

4. TRAINING SIGNAL DESIGN UNDER GENERAL MODEL

In this section, we consider the general case where the correlation matrix $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}$ are not Kronecker structured. In the general case, it can be shown that problem (9) is equivalent to[1]

$$\begin{aligned} \min_{\mathbf{X},\mathbf{G}} \; &\operatorname{Tr}\left\{\left(\mathbf{I} - \mathbf{G}^H \left(\mathbf{I}_{N_r} \otimes \mathbf{X}^H\right)\right) \bar{\mathbf{R}} \left(\mathbf{I} - \mathbf{G}^H \left(\mathbf{I}_{N_r} \otimes \mathbf{X}^H\right)\right)^H\right\} \\ &+ \operatorname{Tr}\left\{\mathbf{G}^H \bar{\mathbf{S}} \mathbf{G}\right\} \\ \text{s.t.} \; &\operatorname{Tr}(\mathbf{X}\mathbf{X}^H) \leq P_{tr} \end{aligned} \tag{17}$$

[1] By canceling $\mathbf{G}$ in (17) (i.e., substituting (18) into the objective of problem (17)), we can reduce problem (17) to (9).

$$
\begin{aligned}
&\operatorname{Tr}\left\{\left(\mathbf{R}_r^{-T}\otimes\mathbf{R}_t^{-1}+\mathbf{S}_r^{-T}\otimes\left(\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H\right)\right)^{-1}\right\}\\
=&\operatorname{Tr}\left\{\left(\mathbf{S}_r^{\frac{T}{2}}\otimes\mathbf{I}_{N_t}\right)\left(\mathbf{S}_r^{\frac{T}{2}}\mathbf{R}_r^{-T}\mathbf{S}_r^{\frac{T}{2}}\otimes\mathbf{R}_t^{-1}+\mathbf{I}_{N_r}\otimes\left(\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H\right)\right)^{-1}\left(\mathbf{S}_r^{\frac{T}{2}}\otimes\mathbf{I}_{N_t}\right)\right\}\\
\overset{(a)}{=}&\operatorname{Tr}\left\{\left(\mathbf{S}_r^{\frac{T}{2}}\otimes\mathbf{I}_{N_t}\right)\left(\mathbf{U}_r\otimes\mathbf{U}_t\right)\left(\mathbf{\Lambda}_r\otimes\mathbf{\Lambda}_t^{-1}+\mathbf{I}_{N_r}\otimes\left(\mathbf{U}_t^H\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H\mathbf{U}_t\right)\right)^{-1}\left(\mathbf{U}_r^H\otimes\mathbf{U}_t^H\right)\left(\mathbf{S}_r^{\frac{T}{2}}\otimes\mathbf{I}_{N_t}\right)\right\}\\
\overset{(b)}{=}&\operatorname{Tr}\left\{\left(\mathbf{U}_r^H\mathbf{S}_r^T\mathbf{U}_r\otimes\mathbf{I}_{N_t}\right)\left(\mathbf{\Lambda}_r\otimes\mathbf{\Lambda}_t^{-1}+\mathbf{I}_{N_r}\otimes\left(\mathbf{U}_t^H\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H\mathbf{U}_t\right)\right)^{-1}\right\}\\
\overset{(c)}{=}&\sum_{i=1}^{N_r}\rho_i\operatorname{Tr}\left\{\left(\lambda_i^{(r)}\mathbf{\Lambda}_t^{-1}+\left(\mathbf{U}_t^H\mathbf{X}\mathbf{S}_q^{-1}\mathbf{X}^H\mathbf{U}_t\right)\right)^{-1}\right\}
\end{aligned}
\tag{14}
$$

in the sense that the two problems has the same optimal solution $\mathbf{X}$. Actually, the objective of problem (17) is the MSE when Linear MMSE is used for estimation of $\operatorname{vec}(\mathbf{H}^H)$ based on (6). As compared to problem (5) or (9), problem (17) is easier to solve by using block coordinate descent method [11].

In the BCD method, we alternately run the following two steps. First, we solve problem (17) for $\mathbf{G}$ while fixing $\mathbf{X}$, yielding the optimal $\mathbf{G}$

$$
\mathbf{G}=\left(\left(\mathbf{I}_{N_r}\otimes\mathbf{X}^H\right)\bar{\mathbf{R}}\left(\mathbf{I}_{N_r}\otimes\mathbf{X}\right)+\bar{\mathbf{S}}\right)^{-1}\left(\mathbf{I}_{N_r}\otimes\mathbf{X}^H\right)\bar{\mathbf{R}} \tag{18}
$$

Second, we solve problem (17) for $\mathbf{X}$ while fixing $\mathbf{G}$, equivalently, solve the following convex quadratic optimization problem

$$
\begin{aligned}
&\min_{\mathbf{X}}\ \operatorname{Tr}\left\{\left(\mathbf{I}-\mathbf{G}^H\left(\mathbf{I}_{N_r}\otimes\mathbf{X}^H\right)\right)\bar{\mathbf{R}}\left(\mathbf{I}-\mathbf{G}^H\left(\mathbf{I}_{N_r}\otimes\mathbf{X}^H\right)\right)^H\right\}\\
&\text{s.t. } \operatorname{Tr}\left\{\mathbf{X}\mathbf{X}^H\right\}\le P_{tr}
\end{aligned}
\tag{19}
$$

The difficulty in solving problem (19) is the Kronecker product term $\mathbf{I}_{N_r}\otimes\mathbf{X}^H$. Fortunately, by block matrix computation, we can reformulate problem (19) as follows

$$
\begin{aligned}
&\min_{\boldsymbol{x}} \boldsymbol{x}^H\mathbf{A}\boldsymbol{x}-\boldsymbol{x}^H\boldsymbol{b}-\boldsymbol{b}^H\boldsymbol{x}+\operatorname{Tr}\{\bar{\mathbf{R}}\}\\
&\text{s.t. } \boldsymbol{x}^H\boldsymbol{x}\le P_{tr}
\end{aligned}
\tag{20}
$$

where $\boldsymbol{x}\triangleq\operatorname{vec}(\mathbf{X})$, $\boldsymbol{b}\triangleq\sum_{i=1}^{N_r}\operatorname{vec}\left(\left(\bar{\mathbf{R}}\mathbf{G}^H\right)_{blk,N_r}^{ii}\right)$, and

$$
\mathbf{A}\triangleq\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\left(\left(\mathbf{G}\mathbf{G}^H\right)_{blk,N_r}^{ji}\right)^T\otimes\bar{\mathbf{R}}_{blk,N_r}^{ij}+\gamma\mathbf{I}_{N_tT_p}.
$$

Note that, problem (20) can be easily solved by using Bisection method. Therefore, the BCD algorithm for problem (17) is efficient. Moreover, the proposed iterative algorithm can keep MSE nonincreasing at each iteration and monotonically converge to a stationary point of problem (17) [11].

5. NUMERICAL EXAMPLES

In this section, we numerically evaluate the performance of the proposed training signal design methods. To offer a benchmark for performance evaluation of the proposed BCD algorithm, we only consider Kronecker-structured MIMO channel estimation in our simulations. We set $N_t{=}N_r{=}T_p{=}4$. We use the following exponential model [8] to generate a correlation matrix $\mathbf{C}$: $\mathbf{C}_{i,j}{=}r^{j-i}$, $\forall j{>}i$ where r is the normalized correlation coefficient with $|r|{<}1$. Specifically, we set the correlation coefficients $r_t{=}0.5e^{-j0.42\pi}$, $r_r{=}0.65e^{-j0.83\pi}$, $r_s{=}0.3e^{-j0.22\pi}$ and $r_q{=}0.99e^{-j0.65\pi}$ to generate matrices $\mathbf{R}_t$, $\mathbf{R}_r$, $\mathbf{R}_s$ and $\mathbf{R}_q$, respectively.

Figure 1 shows the normalized MSE of different training signal design methods versus the total training power where the normalized MSE is defined as $\frac{\mathbb{E}\{\|\mathbf{H}-\hat{\mathbf{H}}_{\mathrm{MMSE}}\|^2\}}{\operatorname{Tr}\{\mathbf{R}\}}$. It is seen that, there exists a substantial performance gap between the optimum training signal design method and the heuristic method (see remark 3.2 or *Heuristic 1* in [5]) when the total training power is small. Moreover, it is observed that, the performance of the proposed iterative algorithm coincides with that of the optimal training signal design method, implying good performance of the proposed iterative algorithm.

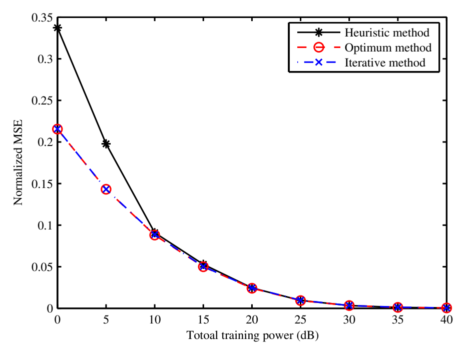

Fig. 1. Normalized MSE Vs. total training power P_{tr}.

6. CONCLUSION

This paper considers training signal design for MIMO channel estimation with disturbance. We have proposed optimal training signal design method for the case of Kronecker–structured MIMO channel estimation and an iterative algorithm for the case of general channel and disturbance model. Simulations show that the proposed iterative method could work as well as the optimal training signal design method.

7. REFERENCES

[1] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453-466, 2005.

[2] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, 2004.

[3] Y. Liu, T. Wong, and W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.

[4] T. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1939-1947, Nov. 2004.

[5] E. Björnson and B. Otterson, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp.1807-1820, Mar. 2010.

[6] M. Biguesh, S. Gazor, and H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, Aug. 2009.

[7] G. B. Giannakis and S. Zhou, "Optimal transmit-diversity precoders for random fading channels,", *IEEE GLOBECOM*, vol. 3, pp. 1839-1843, Dec. 2000.

[8] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369-371, Sep. 2001.

[9] K. B. Petersern and M. S. Pedersern, *The Matrix Cookbook*. http://matrixcookbook.com, Nov. 2008.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[11] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.