Superimposed Channel Training for MIMO Relay Systems

Yue Rong
Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia

***Abstract*—Based on the knowledge of instantaneous channel state information (CSI), the optimal source and relay precoding matrices have been developed recently for multiple-input multiple-output (MIMO) relay communication systems. However, in real communication systems, the instantaneous CSI is unknown and needs to be estimated at the destination node. In this paper, we propose a superimposed channel training method for MIMO relay communication systems. It is shown that to minimize the mean-squared error (MSE) of channel estimation, the optimal training sequence at each node matches the eigenvector matrix of the transmitter correlation matrix of the forward MIMO channel. Then we optimize the power allocation among different streams of the training sequence at the source node and the relay node. Simulation results show that the proposed algorithm leads to a smaller MSE of channel estimation compared with the conventional MIMO relay channel estimation algorithm.**

I. Introduction

Recently, multiple-input multiple-output (MIMO) relay systems have attracted many research interests [1]-[6]. For three-node two-hop MIMO relay systems where the direct source-destination link is omitted, the optimal relay precoding matrix is derived in [2]-[3] to maximize the source-relay-destination channel mutual information. For two-hop MIMO relay systems with multiple parallel relay nodes, the optimal relay precoding matrices are derived in [4] to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination node. A unified framework has been developed in [5] for optimizing the source and relay precoding matrices of two-hop MIMO relay systems with a broad class of commonly used objective functions. Recently, the optimal source and relay precoding matrices have been derived in [6] for MIMO relay systems when a nonlinear decision feedback equalizer (DFE) is applied at the destination node.

For the MIMO relay systems [1]-[6] mentioned above, the instantaneous channel state information (CSI) knowledge of both the source-relay and relay-destination links is required at the destination node to estimate the signals transmitted by the source node. Moreover, in order to optimize the source and/or relay precoding matrices, the instantaneous CSI knowledge of both links is necessary to implement the optimization algorithm. However, in real communication systems, the instantaneous CSI is unknown and needs to be estimated. Recently, a tensor-based channel estimation algorithm has been developed in [7] for two-way MIMO relay systems. However, since the algorithm in [7] exploits the channel reciprocity in two-way relay systems, it cannot be straightforwardly applied to one-way relay systems. In [8], a least-squares (LS) fitting-based relay channel estimation algorithm is proposed. The performance of the algorithm in [8] is further analyzed and improved by using the weighted least-squares (WLS) fitting in [9]. However, the number of training symbol blocks required by the algorithms in [8] and [9] is at least equal to the number of relay antennas, resulting in a low system spectral efficiency, particularly for systems with a large number of relay antennas. For amplify-and-forward relay networks with single-antenna source, relay, and destination nodes, the optimal training sequence is developed in [10]. The optimal training sequence is derived in [11] for a MIMO relay system with one multi-antenna relay node. However, two stages are required in [11] to estimate the CSI of the source-relay and relay-destination links, resulting in a low system spectral efficiency.

In this paper, we propose a superimposed channel training algorithm for MIMO relay communication systems. In particular, the source node first transmits a training block to the relay node. After receiving the training block sent by the source node, the relay node amplifies it, superimposes its own training matrix, and transmits the superimposed signal to the destination node. Finally, the destination node estimates both the source-relay and relay-destination channels based on the training sequences from the source node and the relay node.

Compared with existing methods (for example [8] and [9]), the proposed algorithm has a higher spectral efficiency, since the number of channel training blocks can be much smaller than the number of relay antennas. We prove that in order to minimize the MSE of channel estimation, the optimal training matrix at each node matches the eigenvector matrix of the transmitter correlation matrix of the forward MIMO channel. Then we optimize the power allocation among different streams of the training sequence at the source and relay nodes. Simulation results show that the proposed algorithm leads to a smaller MSE of channel estimation compared with the conventional MIMO relay channel estimation algorithm.

The rest of this paper is organized as follows. In Section II, we introduce the model of a two-hop MIMO relay communication system where superimposed channel training technique is applied. The optimal training sequence and power loading are developed in Sections III. In Section IV, we show some numerical examples. Conclusions are drawn in Section V.

II. System Model

We consider a three-node two-hop MIMO communication system where the source node (node 1) transmits information to the destination node (node 3) with the aid of one relay node (node 2). The ith node is equipped with N_i, $i = 1, 2, 3$, antennas. We focus on the case where the direct link between the source and destination nodes is sufficiently weak to be ignored as in [2]-[5]. This scenario occurs when the direct link is blocked by an obstacle such as a mountain. In fact, a relay plays a much more important role when the direct link is weak than when it is strong.

Training sequences are employed to estimate both the $N_2 \times N_1$ source-relay MIMO channel matrix $\mathbf{H}_1$ and the $N_3 \times N_2$ relay-destination MIMO channel matrix $\mathbf{H}_2$ at the destination node. In the first time block, the source node transmits an $N_1 \times T$ training signal matrix $\mathbf{S}_1$, where T is the length of the training sequence and will be determined later. The $N_2 \times T$ received signal matrix $\mathbf{Y}_2$ at the relay node is given by

$$\mathbf{Y}_2 = \mathbf{H}_1\mathbf{S}_1 + \mathbf{V}_2 \tag{1}$$

where $\mathbf{V}_2$ is an $N_2 \times T$ noise matrix at the relay node.

In the second time block, the relay node amplifies $\mathbf{Y}_2$ and superimposes its own training matrix $\mathbf{S}_2$. Thus, the $N_2 \times T$ signal matrix transmitted by the relay node can be written as

$$\mathbf{X}_2 = \sqrt{\alpha}\mathbf{Y}_2 + \mathbf{S}_2 \tag{2}$$

where $\alpha > 0$ is the relay amplifying factor. From (1) and (2), the $N_3 \times T$ received signal matrix at the destination node is given by

$$\begin{aligned}\mathbf{Y}_3 &= \mathbf{H}_2\mathbf{X}_2 + \mathbf{V}_3\\ &= \sqrt{\alpha}\mathbf{H}_2\mathbf{H}_1\mathbf{S}_1 + \mathbf{H}_2\mathbf{S}_2 + \sqrt{\alpha}\mathbf{H}_2\mathbf{V}_2 + \mathbf{V}_3 \end{aligned}\tag{3}$$

where $\mathbf{V}_3$ is an $N_3 \times T$ noise matrix at the destination node. We assume that all noises are independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and unit variance.

We assume that both channel matrices satisfy the well-known Gaussian-Kronecker model [12], i.e., $\mathbf{H}_i$ is a complex-valued Gaussian random matrix with

$$\mathbf{H}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta}_i \otimes \mathbf{\Phi}_i), \qquad i = 1, 2. \tag{4}$$

Here $\mathbf{\Theta}_i$ denotes the $N_i \times N_i$ covariance matrix at the transmitter side, while $\mathbf{\Phi}_i$ is the $N_{i+1} \times N_{i+1}$ covariance matrix from the receiver side, and $\otimes$ stands for the matrix Kronecker product [13]. In other words, we have $\mathbf{H}_i = \mathbf{A}_i\mathbf{H}_{\mathrm{w},i}\mathbf{B}_i^H$, $i = 1, 2$, where $\mathbf{A}_i\mathbf{A}_i^H = \mathbf{\Phi}_i$, $\mathbf{B}_i\mathbf{B}_i^H = \mathbf{\Theta}_i^T$, and $\mathbf{H}_{\mathrm{w},i}$ is an $N_{i+1} \times N_i$ Gaussian random matrix with i.i.d. zero mean and unit variance entries. Here $(\cdot)^T$ and $(\cdot)^H$ denotes matrix (vector) transpose and Hermitian transpose, respectively. We assume that $\mathbf{H}_{\mathrm{w},1}$ is independent of $\mathbf{H}_{\mathrm{w},2}$. The following lemma is important in deriving the optimal training matrices in the next section.

LEMMA 1 [14]: For $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$, there is $\mathrm{E}[\mathbf{H}\mathbf{A}\mathbf{H}^H] = \mathrm{tr}(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$, and $\mathrm{E}[\mathbf{H}^H\mathbf{A}\mathbf{H}] = \mathrm{tr}(\mathbf{\Phi}\mathbf{A})\mathbf{\Theta}^T$. Here $\mathrm{E}[\cdot]$ stands for statistical expectation.

III. Optimal Training Matrices

Let us introduce the eigenvalue decomposition (EVD) of $\mathbf{\Theta}_i^T$ as $\mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^H$, $i = 1, 2$. Then we have $\mathbf{B}_i^H = \mathbf{\Pi}_i\mathbf{\Lambda}_i^{\frac{1}{2}}\mathbf{U}_i^H$, where $\mathbf{\Pi}_i$ is an arbitrary $N_i \times N_i$ unitary matrix. Using (4), we can equivalently rewrite (3) as

$$\mathbf{Y}_3 = \mathbf{G}\sqrt{\alpha}\tilde{\mathbf{S}}_1 + \tilde{\mathbf{H}}_2\tilde{\mathbf{S}}_2 + \bar{\mathbf{V}} \tag{5}$$

where $\mathbf{G} \triangleq \mathbf{H}_2\tilde{\mathbf{H}}_1$, $\tilde{\mathbf{S}}_i \triangleq \mathbf{U}_i^H\mathbf{S}_i$, $\tilde{\mathbf{H}}_i \triangleq \mathbf{H}_i\mathbf{U}_i$, $i = 1, 2$, and $\bar{\mathbf{V}} \triangleq \sqrt{\alpha}\mathbf{H}_2\mathbf{V}_2 + \mathbf{V}_3$ is the equivalent noise matrix at the destination node. In the following, we develop a novel algorithm to estimate $\tilde{\mathbf{H}}_2$ and $\mathbf{G}$. Then an estimation of $\mathbf{H}_2$ and $\mathbf{H}_1$ can be obtained as $\hat{\mathbf{H}}_2 = \check{\mathbf{H}}_2\mathbf{U}_2^H$ and $\hat{\mathbf{H}}_1 = \hat{\mathbf{H}}_2^{\dagger}\check{\mathbf{G}}\mathbf{U}_1^H$, where $(\cdot)^{\dagger}$ stands for matrix pseudo-inverse, $\check{\mathbf{H}}_2$ and $\check{\mathbf{G}}$ are estimation of $\tilde{\mathbf{H}}_2$ and $\mathbf{G}$, respectively.

By vectorizing both sides of (5), we obtain

$$\begin{aligned}\mathbf{y}_3 &= \left[\sqrt{\alpha}\tilde{\mathbf{S}}_1^T \otimes \mathbf{I}_{N_3}, \;\; \tilde{\mathbf{S}}_2^T \otimes \mathbf{I}_{N_3}\right]\left[\mathbf{g}^T, \;\; \tilde{\mathbf{h}}_2^T\right]^T + \bar{\mathbf{v}}\\ &= \mathbf{M}\boldsymbol{\gamma} + \bar{\mathbf{v}}\end{aligned}\tag{6}$$

where $\mathbf{y}_3 \triangleq \mathrm{vec}(\mathbf{Y}_3)$, $\mathbf{g} \triangleq \mathrm{vec}(\mathbf{G})$, $\tilde{\mathbf{h}}_2 \triangleq \mathrm{vec}(\tilde{\mathbf{H}}_2)$, and $\bar{\mathbf{v}} \triangleq \mathrm{vec}(\bar{\mathbf{V}})$. Here for a matrix $\mathbf{A}$, $\mathrm{vec}(\mathbf{A})$ stacks up the columns of matrix $\mathbf{A}$ into a single column vector, $\mathbf{I}_n$ denotes an $n \times n$ identity matrix, $\mathbf{M} \triangleq \left[\sqrt{\alpha}\tilde{\mathbf{S}}_1^T \otimes \mathbf{I}_{N_3}, \;\; \tilde{\mathbf{S}}_2^T \otimes \mathbf{I}_{N_3}\right]$, and $\boldsymbol{\gamma} \triangleq \left[\mathbf{g}^T, \;\; \tilde{\mathbf{h}}_2^T\right]^T$ is the vector of unknown variables.

Due to its simplicity, a linear MMSE estimator [15] is applied to estimate $\boldsymbol{\gamma}$. We have

$$\hat{\boldsymbol{\gamma}} = \mathbf{W}^H\mathbf{y}_3 \tag{7}$$

where $\hat{\boldsymbol{\gamma}}$ stands for an estimation of $\boldsymbol{\gamma}$ and $\mathbf{W}$ is the weight matrix of the MMSE estimator and given by

$$\mathbf{W} = \left(\mathbf{M}\mathbf{R}_{\gamma}\mathbf{M}^H + \mathbf{R}_{\bar{v}}\right)^{-1}\mathbf{M}\mathbf{R}_{\gamma}. \tag{8}$$

Here $(\cdot)^{-1}$ denotes matrix inversion. From (6), we find that since a linear estimator is used, there is $T \geq N_1 + N_2$. Using (6)-(8), the MSE of estimating $\boldsymbol{\gamma}$ can be obtained as

$$\begin{aligned}\mathrm{MSE} &= \mathrm{E}\left[\mathrm{tr}\left((\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^H\right)\right]\\ &= \mathrm{tr}\left(\left[\mathbf{R}_{\gamma}^{-1} + \mathbf{M}^H\mathbf{R}_{\bar{v}}^{-1}\mathbf{M}\right]^{-1}\right)\end{aligned}\tag{9}$$

where $\mathrm{tr}(\cdot)$ denotes matrix trace. In (9), $\mathbf{R}_{\bar{v}} \triangleq \mathrm{E}[\bar{\mathbf{v}}\bar{\mathbf{v}}^H]$ is the noise covariance matrix which can be calculated using Lemma 1 and is given by

$$\begin{aligned}\mathbf{R}_{\bar{v}} &= \mathbf{I}_T \otimes \left(\alpha\mathrm{tr}(\mathbf{B}_2^H\mathbf{B}_2)\mathbf{A}_2\mathbf{A}_2^H + \mathbf{I}_{N_3}\right)\\ &= \mathbf{I}_T \otimes \left(\alpha\mathrm{tr}(\mathbf{\Theta}_2^T)\mathbf{\Phi}_2 + \mathbf{I}_{N_3}\right).\end{aligned}\tag{10}$$

In (9), $\mathbf{R}_{\gamma} \triangleq \mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^H]$ is the covariance matrix of $\boldsymbol{\gamma}$ and can be calculated in the following. First, the ith column of $\mathbf{G}$ is given by $\mathbf{g}_i = \lambda_{1,i}^{\frac{1}{2}}\mathbf{A}_2\mathbf{H}_{\mathrm{w},2}\mathbf{B}_2^H\mathbf{A}_1\mathbf{H}_{\mathrm{w},1}\boldsymbol{\pi}_{1,i}$, $i = 1, \cdots, N_1$, where $\lambda_{1,i}$ is the ith diagonal element of $\mathbf{\Lambda}_1$, and $\boldsymbol{\pi}_{1,i}$ is the ith column of $\mathbf{\Pi}_1$. Since $\mathbf{H}_{\mathrm{w},1}$ and $\mathbf{H}_{\mathrm{w},2}$ are independent, the covariance matrix of $\mathbf{g}_i$ can be calculated using Lemma 1 and is given by

$$\begin{aligned}\mathrm{E}\left[\mathbf{g}_i\mathbf{g}_i^H\right] &= \lambda_{1,i}\mathrm{tr}(\mathbf{B}_2^H\mathbf{A}_1\mathbf{A}_1^H\mathbf{B}_2)\mathbf{A}_2\mathbf{A}_2^H\\ &= \lambda_{1,i}c_1\mathbf{\Phi}_2, \qquad i = 1, \cdots, N_1\end{aligned}\tag{11}$$

where $c_1 \triangleq \mathrm{tr}(\mathbf{\Phi}_1\mathbf{\Theta}_2^T)$. Second, the covariance matrix of the ith column of $\tilde{\mathbf{H}}_2$, denoted as $\tilde{\mathbf{h}}_{2,i}$, is given by

$$\mathrm{E}\big[\tilde{\mathbf{h}}_{2,i}\tilde{\mathbf{h}}_{2,i}^H\big] = \lambda_{2,i}\mathbf{\Phi}_2, \qquad i = 1, \cdots, N_2 \tag{12}$$

where $\lambda_{2,i}$ is the ith diagonal element of $\mathbf{\Lambda}_2$. From (11) and (12), $\mathbf{R}_\gamma$ can be written as

$$\mathbf{R}_\gamma = \mathrm{Bdiag}\big[\mathbf{\Lambda}_1 \otimes c_1\mathbf{\Phi}_2, \;\; \mathbf{\Lambda}_2 \otimes \mathbf{\Phi}_2\big] \tag{13}$$

where $\mathrm{Bdiag}[\cdot]$ denotes a block diagonal matrix.

The transmission power consumed at the source node is

$$\mathrm{tr}(\mathbf{S}_1\mathbf{S}_1^H) = \mathrm{tr}(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H) \tag{14}$$

From (2), the power consumed at the relay node is given by

$$\begin{aligned}&\alpha\mathrm{E}\big[\mathrm{tr}\big(\mathbf{H}_1\mathbf{S}_1\mathbf{S}_1^H\mathbf{H}_1^H + \mathbf{I}_{N_2}\big)\big] + \mathrm{tr}(\mathbf{S}_2\mathbf{S}_2^H)\\ &= \alpha N_2 + \alpha\mathrm{tr}(\mathbf{\Lambda}_1\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H)\mathrm{tr}(\mathbf{\Phi}_1) + \mathrm{tr}(\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}_2^H).\end{aligned} \tag{15}$$

From (9), (14), and (15), the optimal training matrices can be designed by solving the following optimization problem

$$\min_{\alpha, \tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2} \;\; \mathrm{tr}\Big(\big[\mathbf{R}_\gamma^{-1} + \mathbf{M}^H\mathbf{R}_{\bar{v}}^{-1}\mathbf{M}\big]^{-1}\Big) \tag{16}$$

$$\text{s.t. } \mathrm{tr}(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H) \le P_1 \tag{17}$$

$$\alpha N_2 + \alpha\mathrm{tr}(\mathbf{\Lambda}_1\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H)\mathrm{tr}(\mathbf{\Phi}_1) + \mathrm{tr}(\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}_2^H) \le P_2 \tag{18}$$

where P_i is the transmission power available at node i, $i = 1, 2$. The following theorem establishes the optimal structure of $\mathbf{S}_1$ and $\mathbf{S}_2$.

THEOREM 1: The optimal training sequence $\mathbf{S}_i$ satisfies $\mathbf{S}_1\mathbf{S}_2^H = \mathbf{0}$ and $\mathbf{S}_i\mathbf{S}_i^H = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{U}_i^H$, $i = 1, 2$, where $\mathbf{\Sigma}_i$ is an $N_i \times N_i$ diagonal matrix.

PROOF: See Appendix A. □

The optimal structure of $\mathbf{S}_i$ can be obtained from Theorem 1 as $\mathbf{S}_i = \mathbf{U}_i\mathbf{\Sigma}_i^{\frac{1}{2}}\mathbf{Q}_i$, where $\mathbf{Q}_i$ is an $N_i \times T$ semi-unitary matrix satisfying $\mathbf{Q}_i\mathbf{Q}_i^H = \mathbf{I}_{N_i}$, $i = 1, 2$, and $\mathbf{Q}_1\mathbf{Q}_2^H = \mathbf{0}$. Such $\mathbf{Q}_1$ and $\mathbf{Q}_2$ can be easily constructed, for example, from the normalized discrete Fourier transform (DFT) matrix with $T \ge N_1 + N_2$.

Interestingly, it can be seen that the optimal training matrix at node i matches the eigenvector matrix of the transmitter correlation matrix of $\mathbf{H}_i$. Using Theorem 1, the optimization problem (16)-(18) is converted to the following problem

$$\min_{\alpha, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2} \;\; \mathrm{tr}\Big(\big[\mathbf{D}_1 + \alpha\mathbf{\Sigma}_1 \otimes \mathbf{D}_3\big]^{-1} + \big[\mathbf{D}_2 + \mathbf{\Sigma}_2 \otimes \mathbf{D}_3\big]^{-1}\Big) \tag{19}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{\Sigma}_1) \le P_1 \tag{20}$$

$$\alpha N_2 + \alpha\mathrm{tr}(\mathbf{\Lambda}_1\mathbf{\Sigma}_1)\mathrm{tr}(\mathbf{\Phi}_1) + \mathrm{tr}(\mathbf{\Sigma}_2) \le P_2 \tag{21}$$

$$\mathbf{\Sigma}_1 \ge 0, \quad \mathbf{\Sigma}_2 \ge 0, \quad \alpha > 0 \tag{22}$$

where for a matrix $\mathbf{A}$, $\mathbf{A} \ge 0$ means that $\mathbf{A}$ is a positive semi-definite matrix. Using the definition of $\mathbf{D}_i$, $i = 1, 2, 3$ in (36)-(38), the problem (19)-(22) can be equivalently rewritten as the following problem with scalar variables

$$\begin{aligned}\min_{\alpha, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} \quad & \sum_{i=1}^{N_1}\sum_{j=1}^{N_3}\left(\frac{1}{c_1\lambda_{1,i}\delta_{2,j}} + \frac{\alpha\sigma_{1,i}}{1 + \alpha c_2\delta_{2,j}}\right)^{-1}\\ & + \sum_{i=1}^{N_2}\sum_{j=1}^{N_3}\left(\frac{1}{\lambda_{2,i}\delta_{2,j}} + \frac{\sigma_{2,i}}{1 + \alpha c_2\delta_{2,j}}\right)^{-1}\end{aligned} \tag{23}$$

$$\text{s.t. } \sum_{i=1}^{N_1}\sigma_{1,i} \le P_1 \tag{24}$$

$$\alpha N_2 + \alpha\, c_3\sum_{i=1}^{N_1}\lambda_{1,i}\sigma_{1,i} + \sum_{i=1}^{N_2}\sigma_{2,i} \le P_2 \tag{25}$$

$$\begin{aligned}&\alpha > 0, \quad \sigma_{1,i} \ge 0, \quad \sigma_{2,j} \ge 0,\\ &i = 1, \cdots, N_1, \quad j = 1, \cdots, N_2\end{aligned} \tag{26}$$

where $\boldsymbol{\sigma}_i \triangleq [\sigma_{i,1}, \sigma_{i,2}, \cdots, \sigma_{i,N_i}]^T$, $i = 1, 2$, $c_2 \triangleq \mathrm{tr}(\mathbf{\Theta}_2^T)$, $c_3 \triangleq \mathrm{tr}(\mathbf{\Phi}_1)$, and $\lambda_{1,i}, \lambda_{2,i}, \sigma_{1,i}, \sigma_{2,i}, \delta_{2,i}$ are the ith diagonal element of $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Delta}_2$, respectively.

Interestingly, it can be seen from (23) that the terms in the first double summation are monotonically decreasing and convex with respect to α, while the terms in the second double summation are monotonically increasing and concave with respect to α. This reflects that the estimation error of the source-relay channel is decreased when more power at the relay node is assigned to assist the estimation of $\mathbf{G}$. While the estimation of the relay-destination channel is improved if more power at the relay node is spent on the superimposed training sequence $\mathbf{S}_2$. However, the overall objective function (23) is nonconvex with respect to α, thus the problem (23)-(26) is a nonconvex optimization problem. Nevertheless, it can be shown that the problem (23)-(26) can be efficiently solved by the successive geometric programming (GP) technique [16], [17].

In the following, we show some insights of the optimal α by considering the special case where three nodes have the same number of antennas, i.e., $N_i = N$, $i = 1, 2, 3$, and both $\mathbf{H}_1$ and $\mathbf{H}_2$ have i.i.d. entries, i.e., $\mathbf{\Theta}_i = \mathbf{\Psi}_i = \mathbf{I}_N$, $i = 1, 2$. In this case, we have $\lambda_{1,i} = \lambda_{2,i} = \delta_{2,i} = 1$, $i = 1, \cdots, N$, and $c_i = N$, $i = 1, 2, 3$. Thus, the optimization problem (23)-(26) can be equivalently written as

$$\min_{\alpha, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} \; N\sum_{i=1}^{N}\left[\left(\frac{1}{N} + \frac{\alpha\sigma_{1,i}}{1 + \alpha N}\right)^{-1} + \left(1 + \frac{\sigma_{2,i}}{1 + \alpha N}\right)^{-1}\right] \tag{27}$$

$$\text{s.t. } \sum_{i=1}^{N}\sigma_{1,i} \le P_1 \tag{28}$$

$$\alpha N\left(1 + \sum_{i=1}^{N}\sigma_{1,i}\right) + \sum_{i=1}^{N}\sigma_{2,i} \le P_2 \tag{29}$$

$$\alpha > 0, \quad \sigma_{i,j} \ge 0, \quad i = 1, 2, \; j = 1, \cdots, N. \tag{30}$$

Obviously, the optimal $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ for the problem (27)-(30) is

$$\sigma_{1,i} = \frac{P_1}{N}, \quad \sigma_{2,i} = \frac{P_2}{N} - \alpha(P_1 + 1), \quad i = 1, \cdots, N. \tag{31}$$

Substituting (31) back into (27), the MSE objective function is given by

$$N^2\left[\left(\frac{1}{N} + \frac{\alpha P_1}{N(1 + \alpha N)}\right)^{-1} + \left(1 + \frac{\frac{P_2}{N} - \alpha(P_1 + 1)}{(1 + \alpha N)}\right)^{-1}\right]. \tag{32}$$

Fig. 1 shows the MSE value in (32) versus α for different P_1 where $N = 4$ and P_2 is set to be 20dB. We observe from

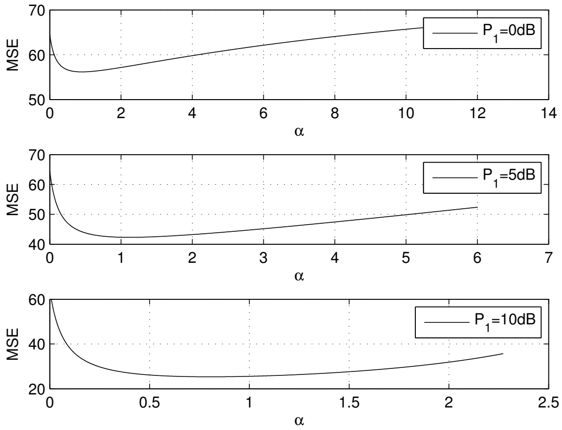

Fig. 1. MSE versus α for different P_1 with $N = 4$ and $P_2 = 20$dB.

Fig. 1 that (32) is a unimodal (quasiconvex) function of α. Moreover, it can also been seen from Fig. 1 that the convexity of (32) with respect to α indeed depends on the value of P_1 and P_2.

IV. Numerical Examples

In this section, we study the performance of the proposed channel estimation algorithm through numerical simulations. In particular, we compare the proposed superimposed channel training algorithm with the conventional channel training algorithm, where the channel estimation is completed in two stages [11]. In particular, in the first stage, the relay node sends training sequence to the destination node to enable the estimation of the relay-destination channel. In the second stage, the training sequence is sent from the source node via the relay node to the destination node, where the source-relay channel is estimated. All simulation results are averaged over 10^4 channel realizations. For each channel realization, the normalized MSE (NMSE) of channel estimation for both algorithms is calculated as

$$\frac{\|\mathbf{H}_1 - \hat{\mathbf{H}}_1\|_F^2}{N_1 N_2} + \frac{\|\mathbf{H}_2 - \hat{\mathbf{H}}_2\|_F^2}{N_2 N_3}$$

where $\|\cdot\|_F^2$ stands for the matrix Frobenius norm.

We consider a two-hop MIMO relay communication system where the number of antennas at each node is $N_1 = 4$, $N_2 = 3$, and $N_3 = 4$. Throughout the simulations, we use the minimal T, i.e., $T = N_1 + N_2$. Based on [12], we assume that $\mathbf{\Theta}_i$ and $\mathbf{\Phi}_i$, $i = 1, 2$, have the commonly used exponential Toeplitz structure such that

$$[\mathbf{\Theta}_i]_{m,n} = \mathcal{J}_0\Big(\frac{2\pi|m-n|}{\theta_i}\Big), \ \ [\mathbf{\Phi}_i]_{m,n} = \mathcal{J}_0\Big(\frac{2\pi|m-n|}{\phi_i}\Big)$$

where $\mathcal{J}_0(\cdot)$ is the zeroth order Bessel function of the first kind, and θ_i and ϕ_i stand for the correlation coefficient which depends on physical factors such as the angle of arrival spread, spacing between antenna elements, and the wavelength at the center frequency [12]. In the simulations, we choose $\theta_i = 10$ and $\phi_i = 5$, $i = 1, 2$.

In the first example, we set the transmission power at the relay node P_2 to be 20dB above the noise level. The NMSE of both the proposed and the conventional MIMO relay channel estimation algorithms is shown in Fig. 2 versus the transmission power at the source node P_1. In the second example, we fix $P_1 = 20$dB. The NMSE performance of both algorithms versus P_2 is displayed in Fig. 3. It can be seen from both Figs. 2 and 3 that the proposed algorithm yields a smaller NMSE than the conventional algorithm.

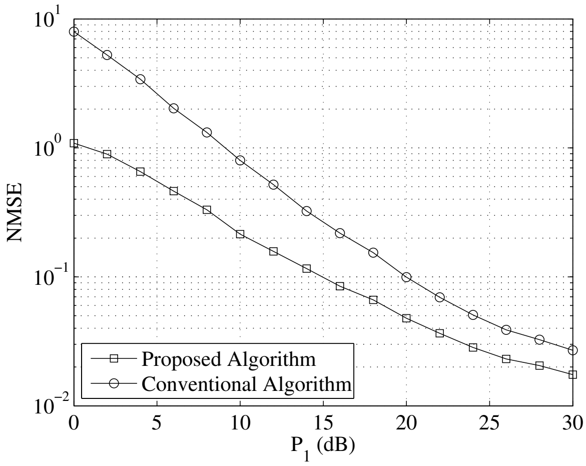

Fig. 2. Example 1: Normalized MSE versus P_1, $P_2 = 20$dB.

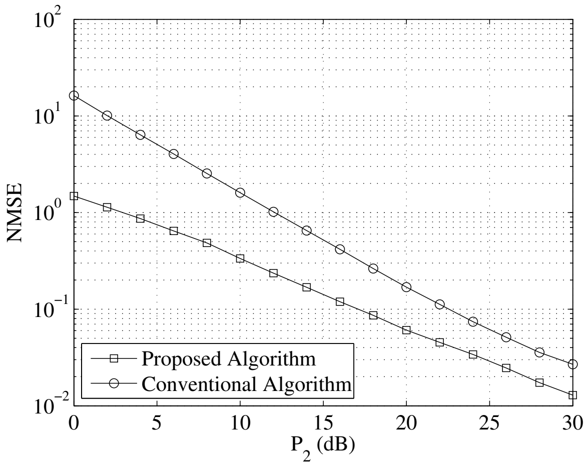

Fig. 3. Example 2: Normalized MSE versus P_2. $P_1 = 20$dB.

V. Conclusions

We have proposed a superimposed channel training algorithm for MIMO relay communication systems. The optimal structure of the training sequence and the optimal power loading among different streams are derived. Simulation results show a better performance of the proposed algorithm compared with the conventional MIMO relay channel estimation algorithm.

Acknowledgment

This work was supported under the Australian Research Council's Discovery Projects funding scheme (project numbers DP110100736, DP110102076).

Appendix A
Proof of Theorem 1

PROOF: Let us introduce the EVD of $\mathbf{\Phi}_2 = \mathbf{V}_2\mathbf{\Delta}_2\mathbf{V}_2^H$. We can equivalently rewrite (10) and (13) as

$$\mathbf{R}_{\bar{v}} = \mathbf{I}_T \otimes \big(\mathbf{V}_2(\alpha\mathrm{tr}(\mathbf{\Theta}_2^T)\mathbf{\Delta}_2 + \mathbf{I}_{N_3})\mathbf{V}_2^H\big) \tag{33}$$

$$\begin{aligned}\mathbf{R}_{\gamma} = {} & \mathrm{Bdiag}\big[\mathbf{I}_{N_1}\otimes\mathbf{V}_2,\ \mathbf{I}_{N_2}\otimes\mathbf{V}_2\big]\mathrm{Bdiag}\big[\mathbf{\Lambda}_1\otimes c_1\mathbf{\Delta}_2,\\ & \mathbf{\Lambda}_2\otimes\mathbf{\Delta}_2\big]\mathrm{Bdiag}\big[\mathbf{I}_{N_1}\otimes\mathbf{V}_2^H,\ \mathbf{I}_{N_2}\otimes\mathbf{V}_2^H\big].\end{aligned} \tag{34}$$

Substituting (33) and (34) back into (9), the MSE can be rewritten as

$$\begin{aligned}\mathrm{MSE} = \mathrm{tr}\Bigg(\Bigg[\begin{pmatrix}\mathbf{D}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_2\end{pmatrix} + \begin{pmatrix}\sqrt{\alpha}\tilde{\mathbf{S}}_1^*\otimes\mathbf{I}_{N_3}\\ \tilde{\mathbf{S}}_2^*\otimes\mathbf{I}_{N_3}\end{pmatrix}\mathbf{I}_T\otimes\mathbf{D}_3\\ \times\Big(\sqrt{\alpha}\tilde{\mathbf{S}}_1^T\otimes\mathbf{I}_{N_3},\ \ \tilde{\mathbf{S}}_2^T\otimes\mathbf{I}_{N_3}\Big)\Bigg]^{-1}\Bigg)\end{aligned} \tag{35}$$

where $(\cdot)^*$ denotes complex conjugate and

$$\mathbf{D}_1 \triangleq \mathbf{\Lambda}_1^{-1}\otimes(c_1\mathbf{\Delta}_2)^{-1} \tag{36}$$

$$\mathbf{D}_2 \triangleq \mathbf{\Lambda}_2^{-1}\otimes\mathbf{\Delta}_2^{-1} \tag{37}$$

$$\mathbf{D}_3 \triangleq \big(\alpha\mathrm{tr}(\mathbf{\Theta}_2^T)\mathbf{\Delta}_2 + \mathbf{I}_{N_3}\big)^{-1} \tag{38}$$

are all diagonal matrices.

It can be seen from (35) that MSE is minimized only if

$$(\tilde{\mathbf{S}}_1^*\otimes\mathbf{I}_{N_3})\mathbf{I}_T\otimes\mathbf{D}_3(\tilde{\mathbf{S}}_2^T\otimes\mathbf{I}_{N_3}) = \big(\tilde{\mathbf{S}}_1^*\tilde{\mathbf{S}}_2^T\big)\otimes\mathbf{D}_3 = \mathbf{0}. \tag{39}$$

Equation (39) holds if and only if $\tilde{\mathbf{S}}_1^*\tilde{\mathbf{S}}_2^T = \mathbf{0}$, or equivalently $\mathbf{S}_1\mathbf{S}_2^H = \mathbf{0}$. Then the MSE in (35) can be written as

$$\mathrm{MSE} = \mathrm{tr}\Big(\big[\mathbf{D}_1 + \alpha\tilde{\mathbf{S}}_1^*\tilde{\mathbf{S}}_1^T\otimes\mathbf{D}_3\big]^{-1} + \big[\mathbf{D}_2 + \tilde{\mathbf{S}}_2^*\tilde{\mathbf{S}}_2^T\otimes\mathbf{D}_3\big]^{-1}\Big). \tag{40}$$

Since $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are all diagonal, to minimize (40), $\tilde{\mathbf{S}}_1^*\tilde{\mathbf{S}}_1^T$ and $\tilde{\mathbf{S}}_2^*\tilde{\mathbf{S}}_2^T$ must be diagonal. Note that the diagonality of $\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H$ does not change $\mathrm{tr}(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H)$ and $\mathrm{tr}(\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}_2^H)$ in the constraints (17) and (18). Moreover, $\mathrm{tr}(\mathbf{\Lambda}_2\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H)$ is minimized if $\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_1^H$ is diagonal and its diagonal entries are in the inverse order of that of $\mathbf{\Lambda}_2$ [18, 9.H.1.h]. Let us denote $\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H = \mathbf{\Sigma}_i$, $i = 1, 2$. Then we have $\mathbf{S}_i\mathbf{S}_i^H = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{U}_i^H$, $i = 1, 2$. □

References

[1] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1774-1786, May 2007.

[2] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398-1407, Apr. 2007.

[3] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798-2802, Aug. 2007.

[4] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pp. 5062-5073, Oct. 2008.

[5] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 4837-4851, Dec. 2009.

[6] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2268-2279, Jul. 2010.

[7] F. Roemer and M. Haardt, "Tensor-Based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Tran. Signal Process.*, vol. 58, pp. 5720-5735, Nov. 2010.

[8] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," in *Proc. IEEE WSA*, Darmstadt, Germany, Feb. 2008, pp. 90-95.

[9] P. Lioliou, M. Viberg, and M. Coldrey "Performance analysis of relay channel estimation," in *Proc. IEEE Asilomar*, Pacific Grove, CA, USA, Nov. 2009, pp. 1533-1537.

[10] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1907-1916, May 2008.

[11] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, pp. 4438-4446, Sep. 2011.

[12] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 503-513, Mar. 2000.

[13] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772-781, Sep. 1978.

[14] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC, 2000.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cilffs, NJ: Prentice Hall, 1993.

[16] C. S. Beightler and D. T. Philips, *Applied Geometric Programming*. Wiley, 1976.

[17] Y. Rong, "Multi-hop non-regenerative MIMO relays: QoS considerations," *IEEE Trans. Signal Process.*, vol. 59, pp. 290-303, Jan. 2011.

[18] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. 2nd Ed., Springer, 2009.