

# Superimposed Channel Training for Two-Way MIMO Relay Systems

Choo W. R. Chiong

Dept. Electrical and Computer Engineering Dept. Electrical and Computer Engineering School of Information Technology  
Curtin University Curtin University Deakin University  
Bentley, WA, 6102, Australia Bentley, WA, 6102, Australia Melbourne, VIC 3125, Australia  
choowee.chiong@student.curtin.edu.au y.rong@curtin.edu.au yong.xiang@deakin.edu.au

Yue Rong

Yong Xiang

**Abstract**—Two-way relaying systems are known to be capable of providing higher spectral efficiency compared with one-way relaying systems. However, the channel estimation problem for two-way relaying systems becomes more complicated. In this paper, we propose a superimposed channel training scheme for two-way MIMO relay communication systems, where the individual channel information for users-relay and relay-users links are estimated. The optimal structure of the source and relay training sequences are derived when the mean-squared error (MSE) of channel estimation is minimized. We also optimize the power allocation between the source and relay training sequences to improve the performance of the algorithm. Numerical examples are shown to demonstrate the performance of the proposed channel training algorithm.

**Index Terms**—Channel estimation, superimposed training, MIMO relay, two-way relay, MMSE.

## I. INTRODUCTION

In recent years, multiple-input multiple-output (MIMO) relay communication systems have attracted many research interests due to the significant growth in the demand for fast and reliable wireless communications [1]. Many research works have been done to estimate the instantaneous channel state information (CSI) that is required to retrieve the transmitted information in a MIMO relay communication system [1]-[2].

However, the MIMO relay systems and channel estimation algorithms mentioned in [1]-[2] are one-way relaying systems, where the direction of the transmission is fixed to one direction, i.e., the source node transmits to the destination node through a relay node. In contrast to one-way relaying systems, transmission of information occurs in both directions for two-way relaying systems, which are getting more attention recently as it can provide higher spectral efficiency.

The channel estimation problem becomes more complicated in two-way relaying systems and several algorithms have been proposed in [3]-[5]. Maximum likelihood (ML) and linear maximum signal-to-noise ratio (SNR) channel estimation techniques have been introduced in [3], while block-based training and pilot-tone based training algorithms are presented in [4]. All these algorithms are developed based on the assumption that each node is equipped with single antenna only, and extension to the MIMO case is not straightforward.

Two algorithms have been proposed in [5], namely, cascaded channel estimation and individual channel estimation. In the

cascaded channel estimation algorithm, individual relay-user channel information is not known, hence, optimization over the MIMO relay network would be difficult. Consequently, the authors of [5] suggested the individual channel estimation algorithm, where the user-relay link channel information is estimated at the relay node and feed-forward to the user nodes. However, this algorithm requires the relay node to be capable of performing signal processing, thus, increasing the cost and complexity at the relay node.

In this paper, we propose a superimposed channel training algorithm for two-way MIMO relay communication systems. In particular, for a three-node MIMO relay communication system, both source nodes transmit the training sequence simultaneously to the relay node in the first time block. The relay then amplifies the received signal and superimposes its own training sequence, before transmitting the superimposed signal to both user nodes. Using the training sequences from the user and relay nodes, the individual channel information for users-relay and relay-users links can be estimated. The individual channel information for users-relay is important for optimization of MIMO relay system, such as power allocation and precoding [6]. The optimal training sequences are derived. In particular, we show that the optimal training matrix matches the eigenvector matrix of the correlation matrix of the MIMO channel. We also optimize the power allocation between the sources and relay training sequence.

The rest of this paper is organized as follows. In Section II, we introduce the model of a two-way three-node wireless MIMO relay communication system where the superimposed channel estimation algorithm is applied. The optimal training sequences and power allocation are derived 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-way MIMO communication system where node 1 and node 2 exchange information through a relay node. Nodes 1 and 2 are equipped with  $N_1$  and  $N_2$  antennas, respectively, while the relay node has  $N_r$  antennas. We concentrate on the case where the direct link between both users, i.e., node 1 and 2, is sufficiently weak and thus can be omitted [1]-[2]. This scenario occurs when the

direct link is blocked by an obstacle, such as tall buildings, and the relay plays an important role in this case as the direct link is weak.

The channel estimation is completed in two time blocks. In the first time block, the source node  $i$  transmits an  $N_i \times T$  training signal matrix  $\mathbf{S}_i$ , where  $T$  is the length of the training sequence. The  $N_r \times T$  received signal matrix  $\mathbf{Y}_r$  at the relay node is given by

$$\mathbf{Y}_r = \sum_{i=1}^2 \mathbf{H}_{ri} \mathbf{S}_i + \mathbf{V}_r \quad (1)$$

where  $\mathbf{H}_{ri}$  is the  $N_r \times N_i$  channel matrix from node  $i$  to the relay node, and  $\mathbf{V}_r$  is an  $N_r \times T$  noise matrix at the relay node.

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

$$\mathbf{X}_r = \sqrt{\alpha} \mathbf{Y}_r + \mathbf{S}_r \quad (2)$$

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

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{H}_{ir} \mathbf{X}_r + \mathbf{V}_i \\ &= \sqrt{\alpha} \mathbf{H}_{ir} \mathbf{H}_{ri} \mathbf{S}_i + \sqrt{\alpha} \mathbf{H}_{ir} \mathbf{H}_{ri} \mathbf{S}_i + \mathbf{H}_{ir} \mathbf{S}_r \\ &\quad + \sqrt{\alpha} \mathbf{H}_{ir} \mathbf{V}_r + \mathbf{V}_i, \quad i = 1, 2 \end{aligned} \quad (3)$$

where  $\mathbf{H}_{ir}$  is the  $N_i \times N_r$  channel matrix from the relay node to node  $i$ , and  $\mathbf{V}_i$  is an  $N_i \times T$  noise matrix at node  $i$ . Here,  $\bar{i} = 2$  for  $i = 1$ , and  $\bar{i} = 1$  for  $i = 2$ .

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 also assume that the channel matrices satisfy the well-known Gaussian-Kronecker model [2], where  $\mathbf{H}_{ri}$  and  $\mathbf{H}_{ir}$  are complex-valued Gaussian random matrix with

$$\begin{aligned} \mathbf{H}_{ri} &\sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_{ri} \otimes \mathbf{R}_{ri}), \\ \mathbf{H}_{ir} &\sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_r \otimes \mathbf{R}_{ir}), \quad i = 1, 2. \end{aligned} \quad (4)$$

Here  $\mathbf{T}_{ri}$  and  $\mathbf{R}_{ri}$  denote the  $N_i \times N_i$  and  $N_r \times N_r$  covariance matrix at the transmit and receive side of  $\mathbf{H}_{ri}$ , respectively, while  $\mathbf{C}_r$  and  $\mathbf{R}_{ir}$  stand for the  $N_r \times N_r$  and  $N_i \times N_i$  covariance matrix at the transmit and receive side of  $\mathbf{H}_{ir}$ , respectively. Here  $\otimes$  stands for the matrix Kronecker product. In other words, from (4) we have

$$\begin{aligned} \mathbf{H}_{ri} &= \mathbf{A}_{ri} \mathbf{H}_{ri,w} \mathbf{B}_{ri}^H, \\ \mathbf{H}_{ir} &= \mathbf{A}_{ir} \mathbf{H}_{ir,w} \mathbf{K}_r^H, \quad i = 1, 2 \end{aligned} \quad (5)$$

where  $\mathbf{A}_{ri} \mathbf{A}_{ri}^H = \mathbf{R}_{ri}$ ,  $\mathbf{B}_{ri} \mathbf{B}_{ri}^H = \mathbf{T}_{ri}$ ,  $\mathbf{A}_{ir} \mathbf{A}_{ir}^H = \mathbf{R}_{ir}$ ,  $\mathbf{K}_r \mathbf{K}_r^H = \mathbf{C}_r$ ,  $i = 1, 2$ ,  $\mathbf{H}_{ri,w}$  and  $\mathbf{H}_{ir,w}$  are  $N_r \times N_i$  and  $N_i \times N_r$  Gaussian random matrix with i.i.d. zero mean and unit variance entries. Here  $(\cdot)^T$  and  $(\cdot)^H$  denote matrix (vector) transpose and Hermitian transpose, respectively. We assume that  $\mathbf{H}_{ri,w}$  and  $\mathbf{H}_{ir,w}$ ,  $i = 1, 2$ , are statistically independent of each other. The following lemma is important for the derivation of optimal training matrices in the next section.

LEMMA 1 [7]: For  $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$ , there is  $E[\mathbf{H} \mathbf{A} \mathbf{H}^H] = \text{tr}(\mathbf{A} \mathbf{\Theta}^T) \mathbf{\Phi}$ , and  $E[\mathbf{H}^H \mathbf{A} \mathbf{H}] = \text{tr}(\mathbf{\Phi} \mathbf{A}) \mathbf{\Theta}^T$ . Here  $E[\cdot]$  stands for statistical expectation, and  $\text{tr}(\cdot)$  denotes matrix trace.

### III. OPTIMAL TRAINING MATRICES

Let us introduce the eigenvalue decomposition (EVD) of  $\mathbf{T}_{ri}^T$  as  $\mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i^H$ ,  $i = 1, 2$ , and the EVD of  $\mathbf{C}_r$  as  $\mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$ . Then we have  $\mathbf{B}_{ri}^H = \mathbf{\Pi}_i \mathbf{\Lambda}_i^{\frac{1}{2}} \mathbf{U}_i^H$ ,  $i = 1, 2$ , and  $\mathbf{K}_r^H = \mathbf{\Pi}_r \mathbf{\Lambda}_r^{\frac{1}{2}} \mathbf{U}_r^H$ , where  $\mathbf{\Pi}_i$  and  $\mathbf{\Pi}_r$  are arbitrary  $N_i \times N_i$  and  $N_r \times N_r$  unitary matrix, respectively. Using (5), we can rewrite (3) as

$$\mathbf{Y}_i = \sqrt{\alpha} \mathbf{G}_{ii} \tilde{\mathbf{S}}_i + \sqrt{\alpha} \mathbf{G}_{i\bar{i}} \tilde{\mathbf{S}}_{\bar{i}} + \tilde{\mathbf{H}}_{ir} \tilde{\mathbf{S}}_r + \tilde{\mathbf{V}}_i \quad (6)$$

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

By vectorizing both sides of (6), we obtain

$$\begin{aligned} \mathbf{y}_i &= [\sqrt{\alpha} \tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i}, \sqrt{\alpha} \tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}, \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}] [\mathbf{g}_{ii}^T, \mathbf{g}_{i\bar{i}}^T, \tilde{\mathbf{h}}_{ir}^T]^T \\ &\quad + \tilde{\mathbf{v}}_i \\ &= \mathbf{M}_i \boldsymbol{\gamma}_i + \tilde{\mathbf{v}}_i, \quad i = 1, 2 \end{aligned} \quad (7)$$

where for  $i = 1, 2$ ,  $\mathbf{y}_i \triangleq \text{vec}(\mathbf{Y}_i)$ ,  $\mathbf{g}_{ij} \triangleq \text{vec}(\mathbf{G}_{ij})$ ,  $j = i, \bar{i}$ ,  $\tilde{\mathbf{h}}_{ir} \triangleq \text{vec}(\tilde{\mathbf{H}}_{ir})$ , and  $\tilde{\mathbf{v}}_i \triangleq \text{vec}(\tilde{\mathbf{V}}_i)$ . Here for a matrix  $\mathbf{A}$ ,  $\text{vec}(\mathbf{A})$  stacks up the columns of matrix  $\mathbf{A}$  into a single column vector,  $\mathbf{I}_n$  denotes an  $n \times n$  identity matrix,  $\mathbf{M}_i \triangleq [\sqrt{\alpha} \tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i}, \sqrt{\alpha} \tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}, \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}]$ , and  $\boldsymbol{\gamma}_i \triangleq [\mathbf{g}_{ii}^T, \mathbf{g}_{i\bar{i}}^T, \tilde{\mathbf{h}}_{ir}^T]^T$  is the vector of unknown variables at node  $i$ .

Due to its simplicity, a linear MMSE estimator [8] is applied at node  $i$  to estimate  $\boldsymbol{\gamma}_i$ . We have

$$\hat{\boldsymbol{\gamma}}_i = \mathbf{W}_i^H \mathbf{y}_i, \quad i = 1, 2 \quad (8)$$

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

$$\mathbf{W}_i = (\mathbf{M}_i \mathbf{R}_{\boldsymbol{\gamma}_i} \mathbf{M}_i^H + \mathbf{R}_{\tilde{\mathbf{v}}_i})^{-1} \mathbf{M}_i \mathbf{R}_{\boldsymbol{\gamma}_i}, \quad i = 1, 2. \quad (9)$$

Here  $(\cdot)^{-1}$  denotes matrix inversion. From (7), we find that since a linear estimator is used, there is  $T \geq N_1 + N_2 + N_r$ . Using (7)-(9), the MSE of estimating  $\boldsymbol{\gamma}_i$  can be obtained as

$$\begin{aligned} \text{MSE}_i &= E[\text{tr}((\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^H)] \\ &= \text{tr}([\mathbf{R}_{\boldsymbol{\gamma}_i}^{-1} + \mathbf{M}_i^H \mathbf{R}_{\tilde{\mathbf{v}}_i}^{-1} \mathbf{M}_i]^{-1}), \quad i = 1, 2. \end{aligned} \quad (10)$$

In (10),  $\mathbf{R}_{\tilde{\mathbf{v}}_i} \triangleq E[\tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^H]$  is the noise covariance matrix which can be calculated using Lemma 1 and is given by

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{v}}_i} &= \mathbf{I}_T \otimes (\alpha \text{tr}(\mathbf{K}_r \mathbf{K}_r^H) \mathbf{A}_{ir} \mathbf{A}_{ir}^H + \mathbf{I}_{N_i}) \\ &= \mathbf{I}_T \otimes (\alpha \text{tr}(\mathbf{C}_r^T) \mathbf{R}_{ir} + \mathbf{I}_{N_i}), \quad i = 1, 2. \end{aligned} \quad (11)$$

In (10),  $\mathbf{R}_{\gamma_i} \triangleq \mathbb{E}[\gamma_i \gamma_i^H]$  is the covariance matrix of  $\gamma_i$  and can be calculated as follows. First, the  $m$ th column of  $\mathbf{G}_{ij}$  is given by  $[\mathbf{G}_{ij}]_m = \lambda_{j,m}^{\frac{1}{2}} \mathbf{A}_{ir} \mathbf{H}_{ir,w} \mathbf{K}_r^H \mathbf{A}_{rj} \mathbf{H}_{rj,w} [\mathbf{\Pi}_j]_m$ ,  $m = 1, \dots, N_j$ , where  $\lambda_{j,m}$  is the  $m$ th diagonal element of  $\mathbf{\Lambda}_j$ , and  $[\mathbf{\Pi}_j]_m$  is the  $m$ th column of  $\mathbf{\Pi}_j$ . Since  $\mathbf{H}_{ir,w}$  and  $\mathbf{H}_{rj,w}$  are independent, the covariance matrix of  $[\mathbf{G}_{ij}]_m$  can be calculated using Lemma 1 and is given by

$$\mathbb{E}[[\mathbf{G}_{ij}]_m [\mathbf{G}_{ij}]_m^H] = \lambda_{j,m} \text{tr}(\mathbf{K}_r^H \mathbf{A}_{rj} \mathbf{A}_{rj}^H \mathbf{K}_r) \mathbf{A}_{ir} \mathbf{A}_{ir}^H = \lambda_{j,m} b_j \mathbf{R}_{ir}, \quad m=1, \dots, N_j, \quad j=i, \bar{i} \quad (12)$$

where  $b_j \triangleq \text{tr}(\mathbf{R}_{rj} \mathbf{C}_r^T)$ . Second, the covariance matrix of the  $m$ th column of  $\tilde{\mathbf{H}}_{ir}$ , denoted as  $[\tilde{\mathbf{H}}_{ir}]_m$ , is given by

$$\mathbb{E}[[\tilde{\mathbf{H}}_{ir}]_m [\tilde{\mathbf{H}}_{ir}]_m^H] = \lambda_{r,m} \mathbf{R}_{ir}, \quad m = 1, \dots, N_r \quad (13)$$

where  $\lambda_{r,m}$  is the  $m$ th diagonal element of  $\mathbf{\Lambda}_r$ . From (12) and (13),  $\mathbf{R}_{\gamma_i}$  can be written as

$$\mathbf{R}_{\gamma_i} = \text{Bdiag}[\mathbf{\Lambda}_i \otimes b_i \mathbf{R}_{ir}, \quad \mathbf{\Lambda}_{\bar{i}} \otimes b_{\bar{i}} \mathbf{R}_{ir}, \quad \mathbf{\Lambda}_r \otimes \mathbf{R}_{ir}], \quad i=1, 2 \quad (14)$$

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

The transmission power consumed at nodes 1 and 2 is

$$\text{tr}(\mathbf{S}_i \mathbf{S}_i^H) = \text{tr}(\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H), \quad i = 1, 2. \quad (15)$$

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

$$\begin{aligned} & \alpha \mathbb{E} \left[ \text{tr} \left( \sum_{i=1}^2 \mathbf{H}_{ri} \mathbf{S}_i \mathbf{S}_i^H \mathbf{H}_{ri}^H + \mathbf{I}_{N_r} \right) \right] + \text{tr}(\mathbf{S}_r \mathbf{S}_r^H) \\ & = \alpha N_r + \alpha \sum_{i=1}^2 \text{tr}(\mathbf{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H) \text{tr}(\mathbf{R}_{ri}) + \text{tr}(\tilde{\mathbf{S}}_r \tilde{\mathbf{S}}_r^H). \end{aligned} \quad (16)$$

From (10), (15), and (16), the optimal training matrices can be designed by solving the following optimization problem

$$\min_{\alpha, \tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2, \tilde{\mathbf{S}}_r} \sum_{i=1}^2 \text{tr} \left( [\mathbf{R}_{\gamma_i}^{-1} + \mathbf{M}_i^H \mathbf{R}_{\bar{v}_i}^{-1} \mathbf{M}_i]^{-1} \right) \quad (17)$$

$$\text{s.t.} \quad \text{tr}(\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H) \leq P_i, \quad i = 1, 2 \quad (18)$$

$$\alpha \left[ N_r + \sum_{i=1}^2 \text{tr}(\mathbf{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H) \text{tr}(\mathbf{R}_{ri}) \right] + \text{tr}(\tilde{\mathbf{S}}_r \tilde{\mathbf{S}}_r^H) \leq P_r \quad (19)$$

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

**THEOREM 1:** The optimal training sequence  $\mathbf{S}_1$ ,  $\mathbf{S}_2$ , and  $\mathbf{S}_r$  satisfies  $\mathbf{S}_r \mathbf{S}_i^H = \mathbf{0}$ ,  $i = 1, 2$ ,  $\mathbf{S}_1 \mathbf{S}_2^H = \mathbf{0}$ , and  $\mathbf{S}_r \mathbf{S}_r^H = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{U}_r^H$ ,  $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H$ ,  $i = 1, 2$ , where  $\mathbf{\Sigma}_r$  and  $\mathbf{\Sigma}_i$  are  $N_r \times N_r$  and  $N_i \times N_i$  diagonal matrices, respectively.

**PROOF:** See Appendix A.  $\square$

The optimal structure of  $\mathbf{S}_r$  and  $\mathbf{S}_i$  can be obtained from Theorem 1 as  $\mathbf{S}_r = \mathbf{U}_r \mathbf{\Sigma}_r^{\frac{1}{2}} \mathbf{\Omega}_r$  and  $\mathbf{S}_i = \mathbf{U}_i \mathbf{\Sigma}_i^{\frac{1}{2}} \mathbf{\Omega}_i$ , where  $\mathbf{\Omega}_r$  and  $\mathbf{\Omega}_i$  are  $N_r \times T$  and  $N_i \times T$  semi-unitary matrix, respectively, satisfying  $\mathbf{\Omega}_r \mathbf{\Omega}_r^H = \mathbf{I}_{N_r}$ ,  $\mathbf{\Omega}_i \mathbf{\Omega}_i^H = \mathbf{I}_{N_i}$ ,  $i = 1, 2$ , and  $\mathbf{\Omega}_1 \mathbf{\Omega}_2^H = \mathbf{0}$ ,  $\mathbf{\Omega}_i \mathbf{\Omega}_r^H = \mathbf{0}$ ,  $i = 1, 2$ . Such  $\mathbf{\Omega}_r$ ,  $\mathbf{\Omega}_1$ , and  $\mathbf{\Omega}_2$  can be easily constructed, for example, from the

normalized discrete Fourier transform (DFT) matrix with  $T \geq N_1 + N_2 + N_r$ .

Interestingly, it can be seen that the optimal training matrix at node  $i$  matches the eigenvector matrix of the transmitter correlation matrix of  $\mathbf{H}_{ri}$ , and the optimal training matrix at the relay node matches the eigenvector matrix of  $\mathbf{C}_r$ . Using Theorem 1 and the definition of  $\mathbf{D}_{ij}$ ,  $\mathbf{D}_{si}$  and  $\mathbf{D}_{ri}$  in (32), the optimization problem (17)-(19) is converted to the following problem with scalar variables

$$\begin{aligned} & \min_{\alpha, \sigma_1, \sigma_2, \sigma_r} \sum_{i=1}^2 \left( \sum_{m=1}^{N_i} \sum_{n=1}^{N_i} \left[ \frac{1}{b_i \lambda_{i,m} \delta_{i,n}} + \frac{\alpha \sigma_{i,m}}{\alpha \text{tr}(\mathbf{C}_r^T) \delta_{i,n} + 1} \right]^{-1} \right. \\ & \quad + \sum_{m=1}^{N_i} \sum_{n=1}^{N_i} \left[ \frac{1}{b_{\bar{i}} \lambda_{\bar{i},m} \delta_{i,n}} + \frac{\alpha \sigma_{\bar{i},m}}{\alpha \text{tr}(\mathbf{C}_r^T) \delta_{i,n} + 1} \right]^{-1} \\ & \quad \left. + \sum_{m=1}^{N_r} \sum_{n=1}^{N_r} \left[ \frac{1}{\lambda_{r,m} \delta_{i,n}} + \frac{\sigma_{r,m}}{\alpha \text{tr}(\mathbf{C}_r^T) \delta_{i,n} + 1} \right]^{-1} \right) \quad (20) \\ & \text{s.t.} \quad \sum_{m=1}^{N_i} \sigma_{i,m} \leq P_i, \quad i = 1, 2 \quad (21) \end{aligned}$$

$$\begin{aligned} & \alpha N_r + \alpha \sum_{i=1}^2 \left( \sum_{m=1}^{N_i} \lambda_{i,m} \sigma_{i,m} \text{tr}(\mathbf{R}_{ri}) \right) \\ & \quad + \sum_{m=1}^{N_r} \sigma_{r,m} \leq P_r \quad (22) \\ & \alpha > 0, \quad \sigma_{1,m} \geq 0, \quad \sigma_{2,m} \geq 0, \quad \sigma_{r,m} \geq 0 \quad (23) \end{aligned}$$

where  $\sigma_r \triangleq [\sigma_{r,1}, \dots, \sigma_{r,N_r}]^T$ ,  $\sigma_i \triangleq [\sigma_{i,1}, \dots, \sigma_{i,N_i}]^T$ ,  $i = 1, 2$ , and  $\lambda_{1,m}$ ,  $\lambda_{2,m}$ ,  $\lambda_{r,m}$ ,  $\sigma_{1,m}$ ,  $\sigma_{2,m}$ ,  $\sigma_{r,m}$ ,  $\delta_{1,m}$ ,  $\delta_{2,m}$  are the  $m$ th diagonal element of  $\mathbf{\Lambda}_1$ ,  $\mathbf{\Lambda}_2$ ,  $\mathbf{\Lambda}_r$ ,  $\mathbf{\Sigma}_1$ ,  $\mathbf{\Sigma}_2$ ,  $\mathbf{\Sigma}_r$ ,  $\mathbf{\Delta}_1$ ,  $\mathbf{\Delta}_2$ , respectively.

The problem (20)-(23) is a non-convex optimization problem. However, the optimization of the problem (20)-(23) with respect to  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_r$  is convex when  $\alpha$  is fixed. In particular, for a given  $\alpha$ , the optimal  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_r$  can be efficiently obtained through the Karush-Kuhn-Tucker (KKT) optimality conditions of the problem (20)-(23). The gradient conditions are given by

$$\sum_{n=1}^{N_1} \frac{\alpha d_1^{(n)}}{[c_{11}^{(mn)} + \alpha d_1^{(n)} \sigma_{1,m}]^2} + \sum_{n=1}^{N_2} \frac{\alpha d_2^{(n)}}{[c_{12}^{(mn)} + \alpha d_2^{(n)} \sigma_{1,m}]^2} = \mu_1 + \mu_3 e_1^{(m)}, \quad m = 1, \dots, N_1 \quad (24)$$

$$\sum_{n=1}^{N_1} \frac{\alpha d_1^{(n)}}{[c_{21}^{(mn)} + \alpha d_1^{(n)} \sigma_{2,m}]^2} + \sum_{n=1}^{N_2} \frac{\alpha d_2^{(n)}}{[c_{22}^{(mn)} + \alpha d_2^{(n)} \sigma_{2,m}]^2} = \mu_2 + \mu_3 e_2^{(m)}, \quad m = 1, \dots, N_2 \quad (25)$$

$$\sum_{n=1}^{N_1} \frac{d_1^{(n)}}{[c_{r1}^{(mn)} + d_1^{(n)} \sigma_{r,m}]^2} + \sum_{n=1}^{N_2} \frac{d_2^{(n)}}{[c_{r2}^{(mn)} + d_2^{(n)} \sigma_{r,m}]^2} = \mu_3 \quad m = 1, \dots, N_r \quad (26)$$

where

$$\begin{aligned} c_{11}^{(mn)} & \triangleq 1/(b_1 \lambda_{1,m} \delta_{1,n}), & c_{12}^{(mn)} & \triangleq 1/(b_1 \lambda_{1,m} \delta_{2,n}) \\ c_{22}^{(mn)} & \triangleq 1/(b_2 \lambda_{2,m} \delta_{2,n}), & c_{21}^{(mn)} & \triangleq 1/(b_2 \lambda_{2,m} \delta_{1,n}) \end{aligned}$$

$$\begin{aligned}
 c_{r1}^{(mn)} &\triangleq 1/(\lambda_{r,m}\delta_{1,n}), & c_{r2}^{(mn)} &\triangleq 1/(\lambda_{r,m}\delta_{2,n}) \\
 d_1^{(n)} &\triangleq 1/(\alpha \text{tr}(\mathbf{C}_r^T)\delta_{1,n} + 1), & d_2^{(n)} &\triangleq 1/(\alpha \text{tr}(\mathbf{C}_r^T)\delta_{2,n} + 1) \\
 e_1^{(m)} &\triangleq \alpha \text{tr}(\mathbf{R}_{r1})\lambda_{1,m}, & e_2^{(m)} &\triangleq \alpha \text{tr}(\mathbf{R}_{r2})\lambda_{2,m}
 \end{aligned}$$

and  $\mu_1 \geq 0, \mu_2 \geq 0, \mu_3 \geq 0$  are Lagrange multipliers such that the complementary slackness conditions given by

$$\mu_i \left( P_i - \sum_{m=1}^{N_i} \sigma_{i,m} \right) = 0, \quad i = 1, 2 \quad (27)$$

$$\mu_3 \left( P_r - \alpha N_r - \alpha \sum_{i=1}^2 \sum_{m=1}^{N_i} \lambda_{i,m} \sigma_{i,m} \text{tr}(\mathbf{R}_{ri}) - \sum_{m=1}^{N_r} \sigma_{r,m} \right) = 0 \quad (28)$$

are satisfied.

With any fixed  $\alpha$ ,  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$ , the non-negative  $\sigma_{1,m}$ ,  $\sigma_{2,m}$ , and  $\sigma_{r,m}$  can be found by using the bi-section search for all  $m$ , since the left-hand-side (LHS) of (24), (25), and (26) are monotonically decreasing function of  $\sigma_{1,m}$ ,  $\sigma_{2,m}$ , and  $\sigma_{r,m}$ , respectively. To find the optimal  $\mu_1$ , an outer bi-section search is used as the LHS of (21) is increasing function of  $\sigma_{1,m}$ , while in (24),  $\sigma_{1,m}$  is monotonically decreasing function of  $\mu_1$ . Similar method can be applied to find optimal  $\mu_2$  and  $\mu_3$ .

Now we show some insights on the optimal  $\alpha$  by considering the MSE objective function in (20). Interestingly, it can be observed from (20) that the terms in the first and second double summation are monotonically decreasing and convex with respect to  $\alpha$ , while the terms in the third double summation are monotonically increasing and concave with respect to  $\alpha$ . This indicates that the estimation errors of both sources-to-relay channels are improved when more power at the relay node is allocated to assist the estimation of  $\mathbf{G}_{ii}$  and  $\mathbf{G}_{i\bar{i}}$  in (6). On the other hand, the estimation errors of the channels from the relay to both sources are decreased when more power at the relay node is assigned to the superimposed training sequence  $\mathbf{S}_r$ .

A plot of the MSE value over a range of feasible values of  $\alpha$  is generated in Fig. 1 for the case where all nodes have the same number of antennas, i.e.,  $N_i = N_r = N = 4$ ,  $i = 1, 2$ , and the channel matrices have i.i.d. entries, i.e.,  $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$ ,  $i = 1, 2$ . Fig. 1 shows the MSE value versus  $\alpha$  for different  $P_1 = P_2$ , and  $P_r$  is set to be 20dB.

It can be observed from Fig. 1 that the MSE objective function (20) is a unimodal (quasiconvex) function of  $\alpha$ , i.e., the function has only one local minimum. For a unimodal function, the minimum value can be efficiently found by the golden section search (GSS) algorithm [10]. Hence, at least the local optimal  $\alpha$ , for the problem (20)-(23) can be obtained by applying the GSS technique, and the procedure is listed in Table I. Note that  $|\cdot|$  denotes the absolute value,  $\varepsilon$  is a positive constant close to 0, and  $\phi$  is a positive constant reduction factor. It has been proven that the optimal  $\phi = 1.618$  [10]. This algorithm requires reasonable computational complexity as only few iterations are needed to find optimal  $\alpha$ .

#### IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel estimation algorithm through numerical

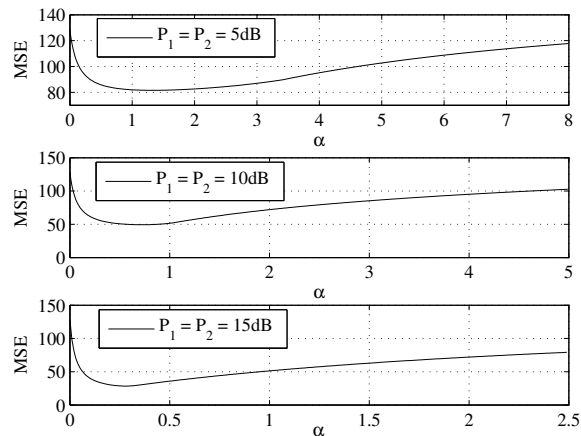


Fig. 1. MSE versus  $\alpha$  for different  $P_1 = P_2$  with  $N = 4$  and  $P_r = 20\text{dB}$ .

TABLE I

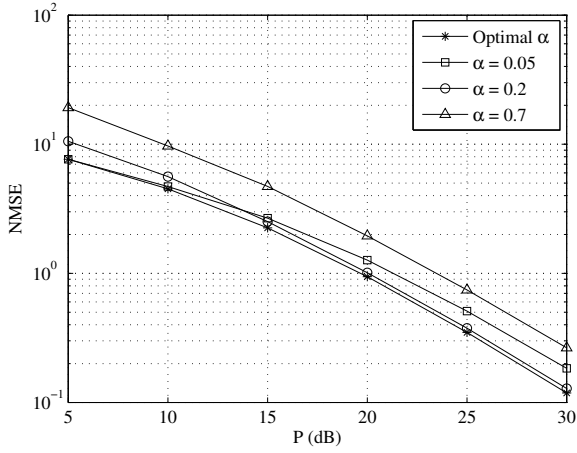
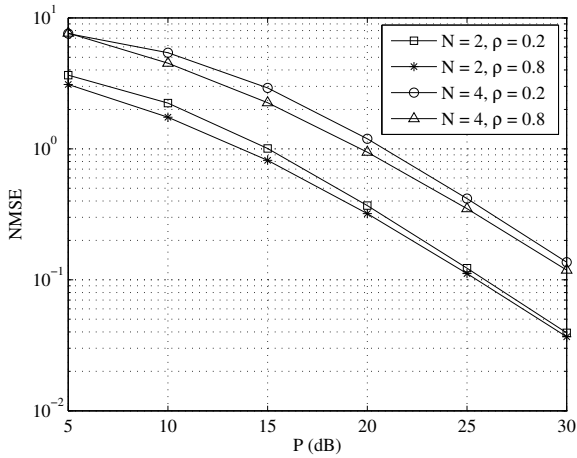
PROCEDURE OF APPLYING THE GOLDEN SECTION SEARCH (GSS) TO FIND THE OPTIMAL  $\alpha$  IN PROBLEM (20)-(23).

- 1) Set a feasible bound  $[a, b]$  on  $\alpha$ .
- 2) Define  $c_1 = (\phi - 1)a + (2 - \phi)b$  and  $c_2 = (2 - \phi)a + (\phi - 1)b$ .
- 3) Solve problem (24) - (26) for  $\alpha = c_1$ ;  
Compute the MSE value defined in (20),  $f_{MSE}(c_1)$  for  $\alpha = c_1$ .
- 4) Repeat Step 3 for  $\alpha = c_2$ .
- 5) If  $f_{MSE}(c_1) < f_{MSE}(c_2)$ , then assign  $b = c_2$ .  
Otherwise, assign  $a = c_1$ .
- 6) If  $|b - a| \leq \varepsilon$ , then end.  
Otherwise, go to step 2.

simulations. We consider a three-node two-way MIMO relay communication system where each node is equipped with the same number of antennas, i.e.,  $N_1 = N_2 = N_r = N$ . For simplicity, we assume the power at the source nodes are the same as the power at the relay node, i.e.,  $P_1 = P_2 = P_r = P$ . The channel covariance matrices are modeled as  $[\mathbf{T}_{ri}]_{m,n} = \rho^{|m-n|}$ ,  $i = 1, 2$ ,  $[\mathbf{R}_{ri}]_{m,n} = \rho^{|m-n|}$ ,  $i = 1, 2$ ,  $[\mathbf{R}_{ir}]_{m,n} = \rho^{|m-n|}$ ,  $i = 1, 2$ , and  $[\mathbf{C}_r]_{m,n} = \rho^{|m-n|}$ , where  $\rho$  is the correlation coefficient with magnitude  $|\rho| < 1$ . For all scenarios, the normalized MSE (NMSE) of channel estimation at node 1 and 2 are computed.

Fig. 2 shows the NMSE of channel estimation at node 1 for different  $\alpha$  when  $N = 4$  and  $\rho = 0.8$ . The optimal  $\alpha$  curve is obtained by applying the GSS algorithm on the proposed superimposed channel estimation algorithm to find the optimal  $\alpha$  for different  $P$ . With these optimal  $\alpha$ , different number of antennas and normalized correlation coefficient are used for the next scenario to show the impact of these parameters on the performance of the proposed algorithm and the results are shown in Fig. 3. Note that the NMSE of channel estimation at node 1 is the same as NMSE at node 2 for both scenarios.

From the simulation results, it can be seen that the optimal  $\alpha$  changes when  $P$  changes. At a smaller  $P$ , the optimal  $\alpha$  is smaller, while at a greater  $P$ , the optimal  $\alpha$  is greater. It is proven as well that GSS algorithm can be used to obtain the optimal  $\alpha$  at different  $P$ , hence, yields the best performance


 Fig. 2. MSE versus  $P$  for different  $\alpha$  with  $N = 4$  and  $\rho = 0.8$ 

 Fig. 3. MSE versus  $P$  for different  $N$  and  $\rho$  with optimal  $\alpha$ 

curve compared to simulations with constant  $\alpha$ . Similar to [5], the NMSE for the case when  $N = 4$  is greater compared to the case when  $N = 2$  as there are more unknowns to estimate.

## V. CONCLUSIONS

We have proposed a superimposed channel training algorithm for two-way MIMO relay systems. This algorithm can efficiently estimate the individual CSI for two-way MIMO relay systems. The optimal training sequences and the power allocation between these training sequences are derived.

### APPENDIX A PROOF OF THEOREM 1

PROOF: Let us introduce the EVD of  $\mathbf{R}_{ir} = \mathbf{Q}_i \mathbf{\Delta}_i \mathbf{Q}_i^H$ . We can equivalently rewrite (11) and (14) as

$$\mathbf{R}_{\bar{v}_i} = \mathbf{I}_T \otimes (\mathbf{Q}_i (\alpha \text{tr}(\mathbf{C}_r^T) \mathbf{\Delta}_i + \mathbf{I}_{N_i}) \mathbf{Q}_i^H), \quad (29)$$

$$\mathbf{R}_{\gamma_i} = \mathbf{U}_{\gamma_i} \mathbf{B} \text{diag}[\mathbf{\Lambda}_i \otimes b_i \mathbf{\Delta}_i, \mathbf{\Lambda}_{\bar{i}} \otimes b_{\bar{i}} \mathbf{\Delta}_i, \mathbf{\Lambda}_r \otimes \mathbf{\Delta}_i] \mathbf{U}_{\gamma_i}^H \quad (30)$$

where  $\mathbf{U}_{\gamma_i} \triangleq \mathbf{B} \text{diag}[\mathbf{I}_{N_i} \otimes \mathbf{Q}_i, \mathbf{I}_{N_{\bar{i}}} \otimes \mathbf{Q}_i, \mathbf{I}_{N_r} \otimes \mathbf{Q}_i]$ ,  $i = 1, 2$ . Substituting (29) and (30) back into (10),  $\text{MSE}_i$  can be

rewritten as

$$\begin{aligned} \text{MSE}_i = \text{tr} \left( \left( \begin{bmatrix} \mathbf{D}_{ii} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\bar{i}\bar{i}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{si} \end{bmatrix} + \begin{pmatrix} \sqrt{\alpha} \tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i} \\ \sqrt{\alpha} \tilde{\mathbf{S}}_{\bar{i}}^* \otimes \mathbf{I}_{N_i} \\ \tilde{\mathbf{S}}_r^* \otimes \mathbf{I}_{N_i} \end{pmatrix} (\mathbf{I}_T \otimes \mathbf{D}_{ri}) \right. \right. \\ \left. \left. \times \left( \sqrt{\alpha} \tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i}, \sqrt{\alpha} \tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}, \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i} \right) \right)^{-1} \right) \quad (31) \end{aligned}$$

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

$$\begin{aligned} \mathbf{D}_{ij} &\triangleq \mathbf{\Lambda}_j^{-1} \otimes (b_j \mathbf{\Delta}_i)^{-1}, \quad j = i, \bar{i}, & \mathbf{D}_{si} &\triangleq \mathbf{\Lambda}_r^{-1} \otimes \mathbf{\Delta}_i^{-1} \\ \mathbf{D}_{ri} &\triangleq (\alpha \text{tr}(\mathbf{C}_r^T) \mathbf{\Delta}_i + \mathbf{I}_{N_i})^{-1} \end{aligned} \quad (32)$$

are all diagonal matrices. It can be seen from (31) that the objective function (17) is minimized only if

$$(\tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i}) \mathbf{I}_T \otimes \mathbf{D}_{ri} (\tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i}) = (\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_i^T) \otimes \mathbf{D}_{ri} = \mathbf{0} \quad (33)$$

$$(\tilde{\mathbf{S}}_j^* \otimes \mathbf{I}_{N_i}) \mathbf{I}_T \otimes \mathbf{D}_{ri} (\tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}) = (\tilde{\mathbf{S}}_j^* \tilde{\mathbf{S}}_r^T) \otimes \mathbf{D}_{ri} = \mathbf{0} \quad (34)$$

where for  $i = 1, 2$ , and  $j = i, \bar{i}$ . Equations (33) and (34) hold if and only if  $\tilde{\mathbf{S}}_1^* \tilde{\mathbf{S}}_2^T = \mathbf{0}$  and  $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_r^T = \mathbf{0}$ ,  $i = 1, 2$ , or equivalently  $\mathbf{S}_1 \mathbf{S}_2^H = \mathbf{0}$  and  $\mathbf{S}_i \mathbf{S}_r^H = \mathbf{0}$ ,  $i = 1, 2$ . Then the objective function (17) can be written as

$$\begin{aligned} \sum_{i=1}^2 \text{tr} \left( [\mathbf{D}_{ii} + \alpha \tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_i^T \otimes \mathbf{D}_{ri}]^{-1} + [\mathbf{D}_{\bar{i}\bar{i}} + \alpha \tilde{\mathbf{S}}_{\bar{i}}^* \tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{D}_{ri}]^{-1} \right. \\ \left. + [\mathbf{D}_{si} + \tilde{\mathbf{S}}_r^* \tilde{\mathbf{S}}_r^T \otimes \mathbf{D}_{ri}]^{-1} \right). \end{aligned} \quad (35)$$

Since  $\mathbf{D}_{ii}$ ,  $\mathbf{D}_{\bar{i}\bar{i}}$ ,  $\mathbf{D}_{si}$ , and  $\mathbf{D}_{ri}$  are all diagonal, to minimize (35),  $\tilde{\mathbf{S}}_1^* \tilde{\mathbf{S}}_2^T$ ,  $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_r^T$ , and  $\tilde{\mathbf{S}}_r^* \tilde{\mathbf{S}}_r^T$  must be diagonal. Note that  $\text{tr}(\mathbf{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H)$  in the constraints (19) is minimized if  $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H$  is diagonal and its diagonal entries are in the inverse order of that of  $\mathbf{\Lambda}_i$  [9]. Denote  $\mathbf{S}_r \tilde{\mathbf{S}}_r^H = \mathbf{\Sigma}_r$ ,  $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H = \mathbf{\Sigma}_i$ ,  $i = 1, 2$ . Then we have  $\mathbf{S}_r \mathbf{S}_r^H = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{U}_r^H$ ,  $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H$ ,  $i = 1, 2$ .  $\square$

## REFERENCES

- [1] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," *Proc. International ITG Workshop on Smart Antennas*, pp. 90-95, Feb. 2008.
- [2] 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, Sept. 2011.
- [3] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, pp. 3024-3033, Oct. 2009.
- [4] C. K. Ho, R. Zhang, and Y.-C. Liang, "Two-way relaying over OFDM: Optimized tone permutation and power allocation," in *Proc. IEEE ICC*, pp. 3908-3912, May 2008.
- [5] Z. Fang, J. Shi, and H. Shan, "Comparison of channel estimation schemes for MIMO two-way relaying systems," in *Proc. Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, 2011, vol. 1, pp. 719-722, Jul. 2011.
- [6] Y. Rong, "Joint Source and Relay Optimization for Two-Way Linear Non-Regenerative MIMO Relay Communications," in *IEEE Trans. Signal Process.*, to appear, 2012.
- [7] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC, 2000.
- [8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [9] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. 2nd Ed., Springer, 2009.
- [10] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Spring Street, NY: Springer Science+Business Media, LCC, 2007.