Superimposed Channel Training Algorithm for Time-Varying MIMO Relay Systems

Choo W. R. Chiong
Dept. Electrical and Computer Engineering
Curtin University
Bentley, WA, 6102, Australia
choowee.chiong@student.curtin.edu.au

Yue Rong
Dept. Electrical and Computer Engineering
Curtin University
Bentley, WA, 6102, Australia
y.rong@curtin.edu.au

Yong Xiang
School of Information Technology
Deakin University
Melbourne, VIC 3125, Australia
yong.xiang@deakin.edu.au

***Abstract*—In this paper, we propose a superimposed channel training algorithm for multiple-input multiple-output (MIMO) relay communication systems with time-varying channels. The time-varying characteristic of the channels is described by the complex-exponential basis expansion model (CE-BEM). The proposed algorithm can estimate the individual first-hop and second-hop time-varying channel matrices of MIMO relay systems. To improve the performance of channel estimation, we derive the optimal structure of the source and relay training sequences that minimize the mean-squared error (MSE) of channel estimation. We also optimize the relay amplification factor which determines the power allocation between the source and relay training sequences. Numerical simulations demonstrate that the proposed superimposed channel training algorithm for MIMO relay systems with time-varying channels outperforms the conventional two-stage channel estimation scheme.**

***Index Terms*—Channel estimation, MIMO relay, superimposed training, MMSE, time-varying channel, CE-BEM**

I. Introduction

Multiple-input multiple-output (MIMO) relay communication is one of the promising solutions to high rate wireless communication [1]. Many research works have been carried out to optimize the MIMO relay communication systems [2]-[4]. For MIMO relay systems discussed in [1]-[4], the knowledge of the instantaneous channel state information (CSI) is required for the retrieval of the source information at the destination node and the optimization of the MIMO relay systems by deriving the optimal source and relay precoding matrices. However, in practical MIMO relay communication systems, the instantaneous CSI is not available at any node, and hence, needs to be estimated.

Channel estimation algorithms have been developed in [5]-[6], assuming that the channels in MIMO relay systems are quasi-static block fading. However, with the increasing number of mobile wireless devices, the channels in MIMO relay systems are more likely to be affected by the changing environment caused by the movement of users. It is stated in [7] that in wireless relay systems, the Doppler spread is doubled when there is a relative motion between any two nodes. Thus, it is important to address the channel estimation issues for MIMO relay systems with time-varying channels. In [8], the channel estimation problem has been investigated for doubly selective wireless fading channels, where the time-varying characteristic of channels is represented by the basis expansion model (BEM). In BEM, the channel is represented by the superposition of the time-varying basis functions weighted by time-invariant coefficients [9]. Note that the channel estimation algorithm in [8] was developed for single-hop single-antenna communication networks.

Channel estimation and training sequence design for time-varying relay networks have been discussed in [10], where the complex-exponential BEM (CE-BEM) [8] was adopted to represent the time-varying channels in terms of Fourier bases. However, the algorithm in [10] was developed for single-antenna relay networks, and the extension to MIMO relay networks is not straightforward. In this paper, we propose a superimposed channel training algorithm to estimate the individual first-hop and second-hop time-varying channel matrices for MIMO relay communication systems. We use the CE-BEM adopted in [10] to capture the time-varying characteristic of the channel, and apply the pilot symbol aided modulation (PSAM) technique [11] for channel estimation. We would like to mention that all channel estimation works are implemented at the destination node to minimize the amount of signal processing at the relay node. We derive the optimal structure of the source and relay training sequences that minimize the MSE of the channel estimation. We also optimize the amplification factor at the relay node that governs the power allocation between the source and relay training sequences.

II. System Model

We consider a three-node two-hop time-varying MIMO relay communication system where the source node transmits information to the destination node through a relay node as shown in Fig. 1. The source, relay, and destination nodes are equipped with N_s, N_r, and N_d antennas, respectively. In this paper, the direct link between the source node and destination node is assumed to be sufficiently weak and thus can be neglected. For notational convenience, we consider a narrow-band single-carrier system. However, our results can be straightforwardly generalized to each subcarrier of a broadband multi-carrier MIMO relay system.

The PSAM technique [11] is applied to estimate the time-varying channels, where training symbols are inserted among information-carrying symbols in each data frame for channel

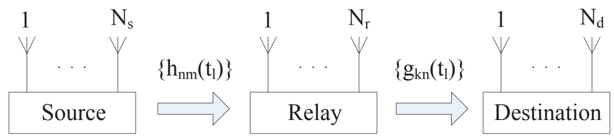

Fig. 1. Block diagram of a three-node two-hop MIMO relay communication system.

estimation. Let us denote t_l, $l = 1, \cdots, L$, as the time indexes of the pilot symbols within one data frame of duration T, where L is the length of the training sequences. The channel estimation process is completed in two time blocks. During the first time block, training signals $s_m(t_l)$ are transmitted from the mth antenna of the source node, $m = 1, \cdots, N_s$, $l = 1, \cdots, L$. The received signals at the nth antenna of the relay node is given by

$$y_{r,n}(t_l) = \sum_{m=1}^{N_s} h_{nm}(t_l)s_m(t_l) + v_{r,n}(t_l), \quad n = 1, \cdots, N_r, \quad l = 1, \cdots, L \tag{1}$$

where $h_{nm}(t_l)$ is the first-hop time-varying channel from the mth antenna at the source node to the nth antenna at the relay node, $v_{r,n}(t_l)$ is the noise at the nth antenna of the relay node. During the second time block, the relay node amplifies $y_{r,n}(t_l)$, $n = 1, \cdots, N_r$, $l = 1, \cdots, L$, and superimposes its own training sequence $r_n(t_l)$ before retransmitting the signal to the destination node. The signal received at the kth antenna of the destination node is given by

$$y_k(t_l) = \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r} g_{kn}(t_l)h_{nm}(t_l)s_m(t_l) + \sum_{n=1}^{N_r} g_{kn}(t_l)r_n(t_l) + \sqrt{\alpha}\sum_{n=1}^{N_r} g_{kn}(t_l)v_{r,n}(t_l) + v_{d,k}(t_l), \quad k = 1, \cdots, N_d, \quad l = 1, \cdots, L. \tag{2}$$

where $\alpha > 0$ is the relay amplifying factor, $g_{kn}(t_l)$ is the second-hop time-varying channel from the nth antenna at the relay node to the kth antenna at the destination node, $v_{d,k}(t_l)$ is the noise at the kth antenna of the destination node. All noises are assumed to be independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and unit variance.

The CE-BEM is used to represent the time-varying characteristic of channels as

$$h_{nm}(t_l) = \sum_{q=0}^{Q} \mu_{nm}(q)e^{j2\pi t_l(q-\frac{Q}{2})/T}, \quad m = 1, \cdots, N_s, \quad n = 1, \cdots, N_r, \quad l = 1, \cdots, L \tag{3}$$

$$g_{kn}(t_l) = \sum_{q=0}^{Q} \lambda_{kn}(q)e^{j2\pi t_l(q-\frac{Q}{2})/T}, \quad n = 1, \cdots, N_r, \quad k = 1, \cdots, N_d, \quad l = 1, \cdots, L. \tag{4}$$

where $j = \sqrt{-1}$, $\mu_{nm}(q)$ and $\lambda_{kn}(q)$ are the BEM coefficients that do not change within the duration of one data frame T, and Q is the number of base functions. The complex exponential terms in (3) and (4) are the Fourier base functions that describe the time-varying characteristic of channels. Note that Q depends on T and the channel bandwidth f, and should be at least $2\lceil fT \rceil$ [8], where $\lceil a \rceil$ denotes the smallest integer greater than a. In this paper, we assume that all channel links have the same number of bases Q, and the generalization to the case where each channel link has a different number of bases is straightforward. We also assume that the channels $h_{nm}(t_l)$ and $g_{kn}(t_l)$ are wide-sense stationary (WSS) zero mean complex Gaussian (ZMCG) random process with constant variances, $\mu_{nm}(q)$ and $\lambda_{kn}(q)$ are variables with zero mean and independent of each other.

The idea of the proposed superimposed channel training algorithm is to estimate the second-hop channels $\{g_{kn}(t_l)\} \triangleq \{g_{kn}(t_l), k = 1, \cdots, N_d, n = 1, \cdots, N_r, l = 1, \cdots, L\}$ using the superimposed relay training sequence $\{r_n(t_l)\} \triangleq \{r_n(t_l), n = 1, \cdots, N_r, l = 1, \cdots, L\}$. Then, the first-hop channels $\{h_{nm}(t_l)\} \triangleq \{h_{nm}(t_l), n = 1, \cdots, N_r, m = 1, \cdots, N_s, l = 1, \cdots, L\}$ are estimated using the source training sequence $\{s_m(t_l)\} \triangleq \{s_m(t_l), m = 1, \cdots, N_s, l = 1, \cdots, L\}$ and the estimated $\{g_{kn}(t_l)\}$. An advantage of the proposed superimposed technique is that the channel estimation process is completed in one transmission cycle.

III. MMSE-Based Optimal Training Matrices

In this section, the optimal training sequences $\{s_m(t_l)\}$ and $\{r_n(t_l)\}$, and the relay amplifying factor α that minimize the MSE of channel estimation are derived. From (3) and (4), we have

$$g_{kn}(t_l)h_{nm}(t_l) = \sum_{p=0}^{Q}\sum_{q=0}^{Q} \lambda_{kn}(p)\mu_{nm}(q)e^{j2\pi t_l(p+q-Q)/T} = \sum_{i=0}^{2Q} x_{knm}(i)e^{j\theta_i t_l} \tag{5}$$

where $m = 1, \cdots, N_s$, $n = 1, \cdots, N_r$, $k = 1, \cdots, N_d$, $l = 1, \cdots, L$, $\theta_i \triangleq 2\pi(i-Q)/T$ and

$$x_{knm}(i) \triangleq \sum_{p=0}^{i} \lambda_{kn}(p)\mu_{nm}(i-p), \quad i = 0, \cdots, 2Q. \tag{6}$$

By substituting (4) and (5) back into (2), we obtain

$$y_k(t_l) = \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r}\sum_{i=0}^{2Q} x_{knm}(i)e^{j\theta_i t_l}s_m(t_l) + \sum_{n=1}^{N_r}\sum_{q=0}^{Q} \lambda_{kn}(q)e^{j\phi_q t_l}r_n(t_l) + \bar{v}_k(t_l), \quad k = 1, \cdots, N_d, \quad l = 1, \cdots, L \tag{7}$$

where $\phi_q \triangleq 2\pi(q - Q/2)/T$ and

$$\bar{v}_k(t_l) \triangleq \sqrt{\alpha}\sum_{n=1}^{N_r} g_{kn}(t_l)v_{r,n}(t_l) + v_{d,k}(t_l), \quad k = 1, \cdots, N_d, \quad l = 1, \cdots, L \tag{8}$$

is the equivalent noise at the kth antenna of the destination node.

Let us denote $\mathbf{y}_k \triangleq [y_k(t_1), y_k(t_2), \cdots, y_k(t_L)]^T$, $\mathbf{S}_m \triangleq \mathrm{diag}[s_m(t_1), s_m(t_2), \cdots, s_m(t_L)]$, $\mathbf{R}_n \triangleq \mathrm{diag}[r_n(t_1), r_n(t_2), \cdots, r_n(t_L)]$, $\bar{\mathbf{v}}_k \triangleq [\bar{v}_k(t_1), \bar{v}_k(t_2), \cdots, \bar{v}_k(t_L)]^T$, and

$$\boldsymbol{\lambda}_{kn} \triangleq [\lambda_{kn}(0), \lambda_{kn}(1), \cdots, \lambda_{kn}(Q)]^T, \tag{9}$$

$$\mathbf{x}_{km} \triangleq \left[\sum_{n=1}^{N_r} x_{knm}(0), \sum_{n=1}^{N_r} x_{knm}(1), \cdots, \sum_{n=1}^{N_r} x_{knm}(2Q)\right]^T, \tag{10}$$

$$\boldsymbol{\Theta} \triangleq \begin{bmatrix} e^{j\theta_0 t_1} & \cdots & e^{j\theta_{2Q} t_1} \\ \vdots & \ddots & \vdots \\ e^{j\theta_0 t_L} & \cdots & e^{j\theta_{2Q} t_L} \end{bmatrix}, \ \boldsymbol{\Phi} \triangleq \begin{bmatrix} e^{j\phi_0 t_1} & \cdots & e^{j\phi_Q t_1} \\ \vdots & \ddots & \vdots \\ e^{j\phi_0 t_L} & \cdots & e^{j\phi_Q t_L} \end{bmatrix} \tag{11}$$

where $(\cdot)^T$ denotes the vector (matrix) transpose and $\mathrm{diag}[\cdot]$ stands for a diagonal matrix. Then we can equivalently rewrite (7) in matrix vector form as

$$\mathbf{y}_k = \sqrt{\alpha} \sum_{m=1}^{N_s} \mathbf{S}_m \boldsymbol{\Theta} \mathbf{x}_{km} + \sum_{n=1}^{N_r} \mathbf{R}_n \boldsymbol{\Phi} \boldsymbol{\lambda}_{kn} + \bar{\mathbf{v}}_k \tag{12}$$

$$= \mathbf{A}\mathbf{b}_k + \bar{\mathbf{v}}_k, \qquad k = 1, \cdots, N_d \tag{13}$$

where

$$\mathbf{A} \triangleq [\sqrt{\alpha}\mathbf{S}_1\boldsymbol{\Theta}, \cdots, \sqrt{\alpha}\mathbf{S}_{N_s}\boldsymbol{\Theta}, \mathbf{R}_1\boldsymbol{\Phi}, \cdots, \mathbf{R}_{N_r}\boldsymbol{\Phi}] \tag{14}$$

$$\mathbf{b}_k \triangleq \left[\mathbf{x}_{k1}^T, \cdots, \mathbf{x}_{kN_s}^T, \boldsymbol{\lambda}_{k1}^T, \cdots, \boldsymbol{\lambda}_{kN_r}^T\right]^T. \tag{15}$$

Here $\mathbf{b}_k$ in (15) is the vector of unknowns that need to be estimated at the destination node.

We apply a linear estimator at the destination node to estimate $\mathbf{b}_k$ due to its simplicity. The estimated $\mathbf{b}_k$ is given by

$$\hat{\mathbf{b}}_k = \mathbf{W}_k^H \mathbf{y}_k, \qquad k = 1, \cdots, N_d \tag{16}$$

where $\mathbf{W}_k$ is the weight matrix of the linear receiver and $(\cdot)^H$ denotes the matrix (vector) Hermitian transpose. From (14), we need $L \geq N_s(2Q+1) + N_r(Q+1)$ since a linear estimator is used. From (13) and (16), the MSE of estimating $\mathbf{b}_k$, $k = 1, \cdots, N_d$, is given by

$$\begin{aligned} \mathrm{MSE} &= \sum_{k=1}^{N_d} \mathrm{tr}\left(\mathrm{E}\left[\left(\hat{\mathbf{b}}_k - \mathbf{b}_k\right)\left(\hat{\mathbf{b}}_k - \mathbf{b}_k\right)^H\right]\right) \\ &= \sum_{k=1}^{N_d} \mathrm{tr}\left(\mathrm{E}\left[\left(\mathbf{W}_k^H\mathbf{A} - \mathbf{I}_D\right)\mathbf{C}_{b,k}\left(\mathbf{W}_k^H\mathbf{A} - \mathbf{I}_D\right)^H \right.\right. \\ &\qquad \left.\left. + \mathbf{W}_k^H \mathbf{C}_{\bar{v},k} \mathbf{W}_k\right]\right) \end{aligned} \tag{17}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace, $\mathrm{E}[\cdot]$ stands for the statistical expectation, $\mathbf{I}_n$ represents the $n \times n$ identity matrix, $D \triangleq N_s(2Q+1) + N_r(Q+1)$, $\mathbf{C}_{b,k} \triangleq \mathrm{E}\left[\mathbf{b}_k\mathbf{b}_k^H\right]$ is the covariance matrix of $\mathbf{b}_k$, and $\mathbf{C}_{\bar{v},k} \triangleq \mathrm{E}\left[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^H\right]$ is the noise covariance matrix.

To find $\mathbf{C}_{b,k}$, we first compute the covariance matrix of $\mathbf{x}_{km}$ using (6) and (10) as

$$\begin{aligned} \mathbf{C}_{b,k}^{x,m} &= \mathrm{E}[\mathbf{x}_{km}\mathbf{x}_{km}^H] \\ &= \sum_{n=1}^{N_r} \mathrm{diag}\left[\boldsymbol{\sigma}_{k,n}^{\lambda} * \boldsymbol{\sigma}_{n,m}^{\mu}\right], \quad m = 1, \cdots, N_s \end{aligned} \tag{18}$$

where $\mathbf{a} * \mathbf{b}$ denotes the linear convolution between vectors $\mathbf{a}$ and $\mathbf{b}$, $\boldsymbol{\sigma}_{k,n}^{\lambda} = [\sigma_{k,n}^{\lambda,0}, \cdots, \sigma_{k,n}^{\lambda,Q}]^T$, $\boldsymbol{\sigma}_{n,m}^{\mu} = [\sigma_{n,m}^{\mu,0}, \cdots, \sigma_{n,m}^{\mu,Q}]^T$, and $\sigma_{k,n}^{\lambda,q} = \mathrm{E}[\lambda_{kn}(q)\lambda_{kn}^*(q)]$ and $\sigma_{n,m}^{\mu,q} = \mathrm{E}[\mu_{nm}(q)\mu_{nm}^*(q)]$ are the variances of $\lambda_{kn}(q)$ and $\mu_{nm}(q)$, respectively. Here, $(\cdot)^*$ denotes complex conjugate. Similarly, from (9), the covariance matrix of $\boldsymbol{\lambda}_{kn}$ can be written as

$$\begin{aligned} \mathbf{C}_{b,k}^{\lambda,n} &= \mathrm{E}[\boldsymbol{\lambda}_{kn}\boldsymbol{\lambda}_{kn}^H] \\ &= \mathrm{diag}\left[\sigma_{k,n}^{\lambda,0}, \cdots, \sigma_{k,n}^{\lambda,Q}\right], \quad n = 1, \cdots, N_r. \end{aligned} \tag{19}$$

Based on (18) and (19), the covariance matrix $\mathbf{C}_{b,k}$ can be rewritten as

$$\mathbf{C}_{b,k} = \mathrm{bd}\left[\mathbf{C}_{b,k}^{x}, \mathbf{C}_{b,k}^{\lambda}\right], \qquad k = 1, \cdots, N_d$$

where $\mathrm{bd}[\cdot]$ represents a block diagonal matrix and

$$\mathbf{C}_{b,k}^{x} = \mathrm{bd}\left[\mathbf{C}_{b,k}^{x,1}, \cdots, \mathbf{C}_{b,k}^{x,N_s}\right], \quad \mathbf{C}_{b,k}^{\lambda} = \mathrm{bd}\left[\mathbf{C}_{b,k}^{\lambda,1}, \cdots, \mathbf{C}_{b,k}^{\lambda,N_r}\right]. \tag{20}$$

Since all noises are assumed to be i.i.d. with zero mean and unit variance, from (8) we have

$$\mathbf{C}_{\bar{v},k} = \left(\alpha \sum_{n=1}^{N_r} \sigma_{k,n}^{g} + 1\right)\mathbf{I}_L, \qquad k = 1, \cdots, N_d$$

where $\sigma_{k,n}^{g} = \mathrm{E}[g_{k,n}(t_l)g_{k,n}^*(t_l)]$ is the variance of $g_{k,n}(t_l)$, $l = 1, \cdots, L$.

The optimal weight matrices $\mathbf{W}_k$, $k = 1, \cdots, N_d$ that minimize the MSE in (17) are given by

$$\mathbf{W}_k = \left(\mathbf{A}\mathbf{C}_{b,k}\mathbf{A}^H + \mathbf{C}_{\bar{v},k}\right)^{-1}\mathbf{A}\mathbf{C}_{b,k}, \quad k = 1, \cdots, N_d \tag{21}$$

where $(\cdot)^{-1}$ stands for matrix inversion. Substituting (21) back into (17), the MSE of channel estimation is given by

$$\mathrm{MSE} = \sum_{k=1}^{N_d} \mathrm{tr}\left(\left[\mathbf{C}_{b,k}^{-1} + \mathbf{A}^H\mathbf{C}_{\bar{v},k}^{-1}\mathbf{A}\right]^{-1}\right). \tag{22}$$

The transmission power constraint at the source node can be written as

$$\sum_{m=1}^{N_s} \mathbf{s}_m^H \mathbf{s}_m \leq P_s \tag{23}$$

where $\mathbf{s}_m \triangleq [s_m(t_1), s_m(t_2), \cdots, s_m(t_L)]^T$ and P_s is the transmission power available at the source node. The transmission power constraint at the relay node is given by

$$\alpha \sum_{m=1}^{N_s}\sum_{n=1}^{N_r} \sigma_{n,m}^{h}\mathbf{s}_m^H\mathbf{s}_m + \sum_{n=1}^{N_r} \mathbf{r}_n^H\mathbf{r}_n + \alpha N_r L \leq P_r \tag{24}$$

where $\sigma_{n,m}^{h} = \mathrm{E}[h_{n,m}(t_l)h_{n,m}^*(t_l)]$ is the variance of $h_{n,m}(t_l)$, $l = 1, \cdots, L$, P_r is the transmission power available at the relay node, and $\mathbf{r}_n \triangleq [r_n(t_1), \cdots, r_n(t_L)]^T$.

The optimal structures of $\{\mathbf{s}_m\} \triangleq \{\mathbf{s}_m, m = 1, \cdots, N_s\}$ and $\{\mathbf{r}_n\} \triangleq \{\mathbf{r}_n, n = 1, \cdots, N_r\}$ that minimize (22) subjected to the power constraints in (23) and (24) are derived in the following theorem.

THEOREM 1: The optimal training matrices $\{\mathbf{s}_m\}$ and $\{\mathbf{r}_n\}$ satisfy the following equations for all $m, p = 1, \cdots, N_s$ and $n, q = 1, \cdots, N_r$

$$(\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{S}_m\boldsymbol{\Theta} = \beta_m\mathbf{I}_{2Q+1} \tag{25}$$
$$(\mathbf{R}_n\boldsymbol{\Phi})^H\mathbf{R}_n\boldsymbol{\Phi} = \gamma_n\mathbf{I}_{Q+1} \tag{26}$$
$$(\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{S}_p\boldsymbol{\Theta} = \mathbf{0}, \quad m \neq p \tag{27}$$
$$(\mathbf{R}_n\boldsymbol{\Phi})^H\mathbf{R}_q\boldsymbol{\Phi} = \mathbf{0}, \quad n \neq q \tag{28}$$
$$(\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{R}_n\boldsymbol{\Phi} = \mathbf{0} \tag{29}$$

where $\beta_m = \mathbf{s}_m^H\mathbf{s}_m$, $m = 1, \cdots, N_s$, and $\gamma_n = \mathbf{r}_n^H\mathbf{r}_n$, $n = 1, \cdots, N_r$.

PROOF: Let us define $\boldsymbol{\Psi}_S \triangleq [\mathbf{S}_1\boldsymbol{\Theta}, \mathbf{S}_2\boldsymbol{\Theta}, \cdots, \mathbf{S}_{N_s}\boldsymbol{\Theta}]$ and $\boldsymbol{\Psi}_R \triangleq [\mathbf{R}_1\boldsymbol{\Phi}, \mathbf{R}_2\boldsymbol{\Phi}, \cdots, \mathbf{R}_{N_r}\boldsymbol{\Phi}]$. From (22), the MSE can be rewritten as

$$\text{MSE} = \sum_{k=1}^{N_d} \text{tr}\left(\left[\begin{pmatrix}\mathbf{C}_{b,k}^x & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{b,k}^\lambda\end{pmatrix}^{-1} + \eta_k\begin{pmatrix}\sqrt{\alpha}\boldsymbol{\Psi}_S^H\\ \boldsymbol{\Psi}_R^H\end{pmatrix}\left(\sqrt{\alpha}\boldsymbol{\Psi}_S, \boldsymbol{\Psi}_R\right)\right]^{-1}\right) \tag{30}$$

where

$$\eta_k \triangleq \left(\alpha\sum_{n=1}^{N_r}\sigma_{k,n}^g + 1\right)^{-1}, \quad k = 1, \cdots, N_d. \tag{31}$$

From (18)-(20), $\mathbf{C}_{b,k}^x$ and $\mathbf{C}_{b,k}^\lambda$ are all diagonal matrices. Hence, the MSE in (30) is minimized only if all off-diagonal matrices of the second term are zero, i.e.,

$$\boldsymbol{\Psi}_S^H\boldsymbol{\Psi}_S = \mathbf{D}_s, \qquad \boldsymbol{\Psi}_R^H\boldsymbol{\Psi}_R = \mathbf{D}_r, \qquad \boldsymbol{\Psi}_S^H\boldsymbol{\Psi}_R = \mathbf{0}$$

where $\mathbf{D}_s \triangleq \text{diag}[\mathbf{D}_{s,1}, \mathbf{D}_{s,2}, \cdots, \mathbf{D}_{s,N_s}]$ and $\mathbf{D}_r \triangleq \text{diag}[\mathbf{D}_{r,1}, \mathbf{D}_{r,2}, \cdots, \mathbf{D}_{r,N_r}]$ are diagonal matrices, and $\mathbf{0}$ denotes a zero matrix. From the definition of $\boldsymbol{\Psi}_S$ and $\boldsymbol{\Psi}_R$, we have that for $m, p = 1, \cdots, N_s$ and $n, q = 1, \cdots, N_r$,

$$\begin{aligned}&(\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{S}_m\boldsymbol{\Theta} = \mathbf{D}_{s,m}, && (\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{S}_p\boldsymbol{\Theta} = \mathbf{0}, \ m \neq p\\ &(\mathbf{R}_n\boldsymbol{\Phi})^H\mathbf{R}_n\boldsymbol{\Phi} = \mathbf{D}_{r,n}, && (\mathbf{R}_n\boldsymbol{\Phi})^H\mathbf{R}_q\boldsymbol{\Phi} = \mathbf{0}, \ n \neq q\\ &(\mathbf{S}_m\boldsymbol{\Theta})^H\mathbf{R}_n\boldsymbol{\Phi} = \mathbf{0}.\end{aligned}$$

Then we can rewrite the MSE in (30) as

$$\text{MSE} = \sum_{k=1}^{N_d}\text{tr}\left(\sum_{m=1}^{N_s}\left[(\mathbf{C}_{b,k}^{x,m})^{-1} + \alpha\eta_k\mathbf{D}_{s,m}\right]^{-1} + \sum_{n=1}^{N_r}\left[(\mathbf{C}_{b,k}^{\lambda,n})^{-1} + \eta_k\mathbf{D}_{r,n}\right]^{-1}\right). \tag{32}$$

Moreover, due to the structure of $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ in (11), we have $\mathbf{D}_{s,m} = \beta_m\mathbf{I}_{2Q+1}$ and $\mathbf{D}_{r,n} = \gamma_n\mathbf{I}_{Q+1}$. This indicates that

$$\mathbf{s}_m^H\mathbf{s}_m = \beta_m, \ m = 1, \cdots, N_s, \qquad \mathbf{r}_n^H\mathbf{r}_n = \gamma_n, \ n = 1, \cdots, N_r. \tag{33}$$

We would like to note that the transmission power constraints in (23) and (24) are not affected by (33). □

Using Theorem 1 and (32), the MSE function in (22) can be written as

$$\text{MSE} = \sum_{k=1}^{N_d}\text{tr}\left(\sum_{m=1}^{N_s}\left[(\mathbf{C}_{b,k}^{x,m})^{-1} + \alpha\beta_m\eta_k\mathbf{I}_{2Q+1}\right]^{-1} + \sum_{n=1}^{N_r}\left[(\mathbf{C}_{b,k}^{\lambda,n})^{-1} + \gamma_n\eta_k\mathbf{I}_{Q+1}\right]^{-1}\right) \tag{34}$$

The optimization problem can be written as the following problem in scalar variables

$$\min_{\{\beta_m\},\{\gamma_n\},\alpha} \sum_{m=1}^{N_s}\sum_{k=1}^{N_d}\sum_{q=1}^{2Q+1}\frac{1}{c_{k,m,q} + \alpha\beta_m\eta_k} + \sum_{n=1}^{N_r}\sum_{k=1}^{N_d}\sum_{p=1}^{Q+1}\frac{1}{d_{k,n,p} + \gamma_n\eta_k} \tag{35}$$

$$\text{s.t.} \ \sum_{m=1}^{N_s}\beta_m \leq P_s \tag{36}$$

$$\alpha\sum_{m=1}^{N_s}\kappa_m\beta_m + \sum_{n=1}^{N_r}\gamma_n + \alpha N_r L \leq P_r \tag{37}$$

$$\alpha > 0, \quad \beta_m \geq 0, \ m = 1, \cdots, N_s$$
$$\gamma_n \geq 0, \ n = 1, \cdots, N_r \tag{38}$$

where $c_{k,m,q} \triangleq [(\mathbf{C}_{b,k}^{x,m})^{-1}]_{q,q}$, $d_{k,n,p} \triangleq [(\mathbf{C}_{b,k}^{\lambda,n})^{-1}]_{p,p}$, $[\cdot]_{j,j}$ denotes the jth diagonal element of a matrix, $\{\beta_m\} \triangleq \{\beta_m, m = 1, \cdots, N_s\}$, $\{\gamma_n\} \triangleq \{\gamma_n, n = 1, \cdots, N_r\}$, and $\kappa_m \triangleq \sum_{n=1}^{N_r}\sigma_{n,m}^h$.

As $c_{k,m,q}$, $d_{k,n,p}$, and η_k are known variables, when α is fixed, it can be seen that the first and second set of fractions in the objective function (35) are monotonically decreasing and convex with respect to β_m and γ_n, respectively. Moreover, the constraints in (36)-(38) are linear inequality constraints when α is fixed. We can conclude that the problem (35)-(38) is a convex optimization problem with respect to β_m and γ_n when α is fixed. The optimal β_m and γ_n can be efficiently derived using the Karush-Kuhn-Tucker (KKT) optimality conditions [12] of the problem (35)-(38).

When α is not fixed and needs to be optimized, the problem (35)-(38) as a whole is not a convex optimization problem. However, it can be proven that (35) is a unimodal function of α. The proof is omitted due to space limit. The minimum value of a unimodal function can be efficiently found by using the golden section search (GSS) method [13].

IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed superimposed channel training algorithm for time-varying MIMO relay systems through numerical simulations. We consider a three-node MIMO relay system where all nodes are equipped with the same number of antennas, i.e., $N_s = N_r = N_d = N$, and the source and relay nodes have the same transmission power, i.e., $P_s = P_r = P$. For simplicity, it is assumed that all BEM coefficients have unit variance. We use the shortest training sequence possible, i.e.,

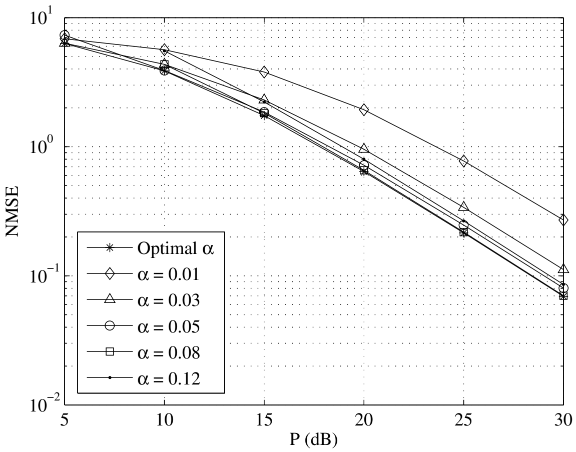

Fig. 2. Example 1: NMSE versus P for different α with $N = 2$ and $Q = 4$.

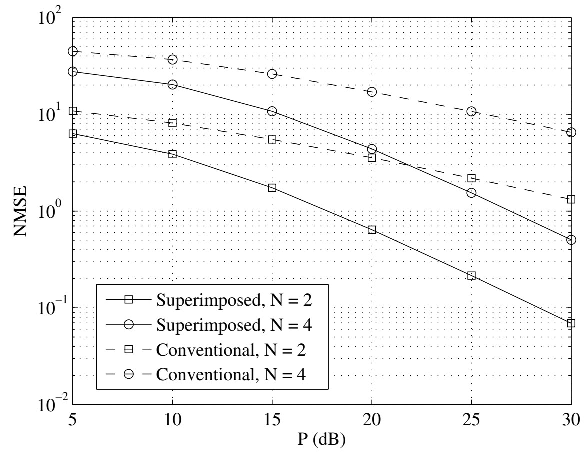

Fig. 3. Example 2: NMSE versus P for different N with $Q = 4$.

$L = N(3Q + 2)$. More training sequences can be transmitted to improve the performance of channel estimation. For all scenarios, we compute the NMSE of channel estimation at the destination node.

In the first example, we study the impact of α on the performance of the proposed superimposed channel training algorithm. Fig. 2 shows the NMSE of the proposed algorithm versus P for different α when $N = 2$ and $Q = 4$. We apply the GSS technique in the superimposed channel training algorithm to obtain the optimal α for different P. It can be observed from Fig. 2 that the optimal α curve always has the lowest MSE value for all P, and hence proves the efficiency of GSS method in obtaining the optimal α for different P. Note that the starting point for the curves associated with $\alpha = 0.08$ and $\alpha = 0.12$ is at $P = 10$dB, as these values of α have exceeded the upper bound limit of α for $P = 5$dB.

In the second example, we compare the performance of the proposed algorithm with the conventional two-stage MMSE-based channel estimation algorithm [6]. We investigate the performance of the two algorithms at various number of antennas. Fig. 3 shows the NMSE performance of both algorithms versus P for different N when $Q = 4$. As expected, the NMSE of the channel estimation increases when the number of antenna increases due to a larger number of unknowns to be estimated. It can also be observed from Fig. 3 that the proposed algorithm consistently has better NMSE performance compared with the conventional two-stage channel estimation algorithm, especially at high P level.

V. Conclusions

We have developed a superimposed training algorithm to estimate the time-varying channels in two-hop MIMO relay communication systems. The proposed algorithm can efficiently estimate the individual first-hop and second-hop CSI for MIMO relay communication systems. The optimal structure of the source and relay training sequences and the optimal power allocation between the source and relay training sequences are derived.

Acknowledgment

This research was supported under the Australian Research Council's Discovery Projects funding scheme (project numbers DP110102076, DP140102131).

References

[1] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, 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, no. 4, pp. 1398-1407, Apr. 2007.

[3] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on transceiver design for amplify-and-forward MIMO relay systems," *IEEE J. Select. Areas Commun.*, vol. 30, no. 8, pp. 1331-1346, Sep. 2012.

[4] M. R. A. Kandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544-3555, Jul. 2013.

[5] 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.

[6] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4438-4446, Sep. 2011.

[7] C. S. Patel and G. L. Stuber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2348-2356, Aug. 2007.

[8] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351-1366, Apr. 2003.

[9] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969-1986, Aug. 2002.

[10] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-way relay networks in time-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681-2691, June 2011.

[11] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686-693, Nov. 1991.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U. K.: Cambridge University Press, 2004.

[13] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Spring Street, NY: Springer Science+Business Media, LCC, 2007.