2011 17th Asia-Pacific Conference on Communications (APCC)
2nd – 5th October 2011 | Sutera Harbour Resort, Kota Kinabalu, Sabah, Malaysia

Channel Estimation of Dual-Hop MIMO Relay Systems Using Parallel Factor Analysis

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

Muhammad R. A. Khandaker
Dept. of Electrical and Computer Engineering
Curtin University
Bentley, WA, 6102, Australia
Email: m.khandaker@postgrad.curtin.edu.au

***Abstract*—The optimal source precoding matrix and relay amplifying matrix have been developed in recent works on multiple-input multiple-output (MIMO) relay communication systems assuming that the instantaneous channel state information (CSI) is available. However, in practical relay communication systems, the instantaneous CSI is unknown, and therefore, has to be estimated at the destination node. In this paper, we develop a novel channel estimation algorithm for two-hop MIMO relay systems using the parallel factor (PARAFAC) analysis. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. Compared with existing approaches, the proposed algorithm requires less number of training data blocks, and is applicable for both one-way and two-way MIMO relay systems with single or multiple relay nodes. Numerical examples demonstrate the effectiveness of the PARAFAC-based channel estimation algorithm.**

I. Introduction

Recently, there have been many research efforts on multiple-input multiple-output (MIMO) relay systems [1]-[6]. For a three-node two-hop MIMO relay system where the direct source-destination link is omitted, the optimal relay amplifying matrix is obtained in [2]-[3] to maximize the mutual information between source and destination. In [4], optimal relay matrices are developed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination node for a two-hop MIMO relay system with multiple parallel relay nodes. A unified framework is established for optimizing the source precoding matrix and the relay amplifying matrix of two-hop linear non-regenerative MIMO relay systems with a broad class of objective functions [5]. Recently, it has been shown in [6] that by using a nonlinear decision feedback equalizer (DFE) based on the minimal MSE (MMSE) criterion at the destination node, the system bit-error-rate (BER) can be significantly reduced.

For the aforementioned MIMO relay systems, the instantaneous channel state information (CSI) knowledge of both the source-relay link and the relay-destination link is required at the destination node to estimate the source signals. Moreover, in order to optimize the source and/or relay matrices in [1]-[6], the instantaneous CSI knowledge of both links is needed to carry out the optimization procedure. However, in practical relay communication systems, the instantaneous CSI is unknown, and therefore, has to be estimated. Recently, a tensor-based channel estimation algorithm is developed in [7] for a two-way MIMO relay system. It is obvious that the algorithm in [7] can not be applied in one-way MIMO relay systems. In [8], a relay channel estimation algorithm using the least-squares (LS) fitting is proposed. However, the number of training data blocks required in [8] is at least equal to the number of relay nodes, which is not spectral efficient. For amplify-and-forward relay networks with single-antenna source, relay, and destination nodes, the optimal training sequence is developed in [9].

In this paper, we propose a novel channel estimation algorithm for two-hop MIMO relay systems using the parallel factor (PARAFAC) analysis [10]. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. Compared with [8], the proposed algorithm requires only two training data blocks in many scenarios, and hence, has a higher spectral efficiency. In contrast to [7], the proposed algorithm is applicable for both one-way and two-way relay systems with single or multiple relay nodes. In particular, for the direct source-destination link in one-way relay systems, the MIMO channel matrix is estimated by the LS approach. For the source-relay-destination link in both one-way and two-way relay systems, we show that under a mild condition of the channel training data block length, the MIMO channel matrices of both hops can be estimated up to permutation and scaling ambiguities, which are inherent to the PARAFAC model. To remove the permutation ambiguity, we exploit the knowledge of relay factors during the channel training period which is designed beforehand and known by both the relay nodes and the destination node. Then by using a bilinear alternating least-squares (BALS) algorithm, the channel matrix of each hop can be estimated up to some scaling ambiguity. Finally, the scaling ambiguity is taken care of by transmitting one complex number from each relay antenna to the destination node over the control channel available in a wireless network. This number contains the channel state between the first antenna at the source node and this relay antenna.

Since during the training period, the noise at the relay nodes is amplified and forwarded to the destination node, the effective noise vector at the destination node is non-white. Taking this fact into account, we propose a weighted least-squares (WLS) approach to further improve the estimate of the source-relay channel, by exploiting the initial estimate of the relay-destination channel. Numerical examples demonstrate the effectiveness of the proposed PARAFAC-based channel estimation algorithm and the WLS approach.

The rest of this paper is organized as follows. In Section II, we introduce the model of a two-hop amplify-and-forward MIMO relay communication system. The proposed channel estimation algorithm is developed in Sections III. In Section IV, we show some numerical examples. Conclusions are drawn in Section V.

II. System Model

We consider a two-hop MIMO communication system where the source node transmits information to the destination node with the aid of M relay nodes as shown in Fig. 1. The source node and the destination node are equipped with $N_s \geq 2$ and $N_d \geq 2$ antennas, respectively. Each relay node has one or multiple antennas. For the simplicity of explanation, we assume that each relay node has one antenna. It will be shown later that the proposed channel estimation algorithm can be extended to the case where each relay node has (different number of) multiple antennas. Considering the practical half-duplex constraint at each relay node, the communication process between the source and destination nodes is completed in two time slots. In the first time slot, the $N_s \times 1$ modulated signal vector $\mathbf{x}_s(t)$ is transmitted to all relay nodes and the destination node, and the received signal vectors are respectively given by

$$\mathbf{y}_r(t) = \mathbf{H}_{sr}\mathbf{x}_s(t) + \mathbf{v}_r(t), \qquad \mathbf{y}_d(t) = \mathbf{H}_{sd}\mathbf{x}_s(t) + \mathbf{v}_d(t) \tag{1}$$

where $\mathbf{y}_r(t)$ is an $M \times 1$ vector stacking the received signals at all relay nodes on top of each other, $\mathbf{y}_d(t)$ is an $N_d \times 1$ received signal vector at the destination node, $\mathbf{H}_{sr}$ is the $M \times N_s$ MIMO fading channel matrix between the source node and all relay nodes, $\mathbf{H}_{sd}$ is the $N_d \times N_s$ MIMO source-destination channel matrix, $\mathbf{v}_r(t)$ is an $M \times 1$ vector stacking the noise at all relay nodes on top of each other, and $\mathbf{v}_d(t)$ is the $N_d \times 1$ noise vector at the destination node.

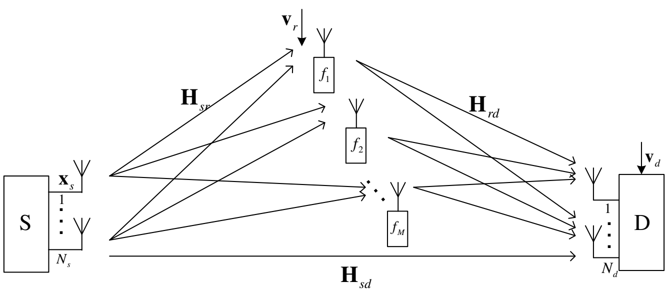

Fig. 1. Two-Hop MIMO relay system with M relay nodes.

In the second time slot, the source node is silent, and each relay node amplifies the received signal with f_i and forwards the amplified signal to the destination node. The received signal vector at the destination node is

$$\mathbf{y}_d(t+1) = \mathbf{H}_{rd}\mathbf{D}_f\mathbf{H}_{sr}\mathbf{x}_s(t) + \mathbf{H}_{rd}\mathbf{D}_f\mathbf{v}_r(t) + \mathbf{v}_d(t+1) \tag{2}$$

where $\mathbf{H}_{rd}$ is the $N_d \times M$ MIMO fading channel matrix between the destination node and all relay nodes, $\mathbf{D}_f \triangleq \mathrm{diag}(f_1, f_2, \cdots, f_M)$, and $\mathbf{v}_d(t+1)$ is an $N_d \times 1$ noise vector at the destination node at time $t+1$. Here $\mathrm{diag}(\cdot)$ stands for a diagonal matrix. We assume that all noises are complex circularly symmetric with zero mean and unit variance. We also assume that $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$ are quasi-static block fading which means they are constant over some time interval before changing to another realization. Combining (1) and (2), the received signals at the destination node over two time slots are given by

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{D}_f\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} \mathbf{x}_s(t) + \begin{bmatrix} \mathbf{H}_{rd}\mathbf{D}_f\mathbf{v}_r(t) + \mathbf{v}_d(t+1) \\ \mathbf{v}_d(t) \end{bmatrix}. \tag{3}$$

Due to its lower computational complexity, linear receiver is used at the destination node to retrieve the transmitted signal vector $\mathbf{x}_s(t)$ [2]-[6]. The estimated signal waveform vector is given by $\hat{\mathbf{x}}_s(t) = \mathbf{W}^H\mathbf{y}(t)$, where $\mathbf{W}$ is the $2N_d \times N_s$ weight matrix. From (3), the MSE of the signal waveform estimation can be written as

$$e = \mathrm{tr}\Big(\mathrm{E}\big[(\hat{\mathbf{x}}_s(t) - \mathbf{x}_s(t))(\hat{\mathbf{x}}_s(t) - \mathbf{x}_s(t))^H\big]\Big) \tag{4}$$

where $\mathrm{E}[\cdot]$ stands for statistical expectation, $\mathrm{tr}(\cdot)$ and $(\cdot)^H$ denote matrix trace, and matrix Hermitian transpose, respectively. The receiver weight matrix which minimizes (4) is the Wiener filter given by [11]

$$\mathbf{W} = \left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \bar{\mathbf{C}}\right)^{-1}\bar{\mathbf{H}} \tag{5}$$

where

$$\bar{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H}_{rd}\mathbf{D}_f\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix}, \quad \bar{\mathbf{C}} \triangleq \begin{bmatrix} \mathbf{H}_{rd}\mathbf{D}_f\mathbf{D}_f^H\mathbf{H}_{rd}^H + \mathbf{I}_{N_d} & \mathbf{0}_{N_d \times N_d} \\ \mathbf{0}_{N_d \times N_d} & \mathbf{I}_{N_d} \end{bmatrix}.$$

Here $(\cdot)^{-1}$ stands for the matrix inversion, $\mathbf{0}_{m \times n}$ denotes an $m \times n$ matrix with all zeros entries, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix.

It can be clearly seen from (5) that in order to compute $\mathbf{W}$, the CSI knowledge of $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$ is required at the destination node. In the following, we develop a PARAFAC analysis based algorithm to estimate all channel matrices ($\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, $\mathbf{H}_{sd}$) at the destination node.

III. Proposed Channel Estimation Algorithm

In order to estimate the channel matrices, training symbols are transmitted from the source node. The overall channel training period is divided into K time blocks (the minimal K required will be determined later). In each time block, the same $N_s \times L$ orthogonal channel training sequence $\mathbf{S}$ with $\mathbf{S}\mathbf{S}^H = \mathbf{I}_{N_s}$ is transmitted by the source node, where L ($L \geq N_s$) is the length of each block. Such $\mathbf{S}$ is optimal in terms of the MSE of channel estimation [12] and can be easily constructed, for example, from the normalized discrete

Fourier transform (DFT) matrix. In the kth time block, the mth relay node amplifies the received signal with $f_{k,m}$ and forwards the amplified signal to the destination node. From (3), the received signal matrices at the destination node over K time blocks are given by

$$\mathbf{Y}_k \triangleq \begin{bmatrix} \mathbf{Y}_k^{(1)} \\ \mathbf{Y}_k^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix}\mathbf{S} + \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{d,k}^{(1)} \\ \mathbf{V}_{d,k}^{(2)} \end{bmatrix} \quad k = 1, \cdots, K \tag{6}$$

where $\mathbf{F}$ is a $K \times M$ matrix whose kth row contains the amplifying factors of all M relay nodes at the kth time block, $\mathcal{D}_k\{\cdot\}$ is the operator that makes a diagonal matrix by selecting the kth row and putting it on the main diagonal while putting zeros elsewhere, $\mathbf{V}_{r,k}$ is the $M \times L$ noise matrix at the relay nodes during the kth time block, $\mathbf{V}_{d,k}^{(i)}$, $i = 1, 2$, is the $N_d \times L$ noise matrix at the destination node at the ith time slot during the kth time block, and $\mathbf{Y}_k^{(1)}$ and $\mathbf{Y}_k^{(2)}$, $k = 1, \cdots, K$, are matrices containing the first and the last N_d rows of $\mathbf{Y}_k$, respectively.

At the destination node, by multiplying both sides of (6) with $\mathbf{S}^H$, we obtain

$$\mathbf{Y}_k\mathbf{S}^H = \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}^H + \mathbf{V}_{d,k}^{(1)}\mathbf{S}^H \\ \mathbf{V}_{d,k}^{(2)}\mathbf{S}^H \end{bmatrix}, k = 1, \cdots, K. \tag{7}$$

From (7), an LS estimate of $\mathbf{H}_{sd}$ is given by

$$\hat{\mathbf{H}}_{sd} = \frac{1}{K}\mathbf{Y}^{(2)}(\mathbf{1}_K \otimes \mathbf{S})^H = \mathbf{H}_{sd} + \frac{1}{K}\mathbf{V}_d^{(2)}(\mathbf{1}_K \otimes \mathbf{S})^H$$

where $\mathbf{Y}^{(2)} \triangleq \left[\mathbf{Y}_1^{(2)}, \mathbf{Y}_2^{(2)}, \cdots, \mathbf{Y}_K^{(2)}\right]$, $\mathbf{V}_d^{(2)} \triangleq \left[\mathbf{V}_{d,1}^{(2)}, \mathbf{V}_{d,2}^{(2)}, \cdots, \mathbf{V}_{d,K}^{(2)}\right]$, $\mathbf{1}_K$ denotes a $1 \times K$ vector with all 1 elements, and $\otimes$ stands for the Kronecker matrix product [13]. In the following, we show how to estimate $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$ at the destination node.

A. PARAFAC model and identifiability of channel matrices

Let us introduce

$$\tilde{\mathbf{X}}_k \triangleq \mathbf{Y}_k^{(1)}\mathbf{S}^H = \mathbf{X}_k + \mathbf{V}_k, \quad k = 1, \cdots, K \tag{8}$$

$$\mathbf{X}_k \triangleq \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{sr}, \quad k = 1, \cdots, K \tag{9}$$

$$\mathbf{V}_k \triangleq \mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}^H + \mathbf{V}_{d,k}^{(1)}\mathbf{S}^H, k = 1, \cdots, K \tag{10}$$

where $\mathbf{X}_k$ is the matrix-of-interest containing both $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$, $\mathbf{V}_k$ is the effective noise matrix, and $\tilde{\mathbf{X}}_k$ is a noisy observation of $\mathbf{X}_k$. We would like to mention that $\mathbf{F}$ is designed beforehand and is known at the destination node. By assembling the set of K matrices (9) together along the direction of the index k (the third dimension), we obtain an $N_d \times N_s \times K$ three-way array $\underline{\mathbf{X}}$, whose (i, j, k)-th element is given by

$$x(i, j, k) = \sum_{m=1}^{M} h_{rd}(i, m) f(k, m) h_{sr}(m, j) \tag{11}$$

for all $i = 1, \cdots, N_d$, $j = 1, \cdots, N_s$, and $k = 1, \cdots, K$. Here $h_{rd}(i, m)$, $f(k, m)$, and $h_{sr}(m, j)$ stand for the (i, m)-th, (k, m)-th, and (m, j)-th elements of $\mathbf{H}_{rd}$, $\mathbf{F}$, and $\mathbf{H}_{sr}$, respectively. Equation (11) expresses $x(i, j, k)$ as a sum of M rank-1 triple products, which is known as the trilinear decomposition, or PARAFAC analysis of $x(i, j, k)$ [10]. Correspondingly, assembling K matrices of $\tilde{\mathbf{X}}_k$ in (8) along the index k leads to a noise-contaminated $\underline{\mathbf{X}}$ given by $\underline{\tilde{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{V}}$.

It can be shown by using the identifiability theorem of the PARAFAC model in [10] and [14] that if

$$k_{\mathbf{H}_{rd}} + k_{\mathbf{F}} + k_{\mathbf{H}_{sr}} \geq 2M + 2 \tag{12}$$

then the triple $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$ is unique up to permutation and scaling ambiguities, i.e., if there exists any other triple $(\bar{\mathbf{H}}_{rd}, \bar{\mathbf{F}}, \bar{\mathbf{H}}_{sr})$ that gives rise to (9), then it is related to $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$ via

$$\bar{\mathbf{H}}_{rd} = \mathbf{H}_{rd}\mathbf{\Pi}\mathbf{\Delta}_1, \quad \bar{\mathbf{F}} = \mathbf{F}\mathbf{\Pi}\mathbf{\Delta}_2, \quad \bar{\mathbf{H}}_{sr}^T = \mathbf{H}_{sr}^T\mathbf{\Pi}\mathbf{\Delta}_3 \tag{13}$$

where $(\cdot)^T$ stands for matrix (vector) transpose, $\mathbf{\Pi}$ is an $M \times M$ permutation matrix, and $\mathbf{\Delta}_i$, $i = 1, 2, 3$, are $M \times M$ diagonal (complex) scaling matrices satisfying

$$\mathbf{\Delta}_1\mathbf{\Delta}_2\mathbf{\Delta}_3 = \mathbf{I}_M. \tag{14}$$

In (12), for a matrix $\mathbf{C}$, $k_{\mathbf{C}}$ denotes the Kruskal rank (or k-rank) [14] of $\mathbf{C}$.

Inequality (12) establishes the sufficient condition for the identifiability of $(\mathbf{H}_{rd}, \mathbf{F}, \mathbf{H}_{sr})$. Since $\mathbf{F}$ is designed beforehand (e.g., based on the DFT matrix), one can guarantee that $\mathbf{F}$ has full k-rank. Moreover, both $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ are random matrices, and hence have full k-rank. Therefore, in such case, condition (12) becomes

$$\min(N_d, M) + \min(K, M) + \min(N_s, M) \geq 2M + 2. \tag{15}$$

From (15), the identifiability condition can be summarized as follows:

THEOREM: The PARAFAC model (11) is identifiable if $N_s \geq 2$, $N_d \geq 2$, and $2 \leq M \leq N_s + N_d - 2$. Moreover, K should satisfy the following:

$$K \geq \begin{cases} 2M + 2 - N_s - N_d & M \geq N_s, N_d \\ M + 2 - N_d & N_d \leq M \leq N_s \\ M + 2 - N_s & N_s \leq M \leq N_d \\ 2 & M \leq N_s, N_d \end{cases}. \tag{16}$$

PROOF: It can be easily proven from (15) by expanding three $\min(\cdot)$ operations. □

Interestingly, it can be seen from (16) that if $N_d \geq M$ and $N_s \geq M$, then as less as two training data blocks ($K = 2$) is sufficient to estimate both $\mathbf{H}_{rd}$ and $\mathbf{H}_{sr}$ at the destination node. While in [8], $K = M$ training data blocks are required to perform the channel estimation. We also observe that if (16) is satisfied, then it holds that $KN_d > M$ and $KN_s > M$.

B. Bilinear alternating least-squares (BALS) fitting

Since $\mathbf{F}$ is known at the destination node, this information can be exploited to solve the permutation ambiguity $\mathbf{\Pi}$ and the scaling ambiguity $\mathbf{\Delta}_2$ in (13), i.e., $\mathbf{\Delta}_2 = \mathbf{I}_M$. In this subsection, we develop a BALS algorithm to estimate the channel $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ by carrying out the PARAFAC model fitting. First we show some rearrangements of three-way arrays $\underline{\mathbf{X}}$, $\underline{\mathbf{V}}$, and $\underline{\tilde{\mathbf{X}}}$ which will be used later.

By stacking K matrices of $\mathbf{X}_k$ in (9) on top of each other, we obtain

$$\mathbf{X} \triangleq \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{rd}\mathcal{D}_1\{\mathbf{F}\} \\ \vdots \\ \mathbf{H}_{rd}\mathcal{D}_K\{\mathbf{F}\} \end{bmatrix} \mathbf{H}_{sr} = (\mathbf{F} \odot \mathbf{H}_{rd})\mathbf{H}_{sr} \tag{17}$$

where $\odot$ stands for the Khatri-Rao (column-wise Kronecker) matrix product [13]. Correspondingly, stacking matrices $\tilde{\mathbf{X}}_k$ in (8) on top of each other gives rise to

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} + \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_K \end{bmatrix} = \mathbf{X} + \mathbf{V}. \tag{18}$$

By slicing $\underline{\mathbf{X}}$ perpendicular to the dimension of j, we obtain a set of N_s matrices $\mathbf{Z}_j = \mathbf{F}\mathcal{D}_j\{\mathbf{H}_{sr}^T\}\mathbf{H}_{rd}^T$, $j = 1, \cdots, N_s$. By stacking N_s matrices of $\mathbf{Z}_j$ on top of each other, we have

$$\mathbf{Z} \triangleq \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_{N_s} \end{bmatrix} = \begin{bmatrix} \mathbf{F}\mathcal{D}_1\{\mathbf{H}_{sr}^T\} \\ \vdots \\ \mathbf{F}\mathcal{D}_{N_s}\{\mathbf{H}_{sr}^T\} \end{bmatrix} \mathbf{H}_{rd}^T = (\mathbf{H}_{sr}^T \odot \mathbf{F})\mathbf{H}_{rd}^T. \tag{19}$$

Similarly, by slicing $\underline{\tilde{\mathbf{X}}}$ perpendicular to the dimension of j and stacking the resulting matrices on top of each other, we have

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_{N_s} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{N_s} \end{bmatrix} \tag{20}$$

where $\mathbf{W}_j$, $j = 1, \cdots, N_s$, are the slabs of $\underline{\mathbf{V}}$ along the dimension of j.

The BALS fitting starts at a random $\hat{\mathbf{H}}_{rd}$. In each iteration, we first update $\mathbf{H}_{sr}$ using the LS fitting with fixed $\mathbf{F}$ and $\hat{\mathbf{H}}_{rd}$. Using (17) and (18), we obtain an updated $\mathbf{H}_{sr}$ as

$$\hat{\mathbf{H}}_{sr} = \arg\min_{\mathbf{H}_{sr}} \left\|\tilde{\mathbf{X}} - (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})\mathbf{H}_{sr}\right\| = (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})^{\dagger}\, \tilde{\mathbf{X}} \tag{21}$$

where $\|\cdot\|$ denotes the matrix Frobenius norm, and for a matrix $\mathbf{C}$, $\mathbf{C}^{\dagger} \triangleq (\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H$. Then we update $\mathbf{H}_{rd}$ through the LS fitting with known $\mathbf{F}$ and $\hat{\mathbf{H}}_{sr}$, and obtain $\hat{\mathbf{H}}_{rd}$ using (19) and (20) as

$$\hat{\mathbf{H}}_{rd} = \arg\min_{\mathbf{H}_{rd}} \left\|\tilde{\mathbf{Z}} - (\hat{\mathbf{H}}_{sr}^T \odot \mathbf{F})\mathbf{H}_{rd}^T\right\| = \left[(\hat{\mathbf{H}}_{sr}^T \odot \mathbf{F})^{\dagger}\, \tilde{\mathbf{Z}}\right]^T. \tag{22}$$

Since the conditional update of matrices in (21) and (22) may either improve or maintain but cannot worsen the current LS fit, a monotonic convergence of the BALS procedure to (at least) a locally optimal solution follows directly from this observation.

C. Scaling ambiguity resolving

After the convergence of the BALS algorithm, the remaining scaling ambiguity $\mathbf{\Delta}_1$ ($\mathbf{\Delta}_3 = \mathbf{\Delta}_1^{-1}$ according to (14)) can be resolved by exploiting the estimation of $\mathbf{H}_{sr}$ at the relay nodes as follows. During the channel training period, the received signal vector at the mth relay node over K time blocks is given by

$$\mathbf{y}_m = \mathbf{h}_{sr,m}(\mathbf{1}_K \otimes \mathbf{S}) + \mathbf{v}_m, \qquad m = 1, \cdots, M \tag{23}$$

where $\mathbf{h}_{sr,m}$ is the mth row of $\mathbf{H}_{sr}$, and $\mathbf{v}_m$ is a $1 \times KL$ noise vector. An easily implementable LS estimate of $\mathbf{h}_{sr,m}$ can be obtained from (23) as $\tilde{\mathbf{h}}_{sr,m} = \frac{1}{K}\mathbf{y}_m(\mathbf{1}_K \otimes \mathbf{S})^H$. Then the first element of $\tilde{\mathbf{h}}_{sr,m}$, denoted as $\tilde{h}_{sr}(m,1)$ is transmitted from the mth relay node to the destination node, for example, over the control channel available in a wireless network. We assume that $\tilde{h}_{sr}(m,1), m = 1, \cdots, M$, can be perfectly received at the destination node as the other information transmitted through the control channel. Now the destination node can resolve the scaling ambiguity by computing $\delta_m = \tilde{h}_{sr}(m,1)/\hat{h}_{sr}(m,1)$, $m = 1, \cdots, M$, where δ_m is the mth main diagonal element of $\mathbf{\Delta}_1$ and $\hat{h}_{sr}(m,1)$ is the $(m,1)$-th element of $\hat{\mathbf{H}}_{sr}$.

D. Weighted least-squares (WLS) algorithm

It can be seen from (10) that the covariance matrix of the effective noise $\mathbf{V}_k$ at the destination node is given by $\mathbf{C}_k \triangleq \mathrm{E}[\mathbf{V}_k\mathbf{V}_k^H] = N_s\mathbf{H}_{rd}\mathcal{D}_k\{\mathbf{F}\}(\mathcal{D}_k\{\mathbf{F}\})^H\mathbf{H}_{rd}^H + N_s\mathbf{I}_{N_d}$, $k = 1, \cdots, K$. Obviously, $\mathbf{V}_k$ is non-white due to the channel $\mathbf{H}_{rd}$ and the relay amplifying factor $\mathbf{F}$. Therefore, after an initial estimation of $\mathbf{H}_{rd}$ is obtained by the BALS algorithm in Section III-B, the initial estimation of $\mathbf{H}_{sr}$ can be improved by the WLS approach with

$$\begin{aligned} \breve{\mathbf{H}}_{sr} &= \arg\min_{\mathbf{H}_{sr}} \mathrm{tr}\Big(\big(\tilde{\mathbf{X}} - (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})\mathbf{H}_{sr}\big)^H \hat{\mathbf{C}}^{-1} \\ &\qquad \times \big(\tilde{\mathbf{X}} - (\mathbf{F} \odot \hat{\mathbf{H}}_{rd})\mathbf{H}_{sr}\big)\Big) \\ &= \big(\hat{\mathbf{C}}^{-\frac{1}{2}}(\mathbf{F} \odot \hat{\mathbf{H}}_{rd})\big)^{\dagger}\hat{\mathbf{C}}^{-\frac{1}{2}}\tilde{\mathbf{X}} \end{aligned} \tag{24}$$

where $\hat{\mathbf{C}} = \mathrm{bd}\big[\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2, \cdots, \hat{\mathbf{C}}_K\big]$, $\hat{\mathbf{C}}^{\frac{1}{2}}\hat{\mathbf{C}}^{\frac{1}{2}} = \hat{\mathbf{C}}$, and $\hat{\mathbf{C}}_k = N_s\hat{\mathbf{H}}_{rd}\mathcal{D}_k\{\mathbf{F}\}(\mathcal{D}_k\{\mathbf{F}\})^H\hat{\mathbf{H}}_{rd}^H + N_s\mathbf{I}_{N_d}$, $k = 1, \cdots, K$, is an estimate of $\mathbf{C}_k$ using $\hat{\mathbf{H}}_{rd}$. Here $\mathrm{bd}[\cdot]$ stands for a block diagonal matrix. It will be seen in Section IV that there is an obvious improvement in estimating $\mathbf{H}_{sr}$ by using (24) after the convergence of the BALS algorithm.

Similarly, we expect that the initial estimation of $\mathbf{H}_{rd}$ can be improved by the WLS approach. It can be shown from (20) that the covariance matrix of the noise $\mathbf{W}_j$, denoted as $\mathbf{\Theta}_j \triangleq \mathrm{E}[\mathbf{W}_j\mathbf{W}_j^H]$, $j = 1, \cdots, N_s$, is a diagonal matrix whose (k,k)-th diagonal element is given by $\sum_{m=1}^{M}\|f(k,m)\mathbf{h}_{rd,m}\|^2 + N_d$, where $\mathbf{h}_{rd,m}$ is the mth column of $\mathbf{H}_{rd}$. Since $\mathbf{\Theta}_j$ is diagonal, it will be shown in Section IV that performing the WLS algorithm after the BALS fitting has only negligible performance gain in estimating $\mathbf{H}_{rd}$. Thus, in practice, we only need to carry out the WLS step for estimating $\mathbf{H}_{sr}$.

E. Extensions

For relay nodes with (possibly different number of) multiple antennas, we can write

$$\begin{aligned}\mathbf{H}_{rd} &= [\mathbf{H}_{rd,1}, \mathbf{H}_{rd,2}, \cdots, \mathbf{H}_{rd,M}] \\ \mathbf{H}_{sr} &= \left[\mathbf{H}_{sr,1}^T, \mathbf{H}_{sr,2}^T, \cdots, \mathbf{H}_{sr,M}^T\right]^T \\ \mathbf{D}_f &= \mathrm{bd}[\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_M]\end{aligned}$$

where $\mathbf{H}_{rd,m}$, $m = 1, \cdots, M$, is the MIMO channel between the destination node and the mth relay node, $\mathbf{H}_{sr,m}$, $m = 1, \cdots, M$, is the MIMO channel between the source node and the mth relay node, $\mathbf{D}_m$ is the diagonal relay amplifying matrix at the mth relay node. It is obvious that the system model (2) is still valid and the proposed PARAFAC-based channel estimation algorithm developed in Section III-A to Section III-D can be straightforwardly applied to estimate $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$. In the following, we show that the proposed algorithm can also be used for channel estimation in two-way MIMO relay systems.

In a two-way relay system, two users exchange their information through one or multiple relay nodes [15]. The received signal matrices at two nodes during the kth time block of the channel training period are given respectively by

$$\begin{aligned}\mathbf{Y}_{1,k} &= \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2}\mathbf{S}_2 + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1}\mathbf{S}_1 \\ &\quad + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{1,k}, \qquad k = 1, \cdots, K \qquad (25)\\ \mathbf{Y}_{2,k} &= \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1}\mathbf{S}_1 + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2}\mathbf{S}_2 \\ &\quad + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k} + \mathbf{V}_{2,k}, \qquad k = 1, \cdots, K \qquad (26)\end{aligned}$$

where $\mathbf{H}_{r,i}$, $i = 1, 2$, is the MIMO channel from node i to all relay nodes, $\mathbf{H}_{i,r}$, $i = 1, 2$, is the MIMO channel from all relay nodes to node i, and $\mathbf{V}_{i,k}$, $i = 1, 2$, is the noise matrix at node i during the kth time block.

The $N_i \times L$ training sequence $\mathbf{S}_i$ used by node i, $i = 1, 2$, in (25) and (26) is designed such that

$$\mathbf{S}_i\mathbf{S}_i^H = \mathbf{I}_{N_i}, \quad i = 1, 2, \qquad \mathbf{S}_1\mathbf{S}_2^H = \mathbf{0}_{N_1 \times N_2} \qquad (27)$$

where N_i is the number of antennas at node i, and $\mathbf{0}_{N_1 \times N_2}$ is an $N_1 \times N_2$ matrix with all zero elements. Note that $\mathbf{S}_1$ and $\mathbf{S}_2$ satisfying (27) can be easily constructed from the normalized DFT matrix with $L \geq N_1 + N_2$. Multiplying both sides of (25) with $\mathbf{S}_2^H$ and both sides of (26) with $\mathbf{S}_1^H$, we have

$$\begin{aligned}\mathbf{Y}_{1,k}\mathbf{S}_2^H &= \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,2} + \mathbf{H}_{1,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}_2^H + \mathbf{V}_{1,k}\mathbf{S}_2^H \\ & \quad k = 1, \cdots, K \\ \mathbf{Y}_{2,k}\mathbf{S}_1^H &= \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{H}_{r,1} + \mathbf{H}_{2,r}\mathcal{D}_k\{\mathbf{F}\}\mathbf{V}_{r,k}\mathbf{S}_1^H + \mathbf{V}_{2,k}\mathbf{S}_1^H \\ & \quad k = 1, \cdots, K.\end{aligned}$$

Now the proposed PARAFAC-based algorithm developed in Section III-A to Section III-D can be applied at node 1 to estimate $\mathbf{H}_{1,r}$ and $\mathbf{H}_{r,2}$ and at node 2 to estimate $\mathbf{H}_{2,r}$ and $\mathbf{H}_{r,1}$.

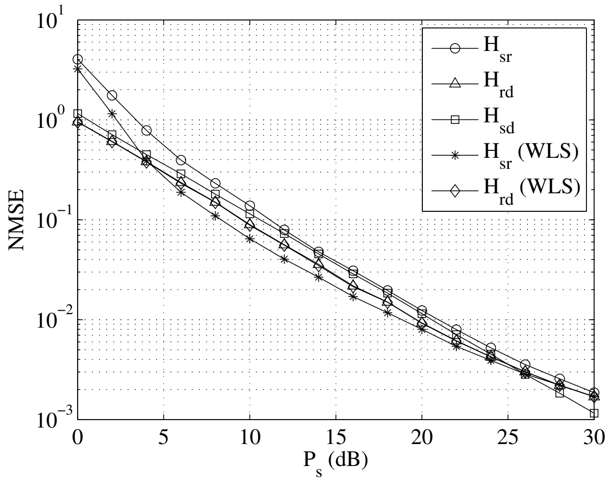

Fig. 2. Example 1: Normalized MSE versus P_s. $K = 3$.

IV. Numerical Examples

In this section, we study the performance of the proposed channel estimation algorithm through numerical simulations. We consider a two-hop MIMO relay communication system where the source and the destination nodes are equipped with $N_s = N_d = 3$ antennas, respectively. For simplicity, there is one relay node in the system with three antennas (equivalent to $M = 3$). All channel matrices have i.i.d. complex Gaussian entries with zero-mean and variances of $1/N_s$, $1/M$, $1/(8N_s)$ for $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$, respectively. Throughout the simulations, we use the minimal L, i.e., $L = N_s = 3$. The transmission power at the relay node is set to be 20dB above the noise level. All simulation results are averaged over 2000 independent channel realizations. For each channel realization, the normalized channel estimation error is calculated as $\|\hat{\mathbf{H}}_{sr} - \mathbf{H}_{sr}\|^2/\|\mathbf{H}_{sr}\|^2$ and $\|\check{\mathbf{H}}_{sr} - \mathbf{H}_{sr}\|^2/\|\mathbf{H}_{sr}\|^2$ for the estimation of $\mathbf{H}_{sr}$ without and with using the additional WLS step, respectively. The channel estimation error of $\mathbf{H}_{rd}$ and $\mathbf{H}_{sd}$ are calculated in a similar way to that of $\mathbf{H}_{sr}$.

In the first example, we study the normalized MSE (NMSE) of channel estimation versus the source node transmission power P_s with $K = 3$. It can be seen from Fig. 2 that the NMSE of channel estimation decreases as P_s increases. As expected, the estimation of $\mathbf{H}_{sr}$ is improved by carrying out the additional WLS step, while the improvement in the estimation of $\mathbf{H}_{rd}$ using the WLS step is negligible. We also observe from Fig. 2 that after the WLS step, the NMSE of $\mathbf{H}_{sr}$ is smaller than that of the $\mathbf{H}_{rd}$ at medium to high P_s. While at low P_s level, the NMSE of $\mathbf{H}_{sr}$ is larger than that of the $\mathbf{H}_{rd}$. This is because $\hat{\mathbf{H}}_{rd}$, rather than the perfect $\mathbf{H}_{rd}$ is used to calculate the weight matrix $\mathbf{C}$ in the WLS step (see (24)), and the mismatch between $\hat{\mathbf{H}}_{rd}$ and $\mathbf{H}_{rd}$ is bigger when P_s is low.

In the second example, we study the performance of the proposed channel estimation algorithm versus K. It can be seen from Fig. 3 that the NMSE decreases with increasing K.

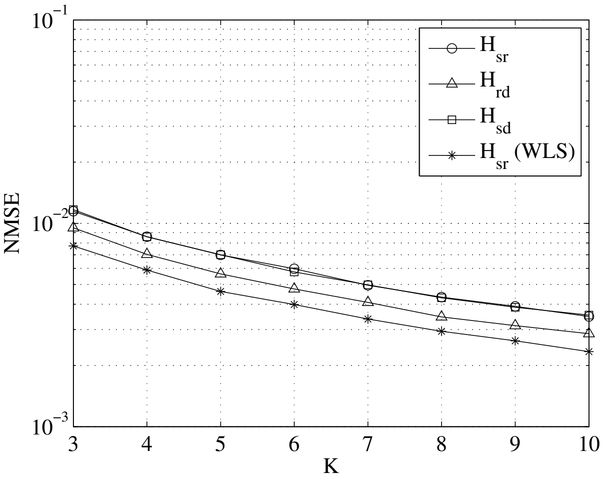

Fig. 3. Example 2: Normalized MSE versus K. $P_s = 20$dB.

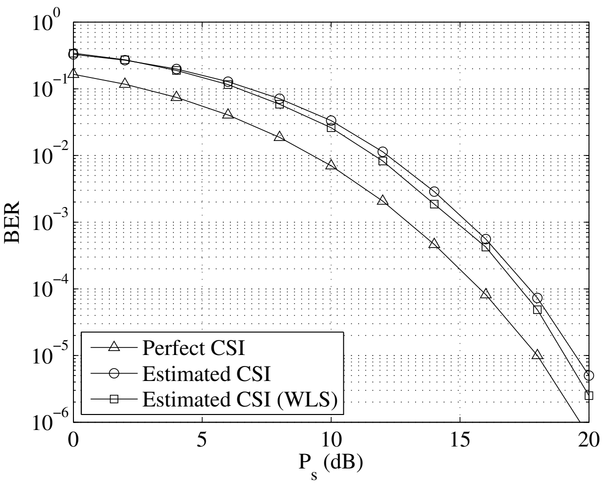

Fig. 4. Example 3: BER versus P_s. $K = 3$.

We would like to mention that the system spectral efficiency is reduced by increasing K. Such performance-efficiency trade-off is important for practical relay communication systems.

In the third example, we investigate the impact of channel estimation on the system BER performance. Fig. 4 shows the comparison of BERs between the system using the CSI estimated by the proposed algorithm and the system with the perfect CSI. The algorithm proposed in [5] is used to optimize the source precoding matrix and the relay amplifying matrix for both systems with the perfect and the estimated CSI. QPSK constellations are used to modulate the source symbols, and 3000 randomly generated bits are transmitted for each channel realization. It can be seen from Fig. 4 that at $P_s = 15$dB, there is only around 1.5dB loss by using the estimated CSI, which is quite reasonable for practical systems. Moreover, we also observe from Fig. 4 that by applying the WLS step, the system BER can be slightly improved.

V. Conclusions

We have developed a novel PARAFAC-based channel estimation method for two-hop MIMO relay communication systems. The proposed algorithm provides the destination node with full knowledge of all channel matrices involved in the communication. Compared with existing approaches, the proposed algorithm requires less number of training data blocks, and is applicable for both one-way and two-way MIMO relay systems. Simulation results demonstrate the effectiveness of the proposed channel estimation algorithm.

Acknowledgment

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

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 (TENCE) for two-way relaying with multiple antennas and spatial reuse," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 3641-3644.

[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] 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.

[10] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, pp. 2377-2388, Aug. 2000.

[11] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Lett.*, vol. 14, pp. 390-392, May 2010.

[12] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884-893, Mar. 2006.

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

[14] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 16, pp. 95-138, 1977.

[15] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 699-712, Jun. 2009.