Robust Channel Estimation Algorithm for Dual-Hop MIMO Relay Channels

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 conventional two-phase channel estimation algorithms for dual-hop multiple-input multiple-output (MIMO) relay systems, the relay-destination channel estimated in the first phase is used for the source-relay channel estimation in the second phase. For these algorithms, the mismatch between the estimated and the true relay-destination channel affects the accuracy of the source-relay channel estimation. In this paper, we investigate the impact of such channel state information (CSI) mismatch on the performance of the two-phase channel estimation algorithm. By explicitly taking into account the CSI mismatch, we develop a robust algorithm to estimate the source-relay channel. Numerical examples demonstrate the improved performance of the proposed algorithm.**

***Index Terms*—Channel estimation, MIMO relay, Robust.**

I. Introduction

In recent years, there is a significant growth in the demand for reliable and high rate wireless communications. This led to great research efforts to improve the overall performance of wireless networks from both industry and academia. Multiple-input multiple-output (MIMO) relay channel has been identified to be one of the promising solutions, as it enhances channel capacity, network reliability and extends the network coverage [1].

For three-node two-hop MIMO relay systems where the direct source-destination link is omitted, the optimal relay precoding matrix is obtained in [2]-[3] to maximize the mutual information between the source and destination nodes. A unified framework has been established recently for optimizing the source and relay precoding matrices of two-hop MIMO relay systems with a broad class of commonly used objective functions [4].

For the MIMO relay systems [1]-[4] mentioned above, the instantaneous channel state information (CSI) knowledge of both the source-relay and relay-destination links is required at the destination node in order to retrieve the signals transmitted from the source node. However, in a real wireless relay system, the instantaneous CSI is unknown, and thus, estimation of channel matrices is required at the destination node. The estimation of channel matrices for single-hop MIMO systems can be found in [5]-[7]. However, the technique used to estimate the channel matrices for single-hop MIMO systems is not applicable for MIMO relay systems.

A novel interim channel estimation technique has been proposed in [8] where the source-relay and relay-destination channels are estimated at the destination node with the help of a known pilot enhancement matrix inserted at the relay node. However, the algorithm in [8] is developed for a MIMO mimicking amplify-and-forward (AF) relay system, and it is proven in [9] that the relay system can never fully mimic a real MIMO relay system as the multiplexing gain is limited. Two algorithms have been proposed in [10], namely, Bayesian-based linear minimum mean-squared error (LMMSE) and expectation-maximization (EM)-based maximum a posteriori (MAP) channel estimation. In the LMMSE channel estimation algorithm, only a sub-optimal solution can be achieved due to the high complexity in the computational of the LMMSE estimator. Consequently, the authors of [10] suggested the EM-based MAP channel estimation algorithm, where the initial estimate of the EM algorithm depends on the LMMSE estimator proposed earlier. However, the training sequences and relay precoding matrix are not optimized in [10]. A parallel factor analysis-based MIMO channel estimator was proposed in [11].

In [12], an algorithm based on the least-squares (LS) method was developed to estimate the channel matrices of MIMO relay networks. In particular, both the source-relay and the relay-destination channel matrices are estimated from the observed composite source-relay-destination channel matrix. A drawback from channel estimation using [12] is the scalar ambiguity of the estimated channel matrices. A two-phase channel estimation scheme based on LMMSE was proposed in [13] for two-hop MIMO relay networks. In particular, in the first phase, the source node is silent while the relay node transmits a pilot matrix to the destination node to estimate the relay-destination channel matrix. In the second phase, the source transmits a source pilot matrix to the relay. The relay node linearly precodes its received signal and forward it to the destination node. Then the source-relay channel is estimated at the destination node making use of the relay-destination channel matrix estimated at the first phase. Compared with the approach in [12], there is no scalar ambiguity in this approach.

However, in practical relay systems, there is always mismatch between the estimated and the true relay-destination channel. Such CSI mismatch affects the accuracy of the

978-1-4673-2569-1/12/$31.00 ©2012 IEEE

source-relay channel estimation in [13]. In this paper, we investigate the impact of this CSI mismatch on the performance of the two-phase channel estimation algorithm [13]. By explicitly taking into account the CSI mismatch, we develop a robust algorithm to estimate the source-relay channel, without the need of greater computation effort. Numerical examples demonstrate the improved performance of the proposed algorithm.

The rest of this paper is organized as follows. In Section II, we introduce the model of a two-hop MIMO relay communication system and the two-phase channel training algorithm. The impact of CSI mismatch on the performance of the two-phase channel estimation algorithm is investigated in Section III. A robust channel estimation algorithm is also developed in Section III. In Section IV, we show some numerical examples. Conclusions are drawn in Section V.

II. Background

We consider a three-node two-hop MIMO relay system where the source node transmits information to the destination node through a relay node. The source, relay, and destination nodes are equipped with n_S, n_R, and n_D antennas, respectively. We focus on the case where the direct link between the source and destination nodes is sufficiently weak to be ignored [12], [13]. This scenario occurs when the direct link is blocked by an obstacle such as a mountain. In fact, a relay plays a much more important role when the direct link is weak than when it is strong.

Similar to [13], the channel matrices are estimated in two phases, where the relay-destination channel matrix $\mathbf{H}_2$ is estimated in phase one while the source-relay channel matrix $\mathbf{H}_1$ is estimated in phase two. In phase one, the signal received by the destination node is given by

$$\mathbf{Y}_D^{(1)} = \mathbf{H}_2\mathbf{S}_R + \mathbf{N}^{(1)} \tag{1}$$

where $\mathbf{S}_R$ is the the $n_R \times n_R$ pilot matrix transmitted by the relay node to the destination node satisfying $\mathbf{S}_R^H\mathbf{S}_R = \mathbf{S}_R\mathbf{S}_R^H = \frac{P_R}{n_R}\mathbf{I}_{n_R}$ [5], and $\mathbf{N}^{(1)}$ is the $n_D \times n_R$ noise matrix at the destination node during phase one. Here P_R is the power budget available at the relay node, $(\cdot)^H$ stands for the matrix (vector) Hermitian transpose, and $\mathbf{I}_n$ denotes an $n\times n$ identity matrix. Note that we choose the length of $\mathbf{S}_R$ to be n_R to maximize the overall system spectral efficiency [14].

A minimal variance unbiased (MVU) estimation [15] of $\mathbf{H}_2$ can be obtained from (1) as

$$\hat{\mathbf{H}}_2 = \frac{n_R}{P_R}\mathbf{Y}_D^{(1)}\mathbf{S}_R^H = \mathbf{H}_2 + \frac{n_R}{P_R}\mathbf{N}^{(1)}\mathbf{S}_R^H. \tag{2}$$

It can be seen from (2) that due to the existence of the noise $\mathbf{N}^{(1)}$, there is a mismatch $\mathbf{\Delta}_2 \triangleq \frac{n_R}{P_R}\mathbf{N}^{(1)}\mathbf{S}_R^H$ between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$. Obviously, $\mathbf{\Delta}_2$ is a complex Gaussian random matrix with zero mean and the variance of its entries is n_R/P_R. Therefore, $\mathbf{H}_2$ is a complex Gaussian matrix with the following distribution

$$\mathbf{H}_2 \sim \mathcal{CN}(\hat{\mathbf{H}}_2, \beta\mathbf{I}_{n_R}\otimes\mathbf{I}_{n_D}) \tag{3}$$

where $\beta \triangleq n_R/P_R$ and $\otimes$ denotes the matrix Kronecker product [16]. It can be seen from (3) that the variance of $\mathbf{H}_2$ decreases when P_R increases.

In phase two, the source node transmits an $n_S \times n_S$ pilot matrix $\mathbf{S}_S$ to the relay node. Here we choose the length of $\mathbf{S}_S$ to be n_S to maximize the overall system spectral efficiency. The relay node applies an $n_R \times n_R$ precoding matrix $\mathbf{F}$ and retransmits the linear precoded signal matrix

$$\mathbf{X}_R = \mathbf{F}\mathbf{H}_1\mathbf{S}_S + \mathbf{F}\mathbf{V} \tag{4}$$

to the destination node, where $\mathbf{V}$ is the $n_R \times n_S$ noise matrix at the relay node. The signal received at the destination node can be written as

$$\mathbf{Y}_D = \mathbf{H}_2\mathbf{F}\mathbf{H}_1\mathbf{S}_S + \mathbf{H}_2\mathbf{F}\mathbf{V} + \mathbf{N} \tag{5}$$

where $\mathbf{N}$ is the $n_D \times n_S$ noise matrix at the destination node during phase two.

By vectorizing both sides of (5), we obtain

$$\mathbf{y}_D = (\mathbf{S}_S^T \otimes \mathbf{H}_2\mathbf{F})\mathbf{h}_1 + (\mathbf{I}_{n_S} \otimes \mathbf{H}_2\mathbf{F})\mathbf{v} + \mathbf{n} \tag{6}$$

where $\mathbf{y}_D \triangleq \mathrm{vec}(\mathbf{Y}_D)$, $\mathbf{h}_1 \triangleq \mathrm{vec}(\mathbf{H}_1)$, $\mathbf{v} \triangleq \mathrm{vec}(\mathbf{V})$, $\mathbf{n} \triangleq \mathrm{vec}(\mathbf{N})$, $(\cdot)^T$ denotes matrix transpose, and $\mathrm{vec}(\cdot)$ denotes the vectorization operator which stacks all column vectors of a matrix on top of each other. To obtain (6) from (5), we use the property of $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [16].

In this paper, we assume that the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfy the well-known Kronecker correlation model [17]

$$\mathbf{H}_i = \mathbf{C}_{r_i}^{\frac{1}{2}}\mathbf{H}_{w,i}\mathbf{C}_{t_i}^{\frac{T}{2}}, \qquad i = 1, 2 \tag{7}$$

where $\mathbf{C}_{t_i}$ and $\mathbf{C}_{r_i}$, $i = 1, 2$, are channel correlation matrices at the transmit side and the receive side of $\mathbf{H}_i$, respectively, and $\mathbf{H}_{w,i}$, $i = 1, 2$, are Gaussian random matrices with independent and identically distributed (i.i.d.) entries having zero mean and unit variance. We also assume that all noises are i.i.d. additive white Gaussian noise (AWGN) with zero mean and unit variance.

III. Robust Channel Estimation Algorithm

In this section, we derive the optimal $\mathbf{S}_S$ and $\mathbf{F}$ that minimize the MSE of estimating $\mathbf{H}_1$. Using a linear estimator, the estimated $\mathbf{h}_1$ is given by

$$\hat{\mathbf{h}}_1 = \mathbf{W}\mathbf{y}_D \tag{8}$$

where $\mathbf{W}$ is the weight matrix of the linear estimator. Using (8), the MSE of estimating $\mathbf{h}_1$ can be written as

$$\begin{aligned} J_1 &= E\big[\mathrm{tr}\big((\mathbf{h}_1 - \hat{\mathbf{h}}_1)(\mathbf{h}_1 - \hat{\mathbf{h}}_1)^H\big)\big] \\ &= \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - \mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}\mathbf{W}^H - \mathbf{W}\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}^H + \mathbf{W}\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}\mathbf{W}^H\big) \end{aligned} \tag{9}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace, $E[\cdot]$ stands for statistical expectation, and from (6) we have

$$\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H} = E[\mathbf{h}_1\mathbf{y}_D^H] = (\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes (\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H) \tag{10}$$

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H} = E[\mathbf{y}_D\mathbf{y}_D^H] &= (\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes (\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H) \\ &+ \mathbf{I}_{n_S} \otimes (\mathbf{H}_2\mathbf{F}\mathbf{F}^H\mathbf{H}_2^H) + \mathbf{I}_{n_S n_D} \end{aligned} \tag{11}$$

$$\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} = E[\mathbf{h}_1\mathbf{h}_1^H] = \mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1}. \tag{12}$$

Here $(\cdot)^*$ stands for complex conjugate, and we use $\mathbf{h}_1 = (\mathbf{C}_{t_1}^{\frac{1}{2}} \otimes \mathbf{C}_{r_1}^{\frac{1}{2}})\mathbf{h}_{w,1}$ with $\mathbf{h}_{w,1} \triangleq \mathrm{vec}(\mathbf{H}_{w,1})$.

From (10)-(12), it can be seen that the CSI of $\mathbf{H}_2$ is needed in order to minimize J_1. However, the exact $\mathbf{H}_2$ is unknown in the second phase. In fact, it is shown in (3) that $\mathbf{H}_2$ is a complex Gaussian random matrix with the mean matrix of $\hat{\mathbf{H}}_2$. Obviously, the mismatch between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$ affects the accuracy of the estimation of $\mathbf{H}_1$. To take such mismatch into account, we adopt a statistically robust objective function through averaging J_1 in (9) with respect to the distribution of $\mathbf{H}_2$ as

$$E_{\mathbf{H}_2}[J_1] = \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}]\mathbf{W}^H - \mathbf{W}E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}^H] + \mathbf{W}E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}]\mathbf{W}^H\big). \tag{13}$$

The estimator $\mathbf{W}$ which minimizes (13) is the linear MMSE estimator [15] given by

$$\mathbf{W} = E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}]\big(E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}]\big)^{-1} \tag{14}$$

where $(\cdot)^{-1}$ denotes matrix inversion. Substituting (14) back into (13), we have

$$E_{\mathbf{H}_2}[J_1] = \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}]\big(E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}]\big)^{-1} \times E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}^H]\big). \tag{15}$$

It can be easily seen from (10) that

$$E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}] = (\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes (\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H). \tag{16}$$

Using the property that for a complex Gaussian random matrix $\mathbf{H} \sim \mathcal{CN}(\bar{\mathbf{H}}, \mathbf{\Theta}\otimes\mathbf{\Phi})$, $E_{\mathbf{H}}[\mathbf{H}\mathbf{A}\mathbf{H}^{\mathbf{H}}] = \bar{\mathbf{H}}\mathbf{A}\bar{\mathbf{H}}^H + \mathrm{tr}(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$ [18], we have from (3) that

$$\begin{aligned} &E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}] \\ &= (\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H)\mathbf{I}_{n_D}\big) \\ &\quad + \mathbf{I}_{n_S} \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{F}^H)\mathbf{I}_{n_D}\big) + \mathbf{I}_{n_S n_D}. \end{aligned} \tag{17}$$

Substituting (16) and (17) back into (15), we obtain that

$$\begin{aligned} &E_{\mathbf{H}_2}[J_1] \\ &= \mathrm{tr}\Big(\mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1} - (\mathbf{S}_S^T\mathbf{C}_{t_1}^H\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes (\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}^H\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H) \\ &\quad \times \big[(\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*) \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H)\mathbf{I}_{n_D}\big) \\ &\quad + \mathbf{I}_{n_S} \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{F}^H)\mathbf{I}_{n_D}\big) + \mathbf{I}_{n_S n_D}\big]^{-1}\Big). \end{aligned} \tag{18}$$

The transmission power consumed at the relay node during phase two can be calculated from (4) as

$$\begin{aligned} p_r &\triangleq E_{\mathbf{H}_1}\big[\mathrm{tr}\big(\mathbf{F}(\mathbf{H}_1\mathbf{S}_S\mathbf{S}_S^H\mathbf{H}_1^H + n_S\mathbf{I}_{n_R})\mathbf{F}^H\big)\big] \\ &= \mathrm{tr}(\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*)\mathrm{tr}(\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H) + n_S\mathrm{tr}(\mathbf{F}\mathbf{F}^H). \end{aligned} \tag{19}$$

Using (18) and (19), the optimal robust $\mathbf{S}_S$ and $\mathbf{F}$ can be found as the solution to the following problem

$$\min_{\mathbf{S}_S,\mathbf{F}} \; E_{\mathbf{H}_2}[J_1] \tag{20}$$

$$\text{s.t.} \; \mathrm{tr}(\mathbf{S}_S\mathbf{S}_S^H) \leq P_S \tag{21}$$

$$\mathrm{tr}(\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*)\mathrm{tr}(\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H) + n_S\mathrm{tr}(\mathbf{F}\mathbf{F}^H) \leq P_R \tag{22}$$

where (21) and (22) are the transmission power constraint at the source and the relay node, respectively, and P_S is the power budget available at the source node. The problem (20)-(22) is complicated with matrices variables. We first show the optimal structure of $\mathbf{S}_S$ and $\mathbf{F}$.

Let us define the following eigenvalue decompositions (EVDs)

$$\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^* = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H \tag{23}$$

$$\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H = \mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H \tag{24}$$

$$\mathbf{C}_{t_1} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}\mathbf{U}_{t_1}^H \tag{25}$$

$$\mathbf{C}_{r_1} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}\mathbf{U}_{r_1}^H \tag{26}$$

where $\mathbf{U}_S$, $\mathbf{U}_F$, $\mathbf{U}_{t_1}$, and $\mathbf{U}_{r_1}$ are the unitary eigenvector matrices, and $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{\Lambda}_{t_1}$, and $\mathbf{\Lambda}_{r_1}$ are the diagonal eigenvalue matrices with descending diagonal elements. From (23)-(24), we can obtain that

$$\mathbf{S}_S^T\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S, \qquad \hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F \tag{27}$$

where $\mathbf{Q}_S$ and $\mathbf{Q}_F$ are unitary matrices. Here $\mathbf{C}_{t_1}^{\frac{1}{2}}$ and $\mathbf{C}_{r_1}^{\frac{1}{2}}$ are defined based on (25) and (26) as

$$\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}^{\frac{1}{2}}, \qquad \mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}^{\frac{1}{2}} \tag{28}$$

Let us introduce the singularvalue decomposition (SVD) of $\hat{\mathbf{H}}_2$ as

$$\hat{\mathbf{H}}_2 = \mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}\mathbf{V}_{H_2}^H \tag{29}$$

where $\mathbf{U}_{H_2}$ and $\mathbf{V}_{H_2}$ are the singular vector matrices and $\mathbf{\Sigma}_{H_2}$ is the singularvalue matrix with descending diagonal elements.

From (27) and (29) we have

$$\mathbf{S}_S^T = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{C}_{t_1}^{-\frac{1}{2}}, \quad \mathbf{F} = \mathbf{V}_{H_2}\mathbf{\Sigma}_{H_2}^{-1}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{C}_{r_1}^{-\frac{1}{2}}. \tag{30}$$

Using (23)-(30), $\bar{J}_1 \triangleq E_{\mathbf{H}_2}[J_1] - \mathrm{tr}(\mathbf{C}_{t_1}\otimes\mathbf{C}_{r_1})$ can be written as

$$\begin{aligned} \bar{J}_1 = &-\mathrm{tr}\big([\mathbf{\Lambda}_S\otimes(\mathbf{\Lambda}_F + a\mathbf{I}_{n_D}) + \mathbf{I}_{n_S}\otimes(\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}) \\ &+ b\mathbf{I}_{n_S n_D}]^{-1}(\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{\Lambda}_{t_1}\mathbf{Q}_S^H\mathbf{\Lambda}_S^{\frac{1}{2}})\otimes(\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}})\big) \end{aligned} \tag{31}$$

where

$$\begin{aligned} a &\triangleq \mathrm{tr}(\beta\mathbf{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F) \\ b &\triangleq \mathrm{tr}(\beta\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{U}_F^H\mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H) + 1. \end{aligned}$$

The power constraints (21) and (22) can be rewritten as

$$\mathrm{tr}(\mathbf{\Lambda}_S\mathbf{Q}_S\mathbf{\Lambda}_{t_1}^{-1}\mathbf{Q}_S^H) \leq P_S \tag{32}$$

$$\begin{aligned} &\mathrm{tr}(\mathbf{\Lambda}_S)\mathrm{tr}(\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}) \\ &+ n_S\mathrm{tr}(\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{U}_F^H\mathbf{U}_{H_2}) \leq P_R \end{aligned} \tag{33}$$

From (31), we see that the mismatch between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$ is considered by matrices $a\mathbf{I}_{n_D}$ and $b\mathbf{I}_{n_S n_D}$. In fact, the objective function in [13] can be viewed as a special case of (31) where $a = b = 0$. It can be proven similar to [13] that if $\mathbf{C}_{r_1} = \alpha\mathbf{I}_{n_R}$, then at the optimal $\mathbf{S}_S$, there is $\mathbf{Q}_S = \mathbf{I}_{n_S}$, $\mathbf{Q}_F = \mathbf{I}_{n_R}$,

$\mathbf{U}_F = \mathbf{U}_{H_2}$, and $\mathbf{U}_S = \mathbf{I}_{n_S}$. Therefore, the optimal structure of $\mathbf{S}_S$ and $\mathbf{F}$ can be written as

$$\mathbf{S}_S^T = \mathbf{\Lambda}_S^{\frac{1}{2}} \mathbf{C}_{t_1}^{-\frac{1}{2}}, \qquad \mathbf{F} = \alpha^{-\frac{1}{2}} \mathbf{V}_{H_2} \mathbf{\Sigma}_{H_2}^{-1} \mathbf{\Lambda}_F^{\frac{1}{2}}. \tag{34}$$

Substituting (34) back into (31)-(33) and let $\lambda_{S,i}$, $\lambda_{F,i}$, $\lambda_{t_1,i}$, $\lambda_{r_1,i}$, and $\sigma_{H_2,i}$ be the ith diagonal element of $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{\Lambda}_{t_1}$, $\mathbf{\Lambda}_{r_1}$, and $\mathbf{\Sigma}_{H_2}$, respectively, the problem (20)-(22) is converted to the following problem with scalar variables

$$\min_{\{\lambda_{S,i}\},\{\lambda_{F,j}\}} \; -\sum_{i=1}^{n_S}\sum_{j=1}^{n_D} \frac{c_{i,j}}{d_{i,j}} \tag{35}$$

$$\text{s.t.} \; \sum_{i=1}^{n_S} \frac{\lambda_{S,i}}{\lambda_{t_1,i}} \leq P_S \tag{36}$$

$$\sum_{i=1}^{n_S} \lambda_{S,i} \sum_{j=1}^{n_D} \frac{\lambda_{F,j}}{\sigma_{H_2,j}^2} + \sum_{j=1}^{n_D} \frac{n_S \lambda_{F,j}}{\sigma_{H_2,j}^2 \lambda_{r_1,j}} \leq P_R \tag{37}$$

$$\lambda_{S,i} \geq 0, \qquad i = 1, \cdots, n_S \tag{38}$$

$$\lambda_{F,j} \geq 0, \qquad j = 1, \cdots, n_D \tag{39}$$

where

$$c_{i,j} \triangleq \lambda_{S,i}\lambda_{t_1,i}\lambda_{F,j}\lambda_{r_1,j}$$

$$d_{i,j} \triangleq \lambda_{S,i}\lambda_{F,j} + \sum_{j=1}^{n_D} \frac{\beta\lambda_{S,i}\lambda_{F,j}}{\sigma_{H_2,j}^2} + \frac{\lambda_{F,j}}{\lambda_{r_1,j}} + \sum_{j=1}^{n_D} \frac{\beta\lambda_{F,j}}{\lambda_{r_1,j}\sigma_{H_2,j}^2} + 1$$

$$\{\lambda_{S,i}\} \triangleq \{\lambda_{S,i}, i = 1, \cdots, n_S\}$$

$$\{\lambda_{F,j}\} \triangleq \{\lambda_{F,j}, j = 1, \cdots, n_D\}$$

The problem (35)-(39) is non-convex. However, as the optimization of $\{\lambda_{F,j}\}$ is convex when $\{\lambda_{S,i}\}$ is fixed, and vice versa, (at least) a local optimum solution can be found by iteratively optimize $\{\lambda_{F,j}\}$ and $\{\lambda_{S,i}\}$. These two sub-optimizations problem are formulated as follows.

1. Optimizing $\{\lambda_{F,j}\}$ *with fixed* $\{\lambda_{S,i}\}$. The power constraint at the source node is irrelevant as $\{\lambda_{S,i}\}$ is fixed. Therefore, the Karush-Kuhn-Tucker (KKT) conditions of optimizing $\{\lambda_{F,j}\}$ can be written as

$$\sum_{i=1}^{n_S} \frac{\lambda_{S,i}\lambda_{t_1,i}\lambda_{r_1,j}}{d_{i,j}^2} \left[\sum_{l=1, l\neq j}^{n_D} \frac{\beta\lambda_{F,l}}{\sigma_{H_2,l}^2} \left(\lambda_{S,i} + \frac{1}{\lambda_{r_1,l}} \right) + 1 \right] = \mu \left[\sum_{i=1}^{n_S} \frac{\lambda_{S,i}}{\sigma_{H_2,j}^2} + \frac{n_S}{\sigma_{H_2,j}^2 \lambda_{r_1,j}} \right] \tag{40}$$

$$\mu \left(\sum_{i=1}^{n_S} \lambda_{S,i} \sum_{j=1}^{n_D} \frac{\lambda_{F,j}}{\sigma_{H_2,j}^2} + \sum_{j=1}^{n_D} \frac{n_S\lambda_{F,j}}{\sigma_{H_2,j}^2 \lambda_{r_1,j}} - P_R \right) = 0 \tag{41}$$

where $\mu \geq 0$ is the Lagrange multiplier such that equation (41) holds. With any fixed $\{\lambda_{S,i}\}$, μ, and $\lambda_{F,l}$, $l = 1, \cdots, n_D, l \neq j$, the non-negative $\lambda_{F,j}$ can be derived using the bi-section search, since the left-hand-side (LHS) of (40) is a monotonically decreasing function of $\lambda_{F,j}$. Note that (40) depends on $\lambda_{F,j}$, $j = 1, \cdots, n_D$, hence, the value of $\{\lambda_{F,j}\}$ needs to be updated each time a new $\lambda_{F,j}$ is obtained. To find the optimal value of μ, an outer bi-section loop is used as the LHS of (37) is an increasing function of $\lambda_{F,j}$, and $\lambda_{F,j}$ is a monotonically decreasing function of μ.

2. Optimizing $\{\lambda_{S,i}\}$ *with fixed* $\{\lambda_{F,j}\}$. The KKT conditions of this subproblem can be written as

$$\sum_{j=1}^{n_D} \frac{\lambda_{t_1,i}\lambda_{F,j}\lambda_{r_1,j}\left(\frac{\lambda_{F,j}}{\lambda_{r_1,j}} + \beta \sum_{l=1}^{n_D} \frac{\lambda_{F,l}}{\lambda_{r_1,l}\sigma_{H_2,l}^2} + 1 \right)}{d_{i,j}^2} = \frac{\nu_1}{\lambda_{t_1,i}} + \nu_2 \sum_{j=1}^{n_D} \frac{\lambda_{F,j}}{\sigma_{H_2,j}^2} \tag{42}$$

$$\nu_1 \left(\sum_{i=1}^{n_S} \frac{\lambda_{S,i}}{\lambda_{t_1,i}} - P_S \right) = 0 \tag{43}$$

$$\nu_2 \left(\sum_{i=1}^{n_S} \lambda_{S,i} \sum_{j=1}^{n_D} \frac{\lambda_{F,j}}{\sigma_{H_2,j}^2} + \sum_{j=1}^{n_D} \frac{n_S\lambda_{F,j}}{\sigma_{H_2,j}^2 \lambda_{r_1,j}} - P_R \right) = 0 \tag{44}$$

where $\nu_1 \geq 0$ and $\nu_2 \geq 0$ are the Lagrange multipliers. For any fixed $\{\lambda_{F,j}\}$, ν_1 and ν_2, the non-negative $\lambda_{S,i}$ can be found by a bi-section search for all i. This is because the LHS of (42) is a monotonically decreasing function of $\lambda_{S,i}$. Note that the LHS of both (36) and (37) are increasing function of $\lambda_{S,i}$, and $\lambda_{S,i}$ is a monotonically decreasing function of both ν_1 and ν_2. Generally, to find the optimal value of ν_1 and ν_2, a 2-D bi-section loop search is required. However, if only one of the constraints is active (i.e. only one of the constraints satisfies the equality), then only 1-D bisection loop search is required to find the corresponding multiplier for the constraint as the other multiplier is zero. If both constraints are inactive, then a 2-D bi-section loop is required to determine the optimal value of ν_1 and ν_2.

IV. Numerical Examples

In this section, we study the performance of the proposed channel estimation algorithm through numerical simulations. We compare the proposed approach with the algorithm developed in [13] (denoted as "imperfect H_2") where $\hat{\mathbf{H}}_2$ is used in the second phase to estimate $\mathbf{H}_1$. As a benchmark, the performance of channel estimation algorithm with exactly known $\mathbf{H}_2$ is also studied.

In the simulations, for simplicity, we set $n_S = n_R = n_D = N$. The channel correlation matrices are modelled as $[\mathbf{C}_{t_i}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{C}_{r_2}]_{m,n} = \rho^{|m-n|}$, where ρ is the correlation coefficient, and $\mathbf{C}_{r_1} = \mathbf{I}_{n_R}$. For each channel realization, the normalized MSE (NMSE) of channel estimation for all three algorithms is calculated as $\|\mathbf{H}_1 - \hat{\mathbf{H}}_1\|_F^2 / n_S n_R$, where $\| \cdot \|_F^2$ stands for the matrix Frobenius norm. All simulation results are averaged over 100 random channel realizations.

Fig. 1 shows the normalized MSE of estimating $\mathbf{H}_1$ when $N = 2$ and $\rho = 0.2$. A different number of antennas $N = 4$ and normalized correlation coefficient $\rho = 0.8$ are used for the next scenario and the results are shown in Fig. 2. Note that for both scenarios, the power at the source node is assumed to be the same as the power at the relay node, i.e. $P_S = P_R = P$.

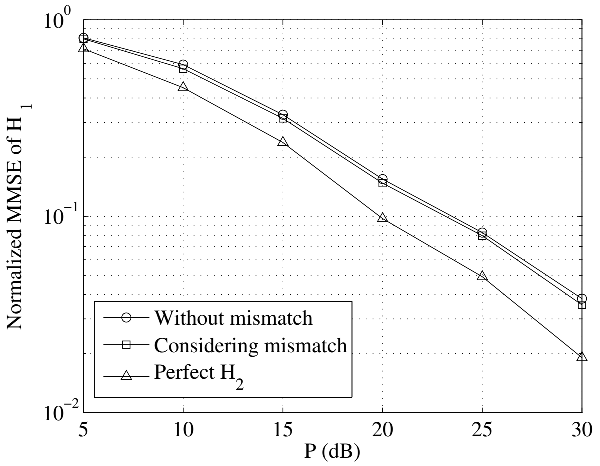

Fig. 1. Normalized MSE versus P. $N = 2$ and $\rho = 0.2$

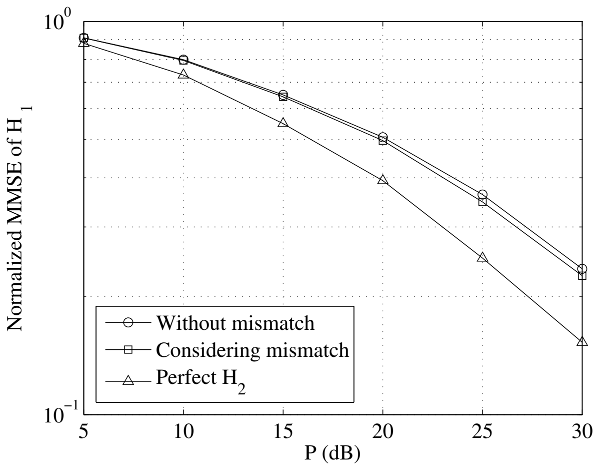

Fig. 2. Normalized MSE of $\mathbf{H}_1$ for $N = 4$ and $\rho = 0.8$

Fig. 3 shows the normalized MSE when the power at the source node P_S is fixed at 20dB while the power at the relay node P_R is varied from 5dB to 30dB. The number of antennas and the normalized correlation coefficient are set to be $N = 2$ and $\rho = 0.8$ respectively.

From the simulation results, it is obvious that by considering the mismatch between $\hat{\mathbf{H}}_2$ and $\mathbf{H}_2$ in the algorithm, the performance of the algorithm has been improved without the need of greater computation effort. The simulations are executed with different parameters to examine the effectiveness of the algorithm, and all results show an improvement in the estimation of channel matrices.

V. Conclusions

The effect of the mismatch between the estimated and true relay-destination channel on the performance of the LMMSE-based MIMO relay channel estimation algorithm has been investigated in this paper. It has been proven that the robust channel estimation algorithm performs better compared to the channel estimation algorithm proposed in [13] that does not take the mismatch into the consideration. Moreover, the robust channel estimation algorithm does not require greater computational effort.

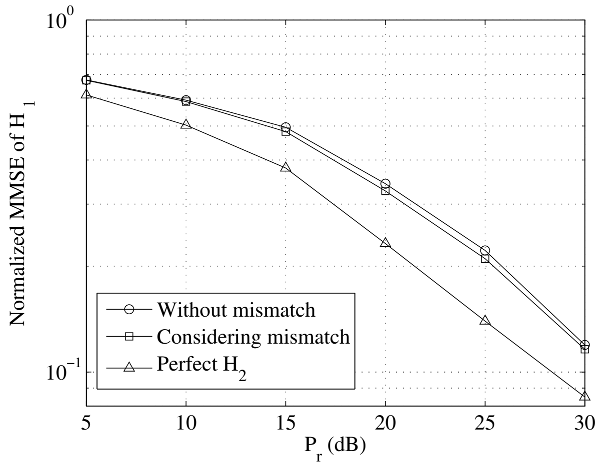

Fig. 3. Normalized MSE versus P_R. $N = 2$, $\rho = 0.8$.

Acknowledgment

This work was supported under the 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] 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.

[5] 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.

[6] M. Biguesh, S. Gazor, and M. H. Shariat,"Optimal training sequence for MIMO wireless systems in colored environment," *IEEE Trans. Signal Process.*, vol. 57, pp. 3144-3153, Aug. 2009.

[7] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, pp. 1807-1820, Mar. 2010.

[8] F. Raphel and S. M. Sameer, "A novel interim channel estimation technique for MIMO mimicking AF cooperative relay systems," *Proc. National Conf. Commun.*, Feb. 2012.

[9] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff in cooperative wireless systems," *Proc. 40th Annual Conf. Inform. Sciences Syst.*, pp. 1062-1067, Mar. 2006.

[10] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian channel estimation techniques for AF MIMO relaying systems," *Proc. IEEE Vehicular Techn. Conf.*, Sep. 2011.

[11] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay systems via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2224-2233, June 2012.

[12] 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.

[13] 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.
[14] Q. Ma and C. Tepedelenlioglu, "Antenna selection for space-time coded systems with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 710-719, Feb. 2007.
[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*. Englewood Cilffs, NJ: Prentice Hall, 1993.
[16] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772-781, Sep. 1978.
[17] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 503-513, Mar. 2000.
[18] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC, 2000.