ML-Based Channel Estimations for Non-Regenerative Relay Networks with Multiple Transmit and Receive Antennas

Yindi Jing, *Member, IEEE* and Xinwei Yu

***Abstract*—This paper investigates the channel estimations in a relay network with multiple transmit and receive antennas, including the estimation of the end-to-end channel matrix and the individual estimation of the transmitter-relay channels and the relay-receiver channels. For the end-to-end channel estimation, instead of directly estimating entries of the channel matrix, we use singular value decomposition (SVD) and estimate its largest singular value and singular vectors, which are then combined to form an estimation of the channel matrix. An approximate maximum-likelihood (ML) estimation is proposed, which is shown to become the exact ML estimation when the time duration of each training step equals the number of antennas at the transmitter. Simulation on the mean square error (MSE) shows that the SVD-based approximate ML estimation performs about the same as the exact ML estimation and is superior to entry-based estimations. For the individual channel estimation, we decompose each channel vector into the product of its length and direction, and find the ML estimation of each. By using an approximation on the probability density function (PDF) of the observations during training, an analytical ML estimation is derived. The ML estimation with the exact PDF is also investigated and a solution is obtained numerically. Simulation on the MSE shows that the two have similar performance. Compared with cascade channel estimations, its performance is superior for the relay-receiver channel estimation and comparable for the transmitter-relay channel estimation. Extension to the general multiple-antenna multiple-relay network is also provided.**

***Index Terms*—Relay network, channel training, channel estimation, maximum-likelihood (ML) estimation, mean square error (MSE).**

I. Introduction

COOPERATIVE relay network, where relay nodes collaborate to establish a virtual MIMO communication link between a transmitter and a receiver, has been shown to be a promising infrastructure to achieve spatial diversity. Early researches in cooperative relay network focus on coherent cooperative schemes, e.g., decode-and-forward (DF) [1]–[3], amplify-and-forward (AF) [2]–[5], distributed space-time coding (DSTC) [6]–[8], and beamforming [9]–[11], in which perfect channel state information (CSI) is assumed at some or all network nodes. In reality, a training process, sometimes, a feedback process as well, is required for a network node to obtain an estimation of its required CSI. Such estimation is always imperfect due to the existence of noise. As a consequence, the training design and channel estimation are important and practical problems in wireless relay networks.

These problems have drawn increasing attention recently, for both regenerative and non-regenerative relay networks [12]–[22]. For non-regenerative relay network [13]–[22], the transmitter-relay channels (channels between the transmitter and the relay) and relay-receiver channels (channels between the relay and the receiver) concatenate with each other in transmissions. Its estimation problem is different from that of a point-to-point multi-input-multi-output (MIMO) system [23]–[25], thus is particularly interesting. [13]–[18] are on single-antenna network with a single relay. In [13]–[15], the channel training and the diversity performance with channel estimation error using mismatched maximum likelihood (ML) receiver are studied. In both [13] and [14], channel estimation is performed at the receiver. In [13], ML and linear-minimum-mean-square-error (LMMSE) estimators are employed, and DSTC is used for data transmission. It is shown that full diversity can be achieved with both estimators. A similar conclusion is drawn in [14] where Golden code is used for data transmission. In [15], a scheme, in which the transmitter-relay channel is estimated at the relay and forwarded to the receiver, is proposed. In [16], the signal-to-noise ratio (SNR) maximizing power allocations between training and data transmission and between the broadcasting and relaying phases are investigated. In [17], the joint optimization of the training time, the power allocation between training and data transmission, and the power allocation between the transmitter and the relay that maximizes a mutual information lower bound is studied. [18] studies the capacity lower bound with channel estimation error and shows that minimum mean square error (MMSE) estimation is optimal. For a relay network with multiple relays but a single antenna at each node, the training problem is studied in [19]–[21]. For an AF protocol with matched filter, [19] proposes a scheme for the receiver to estimate the end-to-end channel. [20], [21] are also on the estimation of the end-to-end channels at the receiver but DSTC is used.

The aforementioned papers are on the training problem for relay networks with single antenna at both the transmitter and the receiver only. There are limited publications on training for

Manuscript received 27 August 2011; revised 7 March 2012. This work was supported in part by the National Science and Engineering Research Council (NSERC) of Canada.

Y. Jing is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada, T6G 2V4 (email: yindi@ualberta.ca).

X. Yu is with the Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada, T6G 2G1 (email: xinweiyu@math.ualberta.ca).

Digital Object Identifier 10.1109/JSAC.2012.1209xx.

relay networks with multiple antennas at the transmitter and/or the receiver [26], [27]. In [26], for the general MIMO relay network, schemes for the receiver to estimate the transmitter-relay channels and the relay-receiver channels individually are proposed. The LMMSE estimations and the optimal pilot designs that minimize the mean square error (MSE) are derived. The requirement on the training time for full diversity in data transmission with mismatched ML decoding is also derived. In [27], both mismatched and matched decodings for networks with multiple relays and multiple receive antennas under channel estimation error are analyzed.

In this paper, for relays network with multiple transmit and receive antennas, we study both the estimation of the end-to-end channel matrix and the individual estimation of the transmitter-relay and relay-receiver channels, all at the receiver. We consider two channel estimation problems because different CSI is required for different cooperative schemes. For non-regenerative DSTC [6], [7] and fixed gain AF [2], [28], the end-to-end channel matrix is required to be known at the receiver, while individual values of the transmitter-relay and relay-receiver channels are not needed. On the other hand, for relay/antenna selection [29]–[32] and beamforming [9]–[11], the receiver needs to know the transmitter-relay and relay-receiver channels individually. In what follows, we clarify the difference of this work to previous ones.

The end-to-end channel estimation is investigated in [20], [21] for relay networks with single transmit and receive antenna only. In the proposed schemes, entries of the end-to-end channel vector are estimated directly. In our paper, we investigate networks with multiple transmit and receive antennas via a different approach. Instead of directly estimating entries of the end-to-end channel matrix, we take into consideration the special structure of the channel matrix and propose to estimate the largest singular value of the matrix as well as its corresponding left and right singular vectors based on the ML criterion. An estimation of the channel matrix is then constructed from the estimated singular value and singular vectors. In other words, we parameterize the end-to-end channel matrix by its eigenvalue and eigenvectors. An approximate ML estimation is proposed, which is shown by simulation to achieve about the same MSE as the exact ML estimation. Comparing with the entry-based estimation that ignores the special structure of the end-to-end channel matrix, the proposed singular value decomposition (SVD) based estimation is superior in MSE. Via simulation, the MSE is also shown to have a linear decreasing rate in the transmit power.

For the individual estimation of the transmitter-relay and relay-receiver channels, we propose a joint estimation of the channel vectors. This is different from the cascade estimation ideas in [26], [27], where the relay-receiver channels are estimated first, the results of which are then used to estimate the transmitter-relay channels. In this work, we decompose each channel vector into the product of its length and direction. The estimations on the length and direction are then combined to obtain an estimation of the channel vector. First, by using an approximation on the conditional probability density function (PDF) of the received signals during training, ML estimation of the individual channels is derived in analytical form. Then, we investigate the ML estimation based on the exact PDF. A solution for the estimation of the relay-receive channels is obtained via numerical alternative optimization. Based on this result, the estimation of the transmitter-relay channels is obtained in analytical form. Compared with cascade estimations, the proposed schemes achieve lower MSE on the relay-receiver channels, and comparable MSE on the transmitter-relay channels.

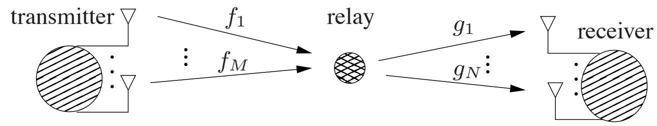


Fig. 1. Relay network with multiple transmit and multiple receive antennas.

The rest of the paper is organized as follows. In Section II, the cooperative relay network model, the channel estimation problems, and the training procedure are explained. Section III is on ML-based estimation of the end-to-end channel matrix; while Section IV is on ML-based estimation of the transmitter-relay channels and the relay-receiver channels individually. Numerically simulated MSEs are shown in Section V. Section VI discusses the extension of the proposed schemes to the general MIMO relay network. Section VII contains the conclusions. Involved proofs are enclosed in the appendix.

II. Channel Estimation Problems and Training Model

First, we explain the notation used in this paper. We use bold upper case letters to denote matrices and bold lower case letters to denote vectors, which can be either row vectors or column vectors. For a matrix $\mathbf{A}$, its conjugate, transpose, Hermitian, Frobenius norm, determinant, trace, and inverse are denoted by $\overline{\mathbf{A}}$, $\mathbf{A}^t$, $\mathbf{A}^*$, $\|\mathbf{A}\|_F$, $\det \mathbf{A}$, $\mathrm{tr}\mathbf{A}$, and $\mathbf{A}^{-1}$, respectively. $\mathrm{vec}(\mathbf{A})$ denotes the column vector formed by stacking the columns of $\mathbf{A}$. $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\otimes$ denotes the Kronecker product. For a complex scalar x, $|x|$ denotes its magnitude and $\angle x$ denotes its angle. $g(x) = O(f(x))$ means $|g(x)| \leq c|f(x)|$ with c a non-zero constant. $\mathbb{E}(\cdot)$ is the average operator for a random variable, a random vector, or a random matrix. We use $\hat{a}$ to denote the estimation of a.

A. Network Model and Channel Estimation Problems

Now we illustrate the relay network model. Consider a network with one transmitter equipped with M antennas, one relay equipped with single antenna, and one receiver equipped with N antennas, as shown in Figure 1. Denote the channel from the mth transmit antenna to the relay antenna as f_m and the channel from the relay antenna to the nth receive antenna as g_n. The channels are assumed to be i.i.d. circularly symmetric complex Gaussian (CSCG) with zero-mean and unit-variance, i.e., $f_m, g_n \sim \mathcal{CN}(0,1)$, so the magnitudes follow Rayleigh distribution. The channels are also assumed to remain constant during training. Define

$$\mathbf{f} \triangleq \left[\, f_1 \;\; f_2 \;\; \cdots \;\; f_M \,\right]^t \quad \text{and} \quad \mathbf{g} \triangleq \left[\, g_1 \;\; g_2 \;\; \cdots \;\; g_N \,\right],$$

which are the $M \times 1$ transmitter-relay channel vector and the $1 \times N$ relay-receiver channel vector, respectively.

For some cooperative schemes such as non-regenerative DSTC [6], [7] and AF with fixed gain relaying [2], [28], where the relay antenna conducts linear transformation and simple amplification respectively, the end-to-end channel values are required to be known at the receiver. The end-to-end channel from the mth transmit antenna to the nth receive antenna via the relay antenna is $f_m g_n$. The individual values of f_m's and g_n's are not needed. By using matrix and vector representation, the $M \times N$ end-to-end channel matrix of the network is

$$\mathbf{H} \triangleq \begin{bmatrix} f_1 g_1 & f_1 g_2 & \cdots & f_1 g_N \\ f_2 g_1 & f_2 g_2 & \cdots & f_2 g_N \\ \vdots & \vdots & \ddots & \vdots \\ f_M g_1 & f_M g_2 & \cdots & f_M g_N \end{bmatrix} = \mathbf{fg}. \tag{1}$$

The first problem to be investigated in this paper is the estimation of $\mathbf{H}$ at the receiver, which we address as the end-to-end channel estimation problem.

For some other cooperative schemes such as relay/antenna selection [29]–[32] and beamforming [9]–[11], the receiver needs to know f_m and g_n for all $m = 1, \cdots, M$ and $n = 1, \cdots, N$, to find the relay/antenna to be selected and calculate the beamforming coefficients (or feedback the CSI to the transmitter and relay). Thus, the channel vectors $\mathbf{f}$ and $\mathbf{g}$ need to be estimated at the receiver. The second problem to be investigated in this paper is the estimation of $\mathbf{f}$ and $\mathbf{g}$ at the receiver, which we address as the individual channel estimation problem.

B. Training Model

We conduct a two-step training process, where in the first step, space-time coding is used at the transmitter and in the second step, the relay antenna conducts AF with fixed gain [2], [28]. Each training step takes T symbol transmissions. Thus, the total time duration for training is $2T$. During the first step, the transmitter sends $\sqrt{P_1 T/M}\mathbf{S}$, where the $T \times M$ matrix $\mathbf{S}$ is the pilot, normalized as $\mathrm{tr}\{\mathbf{S}^*\mathbf{S}\} = M$. This normalization implies that the average transmit power of the transmitter is P_1 per transmission. The relay receives a $T \times 1$ vector:

$$\mathbf{r} = \sqrt{\frac{P_1 T}{M}}\mathbf{Sf} + \mathbf{u},$$

where $\mathbf{u}$ is the noise vector at the relay. We assume that the noises are i.i.d. CSCG with zero-mean and unit-variance, i.e., $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_T)$. During the second step, the relay amplifies its received signal and forwards to the receiver with power P_2. The fixed gain $\sqrt{P_2/(P_1+1)}$ is used. Define

$$\alpha \triangleq \frac{P_2}{P_1+1} \qquad \beta \triangleq \frac{P_1 P_2 T}{(P_1+1)M}.$$

The signal matrix received at the receiver, denoted as $\mathbf{X}$, can be calculated to be:

$$\mathbf{X} = \sqrt{\beta}\mathbf{SH} + \mathbf{W}, \tag{2}$$

where $\mathbf{W} \triangleq \sqrt{\alpha}\mathbf{ug} + \mathbf{V}$ with $\mathbf{V}$ the $T \times N$ noise matrix at the receiver. Again, we assume that entries of $\mathbf{V}$ are i.i.d. CSCG with zero-mean and unit-variance, i.e., $\mathrm{vec}(\mathbf{V}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NT})$. For a given realization of $\mathbf{g}$, $\mathrm{vec}(\mathbf{W})$ can be shown to be a CSCG random vector with zero mean. Its covariance matrix can be calculated to be

$$\mathbf{R}_\mathbf{W} \triangleq \left(\mathbf{I}_N + \alpha \mathbf{g}^t \bar{\mathbf{g}}\right) \otimes \mathbf{I}_T. \tag{3}$$

From the training equation (2), we have the following conditional probability density function (PDF):

$$p(\mathbf{X}|\mathbf{f},\mathbf{g}) = (2\pi)^{-TN} \det{}^{-1}(\mathbf{R}_\mathbf{W}) e^{-\mathrm{vec}(\mathbf{X}-\sqrt{\beta}\mathbf{SH})^* \mathbf{R}_\mathbf{W}^{-1} \mathrm{vec}(\mathbf{X}-\sqrt{\beta}\mathbf{SH})}. \tag{4}$$

$\mathbf{R}_\mathbf{W}$ is defined in (3), which depends on $\mathbf{g}$.

III. Estimation of the End-to-End Channels

In this section, we consider the estimation of the end-to-end channel matrix given in (1) at the receiver. An approximate ML estimation is proposed, which becomes the exact ML estimation when $T = M$. When $T \neq M$, simulation on the MSE in Section V implies that the proposed approximate ML estimation has close performance to the exact one.

The goal is to estimate $\mathbf{H}$ based on the observation $\mathbf{X}$ in (2). The training equation (2) has the same format as the traditional Gaussian observation model [33] or the training equation for a multiple-antenna direct communication system without relaying [23]. However, the end-to-end channel estimation for the relay network has two main differences to the traditional Gaussian estimation. First, in a traditional Gaussian model, the noise term is independent of the parameters to be estimated. Here, however, $\mathbf{R}_\mathbf{W}$, the noise covariance matrix, is a function of $\mathbf{g}$, which is related to $\mathbf{H}$, the matrix to be estimated. Second, in the traditional Gaussian model, entries of the vector/matrix to be estimated are assumed to be independent. Here, for the relay network, due to the special structure of $\mathbf{H}$ defined in (1), its rank is 1. Thus, entries of $\mathbf{H}$ are related. Of the total MN entries in $\mathbf{H}$, only $M+N-1$ of them are independent. These make the end-to-end channel estimation problem of the relay network with multiple transmit and receive antennas different and more challenging.

In this paper, instead of representing $\mathbf{H}$ by its all MN entries or $M+N-1$ independent entries, we decompose $\mathbf{H}$ as

$$\mathbf{H} = a\tilde{\mathbf{f}}\tilde{\mathbf{g}}, \tag{5}$$

where

$$a \triangleq \|\mathbf{f}\|_F \|\mathbf{g}\|_F, \quad \tilde{\mathbf{f}} \triangleq \frac{\mathbf{f}}{\|\mathbf{f}\|_F}, \quad \text{and} \quad \tilde{\mathbf{g}} \triangleq \frac{\mathbf{g}}{\|\mathbf{g}\|_F}. \tag{6}$$

$\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ are the directions of $\mathbf{f}$ and $\mathbf{g}$, respectively. They have unit-norm. For the simplicity of the presentation, define

$$a_\mathbf{f} \triangleq \|\mathbf{f}\|_F, \quad a_\mathbf{g} \triangleq \|\mathbf{g}\|_F. \tag{7}$$

We can see that (5) provides a map between $\mathbf{H}$ and the 3-tuple $(a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}})$. Also, $a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}}$ are mutually independent since entries of $\mathbf{f}$ and $\mathbf{g}$ are i.i.d. $\mathcal{CN}(0,1)$. The estimation of $\mathbf{H}$ can thus be transformed to the estimation of $(a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}})$. Since $\mathbf{H}$ is rank-1, it has a single non-zero singular value. The decomposition in (5) is the SVD of $\mathbf{H}$, where a is the non-zero singular value, and $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ are the corresponding left and right singular vectors. The proposed estimation is thus called SVD-based estimation.

Theorem 1: With the observation model in (2), define

$$\eta \triangleq \frac{\alpha a_{\mathbf{g}}^2}{1+\alpha a_{\mathbf{g}}^2}, \quad (8)$$

$$\mathbf{P} \triangleq \mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*, \quad (9)$$

$$\mathbf{Z} \triangleq [\mathbf{P} + \sqrt{\eta}(\mathbf{I}_T - \mathbf{P})]\mathbf{X}. \quad (10)$$

Let $\mathbf{Z} = \sum_{i=1}^{\min\{T,N\}} \sigma_{\mathbf{Z},i}\mathbf{u}_i\mathbf{v}_i^*$ be the singular value decomposition, where $\sigma_{\mathbf{Z},1} \geq \sigma_{\mathbf{Z},2} \geq \cdots \geq \sigma_{\mathbf{Z},\min\{T,N\}} \geq 0$ are the ordered singular values of $\mathbf{Z}$ and $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors corresponding to the singular value $\sigma_{\mathbf{Z},i}$. If $a_{\mathbf{g}}$ is known, the ML estimation of $(a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ is

$$\hat{a} = \frac{1}{\sqrt{\beta}}\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\mathbf{v}_1\|_F, \quad (11)$$

$$\hat{\tilde{\mathbf{f}}} = \frac{(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\mathbf{v}_1}{\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\mathbf{v}_1\|_F}, \quad \hat{\tilde{\mathbf{g}}} = \mathbf{v}_1^*. \quad (12)$$

The ML estimation of the end-to-end channel matrix is thus

$$\hat{\mathbf{H}} = \hat{a}\hat{\tilde{\mathbf{f}}}\hat{\tilde{\mathbf{g}}} = \frac{1}{\sqrt{\beta}}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\mathbf{v}_1\mathbf{v}_1^*. \quad (13)$$

Proof: See Appendix A. ■

In what follows, we discuss the results in Theorem 1.

The estimation in Theorem 1 is based on the assumption that $a_{\mathbf{g}}$ is known. In reality, $a_{\mathbf{g}}$, the norm of $\mathbf{g}$, is unknown. Thus η is unknown and the estimations given in (11)-(13) cannot be calculated. To obtain an estimation, we replace $a_{\mathbf{g}}^2$ with its mean, i.e., $a_{\mathbf{g}}^2 \approx \mathbb{E}(a_{\mathbf{g}}^2) = N$. With this approximation,

$$\eta \approx \frac{N\alpha}{1+N\alpha}.$$

An estimation of $\mathbf{H}$ can thus be found using (9)-(13). Generally speaking, this estimation is not the exact ML estimation but an approximate ML estimation.

The estimation $\hat{\mathbf{H}}$ in (13) is rank-1, which conforms with the structure of the end-to-end channel matrix in (1).

For the proposed estimation to be valid, from (11)-(13), $\mathbf{S}^*\mathbf{S}$ must be invertible, which implies a condition on the training time: $T \geq M$. The same requirement on the training time was derived for the ML channel estimation in a multiple-antenna system [23].

For the special case that $T = M$ and $\mathbf{S}$ is nonsingular, we have $\mathbf{P} = \mathbf{I}_T$ and $\mathbf{Z} = \mathbf{X}$ regardless of the $a_{\mathbf{g}}$ value. In this case, the estimation results in Theorem 1 are independent of $a_{\mathbf{g}}$. Thus the approximation on $a_{\mathbf{g}}^2$ is not needed and the estimation becomes the exact ML estimation. This is demonstrated in the following corollary.

Corollary 1: When $T = M$ and the pilot matrix $\mathbf{S}$ is nonsingular, let $\mathbf{X} = \sum_{i=1}^{\min\{T,N\}} \sigma_{\mathbf{X},i}\mathbf{u}_i\mathbf{v}_i^*$ be the singular value decomposition, where $\sigma_{\mathbf{X},1} \geq \sigma_{\mathbf{X},2} \geq \cdots \geq \sigma_{\mathbf{X},\min\{T,N\}} \geq 0$ are the ordered singular values of $\mathbf{X}$ and $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors corresponding to the singular value $\sigma_{\mathbf{X},i}$. The ML estimation of $(a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ is

$$\hat{a} = \frac{\sigma_{\mathbf{X},1}\|\mathbf{S}^{-1}\mathbf{u}_1\|_F}{\sqrt{\beta}}, \quad \hat{\tilde{\mathbf{f}}} = \frac{\mathbf{S}^{-1}\mathbf{u}_1}{\|\mathbf{S}^{-1}\mathbf{u}_1\|_F}, \quad \hat{\tilde{\mathbf{g}}} = \mathbf{v}_1^*. \quad (14)$$

The ML estimation of the end-to-end channel matrix is thus

$$\hat{\mathbf{H}} = \frac{\sigma_{\mathbf{X},1}}{\sqrt{\beta}}\mathbf{S}^{-1}\mathbf{u}_1\mathbf{v}_1^*. \quad (15)$$

Proof: The results in this corollary can be obtained directly from Theorem 1. ■

We can see from Corollary 1 that when $T = M$ and the pilot matrix is nonsingular, the ML estimations on a is a scaled version of the largest singular value of $\mathbf{X}$, the ML estimation on the direction of $\mathbf{g}$ is the Hermitian of the corresponding right singular vector, and the ML estimation on the direction of $\mathbf{f}$ is the corresponding left singular vector transformed by the inverse of the pilot.

It is noteworthy that the proposed SVD-based estimation uses the rank-1 property of the end-to-end channel matrix. For networks with single transmit and single receive antenna, the end-to-end channel is 1-dimensional, which is always rank-1. Thus we expect the proposed scheme to be favorable for networks with multiple transmit and/or multiple receive antennas.

IV. Estimation of the Transmitter-Relay and Relay-Receiver Channels

This section is on the individual estimation of the transmitter-relay channels, $\mathbf{f}$, and the relay-receiver channels, $\mathbf{g}$, both at the receiver. With the decompositions $\mathbf{f} = a_{\mathbf{f}}\tilde{\mathbf{f}}$, $\mathbf{g} = a_{\mathbf{g}}\tilde{\mathbf{g}}$, where $a_{\mathbf{f}}, \tilde{\mathbf{f}}, a_{\mathbf{g}}, \tilde{\mathbf{g}}$ are defined in (6) and (7), the problem is equivalent to estimating $a_{\mathbf{f}}, \tilde{\mathbf{f}}, a_{\mathbf{g}}, \tilde{\mathbf{g}}$. The training scheme is explained first. Then an approximate ML estimation is provided in closed-form. Finally, the exact ML estimation is investigated, which can be obtained numerically.

A. Training Scheme

If we use the training scheme proposed in Section II, for the tractability of analysis, we approximate $\mathbf{R}_{\mathbf{W}}$ with its average, which is $(1+\alpha)\mathbf{I}_{TN}$, to get, from (4),

$$p(\mathbf{X}|\mathbf{f},\mathbf{g}) \approx p_{appro}(\mathbf{X}|\mathbf{f},\mathbf{g}) \triangleq (2\pi)^{-TN}(1+\alpha)^{-TN}e^{-\frac{1}{1+\alpha}\|\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\|_F^2}. \quad (16)$$

In $p_{appro}(\mathbf{X}|\mathbf{f},\mathbf{g})$, $a_{\mathbf{f}}, a_{\mathbf{g}}$ appear as the product. It is thus impossible to estimate $a_{\mathbf{f}}$ and $a_{\mathbf{g}}$ separately due to the violation of the observability condition. Thus, extra observations other than the $\mathbf{X}$ in (2) are in need.

For this reason, in addition to the training explained in Section II, where the pilot $\mathbf{S}$ is sent by the transmitter and forwarded by the relay, an extra training stage is conducted, where the relay sends a $\check{T} \times 1$ unit-norm pilot vector $\check{\mathbf{s}}$, i.e., $\|\check{\mathbf{s}}\|_F = 1$, with power P_2. This takes $\check{T}$ symbol transmissions. Denote the received signal at the receiver of this extra stage as $\check{\mathbf{X}}$. We have

$$\check{\mathbf{X}} = \sqrt{P_2\check{T}}\check{\mathbf{s}}\mathbf{g} + \check{\mathbf{V}}, \quad (17)$$

where $\check{\mathbf{V}}$ is the noise whose entries are assumed to be i.i.d. CSCG with zero-mean and unit-variance. We call this training stage the second training stage and the one represented in (2) the first training stage. The total training time duration for the individual channel estimation is thus $2T + \check{T}$.

Equation (17) is the same as that of the training in a multiple-antenna system with 1 transmit antenna and N receive antennas, in which the ML estimation of $\mathbf{g}$ requires

$$p(\mathbf{X},\check{\mathbf{X}}|a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}) = p(\mathbf{X},\check{\mathbf{X}}|\mathbf{f},\mathbf{g}) = p(\mathbf{X}|\mathbf{f},\mathbf{g})p(\check{\mathbf{X}}|\mathbf{g}) = (2\pi)^{-TN}\left(1+\alpha a_{\mathbf{g}}^2\right)^{-T} e^{-\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})^*\mathbf{R}_{\mathbf{W}}^{-1}\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})}(2\pi)^{-\check{T}N}e^{-\left\|\check{\mathbf{X}}-\sqrt{P_2\check{T}}a_{\mathbf{g}}\check{\mathbf{s}}\tilde{\mathbf{g}}\right\|_F^2}. \tag{18}$$

$$\begin{aligned}
&\left(\hat{a}_{\mathbf{f}},\hat{\tilde{\mathbf{f}}},\hat{a}_{\mathbf{g}},\hat{\tilde{\mathbf{g}}}\right)\\
=&\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)}\left[\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})^*\mathbf{R}_{\mathbf{W}}^{-1}\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})+\left\|\check{\mathbf{X}}-\sqrt{P_2\check{T}}a_{\mathbf{g}}\check{\mathbf{s}}\tilde{\mathbf{g}}\right\|_F^2+T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\right]\\
=&\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)}\left(\mathrm{tr}\left[\left(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\left(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)(\mathbf{I}-\eta\tilde{\mathbf{g}}^*\tilde{\mathbf{g}})\right]\right.\\
&\left.\qquad+\left\|\check{\mathbf{X}}-\sqrt{P_2\check{T}}a_{\mathbf{g}}\check{\mathbf{s}}\tilde{\mathbf{g}}\right\|_F^2+T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\right)\\
=&\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)}\left[(1-\eta)\,\beta a_{\mathbf{f}}^2a_{\mathbf{g}}^2\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2-2\,(1-\eta)\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\Re\left(\tilde{\mathbf{f}}^*\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\right)-\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\right.\\
&\left.\qquad+\left(P_2\check{T}a_{\mathbf{g}}^2-2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*)\right)+T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\right]\\
=&\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)}\left[(1-\eta)\left(\sqrt{\beta}\|\mathbf{S}\tilde{\mathbf{f}}\|_Fa_{\mathbf{f}}a_{\mathbf{g}}-\frac{\Re\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F}\right)^2-(1-\eta)\frac{\Re^2\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}\right.\\
&\left.\qquad-\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+\left(P_2\check{T}a_{\mathbf{g}}^2-2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right)+T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\right].
\end{aligned}\tag{19}$$

$\check{T}\geq 1$. This is also sufficient for our relay network, where only extra information on $\|\mathbf{g}\|_F$ is required. Since $\check{T}$ is at least 1, for the individual chanel estimation, the time requirement for the second stage of training is $\check{T}\geq 1$.

B. ML Estimation Using an Approximate PDF

With both observations $\mathbf{X}$ and $\check{\mathbf{X}}$, we can investigate the ML estimations of $\mathbf{f},\mathbf{g}$, or equivalently, estimations of $a_{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{f}},\tilde{\mathbf{g}}$.

Theorem 2: With the observations $\mathbf{X}$ and $\check{\mathbf{X}}$, and the pilots $\mathbf{S}$ and $\check{\mathbf{s}}$, define

$$\mathbf{\Gamma}_1 \triangleq \mathbf{X}^*\mathbf{P}^*\mathbf{P}\mathbf{X}+(1+\alpha)\check{\mathbf{X}}^*\check{\mathbf{s}}\check{\mathbf{s}}^*\check{\mathbf{X}}. \tag{20}$$

Note that $\mathbf{\Gamma}_1$ is a positive semi-definite matrix. Let $\mathbf{\Gamma}_1=\sum_{i=1}^N\sigma_{\mathbf{\Gamma}_1,i}\mathbf{q}_i\mathbf{q}_i^*$ be the singular value decomposition, where $\sigma_{\mathbf{\Gamma}_1,1}\geq\sigma_{\mathbf{\Gamma}_1,2}\geq\cdots\geq\sigma_{\mathbf{\Gamma}_1,N}\geq 0$ are the ordered singular values and $\mathbf{q}_i$ is the singular vector corresponding to $\sigma_{\mathbf{\Gamma}_1,i}$. With the approximation in (16), the ML estimations of $a_{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{f}},\tilde{\mathbf{g}}$ are

$$\hat{\tilde{\mathbf{g}}}=\mathbf{q}_1^*e^{j\angle\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\mathbf{q}_1\right)},\quad \hat{\tilde{\mathbf{f}}}=\frac{(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*}{\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*\|_F}, \tag{21}$$

$$\hat{a}_{\mathbf{g}}=\frac{\check{\mathbf{s}}^*\check{\mathbf{X}}\hat{\tilde{\mathbf{g}}}^*}{\sqrt{P_2\check{T}}},\quad \hat{a}_{\mathbf{f}}=\sqrt{\frac{P_2\check{T}}{\beta}}\frac{\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*\|_F}{\check{\mathbf{s}}^*\check{\mathbf{X}}\hat{\tilde{\mathbf{g}}}^*}. \tag{22}$$

The ML estimations of $\mathbf{f},\mathbf{g}$ are thus

$$\hat{\mathbf{f}}=\hat{a}_{\mathbf{f}}\hat{\tilde{\mathbf{f}}},\quad \hat{\mathbf{g}}=\hat{a}_{\mathbf{g}}\hat{\tilde{\mathbf{g}}}. \tag{23}$$

Proof: See Appendix B. ■

Due to the approximation in (16), Theorem 2 provides approximate ML estimations of $\mathbf{f}$ and $\mathbf{g}$.

C. ML Estimation Using the Exact PDF

In the previous subsection, the approximate PDF in (16) is used. In this subsection, we work on the exact PDF (4) to obtain the exact ML estimation. Notice that

$$\det\mathbf{R}_{\mathbf{W}}=\det{}^T\left(\mathbf{I}_N+\alpha\mathbf{g}^t\bar{\mathbf{g}}\right)\det{}^N\mathbf{I}_T=\left(1+\alpha a_{\mathbf{g}}^2\right)^T.$$

Thus from (4),

$$p(\mathbf{X}|\mathbf{f},\mathbf{g})=(2\pi)^{-TN}(1+\alpha a_{\mathbf{g}}^2)^{-T} e^{-\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})^*\mathbf{R}_{\mathbf{W}}^{-1}\mathrm{vec}(\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}})}. \tag{24}$$

Given $\mathbf{f}$ and $\mathbf{g}$, $\mathbf{X}$ and $\check{\mathbf{X}}$ are independent Gaussian since $\mathbf{W}$ and $\check{\mathbf{V}}$ are independent Gaussian. Using the fact that $\check{\mathbf{X}}$ is also independent of $\mathbf{f}$, we have the equalities in (18) at the top of this page. By using the calculations in (34) and (37) in Appendix A and the definition of η in (8), the ML estimation problem can be analyzed using the equalities in (19) at the top of this page.

Notice that in the objective function, only the first term depends on $a_{\mathbf{f}}$. For the minimization, we need

$$\sqrt{\beta}\|\mathbf{S}\hat{\tilde{\mathbf{f}}}\|_F\hat{a}_{\mathbf{f}}\hat{a}_{\mathbf{g}}=\frac{\Re\left[\left(\mathbf{S}\hat{\tilde{\mathbf{f}}}\right)^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*\right]}{\|\mathbf{S}\hat{\tilde{\mathbf{f}}}\|_F},$$

or equivalently,

$$\hat{a}_{\mathbf{f}}=\frac{1}{\sqrt{\beta}\hat{a}_{\mathbf{g}}}\frac{\Re\left[\left(\mathbf{S}\hat{\tilde{\mathbf{f}}}\right)^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*\right]}{\|\mathbf{S}\hat{\tilde{\mathbf{f}}}\|_F^2}. \tag{25}$$

With this choice, only the second term in (19) depends on $\tilde{\mathbf{f}}$. Following the derivation in (39) of Appendix A, we need Equality (40) of Appendix A for the minimization.

The estimation problem thus reduces to

$$\begin{aligned}&\left(\hat{a}_{\mathbf{g}}, \hat{\tilde{\mathbf{g}}}\right)\\ =&\arg\max_{(a_{\mathbf{g}},\tilde{\mathbf{g}})}\Big[(1-\eta)\left\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\right\|_F^2+\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2-P_2\check{T}a_{\mathbf{g}}^2\\ &\qquad +2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)-T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\Big]\\ =&\arg\max_{(a_{\mathbf{g}},\tilde{\mathbf{g}})}\Big[\|\mathbf{Z}\tilde{\mathbf{g}}^*\|_F^2+2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)-P_2\check{T}a_{\mathbf{g}}^2\\ &\qquad -T\ln\left(1+\alpha a_{\mathbf{g}}^2\right)\Big],\end{aligned}\qquad (26)$$

where the second equality is obtained using Equality (42) in Appendix A and $\mathbf{Z}$ is defined in (8). In general, $\mathbf{Z}$ depends on $a_{\mathbf{g}}$. The optimization problem in (26) is non-convex, and analytical solution is unavailable. In the following, we consider the numerical alternative optimization of $\tilde{\mathbf{g}}$ and $a_{\mathbf{g}}$.

1) Finding $\tilde{\mathbf{g}}$ Given $a_{\mathbf{g}}$: For a fixed $a_{\mathbf{g}}$, the optimization over $\tilde{\mathbf{g}}$ is

$$\begin{aligned}&\max_{\tilde{\mathbf{g}}}\left[\|\mathbf{Z}\tilde{\mathbf{g}}^*\|_F^2+2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right]\\ &\text{subject to}\quad \|\tilde{\mathbf{g}}\|_F=1.\end{aligned}\qquad (27)$$

The optimization problem in (27) is a special case of quadratic constrained quadratic programming (QCQP). The general QCQP is NP-hard. In this paper, we use an exhaustive search by searching the $\tilde{\mathbf{g}}$ over a grid inside the unit hyper-sphere that results in the highest objective value. The complexity of this exhaustive search is exponential in N, the dimension of $\mathbf{g}$. For large N, the complexity is too high for its practical employment. Actually, the main purpose of the exact ML estimation in this paper is to serve as a benchmark for the approximate estimation proposed in Section IV-B.

2) Finding $a_{\mathbf{g}}$ Given $\tilde{\mathbf{g}}$: For a fixed $\tilde{\mathbf{g}}$, we can conduct the optimization over $a_{\mathbf{g}}$ in (26) as follows.

For the special case of $T = M$ and $\mathbf{S}$ is nonsingular, we have $\mathbf{Z} = \mathbf{X}$, independent of $a_{\mathbf{g}}$. Taking the derivative of the objective function over $a_{\mathbf{g}}$ and setting it to 0 leads to the following equation for the optimal $a_{\mathbf{g}}$:

$$\sqrt{P_2\check{T}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)-P_2\check{T}\hat{a}_{\mathbf{g}}-\frac{T\alpha\hat{a}_{\mathbf{g}}}{1+\alpha\hat{a}_{\mathbf{g}}^2}=0,$$

from which we get

$$\begin{aligned}&\alpha P_2\check{T}\hat{a}_{\mathbf{g}}^3-\alpha\sqrt{P_2\check{T}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\hat{a}_{\mathbf{g}}^2+(P_2\check{T}+\alpha T)\hat{a}_{\mathbf{g}}\\ &\qquad -\sqrt{P_2\check{T}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)=0.\end{aligned}\qquad (28)$$

If $\tilde{\mathbf{g}}$ is the solution of (27), from the discussion in the proof of Theorem 2, $\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*$ must be a positive real number. Thus Equation (28) must have a real positive root. $\hat{a}_{\mathbf{g}}$ is the positive root of (28) that results in the highest objective value.

For the general case of $T > M$, $\mathbf{Z}$ depends on $a_{\mathbf{g}}$. Efficient numerical methods can be used in finding the optimal $a_{\mathbf{g}}$ since the objective function is explicitly known in (26). For example, we can calculate the derivative of the objective function and find its roots using Newton's method. The optimal $a_{\mathbf{g}}$ is the root that results in the largest objective value.

Once $\hat{a}_{\mathbf{g}}$ and $\hat{\tilde{\mathbf{g}}}$ are solved, the ML estimations of $\tilde{\mathbf{f}}$ and $a_{\mathbf{f}}$ can be calculated using (40) and (25). Based on the above discussion, Algorithm 1 is proposed. It is noteworthy that, in general, the solution Algorithm 1 founds is not guaranteed to be the global optimal, i.e., the exact ML estimation. One method to increase the possibility of finding the optimal solution is to choose multiple initial values and use the best one among the found solutions. During simulation, we also observe that the alternative optimization in Algorithm 1 converges in a few iterations.

Algorithm 1 Algorithm for the exact ML estimations of the transmitter-relay channels and the relay-receiver channels.

1: Initialize $\hat{a}_{\mathbf{g}}$, e.g., $\hat{a}_{\mathbf{g}} = 1$.
2: Search over a grid inside the unit hyper-sphere to find the solution of (27), denoted as $\hat{\tilde{g}}$.
3: Find the $\hat{a}_{\mathbf{g},new}$ that solves (26) using the previously found $\hat{\tilde{g}}$. If $T = M$, this can be achieved by finding the best real root of (28). Otherwise, use Newton's method for the optimization.
4: If $|\hat{a}_{\mathbf{g},new} - \hat{a}_{\mathbf{g}}|$ is larger than a predetermined threshold, $\hat{a}_{\mathbf{g}} = \hat{a}_{\mathbf{g},new}$, and go to Step 2.
5: Calculate $\hat{\tilde{\mathbf{f}}}$, $\hat{a}_{\mathbf{f}}$ using (40) and (25), respectively.

V. Simulation Results

In this section, we simulate the MSEs of the proposed estimations, including the estimation of the end-to-end channel matrix $\mathbf{H}$ in Section III and the individual estimations of $\mathbf{f}$ and $\mathbf{g}$ in Section IV. Since the focus of this paper is on the estimation rules, optimizations of the training pilot, power allocation, and training time are not considered. They are interesting and important issues but are beyond the scope of this paper, and thus left for future work. For the end-to-end channel matrix training, we set $\mathbf{S} = \mathbf{I}_T$ when $T = M$; otherwise, $\mathbf{S}$ is a randomly generated $T \times M$ unitary matrix. For the individual channel training, the same setting is used for the first training stage. For the second training stage, we set $\check{T} = 1$, in which case $\check{\mathbf{s}}$ reduces to a scalar $\check{s}$, and we let $\check{s} = 1$. We further set $P_1 = P_2 = P$ to avoid optimization of the power allocation.

A. MSE on the End-to-End Channel Estimation

In this subsection, we show the simulated MSE on the end-to-end channel estimation defined as $\mathrm{MSE}(\mathbf{H}) \triangleq \mathbb{E}\{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2\}$. Note that in traditional ML estimation, the MSE is the power of the estimation error averaged over the noises. In our estimation model, in addition to the noises, the channels are also random, the average is over not only the noises but also the random channels. In the conducted Monte-Carlo simulation, a distinct channel realization is used for each iteration. The channels are modeled as CSCG random variables with zero-mean and unit variance. The calculated MSE is the average power of the estimation error over a large number of channel realizations.

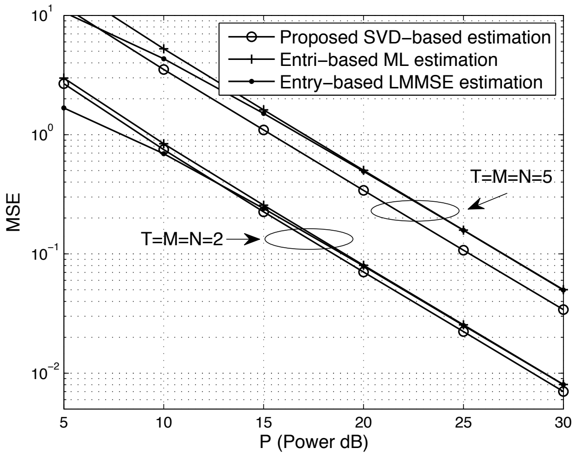


Fig. 2. MSE on $\mathbf{H}$ for networks with $T = M = N = 2$ and $T = M = N = 5$.

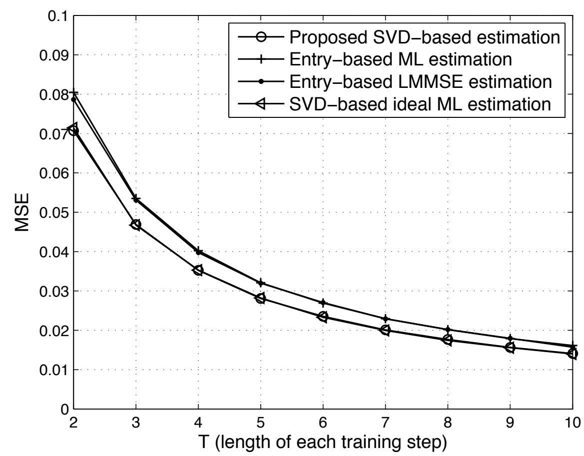


Fig. 3. MSE on $\mathbf{H}$ of a network with $M = N = 2$, $P = 20$dB.

We compare the proposed SVD-based estimation in Theorem 1 with entry-based estimations explained in what follows. From (2), if entries of $\mathbf{H}$ are seen as independent parameters to be estimated and neither the special structure of $\mathbf{H}$ nor the relation between $\mathbf{H}$ and the noise covariance matrix is taken into account, the estimation problem can be seen as a traditional Gaussian one and the ML estimation and LMMSE estimation are, respectively,

$$\hat{\mathbf{H}}_{\text{entry,ML}} = \frac{1}{\sqrt{\beta}}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X},$$
$$\hat{\mathbf{H}}_{\text{entry,LMMSE}} = \sqrt{\beta}\left[(1+\alpha)\mathbf{I}_M + \beta\mathbf{S}^*\mathbf{S}\right]^{-1}\mathbf{S}^*\mathbf{X}. \quad (29)$$

In Figure 2, we show the MSE as a function of P with $T = M = N = 2$ and $T = M = N = 5$. For these two settings, the proposed SVD-based estimation is the exact ML estimation. We can see that when P is large, the proposed SVD-based estimation is about 0.5dB and 1.5dB better than the two entry-based estimations (ML and LMMSE) respectively. At low P values, the proposed scheme is slightly worse than the entry-based LMMSE estimation but still better than the entry-based ML estimation. For all estimations, the MSEs scale as $O(P^{-1})$, that is, the MSEs decrease linearly with respect to the training power P. This is important in achieving full diversity in data transmission [27].

In Figure 3, we show the MSE as a function of the training step size T for a network with $M = N = 2$ and $P = 20$dB. We can see that the MSE decreases and converges as T increases. This shows that increasing the training time will decrease the MSE but only to some positive level. The proposed estimation always has a lower MSE than the two entry-based ones. In this experiment, since $T \neq M$ for $T > 2$, the proposed SVD-based estimation is an approximate ML estimation. To assess the approximation, we also show the simulated MSE of an SVD-based ideal ML estimation, in which, the true value of $a_{\mathbf{g}}$ is used. In reality, $a_{\mathbf{g}}$ is unknown. Thus the true ML estimation is expected to perform no better than this ideal one. Figure 3 shows that the proposed estimation performs about the same as the ideal ML estimation, implying that it is close to the exact ML estimation.

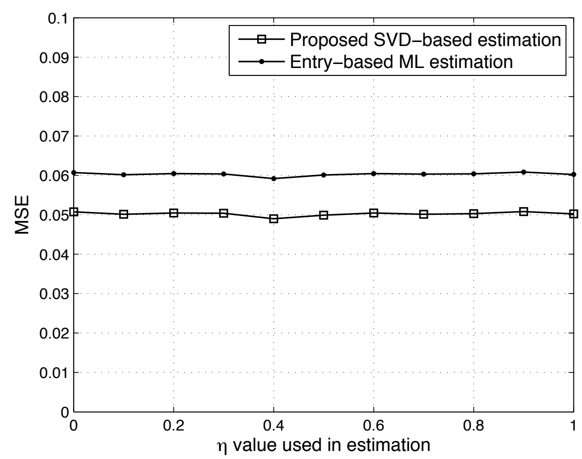


Fig. 4. MSE on $\mathbf{H}$ of a network with $M = 2$, $N = 3$, $T = 4$, $P = 20$dB.

Next, we investigate the effect of η on the proposed estimation. From the definition in (8), $\eta \in (0, 1)$. We simulate a network with $M = 2$, $N = 3$, $T = 4$, $P = 20$dB and draw the MSE of the proposed estimation as η is approximated by different values. We can see from Figure 4 that the effect of the value of η on the MSE is negligible. The maximum MSE fluctuation (ratio of the absolute difference between an MSE and the average MSE to the average MSE) is about 1.1%. The entry-based ML estimation is also shown. We can see that for all possible η values, the proposed SVD-based estimation has lower MSEs. The entry-based ML estimation is independent of η. Its MSE fluctuation, whose largest value is 0.9%, is due to simulation imprecision only. This further shows the unimportance of the value of η in the proposed estimation.

B. MSEs on the Transmitter-Relay and Relay-Receiver Channel Estimations

In this subsection, we simulate the average MSEs on the proposed individual estimations of the transmitter-relay channels and relay-receiver channels in Section IV. Define $\text{MSE}(\mathbf{f}) \triangleq \mathbb{E}\{\|\mathbf{f}-\hat{\mathbf{f}}\|_F^2\}$ and $\text{MSE}(\mathbf{g}) \triangleq \mathbb{E}\{\|\mathbf{g}-\hat{\mathbf{g}}\|_F^2\}$, where the average is over both the noises and the channels.

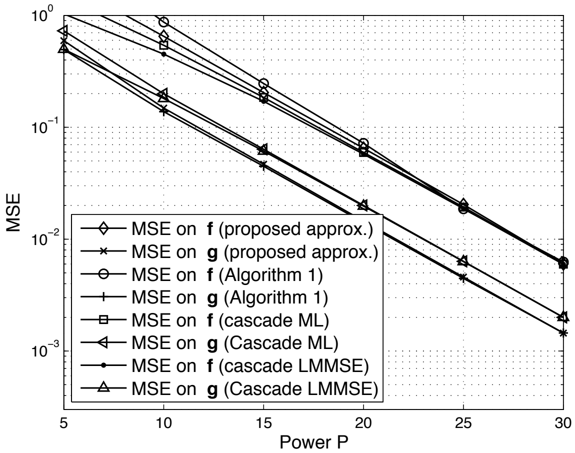


Fig. 5. MSEs on $\mathbf{f}$ and $\mathbf{g}$ of a network with $M = N = T = 2$.

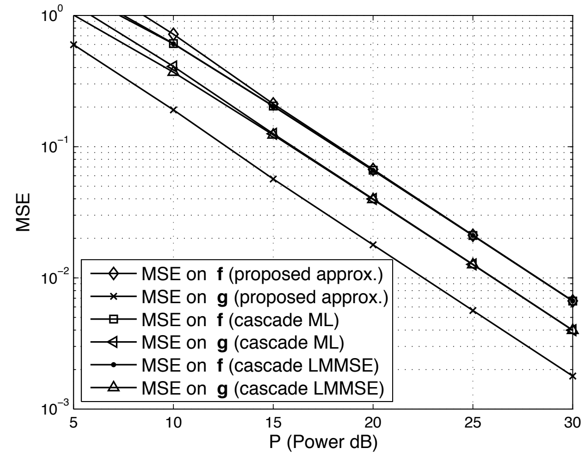


Fig. 6. MSEs on $\mathbf{f}$ and $\mathbf{g}$ of a network with $M = N = T = 4$.

The proposed estimation with the approximate PDF (proposed approximate estimation) and proposed estimation with exact PDF (Algorithm 1) are compared with cascade estimation schemes [26], where $\mathbf{g}$ is estimated first, based on which an estimation on $\mathbf{f}$ is then obtained. In [26], cascade LMMSE estimation is investigated for relay networks using DSTC. Here, we consider both cascade ML and cascade LMMSE estimations for comparison. First, based on the observation $\check{\mathbf{X}}$ in the second training stage only, the ML and LMMSE estimations of $\mathbf{g}$ can be obtained as

$$\hat{\mathbf{g}}_{\text{cas,ML}} = \frac{\check{\mathbf{s}}^*\check{\mathbf{X}}}{\sqrt{P_2\check{T}}}, \quad \hat{\mathbf{g}}_{\text{cas,LMMSE}} = \frac{\sqrt{P_2\check{T}}\check{\mathbf{s}}^*\check{\mathbf{X}}}{1+P_2\check{T}}, \tag{30}$$

respectively. To estimate $\mathbf{f}$, we rewrite the training equation of the first training stage (2) as:

$$\text{vec}(\mathbf{X}) = \sqrt{\beta}\mathbf{S}_g\mathbf{f} + \text{vec}(\mathbf{W}), \tag{31}$$

where $\mathbf{S}_g \triangleq (\mathbf{g}^t \otimes \mathbf{I}_T)\mathbf{S}$. When $\mathbf{g}$ is known, both $\mathbf{S}_g$ and the noise covariance matrix are known. Equation (31) is a traditional Gaussian observation model with $\mathbf{S}_g$ the pilot and $\mathbf{f}$ the vector to be estimated. The ML and LMMSE estimations are, respectively,

$$\hat{\mathbf{f}}_{\text{cas,ML}} = \frac{1}{\sqrt{\beta}}\left(\mathbf{S}_{\mathbf{g}}^*\mathbf{R}_{\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{g}}\right)^{-1}\mathbf{S}_{\mathbf{g}}^*\mathbf{R}_{\mathbf{W}}^{-1}\text{vec}(\mathbf{X}), \tag{32}$$

$$\hat{\mathbf{f}}_{\text{cas,LMMSE}} = \sqrt{\beta}\left(\mathbf{I}_M + \beta\mathbf{S}_{\mathbf{g}}^*\mathbf{R}_{\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{g}}\right)^{-1}\mathbf{S}_{\mathbf{g}}^*\mathbf{R}_{\mathbf{W}}^{-1}\text{vec}(\mathbf{X}), \tag{33}$$

where $\mathbf{R}_{\mathbf{W}}$ is given in (3). In reality, the estimation of $\mathbf{g}$ in (30) is used in (32) and (33) to obtain an estimation on $\mathbf{f}$, i.e., redefining $\mathbf{S}_g$ as $\mathbf{S}_g \triangleq (\hat{\mathbf{g}}^t \otimes \mathbf{I})\mathbf{S}$ and $\mathbf{R}_{\mathbf{W}}$ as $\mathbf{R}_{\mathbf{W}} \triangleq \left(\mathbf{I}_N + \alpha\hat{\mathbf{g}}^t\bar{\hat{\mathbf{g}}}\right) \otimes \mathbf{I}_T$, where $\hat{\mathbf{g}}$ is the cascade ML or cascade LMMSE estimation of $\mathbf{g}$ correspondingly.

Figure 5 shows the MSEs as a function of P for a network with $T = M = N = 2$. We compare the MSEs on both $\mathbf{f}$ and $\mathbf{g}$ of the proposed ML estimation with the approximate PDF (proposed approximate estimation), the proposed estimation with exact PDF (Algorithm 1), the cascade ML estimation, and the cascade LMMSE estimation. For the MSE on $\mathbf{f}$, compared with the two cascade estimations, the two proposed estimations are slightly worse at low values of P. As P increases, the difference diminishes to zero. For the MSE on $\mathbf{g}$, the proposed estimations are about 1.5dB better than the cascade ones. This is because while cascade estimations use only the observation $\check{\mathbf{X}}$ in estimating $\mathbf{g}$, the two proposed estimations use both $\check{\mathbf{X}}$ and $\mathbf{X}$. The estimation using the approximate PDF has close performance to the estimation using the exact PDF (Algorithm 1), especially for large P. Thus, in practice, the approximate ML estimation is more desirable since it has a closed-form solution. Another useful observation is that all MSEs have the scaling $O(P^{-1})$. This is important for full diversity in data transmission [27]. In Figure 6, we simulate a network with $T = M = N = 4$. We only show the MSE of the proposed estimation with approximate PDF (proposed approximate estimation) since simulation of the estimation using the exact PDF (Algorithm 1) is computationally too costly. Similar phenomenon can be observed. With respect to $\mathbf{g}$, the proposed estimation is about 3.5dB better than the two cascade estimations. With respect to $\mathbf{f}$, it has about the same performance.

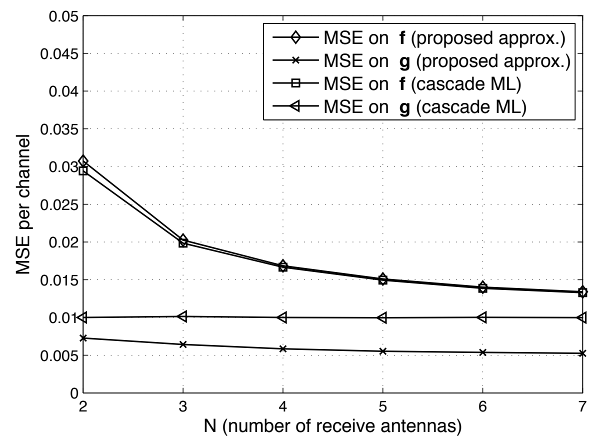


Fig. 7. MSEs on $\mathbf{f}$ and $\mathbf{g}$ of a network with $T = M = 2, P = 20$dB.

Figure 7 shows the MSE as a function of N, the number of receive antennas, while we set $T = M = 2$ and $P = 20$dB. When N changes, the dimension of $\mathbf{g}$ changes, thus for fair comparison, we show the MSE per channel. We also show the cascade ML estimation for comparison. The figure shows

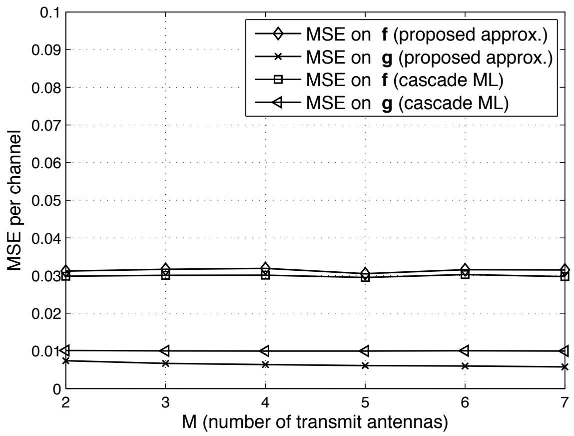


Fig. 8. MSEs on $\mathbf{f}$ and $\mathbf{g}$ of a network with $T = M, N = 2, P = 20$dB.

that for both the proposed and cascade ML estimations, with a fixed training time, the per channel MSE on $\mathbf{f}$ decreases as N increases. For the proposed scheme, the per channel MSE on $\mathbf{g}$ decreases slightly as N increases, while it keeps unchanged for the cascade scheme.

Figure 8 shows the MSE as a function of M, the number of transmit antennas, while we set $N = 2$, $T = M$, and $P =$ 20dB. Thus as M increases, the training time also increases. We can see that for both estimations, the per channel MSEs on both $\mathbf{f}$ and $\mathbf{g}$ keep unchanged with M.

VI. Extension to MIMO Relay Networks

In this section, we discuss the extension of the proposed channel estimations to the general MIMO relay network with multiple relay antennas as shown in Figure 9. The transmitter has M antennas, the receiver has N antennas, and there are in total R antennas at the relays. The relay antennas can be either co-located at one relay or distributively located at different relays. Thus this network contains one transmitter node, one receiver node, and multiple relay nodes, where every node can be equipped with multiple antennas. Denote the channel from the mth transmit antenna to the rth relay antenna as f_{mr} and the channel from the rth relay antenna to the nth receive antenna as g_{rn}. The channel vector from the transmitter to the rth relay antenna and the channel vector from the rth relay antenna to the receiver are thus,

$$\mathbf{f}_r \triangleq \left[\begin{array}{cccc} f_{1r} & f_{2r} & \cdots & f_{Mr} \end{array} \right]^t,$$
$$\mathbf{g}_r \triangleq \left[\begin{array}{cccc} g_{r1} & g_{r2} & \cdots & g_{rN} \end{array} \right],$$

respectively. All channels are assumed to be i.i.d. $\mathcal{CN}(0,1)$.

The end-to-end channel from the mth transmit antenna to the nth receive antenna via the rth relay antenna is $f_{mr}g_{rn}$. By using matrix and vector presentation, the $M \times N$ end-to-end channel matrix from the transmitter to the receiver via the rth relay antenna is $\mathbf{H}_r = \mathbf{f}_r\mathbf{g}_r$. The end-to-end channel estimation problem is the estimation of $\mathbf{H}_1, \cdots, \mathbf{H}_R$. The individual channel estimation problem is the estimation of $\mathbf{f}_1, \cdots, \mathbf{f}_R$ and $\mathbf{g}_1, \cdots, \mathbf{g}_R$.

A straightforward extension of the proposed estimation schemes to the general MIMO relay network is to solve the

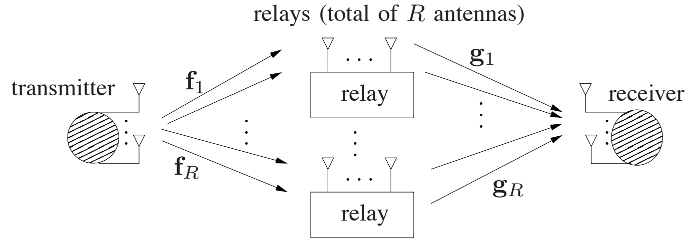


Fig. 9. MIMO multiple-relay network.

estimation problems related to each relay antenna one by one. With respect to each relay antenna, the problems become the same as those for a network with single relay antenna. In other words, for the end-to-end channel estimation, we consider the estimation of $\mathbf{H}_r$ for different r separately and sequentially; for the individual channel estimation, we consider the estimation of $\mathbf{f}_r$ and $\mathbf{g}_r$ separately and sequentially. Since channels related to different relay antennas are independent, this extension does not lose generality or optimality in design[1].

VII. Conclusions

For a relay network with multiple transmit and receive antennas but a single relay antenna, following the maximum likelihood approach, we proposed a scheme for the the receiver to estimate the end-to-end channels, and a scheme for the receiver to individually estimate both the transmitter-relay channels and the relay-receiver channels. The estimation of the end-to-end channels was based on the singular value decomposition and estimations on the singular value and singular vectors. The proposed approximate maximum likelihood estimation was shown to perform close to the exact maximum likelihood estimation and was superior to entry-based estimations. For the individual channel estimation, the lengths and the directions of the channel vectors were estimated and the results were combined to obtain the estimations of the channel vectors. Maximum likelihood estimations were proposed using both the exact and an approximate probability density functions of the received signals during training. The former requires numerical search and the latter is analytical. Simulation on the mean square error showed that the two proposed estimations have similar performance. Compared with previously proposed cascade estimations of the individual channels, the proposed estimations achieved better performance for the relay-receiver channels. Extensions to the general multiple-antenna multi-relay network were discussed.

Appendix

A. Proof of Theorem 1

Using the decomposition of $\mathbf{H}$ in (5), we look for estimations on a, $\tilde{\mathbf{f}}$, and $\tilde{\mathbf{g}}$. From (3), $\mathbf{R}_\mathbf{W} = \left(\mathbf{I}_N + \alpha a_\mathbf{g}^2 \tilde{\mathbf{g}}^t \bar{\tilde{\mathbf{g}}}\right) \otimes \mathbf{I}_T$. For a realization of $(a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}})$, if $a_\mathbf{g}$ is further known, $\mathbf{R}_\mathbf{W}$ is a known matrix. From (4), the PDF of $\mathbf{X}|a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}}, a_\mathbf{g}$ is

$$p(\mathbf{X}|a, \tilde{\mathbf{f}}, \tilde{\mathbf{g}}, a_\mathbf{g}) =$$
$$(2\pi)^{-TN} \det{}^{-1}(\mathbf{R}_\mathbf{W}) e^{-\mathrm{vec}(\mathbf{X}-\sqrt{\beta}\mathbf{S}\mathbf{H})^* \mathbf{R}_\mathbf{W}^{-1} \mathrm{vec}(\mathbf{X}-\sqrt{\beta}\mathbf{S}\mathbf{H})}.$$

[1]It is noteworthy that for the end-to-end channel estimation, the antenna-by-antenna extension with respect to the transmit and receive antennas may cause sub-optimality in design because channel entries corresponding to different transmit and receive antennas are related, instead of independent.

$$\begin{aligned}
&\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^* \mathbf{R}_{\mathbf{W}}^{-1}\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\\
&=\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\left[\left(\mathbf{I}_N+\alpha a_{\mathbf{g}}^2\tilde{\mathbf{g}}^t\bar{\tilde{\mathbf{g}}}\right)\otimes\mathbf{I}_T\right]^{-1}\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\\
&=\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\left[\left(\mathbf{I}_N+\alpha a_{\mathbf{g}}^2\tilde{\mathbf{g}}^t\bar{\tilde{\mathbf{g}}}\right)^{-1}\otimes\mathbf{I}_T\right]\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\\
&=\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\text{vec}\left[\mathbf{I}_T\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\left(\mathbf{I}_N+\alpha a_{\mathbf{g}}^2\tilde{\mathbf{g}}^t\bar{\tilde{\mathbf{g}}}\right)^{-T}\right]\\
&=\text{tr}\left[\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\left(\mathbf{I}_N+\alpha a_{\mathbf{g}}^2\tilde{\mathbf{g}}^*\tilde{\mathbf{g}}\right)^{-1}\right]. \qquad (34)
\end{aligned}$$

$$\begin{aligned}
F\left(a,\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)&\triangleq\text{tr}\left[\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)\left(\mathbf{I}_N-\eta\tilde{\mathbf{g}}^*\tilde{\mathbf{g}}\right)\right]\\
&=(1-\eta)\,a^2\beta\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2-2\,(1-\eta)\,a\sqrt{\beta}\Re\left(\tilde{\mathbf{f}}^*\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\right)+\|\mathbf{X}\|_F^2-\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\\
&=(1-\eta)\,\beta\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2\left(a-\frac{\Re\left(\tilde{\mathbf{f}}^*\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\right)}{\sqrt{\beta}\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}\right)^2-(1-\eta)\frac{\Re^2\left(\tilde{\mathbf{f}}^*\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\right)}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}+\|\mathbf{X}\|_F^2-\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2. \qquad (35)
\end{aligned}$$

The ML estimation finds the parameter value that maximizes the PDF, i.e.,

$$\left(\hat{a},\hat{\tilde{\mathbf{f}}},\hat{\tilde{\mathbf{g}}}\right)=\arg\max_{(a,\tilde{\mathbf{f}},\tilde{\mathbf{g}})}p(\mathbf{X}|a,\tilde{\mathbf{f}},\tilde{\mathbf{g}},a_{\mathbf{g}})=$$
$$\arg\min_{(a,\tilde{\mathbf{f}},\tilde{\mathbf{g}})}\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right)^*\mathbf{R}_{\mathbf{W}}^{-1}\text{vec}\left(\mathbf{X}-a\sqrt{\beta}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right). \qquad (36)$$

We recall the following identities for Kronecker product and vectorization:

$$\begin{aligned}
&(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1},\\
&\text{vec}\left(\mathbf{ABC}\right)=\left(\mathbf{C}^t\otimes\mathbf{A}\right)\text{vec}\left(\mathbf{B}\right),\\
&\text{vec}\left(\mathbf{A}\right)^*\text{vec}\left(\mathbf{B}\right)=\text{tr}\left(\mathbf{A}^*\mathbf{B}\right).
\end{aligned}$$

Applying these identities, we have the derivations in (34) at the top of this page.

Using the Sherman-Morrison-Woodbury formula (Page 48 of [34]), we have

$$\begin{aligned}
&\left(\mathbf{I}_N+\alpha a_{\mathbf{g}}^2\tilde{\mathbf{g}}^*\tilde{\mathbf{g}}\right)^{-1}\\
=&\mathbf{I}_N-\tilde{\mathbf{g}}^*\left[\left(\alpha a_{\mathbf{g}}^2\right)^{-1}+\tilde{\mathbf{g}}\tilde{\mathbf{g}}^*\right]^{-1}\tilde{\mathbf{g}}\\
=&\mathbf{I}_N-\eta\tilde{\mathbf{g}}^*\tilde{\mathbf{g}}, \qquad (37)
\end{aligned}$$

where η is defined in (8) and $\eta\in(0,1)$. By using (34) and (37), the objective function in (36) simplifies to $F\left(a,\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)$ defined in (35) at the top of this page. Thus, to minimize $F\left(a,\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)$, we need

$$a=\frac{\Re\left(\tilde{\mathbf{f}}^*\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\right)}{\sqrt{\beta}\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}. \qquad (38)$$

With this choice of a, the optimization problem in (36) reduces to

$$\left(\hat{\tilde{\mathbf{f}}},\hat{\tilde{\mathbf{g}}}\right)=\arg\max_{(\tilde{\mathbf{f}},\tilde{\mathbf{g}})}G\left(\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right),$$

where

$$G\left(\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)\triangleq(1-\eta)\frac{\Re^2\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}+\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2.$$

From the definition of $\mathbf{P}$ in Theorem 1, we have

$$\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*=\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*.$$

Thus

$$\begin{aligned}
G\left(\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)=&(1-\eta)\frac{\Re^2\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}+\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\\
\le&(1-\eta)\frac{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}+\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\\
=&(1-\eta)\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+\eta\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\triangleq H(\tilde{\mathbf{g}}) \qquad (39)
\end{aligned}$$

with equality when $\mathbf{S}\tilde{\mathbf{f}}=\gamma\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*$ for some real γ. So, when $\tilde{\mathbf{f}}=\gamma(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*$, $G\left(\tilde{\mathbf{f}},\tilde{\mathbf{g}}\right)$ is maximized. Since $\tilde{\mathbf{f}}$ has unit-norm, we have

$$\tilde{\mathbf{f}}=\frac{(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*}{\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\tilde{\mathbf{g}}^*\|_F}. \qquad (40)$$

With this choice of $\tilde{\mathbf{f}}$, the optimization becomes the maximization of $H(\tilde{\mathbf{g}})$ defined in (39). Notice that $\mathbf{P}$ is a projection matrix. Thus $(\mathbf{P}\mathbf{X})^*[(\mathbf{I}_T-\mathbf{P})\mathbf{X}]=\mathbf{0}$ and we have

$$\begin{aligned}
H(\tilde{\mathbf{g}})=&\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+\eta\left(\|\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2-\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\right)\\
=&\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+\eta\|(\mathbf{I}_T-\mathbf{P})\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2\\
=&\|\left[\mathbf{P}+\sqrt{\eta}(\mathbf{I}_T-\mathbf{P})\right]\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2=\|\mathbf{Z}\tilde{\mathbf{g}}^*\|_F^2, \qquad (42)
\end{aligned}$$

where in the last step, the definition of $\mathbf{Z}$ in (10) is used. With the SVD given in Theorem 1, it is clear that $H(\tilde{\mathbf{g}})$ is maximized when $\tilde{\mathbf{g}}^*=\mathbf{v}_1$, the right singular vector of the largest singular value. That is, the ML estimation of $\tilde{\mathbf{g}}$ is $\hat{\tilde{\mathbf{g}}}=\mathbf{v}_1^*$. By using this result in (40) then (38), the ML estimations are obtained as in (11)-(13).

B. Proof of Theorem 2

For given $\mathbf{f}$ and $\mathbf{g}$, $\mathbf{X}$ and $\check{\mathbf{X}}$ are independent Gaussian since $\mathbf{W}$ and $\check{\mathbf{V}}$ are independent Gaussian. Using the approximation

$$\begin{aligned}
\left(\hat{a}_{\mathbf{f}},\hat{\tilde{\mathbf{f}}},\hat{a}_{\mathbf{g}},\hat{\tilde{\mathbf{g}}}\right) &= \arg\max_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)} p\left(\mathbf{X},\check{\mathbf{X}}|a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)\\
&=\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)} \left[\left\|\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right\|_F^2+(1+\alpha)\left\|\check{\mathbf{X}}-\sqrt{P_2\check{T}}a_{\mathbf{g}}\check{\mathbf{s}}\tilde{\mathbf{g}}\right\|_F^2\right]\\
&=\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)} \left[\beta a_{\mathbf{f}}^2a_{\mathbf{g}}^2\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2-2\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\Re\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]+(1+\alpha)\left(P_2\check{T}a_{\mathbf{g}}^2-2\sqrt{P_2\check{T}}a_{\mathbf{g}}\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right)\right]\\
&=\arg\min_{\left(a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}\right)} \left[\left(\sqrt{\beta}\|\mathbf{S}\tilde{\mathbf{f}}\|_Fa_{\mathbf{f}}a_{\mathbf{g}}-\frac{\Re\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F}\right)^2+(1+\alpha)\left(\sqrt{P_2\check{T}}a_{\mathbf{g}}-\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right)^2\right.\\
&\qquad\left.-\frac{\Re^2\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}-(1+\alpha)\Re^2\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right].
\end{aligned} \tag{41}$$

in (16) and the fact that $\check{\mathbf{X}}$ is independent of $\mathbf{f}$, we have

$$\begin{aligned}
&p(\mathbf{X},\check{\mathbf{X}}|a_{\mathbf{f}},\tilde{\mathbf{f}},a_{\mathbf{g}},\tilde{\mathbf{g}}) = p(\mathbf{X}|\mathbf{f},\mathbf{g})p(\check{\mathbf{X}}|\mathbf{g})\\
&\approx(2\pi)^{-TN}(1+\alpha)^{-TN}e^{-\frac{1}{1+\alpha}\left\|\mathbf{X}-\sqrt{\beta}a_{\mathbf{f}}a_{\mathbf{g}}\mathbf{S}\tilde{\mathbf{f}}\tilde{\mathbf{g}}\right\|_F^2}\\
&\qquad(2\pi)^{-\check{T}N}e^{-\left\|\check{\mathbf{X}}-\sqrt{P_2\check{T}}a_{\mathbf{g}}\check{\mathbf{s}}\tilde{\mathbf{g}}\right\|_F^2}.
\end{aligned}$$

The ML estimation problem can thus be analyzed as the equalities in (41) at the top of this page.

To minimize the objective function in (41), we need

$$\sqrt{P_2\check{T}}a_{\mathbf{g}}=\Re\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right). \tag{43}$$

and

$$\sqrt{\beta}\|\mathbf{S}\tilde{\mathbf{f}}\|_Fa_{\mathbf{f}}a_{\mathbf{g}}=\frac{\Re\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F}. \tag{44}$$

With the choices of $a_{\mathbf{f}}$ and $a_{\mathbf{g}}$ that satisfy (44) and (43), the ML estimation problem becomes

$$\left(\hat{\tilde{\mathbf{f}}},\hat{\tilde{\mathbf{g}}}\right)=\arg\max_{(\tilde{\mathbf{f}},\tilde{\mathbf{g}})}\left[\frac{\Re^2\left[\left(\mathbf{S}\tilde{\mathbf{f}}\right)^*\mathbf{X}\tilde{\mathbf{g}}^*\right]}{\|\mathbf{S}\tilde{\mathbf{f}}\|_F^2}+(1+\alpha)\Re^2\left(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\right)\right].$$

Similar to the proof of Theorem 1, the maximum is achieved when $\tilde{\mathbf{f}}$ satisfies (40). With this choice, the ML estimation further reduces to

$$\hat{\tilde{\mathbf{g}}}=\arg\max_{\tilde{\mathbf{g}}}\left[\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+(1+\alpha)\Re^2(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*)\right]. \tag{45}$$

Notice that, replacing $\tilde{\mathbf{g}}$ with $e^{j\theta}\tilde{\mathbf{g}}$ for any θ will not change the first term in the objective function of (45). So the maximum is attained at a $\tilde{\mathbf{g}}$ that makes $\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*$ a positive real number. Thus, to solve (45), we can find the $\tilde{\mathbf{g}}$ that maximizes $\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+(1+\alpha)\|\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\|_F^2$, then adjust $\tilde{\mathbf{g}}$ by the phase $\angle(\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*)$ to make $\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}$ real and positive. This is possible because $\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}$ is a scalar. Since

$$\|\mathbf{P}\mathbf{X}\tilde{\mathbf{g}}^*\|_F^2+(1+\alpha)\|\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}^*\|_F^2=\tilde{\mathbf{g}}\mathbf{\Gamma}_1\tilde{\mathbf{g}}^*,$$

where $\mathbf{\Gamma}_1$ is defined in (20), to maximize it, we need $\tilde{\mathbf{g}}^*$ to be the singular vector of the largest singular value of $\mathbf{\Gamma}_1$. With the SVD of $\mathbf{\Gamma}_1$ in Theorem 2, we need $\tilde{\mathbf{g}}=\mathbf{q}_1^*$. To make $\check{\mathbf{s}}^*\check{\mathbf{X}}\tilde{\mathbf{g}}$ a positive real number, we adjust the angle of $\tilde{\mathbf{g}}$ by multiplying with $e^{j\angle(\check{\mathbf{s}}^*\check{\mathbf{X}}\mathbf{q}_1)}$ to obtain the ML estimation of $\tilde{\mathbf{g}}$ in (21).

Using (40) and (43), we can find the estimations on $\tilde{\mathbf{f}}$ and $a_{\mathbf{g}}$ to be in and (21) and (22).

For the estimation on $a_{\mathbf{f}}$, by using (43),

$$\hat{a}_{\mathbf{f}}=\frac{\Re((\mathbf{S}\hat{\tilde{\mathbf{f}}})^*\mathbf{X}\tilde{\mathbf{g}}^*)}{\sqrt{\beta}\|\mathbf{S}\hat{\tilde{\mathbf{f}}}\|_F^2\hat{a}_{\mathbf{g}}}=\sqrt{\frac{P_2\check{T}}{\beta}}\frac{\|(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}\hat{\tilde{\mathbf{g}}}^*\|_F}{\check{\mathbf{s}}^*\check{\mathbf{X}}\hat{\tilde{\mathbf{g}}}^*},$$

which is the second result in (22).

References

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927-1938, Nov. 2003.

[2] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1099-1109, Aug. 2004.

[3] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless network," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415-2425, Oct. 2003.

[4] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4152-4172, Dec. 2005.

[5] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *IEEE Infocom*, June 2002, pp. 1577-1586.

[6] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524-3536, Dec. 2006.

[7] Y. Jing and B. Hassibi, "Diversity analysis of distributed space-time codes in relay networks with multiple transmit/receive antennas," *EURASIP Journal on Advanced Signal Processing*, 2008.

[8] Y. Jing and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," *IEEE Trans. Inf. Theory,* vol. 53, pp. 4106-4118, Nov. 2007.

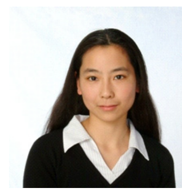
[9] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398-1407, Apr. 2007.

[10] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2499-2517, Jun. 2009.

[11] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, pp. 4306-4316, Sep. 2008.

[12] F. Gao, T. Cui, and A. Nallanathan, "Optimal training design for channel estimation in decode-and-forward relay networks with individual and total power constraints," *IEEE Trans. Signal Process.*, vol. 56, pp. 5937-5949, Dec. 2008.

[13] H. Mheidat and M. Uysal, "Non-coherent and mismatched-coherent receivers for distributed STBCs with amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 4060-4070, Nov. 2007.
[14] B. Gedik and M. Uysal, "Impact of imperfect channel estimation on the performance of amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1468-1479, Mar. 2009.
[15] O. Amin, B. Gedik, and M. Uysal, "Channel estimation for amplify-and-forward relaying: Cascaded against disintegrated estimators," *IET Communications*, vol. 4, pp. 1207-1216, Jan. 2010.
[16] B. Gedik, and M. Uysal, "Training power optimization for amplify-and-forward cooperative systems," *Proc. of IEEE VTC 2008-Fall*, Sept. 2008, pp. 1-5.
[17] Y. Jia, and A. Vosoughi, "Transmission resource allocation for training based amplify-and-forward relay systems," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 450-455, Feb. 2011.
[18] T. A. Lamahewa, P. Sadeghi, and X. Zhou, "On lower bounding the information capacity of amplify and forward wireless relay channels with channel estimation errors," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2075-2079, July 2011.
[19] C. M. Yetis, and A. H. Kayran, "A new training protocol for channel state estimation in wireless relay networks," *IEEE Trans. Signal Process.*, vol. 58, pp. 5950-5955, Nov. 2010.
[20] 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.
[21] G. S. Rajan and B. S. Rajan, "Leveraging coherent distributed space-time codes for noncoherent communication in relay networks via training," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 683-688, Feb. 2009,
[22] F. Gao, R. Zhang, and Y. Liang, "On channel estimation for amplify-and-forward two-way relay networks," *Proc. of IEEE GLOBECOM*, Nov. 2008, pp. 1-5.
[23] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951-963, Apr. 2003.
[24] 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. 884893, Mar. 2006.
[25] A. Vosoughi, and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using the affine precoding framework and the CRB," *IEEE Trans. Signal Process.,* vol. 54, pp. 940954, Mar. 2006.
[26] S. Sun and Y. Jing, "Channel training design in amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 3380-3391, Oct. 2011.
[27] S. Sun and Y. Jing, "Training and decoding for cooperative network with multiple relays and receive antennas," *IEEE Trans. Commun.*, 2011, Accepted.
[28] C. Patel, G. Stüber, and T. Pratt, "Statistical properties of amplify and forward relay fading channels," *IEEE Trans. Veh. Technol.,* vol. 55, pp. 19, Jan. 2006.
[29] A. Bletsas, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 659-672, Mar. 2006.
[30] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3114-3122, Aug. 2007.
[31] A. Bletsas, H. Shin, and M. Z. Win, "Outage optimality of opportunistic amplify-and-forward relaying," *IEEE Commun. Lett.*, vol. 11, pp. 261-263, Mar. 2007.
[32] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their diversity orders," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1414-1423, Mar. 2009.
[33] S. M. Key, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.
[34] Alan J. Laub, *Matrix Analysis for Scientists and Engineers,* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005.



Yindi Jing received the B.Eng. and M.Eng. degrees in automatic control from the University of Science and Technology of China, Hefei, China, in 1996 and 1999, respectively. She received the M.Sc. degree and the Ph.D. in electrical engineering from California Institute of Technology, Pasadena, CA, in 2000 and 2004, respectively. From October 2004 to August 2005, she was a postdoctoral scholar at the Department of Electrical Engineering of California Institute of Technology. Since February 2006 to June 2008, she was a postdoctoral scholar at the Department of Electrical Engineering and Computer Science of the University of California, Irvine. She joined the Electrical and Computer Engineering Department of the University of Alberta as an assistant professor in August 2008. She serves as an Associate Editor for the IEEE Trans. Wireless Commun..

Xinwei Yu received B.S. in Mathematics and M.S. in Computational Mathematics from Peking University in Beijing, China, in 1997 and 2000, respectively. He received Ph.D. in Applied and Computational Mathematics from California Institute of Technology, Pasadena, CA, in 2005. From 2005 to 2008 he was CAM Assistant Professor at the Department of Mathematics of University of California, Los Angeles. He joined the Department of Mathematical and Statistical Sciences of the University of Alberta as an assistant professor in July 2008. He was the recipient of the SIAM Richard C. DiPrima Prize in 2006. His research interest is applied mathematics.