Title
Optimal Design of Source and Relay Pilots for MIMO Relay Channel Estimation

Journal
IEEE Transactions on Signal Processing, 59(9)

ISSN
1053-587X

Authors
Kong, Ting
Hua, Yingbo

Publication Date
2011-06-02

Peer reviewed

Optimal Design of Source and Relay Pilots for MIMO Relay Channel Estimation

Ting Kong and Yingbo Hua, *Fellow, IEEE*

***Abstract*—In this paper, we consider a channel estimation scheme for a two-hop nonregenerative MIMO relay system without the direct link between source and destination. This scheme has two phases. In the first phase, the source does not transmit while the relay transmits and the destination receives. In the second phase, the source transmits, the relay amplifies and forwards, and the destination receives. At the destination, the data received in the first phase are used to estimate the relay-to-destination channel, and the data received in the second phase are used to estimate the source-to-relay channel. The linear minimum mean-square error estimation (LMMSE) is used for channel estimation, which allows the use of prior knowledge of channel correlations. For phase 1, an algorithm is developed to compute the optimal source pilot matrix for use at the relay. For phase 2, an algorithm is developed to compute the optimal source pilot matrix for use at the source and the optimal relay pilot matrix for use at the relay.**

***Index Terms*—Convex optimization, MIMO wireless relays, non-convex optimization, pilot waveform design, relay channel estimation.**

I. Introduction

Multiple-input multiple-output (MIMO) relays have received much attention in recent years, e.g., see [1]–[11]. It is well established that relays can substantially improve the wireless coverage for users subject to limited power and spectral resources. MIMO relays can provide additional power and spectral savings by exploiting the spatial diversity of multiple antennas. It is also known that the channel state information or channel matrices between MIMO nodes can be used to maximize the power and spectral efficiency of MIMO relay systems; e.g., see [10] and [11]. Channel estimation of MIMO relays is clearly important.

The conventional MIMO channel estimation methods can be applied to MIMO relays if every pair of adjacent MIMO nodes can be treated as a conventional pair of transmitter and receiver. The recently developed methods for single-hop MIMO channel estimation can be found in [12]–[14] and the references therein.

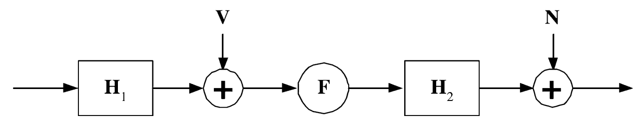


Fig. 1. A two-hop nonregenerative MIMO relay system.

In this paper, we consider alternative channel estimation schemes for MIMO relay channels. In particular, we focus on a two-hop nonregenerative MIMO relay system as illustrated in Fig. 1 where the direct link between source and destination is assumed to be negligible and hence ignored. One such example in practice is when a source and a destination are blocked from each other by a large building and a relay is added in between them around the building. A nonregenerative relay is also known as amplify-and-forward relay. It is subject to limited signal processing functions and can not decode or estimate information. Therefore, a nonregenerative relay may not be able to complete the task of channel estimation by following a single-hop MIMO channel estimation approach. It is because of such a reason that researchers have started to explore non-conventional MIMO channel estimation methods for MIMO relays.

In [15], the authors developed a least square method to search for the source-to-relay channel matrix $\mathbf{H}_1$ and the relay-to-destination channel matrix $\mathbf{H}_2$ from the observed composite source-relay-destination channel matrix $\mathbf{H}_c = \mathbf{H}_2\mathbf{F}\mathbf{H}_1$, where $\mathbf{F}$ is a known transformation matrix applied at the nonregenerative relay. In [16], the authors studied sufficient and necessary conditions on $\mathbf{F}$ to ensure a successful estimation of $\mathbf{H}_1$ and $\mathbf{H}_2$ from $\mathbf{H}_c$. The advantage of using $\mathbf{H}_c$ to estimate $\mathbf{H}_1$ and $\mathbf{H}_2$ is that for channel estimation, the relay node does not need to do anything different from that for data transmission, and the destination node performs all the tasks needed for estimation of $\mathbf{H}_1$ and $\mathbf{H}_2$. But a disadvantage of the above approach is that there is always a scalar ambiguity for the estimates of $\mathbf{H}_1$ and $\mathbf{H}_2$.

In this paper, we consider a different channel estimation scheme for the same type of two-hop MIMO relay system as discussed in [15] and [16]. This system was also a focus in [3]–[6] and [9]. Our scheme has two phases. In the first phase, the source transmits no signal while the relay transmits a source pilot matrix $\mathbf{S}_R \in C^{n_R \times L}$ (using n_R antennas and L time slots) and the destination receives $\mathbf{Y}_D^{(1)} \in C^{n_D \times L}$ (using n_D antennas and L time slots). And $\mathbf{H}_2$ is estimated at the destination by using $\mathbf{S}_R$ and $\mathbf{Y}_D^{(1)}$. In the second phase, the source transmits a source pilot matrix $\mathbf{S}_S \in C^{n_S \times L}$ (using n_S antennas and L time slots), the relay applies a relay pilot matrix $\mathbf{F} \in C^{n_R \times n_R}$ for each time slot, and the destination receives $\mathbf{Y}_D^{(2)} \in C^{n_D \times L}$. And $\mathbf{H}_1$ is estimated at the destination by using $\mathbf{S}_S$, $\mathbf{F}$, $\mathbf{Y}_D^{(2)}$ and $\mathbf{H}_2$.

Manuscript received January 27, 2011; revised April 26, 2011; accepted May 18, 2011. Date of publication June 02, 2011; date of current version August 10, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maja Bystrom. Part of this work was presented in the Asilomar Conference on Signals, Systems and Computers, Monterey, California, November 2010.

T. Kong was with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA. She is now with the VIA Telecom, San Diego, CA 92121 USA (e-mail: betsykong@gmail.com).

Y. Hua is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu).

Digital Object Identifier 10.1109/TSP.2011.2158429

Note that we use n_S, n_R and n_D to denote the number of antennas at source, relay and destination, respectively. We also assume that the number of transmit antennas and the number of receive antennas on the relay are equal. In the first phase, L denotes the number of time slots for channel training. In the second phase, if the relay is a time-division half-duplex relay, $2L$ is the total number of time slots used for channel training. We assume that the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are constant during the two phases of estimation. In fact, we need the coherence time of $\mathbf{H}_1$ and $\mathbf{H}_2$ to be much larger than the time interval needed for subsequent data transmission so that the overhead for channel estimation, pilot computation and feedback of pilot matrices only consumes a small fraction of the spectral resource.

The above channel estimation scheme requires the relay to generate a source pilot in the first phase, which is slightly more complex than the scheme in [15] and [16]. All computations are done at the destination, however, which is similar to that in [15] and [16]. However, unlike [15] and [16], the channel estimates in the new scheme are not subject to any scalar ambiguity. For both phases, we consider LMMSE of channel matrices, which allows the use of prior knowledge of channel correlations.

The estimation of $\mathbf{H}_2$ in the first phase is actually the same type of problem as in [12] and [14]. Therefore, the optimal design of $\mathbf{S}_R$ for the first phase coincides with that in [14]. Yet, the estimate of $\mathbf{H}_1$ in the second phase is a nonconventional problem. We will derive the optimal design of $\mathbf{S}_S$ and $\mathbf{F}$ for estimation of $\mathbf{H}_1$ in the second phase, which is the main focus of this paper.

The channel matrices $\mathbf{H}_i$ for $i = 1, 2$ are associated with a narrow frequency band. We use the well-known Kronecker correlation model for the channel matrices, i.e.,

$$\mathbf{H}_i = \mathbf{C}_{r_i}^{\frac{1}{2}} \mathbf{W}_i \mathbf{C}_{t_i}^{\frac{T}{2}} \tag{1}$$

where $\mathbf{C}_{t_i} = \mathbf{C}_{t_i}^{\frac{1}{2}} \mathbf{C}_{t_i}^{\frac{H}{2}}$, $\mathbf{C}_{r_i} = \mathbf{C}_{r_i}^{\frac{1}{2}} \mathbf{C}_{r_i}^{\frac{H}{2}}$, and the elements in $\mathbf{W}_i$ are uncorrelated random variables with zero mean and unit variance. The matrix $\mathbf{C}_{r_i}$ is known as the receive correlation matrix of $\mathbf{H}_i$, and $\mathbf{C}_{t_i}$ the transmit correlation matrix of $\mathbf{H}_i$. The correlation matrices are assumed to be known. If there is no information on the correlation for a particular application, the correlation matrices should be set as the identity matrices.

The rest of the paper is organized as follows. In Section II, we present the optimal pilot design for channel estimation in phase 1, the result of which is similar to one in [14] although our derivations are different and provide a complementary perspective. In Section III, we show the optimal source-and-relay pilots design for channel estimation in phase 2. It should be noted that for $\mathbf{C}_{r_1}$ proportional to the identity matrix, an optimality of the designed relay pilot is established. For arbitrary $\mathbf{C}_{r_1}$, the optimality of the designed relay pilot remains an open problem. Numerical results are illustrated in Section IV. Section V concludes this paper.

II. Pilot Design for Phase 1

As mentioned earlier, the pilot design problem for phase 1 is the same as for single-hop MIMO system. The result shown in this section is similar to that in [14] although our derivations are different, and were found independently, from [14]. We keep this section as brief as possible.

In phase 1, the source transmits nothing, the relay transmits $\mathbf{S}_R \in C^{n_R \times L}$, and the destination receives

$$\mathbf{Y}_D^{(1)} = \mathbf{H}_2 \mathbf{S}_R + \mathbf{N}^{(1)} \in C^{n_D \times L} \tag{2}$$

where $\mathbf{N}^{(1)}$ is the noise matrix at the destination in phase 1. Without loss of generality, we assume that the noise samples in $\mathbf{Y}_D^{(1)}$ has been (spatially) whitened such that the elements in $\mathbf{N}^{(1)}$ are uncorrelated random variables with zero mean and unit variance.

Let $\mathbf{y}_D^{(1)} = \mathrm{vec}\left(\mathbf{Y}_D^{(1)}\right)$, $\mathbf{w}_2 = \mathrm{vec}(\mathbf{W}_2)$ and $\mathbf{n}^{(1)} = \mathrm{vec}(\mathbf{N}^{(1)})$. Here, vec stacks up the columns of a matrix into a single column vector. Recall the fact $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ where $\otimes$ denotes the Kronecker product [17]. We will also use frequently $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ and $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$. Using $\mathbf{H}_2 = \mathbf{C}_{r_2}^{\frac{1}{2}} \mathbf{W}_2 \mathbf{C}_{t_2}^{\frac{T}{2}}$, we can write (2) as

$$\mathbf{y}_D^{(1)} = \left(\mathbf{S}_R^T \mathbf{C}_{t_2}^{\frac{1}{2}} \otimes \mathbf{C}_{r_2}^{\frac{1}{2}}\right) \mathbf{w}_2 + \mathbf{n}^{(1)}. \tag{3}$$

Since $\mathbf{H}_2$ and $\mathbf{W}_2$ are one-to-one related via $\mathbf{H}_2 = \mathbf{C}_{r_2}^{\frac{1}{2}} \mathbf{W}_2 \mathbf{C}_{t_2}^{\frac{T}{2}}$ or equivalently $\mathbf{h}_2 = \mathrm{vec}(\mathbf{H}_2) = \left(\mathbf{C}_{t_2}^{\frac{1}{2}} \otimes \mathbf{C}_{r_2}^{\frac{1}{2}}\right) \mathbf{w}_2$, we can focus on the estimation of $\mathbf{W}_2$ or equivalently $\mathbf{w}_2$.

We consider the LMMSE of $\mathbf{w}_2$ from $\mathbf{y}_D^{(1)}$, i.e., $\hat{\mathbf{w}}_2 = \mathbf{T}\mathbf{y}_D^{(1)}$ where $\mathbf{T}$ is such that the following cost is minimized:

$$J_2 \doteq E\left\{tr[\mathbf{C}_0(\mathbf{w}_2 - \hat{\mathbf{w}}_2)(\mathbf{w}_2 - \hat{\mathbf{w}}_2)^H]\right\} \tag{4}$$

where E denotes expectation. There are two special choices of the weight matrix $\mathbf{C}_0$. If $\mathbf{C}_0 = \mathbf{I}$, then $J_2 = E(\|\mathbf{w}_2 - \hat{\mathbf{w}}_2\|^2)$ which is the mean squared errors of $\hat{\mathbf{w}}_2$. If $\mathbf{C}_0 = \mathbf{C}_{t_2}^{\frac{H}{2}} \mathbf{C}_{t_2}^{\frac{1}{2}} \otimes \mathbf{C}_{r_2}^{\frac{H}{2}} \mathbf{C}_{r_2}^{\frac{1}{2}}$, then $J_2 = E(\|\mathbf{h}_2 - \hat{\mathbf{h}}_2\|^2)$ which is the mean squared errors of $\hat{\mathbf{h}}_2$. In either case, we can write $\mathbf{C}_0 = \mathbf{C}_1 \otimes \mathbf{C}_2$. It is useful to note that as long as $\mathbf{C}_0$ is positive definite, $\mathbf{C}_0$ does not affect the optimal $\mathbf{T}$ which is given by

$$\mathbf{T} = \mathbf{R}_{\mathbf{w}_2 \mathbf{y}_D^{(1)H}} \mathbf{R}^{-1}_{\mathbf{y}_D^{(1)} \mathbf{y}_D^{(1)H}} \tag{5}$$

where

$$\mathbf{R}_{\mathbf{w}_2 \mathbf{y}_D^{(1)H}} = E\left[\mathbf{w}_2 \mathbf{y}_D^{(1)H}\right]$$

and

$$\mathbf{R}_{\mathbf{y}_D^{(1)} \mathbf{y}_D^{(1)H}} = E\left[\mathbf{y}_D^{(1)} \mathbf{y}_D^{(1)H}\right].$$

Substituting $\hat{\mathbf{w}}_2 = \mathbf{T}\mathbf{y}_D^{(1)}$, (3) and (5) into (4), and using the identity $(\mathbf{I} + \mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}$, it is easy to verify that with the optimal $\mathbf{T}$,

$$J_2 = J_2' \doteq tr\left(\mathbf{C}_0 \left(\mathbf{I} + \mathbf{C}_{t_2}^{\frac{H}{2}} \mathbf{C}_{S_R} \mathbf{C}_{t_2}^{\frac{1}{2}} \otimes \mathbf{C}_{r_2}^{\frac{H}{2}} \mathbf{C}_{r_2}^{\frac{1}{2}}\right)^{-1}\right) \tag{6}$$

where $\mathbf{C}_{S_R} = \mathbf{S}_R^* \mathbf{S}_R^T$. If $\mathbf{T}$ is arbitrary, $J_2 \geq J_2'$.

Once $\mathbf{C}_{S_R}$ is fixed, J_2' is invariant to $\mathbf{S}_R$. So, to find the optimal relay pilot matrix $\mathbf{S}_R$ for phase 1, it suffices to find $\mathbf{C}_{S_R}$ by solving the following problem:

$$\begin{aligned} \min_{\mathbf{C}_{S_R}\geq \mathbf{0}} \quad & J_2' \\ \text{s.t.} \quad & \operatorname{tr}(\mathbf{C}_{S_R}) \leq P_R \end{aligned} \tag{7}$$

where $\mathbf{C}_{S_R} \geq \mathbf{0}$ is the positive semidefinite constraint on $\mathbf{C}_{S_R}$, and P_R is the power bound at the relay.

The problem (7) is convex. In Appendix I, we apply the generalized KKT conditions [18] to arrive at the following optimal solution: $\mathbf{C}_{S_R} = \mathbf{U}_{t_2}\mathbf{C}\mathbf{U}_{t_2}^H$ where $\mathbf{U}_{t_2}$ is the unitary eigenvector matrix of $\mathbf{C}_{t_2}$ and $\mathbf{C}$ is a diagonal matrix with its diagonal elements $c(j)$, $j = 1, 2, \ldots, n_R$, either equal to zero or given by the positive solution of the following equations:

$$\lambda_1(j)\lambda_{t_2}(j)\sum_{i=1}^{n_R}\frac{\lambda_2(i)\lambda_{r_2}(i)}{[1+\lambda_{r_2}(i)\lambda_{t_2}(j)c(j)]^2} = \mu, \quad j = 1, 2, \ldots, n_R \tag{8}$$

where $\mu > 0$ is such that $\operatorname{tr}(\mathbf{C}) = P_R$. Recall $\mathbf{C}_0 = \mathbf{C}_1 \otimes \mathbf{C}_2$. Here, $\lambda_2(i)$ and $\lambda_{r_2}(i)$ are the ith largest eigenvalue of $\mathbf{C}_2$ and $\mathbf{C}_{r_2}$. $\lambda_1(j)$ and $\lambda_{t_2}(j)$ are the jth largest eigenvalue of $\mathbf{C}_1$ and $\mathbf{C}_{t_2}$.

For any given $\mu > 0$, for each j, $c(j)$ can be easily found by the bisection search [19] since the left side expression of (8) is a monotonically decreasing function of $c(j)$. Consequently, $\operatorname{tr}(\mathbf{C}_{S_R}) = \operatorname{tr}(\mathbf{C}) = \sum_{j=1}^{n_R} c(j)$ is a monotonically decreasing function of μ, and hence the optimal μ can be found by an outer-loop bisection search.

The above solution for $\mathbf{C}_{S_R}$ is similar to one in [14] although a different method of derivation was used in [14].

III. Pilot Design for Phase 2

In phase 2, the source transmits the pilot matrix $\mathbf{S}_S \in C^{n_S\times L}$, the relay receives $\mathbf{Y}_R = \mathbf{H}_1\mathbf{S}_S + \mathbf{V} \in C^{n_R\times L}$ and retransmits $\mathbf{X}_R = \mathbf{F}\mathbf{Y}_R \in C^{n_R\times L}$, and the destination receives

$$\mathbf{Y}_D = \mathbf{H}_2\mathbf{F}\mathbf{H}_1\mathbf{S}_S + \mathbf{H}_2\mathbf{F}\mathbf{V} + \mathbf{N} \in C^{n_D\times L} \tag{9}$$

where $\mathbf{F}$ now serves as a co-pilot matrix from the relay, $\mathbf{V}$ is the noise matrix at the relay, and $\mathbf{N}$ is the noise matrix received at the destination. Without loss of generality (for temporally white noise), we assume that the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ (along with $\mathbf{Y}_R$ and $\mathbf{Y}_D$) are normalized in such a way that the elements in both $\mathbf{V}$ and $\mathbf{N}$ are uncorrelated with zero mean and unit variance. Given the result from phase 1, we now assume that for phase 2, $\mathbf{H}_2$ is known. The objective in this section is to derive the optimal pilot matrices $\mathbf{S}_S$ and $\mathbf{F}$ so that the estimation errors of $\mathbf{H}_1$ are minimized. The assumption of known $\mathbf{H}_2$ is justifiable when the training power for phase 1 is high so that the estimate of $\mathbf{H}_2$ is near perfect. Also note that the under the same training power for both phases, the accuracy of $\mathbf{H}_2$ is always much higher than that of $\mathbf{H}_1$, which is illustrated later in Fig. 5.

Using the vec operator, we have from (9) that

$$\mathbf{y}_D = \left(\mathbf{S}_S^T \otimes \mathbf{H}_2\mathbf{F}\right)\mathbf{h}_1 + (\mathbf{I}\otimes\mathbf{H}_2\mathbf{F})\,\mathbf{v} + \mathbf{n} \tag{10}$$

where the bold lower case symbols are the vector forms of the bold upper case symbols, e.g., $\mathbf{y}_D = \operatorname{vec}(\mathbf{Y}_D)$.

For notational convenience, we have chosen not to use the superscript $^{(2)}$ for phase 2 although we have used $^{(1)}$ for some of the variables for phase 1.

A. Channel Estimation

Unlike the discussion in Section II, we now only focus on the mean squared errors of $\hat{\mathbf{h}}_1 = \operatorname{vec}(\hat{\mathbf{H}}_1)$. The choice of the mean squared errors of $\hat{\mathbf{w}}_1 = \operatorname{vec}(\hat{\mathbf{W}}_1)$ would make the optimal pilot design more difficult, which will not be further mentioned. Namely, we define the cost

$$J_1 \doteq E\{tr[(\mathbf{h}_1 - \hat{\mathbf{h}}_1)(\mathbf{h}_1 - \hat{\mathbf{h}}_1)^H]\} \tag{11}$$

without any other weighting.

For fixed $\mathbf{S}_S$ and $\mathbf{F}$, the LMMSE of $\mathbf{h}_1$ is known as $\hat{\mathbf{h}}_1 = \mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}\mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H}^{-1}\mathbf{y}_D$. Using (10) and $\mathbf{h}_1 = (\mathbf{C}_{t_1}^{\frac{1}{2}} \otimes \mathbf{C}_{r_1}^{\frac{1}{2}})\mathbf{w}_1$, we have

$$\begin{aligned} \mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H} &= E\left[\mathbf{h}_1\mathbf{y}_D^H\right] = \mathbf{C}_{t_1}\mathbf{S}_S^* \otimes \mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H & (12)\\ \mathbf{R}_{\mathbf{y}_D\mathbf{y}_D^H} &= E\left[\mathbf{y}_D\mathbf{y}_D^H\right] = \mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^* \otimes \mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H \\ &\quad + \mathbf{I}\otimes\mathbf{H}_2\mathbf{F}\mathbf{F}^H\mathbf{H}_2^H + \mathbf{I}. & (13) \end{aligned}$$

The covariance matrix of the estimation error $\delta\mathbf{h}_1 = \mathbf{h}_1 - \hat{\mathbf{h}}_1$ is well known as

$$\mathbf{R}_{\delta\mathbf{h}_1,\delta\mathbf{h}_1^H} = \mathbf{R}_{\mathbf{h}_1,\mathbf{h}_1^H} - \mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}\mathbf{R}_{\mathbf{y}_D,\mathbf{y}_D^H}^{-1}\mathbf{R}_{\mathbf{h}_1\mathbf{y}_D^H}^H$$

where $\mathbf{R}_{\mathbf{h}_1,\mathbf{h}_1^H} = E\left\{\mathbf{h}_1\mathbf{h}_1^H\right\} = \mathbf{C}_{t_1}\otimes\mathbf{C}_{r_1}$. Therefore, with the LMMSE of $\mathbf{h}_1$ and (12) and (13), we have

$$\begin{aligned} J_1 &= J_1' \\ &\doteq \operatorname{tr}\left\{\mathbf{R}_{\delta\mathbf{h}_1,\delta\mathbf{h}_1^H}\right\} \\ &= \operatorname{tr}\left[\mathbf{C}_{t_1}\otimes\mathbf{C}_{r_1}\right] \\ &\quad - \operatorname{tr}\Bigg[\left(\mathbf{C}_{t_1}\mathbf{S}_S^*\otimes\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H\right) \\ &\qquad \cdot\Big[\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*\otimes\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H \\ &\qquad\quad +\mathbf{I}\otimes\mathbf{H}_2\mathbf{F}\mathbf{F}^H\mathbf{H}_2^H + \mathbf{I}\Big]^{-1}\left(\mathbf{S}_S^T\mathbf{C}_{t_1}^H\otimes\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}^H\right)\Bigg]. \end{aligned} \tag{14}$$

In other words, with any other linear estimator of $\mathbf{h}_1$, $J_1 \geq J_1'$.

B. Pilot Design Problem

We formulate the pilot design problem as minimization of J_1' with respect to $\mathbf{S}_S$ and $\mathbf{F}$ subject to power constraints at both source and relay, i.e.,

$$\begin{aligned} \min_{\mathbf{S}_S,\mathbf{F}} \quad & J_1' \\ \text{s.t.} \quad & \operatorname{tr}(\mathbf{S}_S\mathbf{S}_S^H) \leq P_S \\ & E\left\{\operatorname{tr}\left(\mathbf{X}_R\mathbf{X}_R^H\right)\right\} \leq P_R \end{aligned} \tag{15}$$

where P_S is the power bound at the source and P_R the power bound at the relay.

The power constraint for the relay here has the additional operator E which averages the fluctuations due to the noise $\mathbf{V}$ and the unknown $\mathbf{H}_1$. Although the instantaneous power at the relay is not possible to be bounded exactly due to the unknown nature of noise and $\mathbf{H}_1$, the averaged power bound at the relay is still meaningful and necessary. Clearly, the power constraint at the relay in phase 2 does not have the same meaning as that in phase 1. But for convenience, we will use the same notation P_R as the relay power bound for both phases.

Recall $\mathbf{X}_R = \mathbf{F}\mathbf{Y}_R = \mathbf{F}\mathbf{H}_1\mathbf{S}_S + \mathbf{F}\mathbf{V}$ and $\mathbf{H}_1 = \mathbf{C}_{r_1}^{\frac{1}{2}}\mathbf{W}_1\mathbf{C}_{t_1}^{\frac{T}{2}}$ or equivalently $\mathbf{x}_R = \mathrm{vec}(\mathbf{X}_R) = (\mathbf{S}_S^T\mathbf{C}_{t_1}^{\frac{1}{2}} \otimes \mathbf{F}\mathbf{C}_{r_1}^{\frac{1}{2}})\mathbf{w}_1 + (\mathbf{I}\otimes\mathbf{F})\mathbf{v}$. The power constraint at the relay can be rewritten as

$$\begin{aligned} E\{\mathrm{tr}\left(\mathbf{X}_R\mathbf{X}_R^H\right)\} &= E\left\{\mathrm{tr}\left(\mathbf{x}_R\mathbf{x}_R^H\right)\right\} \\ &= \mathrm{tr}\left(\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*\otimes\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H + \mathbf{I}\otimes\mathbf{F}\mathbf{F}^H\right) \\ &\le P_R. \end{aligned} \tag{16}$$

The problem (15) is not convex. The generalized KKT conditions are not an effective tool for this problem. In the rest of this section, we present methods to simplify the problem (15).

C. Decomposition of Pilots

In this section, we show a decomposition of pilots into two sets of components: unitary components and diagonal components.

Denote the eigenvalue decompositions (EVD) of $\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^*$ and $\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H$, respectively, as

$$\mathbf{S}_S^T\mathbf{C}_{t_1}\mathbf{S}_S^* = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H \tag{17}$$
$$\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H = \mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H \tag{18}$$

where the $\mathbf{U}$ matrices are the unitary eigenvector matrices and the $\mathbf{\Lambda}$ matrices are the diagonal eigenvalue matrices with descending diagonal elements.

Also let $\mathbf{C}_{t_1} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}\mathbf{U}_{t_1}^H$ and $\mathbf{C}_{r_1} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}\mathbf{U}_{r_1}^H$ be the EVDs of $\mathbf{C}_{t_1}$ and $\mathbf{C}_{r_1}$, respectively, with descending eigenvalues. Define $\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}^{\frac{1}{2}}$ and $\mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}^{\frac{1}{2}}$. Then, we can write

$$\mathbf{S}_S^T\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S \tag{19}$$
$$\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F \tag{20}$$

where $\mathbf{Q}_S$ and $\mathbf{Q}_F$ are unitary matrices.

It is important to note here that if $\mathbf{C}_{t_1}$, $\mathbf{C}_{r_1}$ and $\mathbf{H}_2$ are nonsingular, the pilot matrices $\mathbf{S}_S$ and $\mathbf{F}$ are uniquely determined by the unitary components: $\mathbf{U}_S$, $\mathbf{U}_F$, $\mathbf{Q}_S$, $\mathbf{Q}_F$ and the diagonal components: $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$. Namely,

$$\mathbf{S}_S^T = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{C}_{t_1}^{-\frac{1}{2}} \tag{21}$$
$$\mathbf{F} = \mathbf{H}_2^{-1}\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{C}_{r_1}^{-\frac{1}{2}}. \tag{22}$$

On the other hand, if any of $\mathbf{C}_{t_1}$, $\mathbf{C}_{r_1}$ and $\mathbf{H}_2$ is singular, it is optimal in minimized transmission power to choose $\mathbf{S}_S$ and $\mathbf{F}$ as the minimum norm solutions to (19) and (20), respectively, i.e., replacing the inverse in (21) and (22) by *pseudoinverse*. Consequently, as is easy to verify, the inverse in the sequel should be viewed as *pseudoinverse* in each case where a matrix under the inverse operator is singular.

It then follows from (11) that

$$\begin{aligned} &J_1' - \mathrm{tr}(\mathbf{C}_{t_1}\otimes\mathbf{C}_{r_1}) \\ &= -\,\mathrm{tr}\Big\{\big[\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}\otimes\mathbf{H}_2\mathbf{F}\mathbf{F}^H\mathbf{H}_2^H + \mathbf{I}\big]^{-1} \\ &\quad \big(\mathbf{S}_S^T\mathbf{C}_{t_1}^H\mathbf{C}_{t_1}\mathbf{S}_S^*\otimes\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}^H\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H\big)\Big\} \\ &= -\,\mathrm{tr}\Big\{\big[\mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H\otimes\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H + \mathbf{U}_S\mathbf{U}_S^H \\ &\quad \otimes\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{C}_{r_1}^{-\frac{1}{2}}\mathbf{C}_{r_1}^{-\frac{H}{2}}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{H}{2}}\mathbf{U}_F^H + \mathbf{I}\big]^{-1} \\ &\quad \Big(\mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{C}_{t_1}^{-\frac{1}{2}}\mathbf{C}_{t_1}^H\mathbf{C}_{t_1}\mathbf{C}_{t_1}^{-\frac{H}{2}}\mathbf{Q}_S^H\Lambda_S^{\frac{H}{2}}\mathbf{U}_S^H \\ &\quad \otimes\mathbf{U}_F\Lambda_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{C}_{r_1}^{-\frac{1}{2}}\mathbf{C}_{r_1}^H\mathbf{C}_{r_1}\mathbf{C}_{r_1}^{-\frac{H}{2}}\mathbf{Q}_F^H\Lambda_F^{\frac{H}{2}}\mathbf{U}_F^H\Big)\Big\} \\ &= -\,\mathrm{tr}\Big\{\big[\mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H\otimes\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H + \mathbf{U}_S\mathbf{U}_S^H\otimes \\ &\quad \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{H}{2}}\mathbf{U}_F^H + \mathbf{I}\big]^{-1} \\ &\quad \Big(\mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{\Lambda}_{t_1}\mathbf{Q}_S^H\mathbf{\Lambda}_S^{\frac{H}{2}}\mathbf{U}_S^H\otimes\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{H}{2}}\mathbf{U}_F^H\Big)\Big\} \\ &= -\,\mathrm{tr}\Big\{\Big(\mathbf{\Lambda}_S\otimes\mathbf{\Lambda}_F + \mathbf{I}\otimes\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{H}{2}} + \mathbf{I}\Big)^{-1} \\ &\quad \Big(\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{\Lambda}_{t_1}\mathbf{Q}_S^H\mathbf{\Lambda}_S^{\frac{H}{2}}\otimes\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{H}{2}}\Big)\Big\}. \end{aligned} \tag{23}$$

We can see from the above equation that the cost J_1' is invariant to $\mathbf{U}_S$ and $\mathbf{U}_F$ but depends on $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{Q}_S$ and $\mathbf{Q}_F$.

It is easy to verify that the power constraint at the source can now be written as

$$\mathrm{tr}\left\{\mathbf{\Lambda}_S\mathbf{Q}_S\mathbf{\Lambda}_{t_1}^{-1}\mathbf{Q}_S^H\right\} \le P_S \tag{24}$$

which depends on $\mathbf{\Lambda}_S$ and $\mathbf{Q}_S$, and is invariant to all other components of the pilots.

To simplify the power constraint at the relay, we denote the singular value decomposition (SVD): $\mathbf{H}_2 = \mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}\mathbf{V}_{H_2}^H$, with descending singular values, where $\mathbf{U}_{H_2}$ and $\mathbf{V}_{H_2}$ are (square) unitary singular vector matrices. Note that we will use $\mathbf{\Sigma}_{H_2}^2 = \mathbf{\Sigma}_{H_2}\mathbf{\Sigma}_{H_2}^H$ in the case where $\mathbf{H}_2$ is nonsquare. Then, using (19) and (20) and $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = \mathrm{tr}(\mathbf{A})tr(\mathbf{B})$, one can verify that the relay power constraint (16) can be rewritten as

$$\begin{aligned} &\mathrm{tr}\{\mathbf{\Lambda}_S\}tr\left\{\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}\right\} \\ &\quad + L\cdot\mathrm{tr}\left\{\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{H}{2}}\mathbf{U}_F^H\mathbf{U}_{H_2}\right\} \\ &\quad \le P_R \end{aligned} \tag{25}$$

which depends on $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{U}_F$, and $\mathbf{Q}_F$, and is invariant to $\mathbf{U}_S$ and $\mathbf{Q}_S$.

In the following two subsections, we will show how to identify the optimal unitary components and the optimal diagonal components, respectively.

D. Optimal Unitary Components of the Pilots

Among the pilot components, we have the unitary (matrix) components $\mathbf{U}_S$, $\mathbf{Q}_S$, $\mathbf{U}_F$ and $\mathbf{Q}_F$, and the diagonal (matrix) components $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_F$. The optimality of the choices of the unitary components are given by following two theorems.

Theorem 1: For any $\mathbf{C}_{r_1}$ and $\mathbf{C}_{t_1}$, the solution to the problem (15) is such that $\mathbf{Q}_S = \mathbf{I}$ and $\mathbf{U}_S$ is arbitrary unitary.

Proof: See Appendix II.

Theorem 2: If $\mathbf{C}_{r_1} = \alpha\mathbf{I}$, the solution to (15) is such that $\mathbf{U}_F = \mathbf{U}_{H_2}$ and $\mathbf{Q}_F$ is arbitrary unitary.

Proof: See Appendix III.

Namely, if $\mathbf{C}_{r_1} = \alpha\mathbf{I}$, then we can choose $\mathbf{U}_S = \mathbf{I}$ and $\mathbf{Q}_F = \mathbf{I}$ as optimal, and the optimal $\mathbf{S}_S$ and $\mathbf{F}$ have the following structures:

$$\mathbf{S}_S^T = \mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{C}_{t_1}^{-\frac{1}{2}} \tag{26}$$

$$\mathbf{F} = \frac{1}{\sqrt{\alpha}}\mathbf{V}_{H_2}\Sigma_{H_2}^{-1}\mathbf{\Lambda}_F^{\frac{1}{2}}. \tag{27}$$

If $\mathbf{C}_{r_1} \neq \alpha\mathbf{I}$, finding the optimal $\mathbf{Q}_F$ (which also affects the optimal $\mathbf{U}_F$) is a difficult problem.

E. Optimal Diagonal Components of the Pilots

In this section, we apply $\mathbf{Q}_S = \mathbf{I}$, $\mathbf{Q}_F = \mathbf{I}$, $\mathbf{U}_F = \mathbf{U}_{H_2}$ and $\mathbf{U}_S = \mathbf{I}$ to develop efficient algorithms for finding the optimal $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_F$. The above choice of the unitary components is optimal if $\mathbf{C}_{r_1}$ is proportional to the identity matrix, and has no known optimality property otherwise.

Then, the pilot design problem (15) becomes

$$\begin{aligned}\min_{\mathbf{\Lambda}_S\geq 0,\mathbf{\Lambda}_F\geq 0}\quad & -\operatorname{tr}\left\{\left(\mathbf{\Lambda}_S\otimes\mathbf{\Lambda}_F+\mathbf{I}\otimes\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_{r_1}^{-1}\mathbf{\Lambda}_F^{\frac{H}{2}}+\mathbf{I}\right)^{-1}\right.\\ &\left.\left(\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{\Lambda}_{t_1}\mathbf{\Lambda}_S^{\frac{H}{2}}\otimes\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_{r_1}\mathbf{\Lambda}_F^{\frac{H}{2}}\right)\right\}\\ \text{s.t.}\quad & \operatorname{tr}\left\{\mathbf{\Lambda}_S\mathbf{\Lambda}_{t_1}^{-1}\right\}\leq P_S\\ & \operatorname{tr}\{\mathbf{\Lambda}_S\}\operatorname{tr}\left\{\mathbf{\Sigma}_{H_2}^{-2}\mathbf{\Lambda}_F\right\}\\ & +L\operatorname{tr}\left\{\mathbf{\Sigma}_{H_2}^{-2}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_{r_1}^{-1}\mathbf{\Lambda}_F^{\frac{H}{2}}\right\}\leq P_R.\end{aligned} \tag{28}$$

Denote $\lambda_S(i)$, $\lambda_F(i)$, $\lambda_{t_1}(i)$, $\lambda_{r_1}(i)$ and $\sigma_{H_2}(i)$ as the ith diagonal element of Λ_S, Λ_F, Λ_{t_1}, Λ_{r_1} and Σ_{H_2}, respectively. The problem (28) can be further written as

$$\begin{aligned}\min_{\substack{\{\lambda_S(i)\geq 0,\forall\ i\}\\ \{\lambda_F(j)\geq 0,\forall\ j\}}}\quad & -\sum_{i=1}^{L}\sum_{j=1}^{n_D}\frac{\lambda_S(i)\lambda_{t_1}(i)\lambda_F(j)\lambda_{r_1}(j)}{\lambda_S(i)\lambda_F(j)+\lambda_F(j)\lambda_{r_1}^{-1}(j)+1}\\ \text{s.t.}\quad & \sum_{i=1}^{L}\lambda_{t_1}^{-1}(i)\lambda_S(i)\leq P_S\\ & \left(\sum_{i=1}^{L}\lambda_S(i)\right)\left(\sum_{j=1}^{n_D}\sigma_{H_2}^{-2}(j)\lambda_F(j)\right)\\ & +L\sum_{j=1}^{n_D}\sigma_{H_2}^{-2}(j)\lambda_{r_1}^{-1}(j)\lambda_F(j)\leq P_R.\end{aligned} \tag{29}$$

Let $\bar{n}_S = \operatorname{rank}(\mathbf{C}_{t_1})$ and $\bar{n}_F = \min(\operatorname{rank}(\mathbf{C}_{r_1}), \operatorname{rank}(\mathbf{H}_2))$. Under the pseudoinverse condition (or interpretation of inverse) mentioned previously, for $i > \bar{n}_S$, $\lambda_{t_1}^{-1}(i) = 0$, and for $j > \bar{n}_F$, $\sigma_{H_2}^{-2}(j)\lambda_{r_1}^{-1}(j) = 0$. Also, the minimum norm solutions to (19) and (20) ensure that $\lambda_S(i) = 0$ for $L \geq i > \bar{n}_S$ and $\lambda_F(j) = 0$ for $n_D \geq j > \bar{n}_F$. So, we can replace L and n_D in (29) by $\bar{n}_S$ and $\bar{n}_F$, respectively.

The problem (29) is nonconvex. But if we fix $\lambda_S(i)$ for all i subject to the source power constraint [the first constraint in (29)], the optimization over $\lambda_F(j)$ for all j subject to the relay power constraint [the second constraint in (29)] is a convex problem. Similarly, if we fix $\lambda_F(j)$ for all j, the optimization over $\lambda_S(i)$ for all i subject to the source and relay power constraints is also a convex problem. By alternating between the two sub-optimizations, we can find a local optimal solution to (29), which is the algorithm we propose to use. Note that a local convergence of the alternations is guaranteed since the cost is lower bounded and is reduced by each alternation subject to the same power constraints.

The algorithms of the two suboptimizations are shown next.

1) Optimizing $\{\lambda_F(j)\}$ With Fixed $\{\lambda_S(i)\}$: With fixed $\lambda_S(i)$, $i = 1, \ldots, \bar{n}_S$ satisfying $\sum_{i=1}^{\bar{n}_S}\frac{\lambda_S(i)}{\lambda_{t_1}(i)} = P_S$, the first power constraint in the problem (29) is no longer needed. The KKT conditions of this problem can be simplified to

$$\begin{aligned}g_j(\lambda_F(j)) &\doteq \sum_{i=1}^{\bar{n}_S}\frac{\lambda_{t_1}(i)\lambda_{r_1}(j)\lambda_S(i)}{\left\{\left[\lambda_S(i)+\lambda_{r_1}^{-1}(j)\right]\lambda_F(j)+1\right\}^2}\\ &=\mu\left[\sum_{i=1}^{\bar{n}_S}\frac{\lambda_S(i)}{\sigma_{H_2}^2(j)}+\frac{L}{\sigma_{H_2}^2(j)\lambda_{r_1}(j)}\right]\end{aligned} \tag{30}$$

$$\begin{aligned}f(\boldsymbol{\lambda}_F) \doteq{}& \left(\sum_{i=1}^{\bar{n}_S}\lambda_S(i)\right)\left(\sum_{j=1}^{\bar{n}_F}\frac{\lambda_F(j)}{\sigma_{H_2}^2(j)}\right)\\ &+L\sum_{j=1}^{\bar{n}_F}\frac{\lambda_F(j)}{\sigma_{H_2}^2(j)\lambda_{r_1}(j)} = P_R\end{aligned} \tag{31}$$

where $\boldsymbol{\lambda}_F = [\lambda_F(1), \ldots, \lambda_F(\bar{n}_F)]^T$ and $\mu > 0$. And $\lambda_F(j)$ is either zero or a positive value satisfying (30).

Here, $g_j(\lambda_F(j))$ is a monotonically decreasing function of $\lambda_F(j) \geq 0$. So, for any given $\mu > 0$, for each j, either a nonnegative solution for $\lambda_F(j)$ can be found from (30) by the bisection method or we set $\lambda_F(j)$ to zero.

Since $f(\boldsymbol{\lambda}_F)$ is an increasing function of $\lambda_F(j)$ for all j, it is monotonically decreasing with μ. Hence, the optimal $\mu > 0$ can be found by an outer layer bisection search.

2) Optimizing $\{\lambda_S(i)\}$ With Fixed $\{\lambda_F(j)\}$: With fixed $\lambda_F(j)$, $j = 1, 2, \ldots, \bar{n}_F$, the problem (29) still has two power constraints. The Lagrangian condition of the problem (i.e., the derivative of the Lagrangian function set to zero) can be shown to be

$$\begin{aligned}&\sum_{j=1}^{\bar{n}_F}\frac{\lambda_{t_1}(i)\lambda_F(j)\lambda_{r_1}(j)[\lambda_F(j)\lambda_{r_1}^{-1}(j)+1]}{\left[\lambda_F(j)\lambda_S(i)+\lambda_F(j)\lambda_{r_1}^{-1}(j)+1\right]^2}\\ &\qquad=\frac{\mu_1}{\lambda_{t_1}(i)}+\mu_2\sum_{j=1}^{\bar{n}_F}\frac{\lambda_F(j)}{\sigma_{H_2}^2(j)}\end{aligned} \tag{32}$$

where $\mu_1 \geq 0$ is the multiplier for the first constraint and $\mu_2 \geq 0$ is the multiplier for the second constraint.

The left-hand side function in (32) is a monotonically decreasing function of $\lambda_S(i) \geq 0$. So, for any given pair of μ_1

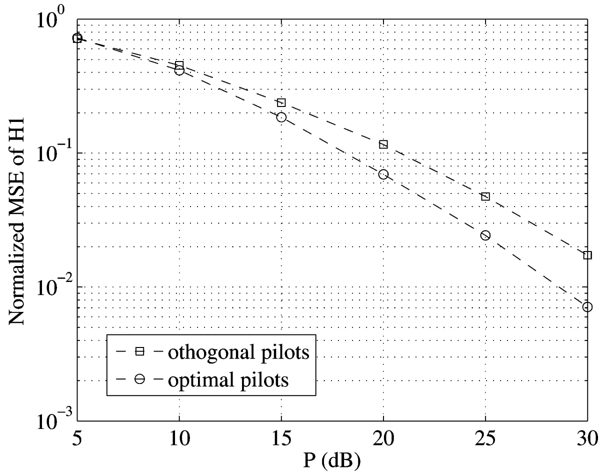


Fig. 2. Normalized MSE of $\mathbf{H}_1$ with optimal pilots and orthogonal pilots.

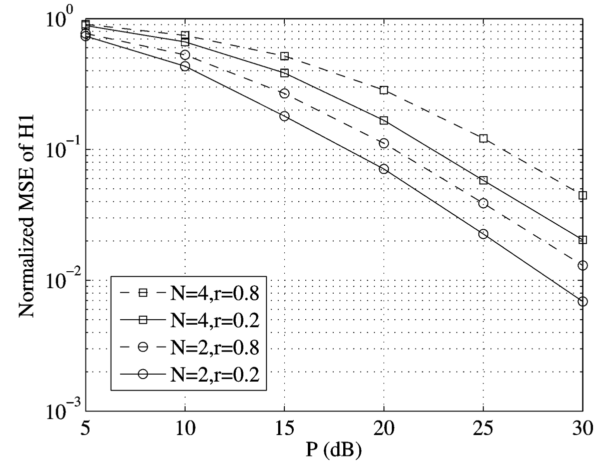


Fig. 3. Normalized MSE of $\mathbf{H}_1$ where $\mathbf{H}_1$ is uncorrelated and $\mathbf{H}_2$ is correlated with the correlation factor r.

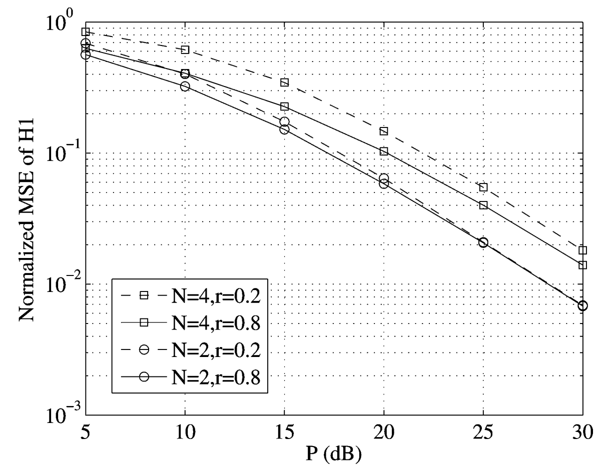


Fig. 4. Normalized MSE of $\mathbf{H}_1$ where $\mathbf{H}_1$ is correlated with the correlation factor r and $\mathbf{H}_2$ is uncorrelated.

and μ_2, for each i, $\lambda_S(i)$ is either zero or a positive solution from (32) by bisection search.

Both of the power constraint functions are increasing functions of $\lambda_S(i)$ for all i, and hence they are also decreasing functions of both μ_1 and μ_2. So, with one of μ_1 and μ_2 fixed, the other (whether or not the solution exists) can be determined by an outer layer bisection search. Such a search for the optimal pair of μ_1 and μ_2 is a 2-D bisection search. In general, we have to conduct this 2-D search. But if only one of the two constraints is active (i.e., satisfied with equality), then only the corresponding multiplier is positive (and the other is zero). In this case, a 1-D bisection search for either μ_1 or μ_2 suffices. A good strategy for efficient implementation is to first consider the two possibilities that $\mu_1 = 0$ or $\mu_2 = 0$. If none of these two cases is the actual solution, we then consider the 2-D search.

IV. Simulation Results

In this section, we present some numerical examples to illustrate the performance of our proposed algorithm. We assume $P_S = P_R = P$ and $n_S = n_R = n_D = L = N$. We define a correlation matrix $\mathbf{C}_r$ with $[\mathbf{C}_r]_{i,j} = r^{|i-j|}$ where r is the normalized correlation coefficient with magnitude $|r| < 1$ [21]. We also define the normalized MSE of $\mathbf{H}_1$ and $\mathbf{H}_2$ as $\frac{E_{\mathbf{H}_2}[J_1']}{N^2}$ and $\frac{J_2'}{N^2}$, respectively. The average over $\mathbf{H}_2$ is computed by using 100 realizations of $\mathbf{H}_2$.

Fig. 2 compares the normalized MSE of $\mathbf{H}_1$ between "optimal source and relay pilots" and "orthogonal source and relay pilots", where $\mathbf{H}_1 = \mathbf{W}_1$, $\mathbf{H}_2 = \mathbf{W}_2$ and $N = 4$. For orthogonal pilots, we use $\mathbf{S}_S = \sqrt{\frac{P_S}{n_S}}\mathbf{I}$ and $\mathbf{F} = \sqrt{\frac{P_R}{(P_S+n_S)n_R}}\mathbf{I}$. Here, both $\mathbf{S}_S$ and $\mathbf{F}$ have orthogonal columns. As expected, the optimal pilots yield a better accuracy than the orthogonal pilots.

From (14), we can see that the channel correlation of both $\mathbf{H}_1$ and $\mathbf{H}_2$ ($\mathbf{C}_{t_1}, \mathbf{C}_{r_1}, \mathbf{C}_{t_2}, \mathbf{C}_{r_2}$) will impact the MSE of $\mathbf{H}_1$. In the following, Figs. 3 and 4 will illustrate how the channel correlation of $\mathbf{H}_1$ and $\mathbf{H}_2$ impact the MSE of $\mathbf{H}_1$ differently. Generally speaking, MSE performance gets better when $\mathbf{H}_1$ is strongly correlated and $\mathbf{H}_2$ is weakly correlated.

Fig. 3 illustrates the normalized MSE of $\mathbf{H}_1$ with the optimal source and relay pilots, where $\mathbf{H}_1 = \mathbf{W}_1$ and $\mathbf{H}_2 = \mathbf{C}_r^{\frac{1}{2}}\mathbf{W}_2\mathbf{C}_r^{\frac{1}{2}}$ with $r = 0.2$ (weak correlation) and $r = 0.8$ (strong correlation). We can see from Fig. 3, as $\mathbf{H}_2$ becomes strongly correlated, the MSE of $\mathbf{H}_1$ degrades, with the difference in performance being more apparent in high power constraint region. We can also observe that with the increase of N, the gap between the performances of weakly correlated $\mathbf{H}_2$ and strongly correlated $\mathbf{H}_2$ becomes larger.

Fig. 4 illustrates the normalized MSE of $\mathbf{H}_1$ with the optimal source and relay pilots, where $\mathbf{H}_2 = \mathbf{W}_2$ and $\mathbf{H}_1 = \mathbf{C}_r^{\frac{1}{2}}\mathbf{W}_1\mathbf{C}_r^{\frac{1}{2}}$ with $r = 0.2$ (weak correlation) and $r = 0.8$ (strong correlation). Different from Fig. 3, Fig. 4 shows that with $\mathbf{H}_1$ getting strongly correlated, the MSE of $\mathbf{H}_1$ improves, with the performance gap more obvious in low power constraint region. We can also observe that with the increase of N, the gap between the performances of weakly correlated $\mathbf{H}_1$ and strongly correlated $\mathbf{H}_1$ becomes larger.

Fig. 5 compares the normalized MSE of $\mathbf{H}_1$ and that of $\mathbf{H}_2$, where $\mathbf{H}_1 = \mathbf{W}_1$ and $\mathbf{H}_2 = \mathbf{W}_2$. We see that the estimation accuracy of $\mathbf{H}_2$ is much higher than that of $\mathbf{H}_1$, which is expected. Recall that the estimation of $\mathbf{H}_2$ in phase 1 is based on a single-hop link while the estimation of $\mathbf{H}_1$ in phase 2 is based on a two-hop relay system where the relay only does "amplify and

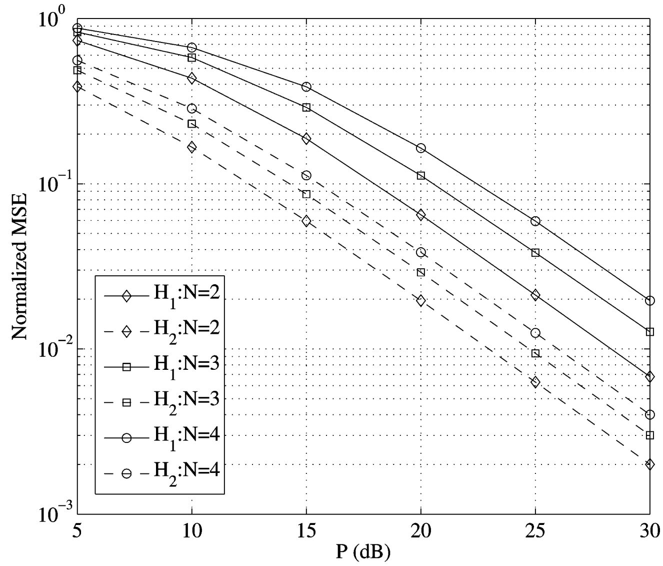


Fig. 5. Normalized MSE of $\mathbf{H}_2$ estimated in phase 1 and the normalized MSE of $\mathbf{H}_1$ estimated in phase 2.

forward". The high accuracy of $\mathbf{H}_2$ is in fact important for the estimation of $\mathbf{H}_1$ in phase 2 where $\mathbf{H}_2$ is assumed to be known. Also note that the method shown in this paper does not have the ambiguity problem suffered by those in [15] and [16].

Unlike the previous figures, for Fig. 5, each estimate of $\mathbf{H}_1$ is based on an estimate of $\mathbf{H}_2$, and the error in $\mathbf{H}_2$ is propagated to $\mathbf{H}_1$.

V. Conclusion

In this paper, we have proposed a two-phase LMMSE-based channel estimation method for a two-hop nonregenerative MIMO relay system. In phase 1, the relay-to-destination channel is estimated for which the relay sends out a source pilot matrix. In phase 2, the source-to-relay channel is estimated for which the source sends out a source pilot matrix and the relay applies a relay pilot matrix. For phase 1, an optimal design of the source pilot has been presented, the result of which is similar to one in [14] while our approach based on generalized KKT conditions provides a complementary perspective. For phase 2, an optimal joint design of the source and relay pilots has been developed, which is a much harder problem than in phase 1. The two-phase channel estimation scheme shown in this paper does not have the ambiguity problem suffered by the schemes in [15] and [16].

The two-phase scheme can be extended to an M-phase scheme for an M-hop nonregenerative relay system. If all nodes are indexed sequentially with the source node being node 0 and the destination node being node M, then in phase m the channel matrix between node $M-m$ and node $M-m+1$ is estimated for which node $M-m$ sends out a source pilot matrix, all other down-stream nodes (except node M) sends out relay pilot matrices and the channel matrices between the adjacent down-stream nodes can be assumed to be known. But a problem with such a scheme is that the estimation errors for the down-stream channels will accumulate and affect the estimation of their upper-stream channels. In practice, such a scheme can be useful only if SNR for each link is sufficiently high.

Appendix I
Proof of Optimal Pilot for Phase 1

Denote the eigenvalue decompositions (EVD) of $\mathbf{C}_{t_2}$ and $\mathbf{C}_{r_2}$ as $\mathbf{C}_{t_2} = \mathbf{U}_{t_2}\mathbf{\Lambda}_{t_2}\mathbf{U}_{t_2}^H$ and $\mathbf{C}_{r_2} = \mathbf{U}_{r_2}\mathbf{\Lambda}_{r_2}\mathbf{U}_{r_2}^H$ with descending eigenvalues. We can then write $\mathbf{C}_{t_2}^{\frac{1}{2}} = \mathbf{U}_{t_2}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}$ and $\mathbf{C}_{r_2}^{\frac{1}{2}} = \mathbf{U}_{r_2}\mathbf{\Lambda}_{r_2}^{\frac{1}{2}}$. We also write $\mathbf{C}_0 = \mathbf{C}_1 \otimes \mathbf{C}_2 = \mathbf{\Lambda}_1 \otimes \mathbf{\Lambda}_2$. If $\mathbf{C}_0 = \mathbf{I}$, both $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are the identity matrices. If $\mathbf{C}_0 = \mathbf{C}_{t_2}^{\frac{H}{2}}\mathbf{C}_{t_2}^{\frac{1}{2}} \otimes \mathbf{C}_{r_2}^{\frac{H}{2}}\mathbf{C}_{r_2}^{\frac{1}{2}}$, we have equivalently $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_{t_2}$ and $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_{r_2}$.

It then follows from (6) that

$$\begin{aligned} L &\doteq J_2' + \mu \cdot \mathrm{tr}(\mathbf{C}_{S_R}) \\ &= \mathrm{tr}\left\{(\mathbf{\Lambda}_1 \otimes \mathbf{\Lambda}_2)\left[\mathbf{I} + \mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{U}_{t_2}^H\mathbf{C}_{S_R}\mathbf{U}_{t_2}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}} \otimes \mathbf{\Lambda}_{r_2}\right]^{-1}\right\} \\ &\quad + \mu \cdot \mathrm{tr}(\mathbf{C}_{S_R}) \\ &= \sum_{i=1}^{n_D} \lambda_2(i)\,\mathrm{tr}\left\{\mathbf{\Lambda}_1\left(\mathbf{I} + \lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\right)^{-1}\right\} \\ &\quad + \mu \cdot \mathrm{tr}(\mathbf{C}) \end{aligned} \tag{33}$$

where $\mathbf{C} = \mathbf{U}_{t_2}^H\mathbf{C}_{S_R}\mathbf{U}_{t_2}$, which has not yet been shown to be diagonal.

It follows from the generalized KKT conditions that the solution to (7) satisfies the sufficient and necessary conditions: $\frac{\partial L}{\partial \mathbf{C}^H} \geq \mathbf{0}$, $\mathbf{C} \geq 0$, $\mu > 0$ and $\mathrm{tr}(\mathbf{C}) = P_R$, which is easy to prove by using (5.95) in [18].

To derive $\frac{\partial L}{\partial \mathbf{C}^H}$, we will use $\partial(\mathbf{AXB}) = \mathbf{A}\partial\mathbf{X}\mathbf{B}$, $\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}\partial\mathbf{X}\mathbf{X}^{-1}$ and $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$. Then, we have

$$\begin{aligned} \partial L = &-\sum_{i=1}^{n_D}\lambda_2(i) \\ &\mathrm{tr}\left\{\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\left(\mathbf{I} + \lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\right)^{-1}\mathbf{\Lambda}_1\right. \\ &\left.\left(\mathbf{I} + \lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\right)^{-1}\lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\partial\mathbf{C}\right\} + \mu \cdot \mathrm{tr}(\partial\mathbf{C}). \end{aligned} \tag{34}$$

Recall that if $\partial L = tr(\mathbf{A}\partial\mathbf{X})$, then $\frac{\partial L}{\partial \mathbf{X}^H} = \mathbf{A}$ [11]. Therefore,

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{C}^H} = &-\sum_{i=1}^{n_D}\lambda_2(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\left(\mathbf{I} + \lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\right)^{-1}\mathbf{\Lambda}_1 \\ &\left(\mathbf{I} + \lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\mathbf{C}\mathbf{\Lambda}_{t_2}^{\frac{1}{2}}\right)^{-1}\lambda_{r_2}(i)\mathbf{\Lambda}_{t_2}^{\frac{1}{2}} + \mu\mathbf{I}. \end{aligned} \tag{35}$$

It is easy to observe from (35) that for any $\mu > 0$, there is always a diagonal $\mathbf{C} \geq 0$ such that $\frac{\partial L}{\partial \mathbf{C}^H} \geq 0$. Therefore, the optimal $\mathbf{C}$ is diagonal.

Let $\mathbf{C} = \mathrm{diag}(c(1), \ldots, c(n_R)) \geq 0$. It follows from (35) that for all $j = 1, 2, \ldots, n_R$,

$$\begin{aligned} \left(\frac{\partial L}{\partial \mathbf{C}^H}\right)_{j,j} &= \mu - \lambda_1(j)\lambda_{t_2}(j)\sum_{i=1}^{n_D}\frac{\lambda_2(i)\lambda_{r_2}(i)}{\left[1 + \lambda_{r_2}(i)\lambda_{t_2}(j)c(j)\right]^2} \\ &\geq 0 \end{aligned} \tag{36}$$

where $\mu > 0$ is such that $\mathrm{tr}(\mathbf{C}) = P_R$.

APPENDIX II
PROOF OF THEOREM 1

From (23), (24), and (25), it is obvious that the cost and the constraints in the problem (15) is invariant to $\mathbf{U}_S$ and hence any unitary $\mathbf{U}_S$ is optimal.

To prove the optimality of the choice of $\mathbf{Q}_S = \mathbf{I}$, we need the following definitions and lemmas from [20].

Definition 1 [20, 1.A.1]: Consider any two real-valued $N \times 1$ vectors $\mathbf{x}$, $\mathbf{y}$, and let $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[N]}$ and $y_{[1]} \geq y_{[2]} \geq \cdots \geq y_{[N]}$ denote the elements of $\mathbf{x}$ and $\mathbf{y}$, respectively, sorted in decreasing order. Then $\mathbf{x}$ is said to be majorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec \mathbf{y}$, if $\sum_{i=1}^{n} x_{[i]} \leq \sum_{i=1}^{n} y_{[i]}, n = 1, 2, \ldots, N-1$ and $\sum_{i=1}^{N} x_{[i]} = \sum_{i=1}^{N} y_{[i]}$.

Definition 2 [20, 1.A.2]: Using the same notations as in DEFINITION 1, $\mathbf{x}$ is said to be weakly majorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec_w \mathbf{y}$, if $\sum_{i=1}^{n} x_{[i]} \leq \sum_{i=1}^{n} y_{[i]}, n = 1, 2, \ldots, N$.

Lemma 1 [20, 9.H.1.h]: For two $N \times N$ positive semidefinite Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\lambda_{a,i}$ and $\lambda_{b,i}$, $i = 1, \ldots, N$, arranged in the descending order respectively, it follows that $\mathrm{tr}(\mathbf{AB}) \geq \sum_{i=1}^{N} \lambda_{a,i}\lambda_{b,N+1-i}$.

Lemma 2 [20, 9.B.1]: For a Hermitian matrix $\mathbf{A}$ with the vector $\mathbf{d}[\mathbf{A}]$ of its main diagonal elements (in descending order for convenience) and the vector $\boldsymbol{\lambda}[\mathbf{A}]$ of its eigenvalues (in descending order for convenience), it follows that $\mathbf{d}[\mathbf{A}] \prec \boldsymbol{\lambda}[\mathbf{A}]$.

Lemma 3 [20, 9.H.2]: For m $N \times N$ complex matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_m$, let $\mathbf{B} = \mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_m$, then $\sigma_b \prec_w \sigma_{a,1} \odot \sigma_{a,2} \odot \cdots \odot \sigma_{a,m}$, where σ_b and $\sigma_{a,i}, i = 1, \ldots, m$, denote $N \times 1$ vectors containing the singular values of $\mathbf{B}$ and $\mathbf{A}_i$ arranged in the same order, respectively, and $\odot$ denotes the Schur (element-wise) product of two vectors.

Lemma 4 [20,3.A.8]: For a real-valued function f, $\mathbf{x} \prec_w \mathbf{y}$ implies $f(\mathbf{x}) \leq f(\mathbf{y})$ if and only if f is increasing with respect to each variable and Schur-convex.

Recall that among the two constraints in the problem (15), only the first, or equivalently (24), depends on $\mathbf{Q}_S$. From Lemma 1, we have

$$\mathrm{tr}\left[\boldsymbol{\Lambda}_S\mathbf{Q}_S\boldsymbol{\Lambda}_{t_1}^{-1}\mathbf{Q}_S^H\right] = \mathrm{tr}\left[\boldsymbol{\Lambda}_S(\mathbf{Q}_S\boldsymbol{\Lambda}_{t_1}^{-1}\mathbf{Q}_S^H)\right] \geq \mathrm{tr}\left[\boldsymbol{\Lambda}_S\boldsymbol{\Lambda}_{t_1}^{-1}\right] \tag{37}$$

where the equality holds when $\mathbf{Q}_S = \mathbf{I}$. Namely, for any given $\boldsymbol{\Lambda}_S$, the source consumes the least amount of power when $\mathbf{Q}_S = \mathbf{I}$.

For the cost J_1' in (15), let us define

$$\begin{aligned}\mathbf{X} &= \left(\boldsymbol{\Lambda}_S \otimes \boldsymbol{\Lambda}_F + \mathbf{I} \otimes \boldsymbol{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\boldsymbol{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\boldsymbol{\Lambda}_F^{\frac{1}{2}} + \mathbf{I}\right)^{-1} \\ &\quad \left(\mathbf{I} \otimes \boldsymbol{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\boldsymbol{\Lambda}_{r_1}\mathbf{Q}_F^H\boldsymbol{\Lambda}_F^{\frac{1}{2}}\right) \\ \mathbf{Y} &= \boldsymbol{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\boldsymbol{\Lambda}_{t_1}\mathbf{Q}_S^H\boldsymbol{\Lambda}_S^{\frac{1}{2}} \otimes \mathbf{I}.\end{aligned}$$

Then, from (23),

$$J_1' - \mathrm{tr}[\mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1}] = -\,\mathrm{tr}[\mathbf{XY}] \tag{38}$$

where $\mathbf{X}$ depends on $\mathbf{Q}_F$, and $\mathbf{Y}$ depends on $\mathbf{Q}_S$.

It follows from Lemma 2 and Lemma 3 that

$$\mathbf{d}[\mathbf{XY}] \prec \boldsymbol{\lambda}[\mathbf{XY}] \prec_w \boldsymbol{\lambda}[\mathbf{X}] \odot \boldsymbol{\lambda}[\mathbf{Y}]. \tag{39}$$

It is known [20] that $\boldsymbol{\lambda}[\mathbf{A}] \prec_w \boldsymbol{\lambda}[\mathbf{A}']$ implies $\boldsymbol{\lambda}[\mathbf{A} \otimes \mathbf{I}] \prec_w \boldsymbol{\lambda}[\mathbf{A}' \otimes \mathbf{I}]$, and $\boldsymbol{\lambda}[\mathbf{B}] \prec_w \boldsymbol{\lambda}[\mathbf{B}']$ implies $\boldsymbol{\lambda}[\mathbf{A}] \odot \boldsymbol{\lambda}[\mathbf{B}] \prec_w \boldsymbol{\lambda}[\mathbf{A}] \odot \boldsymbol{\lambda}[\mathbf{B}']$. It then follows that $\boldsymbol{\lambda}[\mathbf{Y}] \prec_w \boldsymbol{\lambda}[\mathbf{Y}']$ where $\mathbf{Y}' = \boldsymbol{\Lambda}_S^{\frac{1}{2}}\boldsymbol{\Lambda}_{t_1}\boldsymbol{\Lambda}_S^{\frac{1}{2}} \otimes \mathbf{I}$ which is $\mathbf{Y}$ when $\mathbf{Q}_S = \mathbf{I}$. Furthermore,

$$\mathbf{d}[\mathbf{XY}] \prec \boldsymbol{\lambda}[\mathbf{XY}] \prec_w \boldsymbol{\lambda}[\mathbf{X}] \odot \boldsymbol{\lambda}[\mathbf{Y}']. \tag{40}$$

Since $\mathrm{tr}(\cdot)$ is increasing and Schur-convex function of $\mathbf{d}[\mathbf{XY}]$, from Lemma 4, we have

$$tr(\mathbf{XY}) \leq tr(\boldsymbol{\Lambda}[\mathbf{X}]\boldsymbol{\Lambda}[\mathbf{Y}']) \tag{41}$$

where $\boldsymbol{\Lambda}[\mathbf{X}]$ and $\boldsymbol{\Lambda}[\mathbf{Y}']$ are diagonal matrices with the elements of $\boldsymbol{\lambda}[\mathbf{X}]$ and $\boldsymbol{\lambda}[\mathbf{Y}']$ being their diagonal values, respectively. From (38) and (41), J_1' is minimized when $\mathbf{Q}_S = \mathbf{I}$.

The above discussions show that when $\mathbf{Q}_S = \mathbf{I}$, the source consumes the least amount of power and J_1' is minimized. Therefore, we reach the conclusion that $\mathbf{Q}_S = \mathbf{I}$ is optimal.

APPENDIX III
PROOF OF THEOREM 2

It is easy to observe from (23), (24) and (25) that when $\mathbf{C}_{r_1} = \alpha\mathbf{I}$ or equivalently $\boldsymbol{\Lambda}_{r_1} = \alpha\mathbf{I}$, the cost function and both constraints in (15) are invariant to $\mathbf{Q}_F$. Therefore, any unitary $\mathbf{Q}_F$ is optimal.

We know from (23) and (24) that the cost and the first constraint of (15) are invariant to $\mathbf{U}_F$. Using $\mathbf{C}_{r_1} = \alpha\mathbf{I}$ and Lemma 1, the left-hand side of (25) can be written as

$$\begin{aligned}&tr\{\boldsymbol{\Lambda}_S\}tr\left\{\boldsymbol{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\boldsymbol{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}\right\} \\ &\quad + \frac{L}{\alpha} \cdot tr\left\{\boldsymbol{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\boldsymbol{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}\right\} \\ &\quad \geq tr(\Lambda_S)tr\left(\Sigma_{H_2}^{-2}\Lambda_F\right) + \frac{L}{\alpha}tr\left(\Sigma_{H_2}^{-2}\Lambda_F\right)\end{aligned} \tag{42}$$

where the lower bound is achieved when $\mathbf{U}_F = \mathbf{U}_{H_2}$.

REFERENCES

[1] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 29–43, Jan. 2005.

[2] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[3] Z. Fang, Y. Hua, and J. C. Koshy, "Joint source and relay optimization for non-regenerative MIMO relay," in *Proc. IEEE Workshop Sensor Array Multi-Channel Signal Process.*, Waltham, WA, Jul. 2006, pp. 239–243.

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

[5] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.

[6] O. Munoz, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.

[7] C.-B. Chae, T. W. Tang, R. W. Health, and S.-Y. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay netwrks," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 727–738, Feb. 2008.

[8] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pt. Part 2, pp. 5062–5073, Oct. 2008.

[9] 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, no. 12, pp. 4837–4852, Dec. 2009.

[10] Y. Rong and Y. Hua, "Optimality of diagonalization of multihop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.

[11] Y. Yu and Y. Hua, "Power allocation for a MIMO relay system with multiple-antenna users," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2823–2835, May 2010.

[12] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[13] M. Biguesh, S. Gazor, and M. H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 1807–1820, Aug. 2009.

[14] 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, no. 3, pp. 1807–1820, Mar. 2010.

[15] P. Lioliou and M. Viberg, "Least square based channel estimation for MIMO relays," in *Proc. IEEE Int. ITG Workshop Smart Antennas*, Vienna, Austria, Feb. 2008, pp. 90–95.

[16] J. Ma, P. Orlik, J. Zhang, and G. Y. Li, "Pilot matrix design for interim channel estimation in two hop MIMO AF relay systems," in *Proc. Int. Conf. Commun. (ICC)*, Beijing, China, Jun. 2009, pp. 1–5.

[17] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.

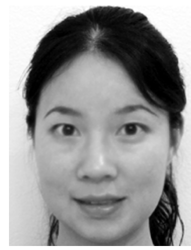
[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[19] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1995.

[20] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[21] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[22] T. Kong and Y. Hua, "Optimal channel estimation and training design for MIMO relays," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2010, pp. 663–667.



Ting Kong received B.S. and M.S. degrees in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2003 and 2006 and the Ph.D. degree from the University of California, Riverside, in 2011.

Her research interests include statistical and array signal processing and signal processing in wireless multihop networks.

Yingbo Hua (S'86–M'88–SM'92–F'02) received the B.S. degree from Southeast University, Nanjing, China, in 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1983 and 1988.

From 1990 to 2001, he was a Professor at the University of Melbourne, Australia. In 2001, he joined the University of California, Riverside, where he is a Senior Full Professor. He was a Visiting Professor with the Hong Kong University of Science and Technology in 1999–2000. He consulted with Microsoft Research, WA, in 2000. He has edited two books and published over 280 articles, with thousands of citations, in the fields of signal processing, sensing, and wireless communications.

Dr. Hua has served as steering and editorial member for five IEEE journals and one EURASIP journal, on several IEEE Signal Processing Society Technical Committees, and other technical, organizing, and advisory committees for numerous international conferences.