Tomlinson-Harashima Precoding Based Transceiver Design for MIMO Relay Systems With Channel Covariance Information

Lenin Gopal, *Member, IEEE,* Yue Rong, *Senior Member, IEEE,* and Zhuquan Zang, *Member, IEEE*

Abstract—In this paper, we investigate the performance of the Tomlinson-Harashima (TH) precoder based nonlinear transceiver design for a nonregenerative multiple-input multiple-output (MIMO) relay system assuming that the full channel state information (CSI) of the source-relay link is known, while only the channel covariance information (CCI) of the relay-destination link is available at the relay node. We first derive the structure of the optimal TH precoding matrix and the source precoding matrix that minimize the mean-squared error (MSE) of the signal waveform estimation at the destination. Then we develop an iterative algorithm to optimize the relay precoding matrix. To reduce the computational complexity of the iterative algorithm, we propose a simplified precoding matrices design scheme. Numerical results show that the proposed precoding matrices design schemes have a better bit-error-rate performance than existing algorithms.

Index Terms—Nonregenerative relay, multiple-input multiple-output (MIMO) relay, Tomlinson-Harashima (TH) precoder, minimum mean-squared error (MMSE).

I. INTRODUCTION

In cooperative communication systems, relay nodes can be deployed between the source and destination nodes to mitigate the channel shadowing effect and provide system spatial diversity. Therefore, cooperative communication has great potential in extending the network coverage and increasing the system throughput with reduced infrastructure cost, and thus, has attracted much research interest recently [1].

Wireless relays can be regenerative or nonregenerative [1], [2], [3]. In the regenerative relay strategy, the relay decodes the received signals from the source node and retransmits the re-encoded information to the destination node. In the nonregenerative relay strategy [3], the relay node simply amplifies (including a possible linear transformation) the received signals from the source node and retransmits the amplified signals to the destination node. Therefore, the complexity and the processing delay of the nonregenerative strategy are generally much smaller than the regenerative strategies.

On the other hand, multiple-input multiple-output (MIMO) systems can provide spatial diversity and multiplexing gains

L. Gopal is with the Department of Electrical and Computer Engineering, Curtin University, Miri, 98009, Sarawak, Malaysia (e-mail: lenin@curtin.edu.my).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, WA 6102, Australia (e-mail: y.rong@curtin.edu.au). The work of Y. Rong was supported in part by the Australian Research Council's Discovery Projects funding scheme under Project DP140102131.

Z. Zang is with the Department of Electrical and Computer Engineering, Curtin University, Miri, 98009, Sarawak, Malaysia (e-mail: zqzang@curtin.edu.my).

to wireless communication systems [4]. When nodes in a relay network have multiple transmit/receive dimensions, we call such system a MIMO relay system. In [5] and [6], relay precoder designs for a two-hop nonregenerative MIMO relay system have been proposed to maximize the mutual information between the source and destination nodes. In [7]-[12], relay precoding algorithms have been developed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination node. The precoder designs in [5]-[12] assume that the full channel state information (CSI) of the source-relay and relay-destination links is available at the relay node.

However, in practical relay communication systems, the exact CSI is unknown and therefore, has to be estimated. There is always mismatch between the true and the estimated CSI due to channel noise, quantization errors and outdated channel estimates. A practical assumption is that only partial information of the relay-destination channel is available at the relay node. In [13] and [14], relay precoding matrix design has been investigated for maximizing the ergodic capacity of the relay system with the channel covariance information (CCI) of the relay-destination channel. Robust broadcasting schemes have been developed in [15] to minimize the transmission power necessary to guarantee that the quality-of-service (QoS) requirements are satisfied for all channels within bounded uncertainty regions around the transmitter's estimate of each user's channel. Minimum MSE (MMSE)-based transceiver designs have been addressed in [16], [18]-[20] with the assumption that the relay knows the CCI of the relay-destination link and the full CSI of the source-relay link.

In the work of [13], [14], [16], [18]-[20], linear transceiver design has been considered for MIMO relay systems, i.e., linear source/relay precoders and linear MMSE receiver. Compared with linear transceivers, nonlinear transceivers have a better bit-error-rate (BER) performance. Recently, nonlinear transceiver based nonregenerative MIMO relay system design has been proposed [21], [22]. Nonlinear transceiver can be implemented at the receiver as a decision-feedback equalizer (DFE) and/or at the transmitter in the form of a Tomlinson-Harashima (TH) precoder. In general, the TH precoding scheme has a better BER performance than the DFE-based transceiver design, as the latter one suffers from error propagation.

The performance of the TH precoding scheme has been well studied for one-hop MIMO systems [23], [24]. In [25], a TH-based pre-filtering algorithm has been designed for multi-

antenna multi-user systems where the base station allocates the transmit power according to the QoS requirement of each active user. A unified approach has been developed in [26] for transceiver optimization in MIMO systems with TH precoding at the transmitter and linear equalization at the receiver. In [27], a multiuser MIMO TH precoding algorithm has been proposed based on quantized CSI at the transmitter side. Recently, the TH precoding scheme has been introduced to nonregenerative MIMO relay systems [28] with the assumption that the full CSI of the entire channel is known at the relay node. In [29]-[32], imperfect CSI has been considered for designing the TH precoding based nonregenerative MIMO relay systems. Due to the nonlinearity nature of the precoding scheme, the TH precoding is highly sensitive to the time-varying nature of the wireless channel [33]. Hence, covariance information based nonlinear transceiver design is more appropriate in such scenario.

In this paper, we propose a TH precoder-based transceiver design for two-hop nonregenerative MIMO relay systems where the full CSI of the source-relay link is known, while only the CCI of the relay-destination link is available at the relay node. To the best of our knowledge, such scenario has not been considered by existing literature in two-hop MIMO relay systems with TH precoder. In particular, we assume that the channel of the relay-destination link is correlated at the transmit antennas and uncorrelated at the receive antennas. This model is suitable for an environment where the relay is not surrounded by local scatterers [34] and the destination node is located amongst rich scatterers [14]. Similar to [7], [9], [16], and [18], we assume that there is no direct link between the source and destination nodes. Moreover, a TH precoder is considered at the source node. The relay precoder is assumed as a linear precoder and the destination node is considered as a linear MMSE receiver.

We propose a transceiver design that minimizes the MSE of the signal waveform estimation at the destination node. We first derive the structure of the optimal TH precoder and the source precoder as a function of the relay precoder. Then we propose an iterative algorithm to optimize the relay precoding matrix by exploiting the link between the mutual information and the weighted MMSE functions [35]. To reduce the computational complexity of the proposed iterative algorithm, we propose a simplified precoding matrices design algorithm. Numerical simulations are carried out to compare the performance of the proposed precoding matrices design algorithms with existing schemes. Simulation results show that both proposed algorithms outperform existing TH precoder based MIMO transceiver optimization algorithms in terms of BER. Moreover, the system BER yielded by the proposed algorithms is very close to that of the system with the perfect CSI. Furthermore, the BER performance of the simplified precoding matrices design algorithm is very close to that of the iterative algorithm. Therefore, the simplified algorithm is very attractive for practical MIMO relay communication systems.

The rest of this paper is organized as follows. The system model and problem formulation are presented in Section II. In Section III, the structure of the optimal TH and source precoding matrix is presented. An iterative algorithm is then proposed to optimize the relay precoding matrix. Finally, a simplified precoding matrices design is proposed to reduce the computational complexity of the proposed iterative algorithm. Numerical examples are shown in Section IV to verify the performance of the proposed algorithms, and conclusions are drawn in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a two-hop nonregenerative MIMO relay system as shown in Fig. 1, where the source, relay, and destination nodes have N_S, N_R, and N_D antennas, respectively. It is assumed that there is no direct link between the source and destination due to the long distance between these two nodes. We also assume that $N_S \leq N_R, N_D$, so that N_S independent data streams can be transmitted.

As shown in Fig. 1, the nonregenerative MIMO relay system has two precoders, i.e, a TH-based source precoder and a relay precoder. At the receiver, we consider a linear MMSE receiver filter. At the transmitter side, the source signal vector $\mathbf{a} \in \mathbb{C}^{N_S \times 1}$ is first fed into the TH precoder. The TH precoder performs a successive cancelation operation which can be implemented through a feedback matrix $\mathbf{B}$ and a modulo operation $\mathrm{MOD}_{\mathrm{m}}(\cdot)$ expressed as

$$\mathrm{MOD}_m(x) = x - 2\sqrt{m}\left\lfloor \frac{x+\sqrt{m}}{2\sqrt{m}} \right\rfloor. \tag{1}$$

Here m is the number of constellation points in the modulation scheme and $\lfloor \cdot \rfloor$ denotes the floor operation.

We denote the signal vector after the modulo operation as $\mathbf{x}$, whose nth element can be written as

$$x_n = a_n - \sum_{l=1}^{n-1}[\mathbf{B}]_{k,l}x_l + e_n, \quad n = 1, \cdots, N_S. \tag{2}$$

where $[\mathbf{B}]_{k,l}$ is the (k,l)-th element of $\mathbf{B}$, $e_n = 2\sqrt{m}\,q_n$, and q_n is a complex-valued quantity with integer real and imaginary components that reduces x_n within the region of $\mathcal{R} = \{x_R + jx_I | x_R, x_I \in (-\sqrt{m}, \sqrt{m})\}$ [26]. By introducing $\mathbf{e} = [e_1, \cdots, e_{N_S}]^T$, (2) can be expressed in matrix-vector form as

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{v} \tag{3}$$

where $\mathbf{C} = \mathbf{B} + \mathbf{I}_{N_S}$ is a lower triangular matrix with unit diagonal elements, $\mathbf{v} = \mathbf{a} + \mathbf{e}$, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, and $\mathbf{x}$ has the covariance matrix of $E\{\mathbf{x}\mathbf{x}^H\} = \sigma_x^2\mathbf{I}_{N_S}$. Here $E\{\cdot\}$ denotes the statistical expectation, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ stand for the matrix transpose, Hermitian transpose, and inversion, respectively.

The data transmission from source to destination is completed in two time slots. At the first time slot, the source node linearly precodes $\mathbf{x}$ as

$$\mathbf{s} = \mathbf{F}\mathbf{x} \tag{4}$$

and transmits $\mathbf{s}$ to the relay node, where $\mathbf{F} \in \mathbb{C}^{N_S \times N_S}$ is the source precoding matrix. The received signal vector at the relay is given by

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{F}\mathbf{x} + \mathbf{n}_1 \tag{5}$$
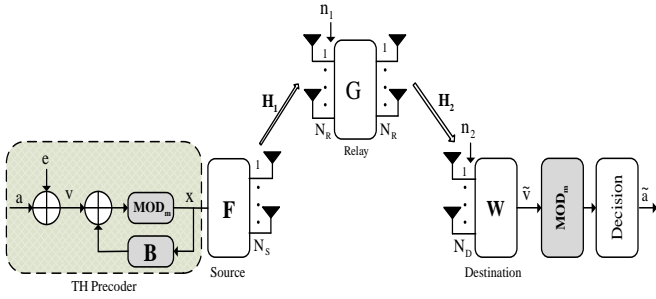
Fig. 1. Two-hop nonregenerative MIMO relay system with TH precoder.

where $\mathbf{H}_1 \in \mathbb{C}^{N_R \times N_S}$ is the channel matrix of the source-relay link, $\mathbf{n}_1 \in \mathbb{C}^{N_R \times 1}$ is the circularly symmetric complex Gaussian noise vector with zero mean and covariance matrix $E\{\mathbf{n}_1 \mathbf{n}_1^H\} = \sigma_1^2 \mathbf{I}_{N_R}$.

At the second time slot, the relay linearly precodes $\mathbf{y}_1$ as

$$\mathbf{x}_2 = \mathbf{G}\mathbf{y}_1 = \mathbf{G}\mathbf{H}_1\mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{n}_1 \quad (6)$$

and forwards $\mathbf{x}_2$ to the destination, where $\mathbf{G} \in \mathbb{C}^{N_R \times N_R}$ is the relay precoding matrix. The received signal vector at the destination is given by

$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{x}_2 + \mathbf{n}_2 = \mathbf{H}_2\mathbf{G}\mathbf{H}_1\mathbf{F}\mathbf{x} + \mathbf{H}_2\mathbf{G}\mathbf{n}_1 + \mathbf{n}_2 \quad (7)$$

where $\mathbf{H}_2 \in \mathbb{C}^{N_D \times N_R}$ is the channel matrix of the relay-destination link, $\mathbf{n}_2 \in \mathbb{C}^{N_D \times 1}$ is the circularly symmetric complex Gaussian noise vector with zero mean and covariance matrix $E\{\mathbf{n}_2 \mathbf{n}_2^H\} = \sigma_2^2 \mathbf{I}_{N_D}$. Let us introduce

$$\mathbf{H} = \mathbf{H}_2\mathbf{G}\mathbf{H}_1\mathbf{F}, \quad \mathbf{n} = \mathbf{H}_2\mathbf{G}\mathbf{n}_1 + \mathbf{n}_2 \quad (8)$$

where $\mathbf{H} \in \mathbb{C}^{N_D \times N_S}$ is the equivalent MIMO channel matrix between the source and destination nodes, and $\mathbf{n} \in \mathbb{C}^{N_D \times 1}$ represents the equivalent noise vector. Now (7) can be rewritten as

$$\mathbf{y}_2 = \mathbf{H}\mathbf{x} + \mathbf{n}. \quad (9)$$

We assume that the relay node knows the instantaneous CSI of $\mathbf{H}_1$, which can be obtained at the relay node through training sequence from the source node. To obtain the instantaneous CSI of $\mathbf{H}_2$ at the relay node, the channel $\mathbf{H}_2$ must be fed back to the relay node from the destination node. When the relay-destination channel varies rapidly, a large signalling overhead for feedback of $\mathbf{H}_2$ is required and this may not be feasible since the rate of feedback link is often limited in practical wireless communication systems. Hence, in the proposed design, we assume that only the covariance information of $\mathbf{H}_2$, which is much more stable than the instantaneous information of $\mathbf{H}_2$, is known at the relay node. In particular, we consider a scenario where the channel of the relay-destination link is correlated at the transmit antennas and uncorrelated at the receive antennas. For example, this scenario can occur in a relay communication system where the relay node is located at the top of a radio mast and a mobile destination node is in an urban area [17]. With this assumption, the channel matrix $\mathbf{H}_2$ can be modelled as

$$\mathbf{H}_2 = \mathbf{H}_\omega \boldsymbol{\Sigma}^{\frac{1}{2}} \quad (10)$$

where $\mathbf{H}_\omega$ is an $N_D \times N_R$ Gaussian matrix having independent and identically distributed (i.i.d.) circularly symmetric complex entries with zero mean and unit variance, and $\boldsymbol{\Sigma}$ is an $N_R \times N_R$ covariance matrix of $\mathbf{H}_2$ at the relay side. Note that the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be known to the relay node and $\mathbf{H}_\omega$ is unown to the relay node.

At the destination node, a linear receiver with weight matrix $\mathbf{W}$ is applied due to its implementation simplicity. Hence, the estimated signal vector at the destination node can be expressed as

$$\tilde{\mathbf{v}} = \mathbf{W}\mathbf{y}_2 = \mathbf{W}\mathbf{H}\mathbf{x} + \mathbf{W}\mathbf{n}. \quad (11)$$

We assume that the average transmission power at the source and relay is upper bounded by p_s and p_r, respectively. Based on (4) and (6), the power constraints at the source and relay can be expressed as

$$P(\mathbf{F}) = \sigma_x^2 tr\{\mathbf{F}\mathbf{F}^H\} \le p_s \quad (12)$$
$$Q_1(\mathbf{F}, \mathbf{G}) = tr\{\mathbf{G}(\sigma_x^2 \mathbf{H}_1\mathbf{F}\mathbf{F}^H\mathbf{H}_1^H + \sigma_1^2 \mathbf{I}_{N_R})\mathbf{G}^H\} \le p_r \quad (13)$$

where $tr\{\cdot\}$ is the trace of a matrix. Our goal is to design $\mathbf{C}$, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{W}$ to obtain the estimated signal $\tilde{\mathbf{v}}$ which minimizes the following MSE cost function subjecting to the power constraints (12) and (13)

$$J_1(\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W}) = tr\{E\{(\tilde{\mathbf{v}} - \mathbf{v})(\tilde{\mathbf{v}} - \mathbf{v})^H\}\}. \quad (14)$$

Note that once $\tilde{\mathbf{v}}$ is obtained, $\mathbf{a}$ can be recovered from (1).

After substituting (11) into (14), the MSE cost function (14) can be written as

$$J_1(\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W}) = tr\{\sigma_x^2(\mathbf{W}\mathbf{H} - \mathbf{C})(\mathbf{W}\mathbf{H} - \mathbf{C})^H$$
$$+ \mathbf{W}\mathbf{R}_n\mathbf{W}^H\} \quad (15)$$

where $\mathbf{R}_n = E\{\mathbf{n}\mathbf{n}^H\}$ is the equivalent noise covariance matrix given by

$$\mathbf{R}_n = \sigma_1^2 \mathbf{H}_2\mathbf{G}\mathbf{G}^H\mathbf{H}_2^H + \sigma_2^2 \mathbf{I}_{N_D}. \quad (16)$$

Based on (12), (13), and (15), the optimal precoding matrices design problem can be formulated as

$$\min_{\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W}} J_1(\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W})$$
$$s.t. \ P(\mathbf{F}) \le p_s$$
$$Q_1(\mathbf{F}, \mathbf{G}) \le p_r. \quad (17)$$

Directly solving the problem (17) is difficult due to the fact that $J_1(\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W})$ is a nonlinear and nonconvex function of $\mathbf{C}$, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{W}$. In the following section, optimal and suboptimal approaches are proposed to solve the problem (17). Firstly, we derive the optimal structure of $\mathbf{C}$ and $\mathbf{F}$ as a function of $\mathbf{G}$. Then an iterative algorithm is proposed to optimize the relay precoding matrix $\mathbf{G}$. Finally, a simplified precoding matrices design is developed to reduce the complexity of the iterative algorithm.

III. PROPOSED TRANSCEIVER DESIGN ALGORITHMS

Since concurrently finding the optimum $\mathbf{C}$, $\mathbf{F}$, and $\mathbf{G}$ in (17) is not possible, hence the optimization problem in (17) is reformulated into three subproblems. In the proposed first subproblem, the lower triangular matrix $\mathbf{C}$ is derived as a function of $\mathbf{F}$, and $\mathbf{G}$, and then, second subproblem optimizes the source precoder matrix $\mathbf{F}$. In the third subproblem, an iterative approach is proposed to obtain the relay precoder matrix $\mathbf{G}$. Due to the computational complexity of the proposed iterative approach, a simplified precoding matrices design is proposed in the subsequent subsection.

A. Tomlinson-Harashima Precoder Design

For any given precoding matrices $\mathbf{C}$, $\mathbf{F}$, and $\mathbf{G}$ which satisfy the power constraints at the source and relay nodes (12) and (13), the weight matrix $\mathbf{W}$ of the optimal linear receiver that minimizes the MSE function $J_1(\mathbf{C}, \mathbf{F}, \mathbf{G}, \mathbf{W})$ is the well known MMSE receiver (Wiener filter) which is given by [36]

$$\mathbf{W} = \sigma_x^2 \mathbf{C} \mathbf{H}^H (\sigma_x^2 \mathbf{H} \mathbf{H}^H + \mathbf{R}_n)^{-1}. \tag{18}$$

After substituting (18) back into (15) and using the matrix inversion lemma [37], the MSE function can be written as

$$J_2(\mathbf{C}, \mathbf{F}, \mathbf{G}) = tr\left\{ \mathbf{C}\left(\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H \tilde{\mathbf{M}}^H \tilde{\mathbf{M}} \mathbf{F}\right)^{-1}\mathbf{C}^H \right\} \tag{19}$$

where

$$\tilde{\mathbf{M}} = \left(\sigma_1^2 \mathbf{H}_2 \mathbf{G} \mathbf{G}^H \mathbf{H}_2^H + \sigma_2^2 \mathbf{I}_{N_D}\right)^{-\frac{1}{2}} \mathbf{H}_2 \mathbf{G} \mathbf{H}_1. \tag{20}$$

To proceed further, we minimize the MSE function (19) with respect to the lower triangular and unit diagonal matrix $\mathbf{C}$. The optimum $\mathbf{C}$ is given in [23] and can be written as

$$\mathbf{C}_{opt} = \mathbf{D} \mathbf{L}^{-1} \tag{21}$$

where

$$\mathbf{L}\mathbf{L}^H = \left(\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H \tilde{\mathbf{M}}^H \tilde{\mathbf{M}} \mathbf{F}\right)^{-1} \tag{22}$$

is the Cholesky factorization. Here $\mathbf{L}$ is a lower triangular matrix and $\mathbf{D}$ is a diagonal matrix which scales the diagonal elements of $\mathbf{C}$ to unit, and given by

$$\mathbf{D} = \text{diag}\{[\mathbf{L}]_{1,1}, \cdots, [\mathbf{L}]_{N_S,N_S}\}. \tag{23}$$

Substituting (21)-(23) back into (19), the MSE function can be written as

$$J_3(\mathbf{F}, \mathbf{G}) = \sum_{i=1}^{N_S}[\mathbf{L}]_{i,i}^2 \geq N_S\left(\prod_{i=1}^{N_S}[\mathbf{L}]_{i,i}\right)^{2/N_S}. \tag{24}$$

Using the arithmetic-geometric inequality (AGI), the inequality in (24) can be obtained and the equality can be achieved when $[\mathbf{L}]_{i,i} = [\mathbf{L}]_{j,j}$, $i \neq j$.

B. Source Precoding Matrix Design

It can be seen from (22) that $[\mathbf{L}]_{i,i}$ depends on the source precoding matrix $\mathbf{F}$. Hence, in this subsection, we determine $\mathbf{F}$ which minimizes the objective function (24). This problem is solved in [24] and [38]. Let us define the eigenvalue decomposition (EVD) of $\tilde{\mathbf{M}}^H \tilde{\mathbf{M}}$ as

$$\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} = \mathbf{V}_{\tilde{M}} \mathbf{\Lambda}_{\tilde{M}} \mathbf{V}_{\tilde{M}}^H \tag{25}$$

where $\mathbf{V}_{\tilde{M}}$ is the eigenvector matrix of $\tilde{\mathbf{M}}$ and $\mathbf{\Lambda}_{\tilde{M}} = \text{diag}\{\Lambda_{\tilde{M},1}, \cdots, \Lambda_{\tilde{M},N_S}\}$ is the diagonal eigenvalue matrix with $\Lambda_{\tilde{M},1} \geq \cdots \geq \Lambda_{\tilde{M},N_S}$.

Lemma 1: [28]. The optimal source precoding matrix as the solution to the problem (17) can be expressed as

$$\mathbf{F}_{opt} = \rho \mathbf{V}_{\tilde{M}} \mathbf{\Phi}_F \tag{26}$$

where $\mathbf{\Phi}_F$ is a unitary matrix and ρ is chosen to satisfy the power constraint (12).

Substituting (25) and (26) back into (22), the Cholesky factorization (22) can be written as

$$\mathbf{L}\mathbf{L}^H = \mathbf{\Phi}_F^H \tilde{\mathbf{\Sigma}}^{\frac{1}{2}} \tilde{\mathbf{\Sigma}}^{\frac{1}{2}} \mathbf{\Phi}_F \tag{27}$$

where

$$\tilde{\mathbf{\Sigma}}^{\frac{1}{2}} = \left(\sigma_x^{-2}\mathbf{I}_{N_S} + \rho^2 \mathbf{\Lambda}_{\tilde{M}}\right)^{-\frac{1}{2}}. $$

Applying the geometric mean decomposition (GMD) [39] to $\tilde{\mathbf{\Sigma}}^{\frac{1}{2}}$, we have

$$\tilde{\mathbf{\Sigma}}^{\frac{1}{2}} = \mathbf{Q}\mathbf{R}\mathbf{P}^H \tag{28}$$

where $\mathbf{Q}$ and $\mathbf{P}$ are semi-unitary matrices and $\mathbf{R}$ is an upper triangular matrix. Substituting (28) back into (27), we have

$$\mathbf{L}\mathbf{L}^H = \mathbf{\Phi}_F^H \mathbf{P} \mathbf{R}^H \mathbf{R} \mathbf{P}^H \mathbf{\Phi}_F. \tag{29}$$

Let us assume $\mathbf{\Phi}_F = \mathbf{P}$ to achieve the lower bound in (24), then (29) can be simplified as

$$\mathbf{L}\mathbf{L}^H = \mathbf{R}^H \mathbf{R}. \tag{30}$$

From (30), we can conclude that $\mathbf{L} = \mathbf{R}^H$. By substituting (30) back into (24), the MSE function can be depicted as

$$J_4(\mathbf{G}) = N_S \prod_{k=1}^{N_S}\left(\sigma_x^{-2} + \rho^2 \Lambda_{\tilde{M},k}\right)^{-1/N_S}. \tag{31}$$

After substituting (26) back into (13), the relay power constraint (13) can be written as

$$Q_2(\mathbf{G}) = tr\left\{\mathbf{G}(\sigma_x^2 \rho^2 \mathbf{H}_1 \mathbf{H}_1^H + \sigma_1^2 \mathbf{I}_{N_R})\mathbf{G}^H\right\} \leq p_r. \tag{32}$$

Now the relay precoding matrix optimization problem can be formulated as

$$\min_{\mathbf{G}} \quad J_4(\mathbf{G}) \qquad s.t. \ Q_2(\mathbf{G}) \leq p_r. \tag{33}$$

C. Relay Precoding Matrix Design

In this subsection, we derive the optimum $\mathbf{G}$. It is worth to note that the eigenvalues of (20) are a nonlinear function of $\mathbf{G}$ and the optimization problem (33) is not convex. To solve the problem (33), we consider the following equivalent function

$$\prod_{k=1}^{N_S}\left(\sigma_x^{-2}+\rho^2\Lambda_{\tilde{M},k}\right)=\left|\sigma_x^{-2}\mathbf{I}_{N_S}+\rho^2\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\right|. \quad (34)$$

Here $|\cdot|$ denotes the matrix determinant. Substituting (34) into (33) and taking the $\log$ operation to the cost function, the optimization problem (33) can be reformulated as

$$\min_{\mathbf{G}}\ -X \qquad s.t.\ Q_2(\mathbf{G})\le p_r \quad (35)$$

where

$$X=\log|\mathbf{A}^{-1}|$$
$$\mathbf{A}=\left(\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S}+\mathbf{H}_1^H\mathbf{G}^H\mathbf{H}_2^H\mathbf{R}_n^{-1}\mathbf{H}_2\mathbf{G}\mathbf{H}_1\right)^{-1}. \quad (36)$$

It is worth noting that if $\mathbf{H}_2$ is known at the relay node, (35) has a closed-form solution [9]. However, as the exact $\mathbf{H}_2$ is unknown, it is impossible to solve the problem (35). To overcome this difficulty, we consider to minimize the mean value of $-X$ as given by the following problem

$$\min_{\mathbf{G}}\ E_{\mathbf{H}_2}\{-X\} \qquad s.t.\ Q_2(\mathbf{G})\le p_r \quad (37)$$

where $E_{\mathbf{H}_2}\{\cdot\}$ denotes the statistical expectation with respect to $\mathbf{H}_2$.

We notice that due to the matrix determinant operator in X, the closed-form expression of the objective function in (37), if possible to obtain, is a very complicated function of $\mathbf{G}$, which makes the problem (37) difficult to solve. To overcome this challenge, we apply the following theorem.

Theorem 1: The problem (35) has the same Karush-Kuhn-Tucker (KKT) conditions on $\mathbf{G}$ as the problem of

$$\min_{\mathbf{G},\boldsymbol{\Omega}}\ tr\{\boldsymbol{\Omega}\mathbf{A}\}-\log|\boldsymbol{\Omega}| \quad (38)$$
$$s.t.\ Q_2(\mathbf{G})\le p_r \quad (39)$$

when the Hermitian weight matrix $\boldsymbol{\Omega}$ takes value of

$$\boldsymbol{\Omega}=\mathbf{A}^{-1}. \quad (40)$$

Moreover, with given $\mathbf{G}$, the weight matrix $\boldsymbol{\Omega}$ minimizing (38) is given by (40).

PROOF: See Appendix A. □

Based on Theorem 1, we can solve the problem (35) using an iterative approach, where in each iteration, with $\boldsymbol{\Omega}$ from the previous iteration, we first optimize $\mathbf{G}$ by solving the problem (38)-(39). Then we update $\boldsymbol{\Omega}$ as (40) using $\mathbf{G}$ obtained in the current iteration. Note that the conditional updates of $\mathbf{G}$ and $\boldsymbol{\Omega}$ may either decrease or maintain but cannot increase the objective function (38). Monotonic convergence of the iterative algorithm towards (at least) a locally optimal solution follows directly from this observation. As $\boldsymbol{\Omega}$ is unknown due to an unknown $\mathbf{H}_2$, we use $\bar{\boldsymbol{\Omega}}$ given by

$$\bar{\boldsymbol{\Omega}}=E_{\mathbf{H}_2}\{\boldsymbol{\Omega}\}$$
$$=E_{\mathbf{H}_2}\{\mathbf{A}^{-1}\}$$
$$=\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S}+E_{\mathbf{H}_2}\{\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\} \quad (41)$$

where

$$E_{\mathbf{H}_2}\{\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\}$$
$$=\mathbf{H}_1^H E_{\mathbf{H}_2}\{\mathbf{G}^H\mathbf{H}_2^H\mathbf{R}_n^{-1}\mathbf{H}_2\mathbf{G}\}\mathbf{H}_1$$
$$\le\sigma_1^{-2}\mathbf{H}_1^H\left[\mathbf{I}_{N_R}-\sigma_2^2(\sigma_1^2\mathbf{G}^H E_{\mathbf{H}_2}\{\mathbf{H}_2^H\mathbf{H}_2\}\mathbf{G}+\sigma_2^2\mathbf{I}_{N_R})^{-1}\right]\mathbf{H}_1$$
$$=\sigma_1^{-2}\mathbf{H}_1^H\left[\mathbf{I}_{N_R}-\sigma_2^2(\sigma_1^2 N_D\mathbf{G}^H\boldsymbol{\Sigma}\mathbf{G}+\sigma_2^2\mathbf{I}_{N_R})^{-1}\right]\mathbf{H}_1. \quad (42)$$

Substituting (36) into (38), for a given $\bar{\boldsymbol{\Omega}}$, the objective function of $\mathbf{G}$ can be expressed as

$$T_1(\mathbf{G})=tr\Big\{\bar{\boldsymbol{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S}+\mathbf{H}_1^H\mathbf{G}^H\mathbf{H}_2^H$$
$$\times\left(\sigma_1^2\mathbf{H}_2\mathbf{G}\mathbf{G}^H\mathbf{H}_2^H+\sigma_2^2\mathbf{I}_{N_D}\right)^{-1}\mathbf{H}_2\mathbf{G}\mathbf{H}_1\Big]^{-1}\Big\}. (43)$$

Now the problem is reduced to find the optimal $\mathbf{G}$ that minimizes $T_1(\mathbf{G})$ subjecting to the relay power constraint (32). Using the matrix inversion lemma [37], (43) can be rewritten as

$$T_1(\mathbf{G})=tr\Big\{\bar{\boldsymbol{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S}+\sigma_1^{-2}\mathbf{H}_1^H\Big[\mathbf{I}_{N_R}$$
$$-\Big(\mathbf{I}_{N_R}+\frac{\sigma_1^2}{\sigma_2^2}\mathbf{G}^H\mathbf{H}_2^H\mathbf{H}_2\mathbf{G}\Big)^{-1}\Big]\mathbf{H}_1\Big]^{-1}\Big\}. \quad (44)$$

Let us introduce the singularvalue decomposition (SVD) of $\mathbf{H}_1$ as

$$\mathbf{H}_1=\mathbf{U}_1\boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H \quad (45)$$

where $\boldsymbol{\Lambda}_1=\text{diag}\{\Lambda_{1,1},\cdots,\Lambda_{1,N_S}\}$ is a diagonal matrix with $\Lambda_{1,1}\ge\cdots\ge\Lambda_{1,N_S}$, $\mathbf{U}_1\in\mathbb{C}^{N_R\times N_S}$ and $\mathbf{V}_1\in\mathbb{C}^{N_S\times N_S}$ are the singular vector matrices of $\mathbf{H}_1$. We also introduce the EVD of $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma}=\mathbf{V}_\Sigma\boldsymbol{\Lambda}_\Sigma\mathbf{V}_\Sigma^H \quad (46)$$

where $\boldsymbol{\Lambda}_\Sigma=\text{diag}\{\Lambda_{\Sigma,1},\cdots,\Lambda_{\Sigma,N_R}\}$ with $\Lambda_{\Sigma,1}\ge\cdots\ge\Lambda_{\Sigma,N_R}$. Substituting (46) back into (10), the channel matrix $\mathbf{H}_2$ can be written as

$$\mathbf{H}_2=\widetilde{\mathbf{H}}_\omega\boldsymbol{\Lambda}_\Sigma^{\frac{1}{2}}\mathbf{V}_\Sigma^H \quad (47)$$

where $\widetilde{\mathbf{H}}_\omega\triangleq\mathbf{H}_\omega\mathbf{V}_\Sigma$ has the same distribution as $\mathbf{H}_\omega$, as the unitary matrix $\mathbf{V}_\Sigma$ does not change the statistical distribution of $\mathbf{H}_\omega$. Thus, $\widetilde{\mathbf{H}}_\omega$ is an $N_D\times N_R$ complex Gaussian matrix having i.i.d. circularly symmetric entries. It can be shown that the optimal $\mathbf{G}$ minimizing (44) can be expressed as

$$\mathbf{G}=\mathbf{V}_\Sigma\widetilde{\mathbf{G}}\mathbf{U}_1^H. \quad (48)$$

It can be seen from (48) that the optimal $\mathbf{G}$ allocates power according to the eigenmodes of $\mathbf{H}_1\mathbf{H}_1^H$ and $\boldsymbol{\Sigma}$, and we need to determine $\widetilde{\mathbf{G}}$.

Substituting (45)-(48) back into (44), we obtain

$$T_1(\widetilde{\mathbf{G}})=tr\Big\{\bar{\boldsymbol{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S}+\sigma_1^{-2}\mathbf{V}_1\boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{U}_1^H$$
$$\times(\mathbf{I}_{N_R}-\mathbf{D}_1)\mathbf{U}_1\boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H\Big]^{-1}\Big\} \quad (49)$$

where

$$\mathbf{D}_1=\Big(\mathbf{I}_{N_R}+\frac{\sigma_1^2}{\sigma_2^2}\mathbf{U}_1\widetilde{\mathbf{G}}^H\boldsymbol{\Lambda}_\Sigma^{\frac{1}{2}}\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\boldsymbol{\Lambda}_\Sigma^{\frac{1}{2}}\widetilde{\mathbf{G}}\mathbf{U}_1^H\Big)^{-1}.$$

Using $\mathbf{U}_1^H \mathbf{U}_1 = \mathbf{I}_{N_S}$, (49) can be simplified to

$$T_1(\widetilde{\mathbf{G}}) = tr\Big\{\bar{\mathbf{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}$$
$$\times(\mathbf{I}_{N_R} - \mathbf{D}_2)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H\Big]^{-1}\Big\} \quad (50)$$

where

$$\mathbf{D}_2 = \Big(\mathbf{I}_{N_R} + \frac{\sigma_1^2}{\sigma_2^2}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_\Sigma^{\frac{1}{2}}\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\mathbf{\Lambda}_\Sigma^{\frac{1}{2}}\widetilde{\mathbf{G}}\Big)^{-1}.$$

It can be seen from (50) that $T_1(\widetilde{\mathbf{G}})$ depends on $\widetilde{\mathbf{H}}_\omega$, which is random and unknown. In the following, we optimize $E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\}$, where $E_{\widetilde{\mathbf{H}}_\omega}\{\cdot\}$ indicates that the expectation is taken with respect to the random matrix $\widetilde{\mathbf{H}}_\omega$. Now $E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\}$ can be expressed as

$$E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\} = E_{\widetilde{\mathbf{H}}_\omega}\Big\{tr\Big\{\bar{\mathbf{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}$$
$$\times(\mathbf{I}_{N_R} - \mathbf{D}_2)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H\Big]^{-1}\Big\}\Big\}. \quad (51)$$

Direct minimization of (51) over $\widetilde{\mathbf{G}}$ is difficult due to the expectation operation. In the following, a lower bound of (51) is exploited together with the power constraint (32) to derive the suboptimal $\widetilde{\mathbf{G}}$ for the precoding matrix $\mathbf{G}$.

Theorem 2: A lower bound of (51) is given by

$$T_2(\widetilde{\mathbf{G}}) = tr\Big\{\bar{\mathbf{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}\Big[\mathbf{I}_{N_R}$$
$$- \Big(\mathbf{I}_{N_R} + \frac{\sigma_1^2 N_D}{\sigma_2^2}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_\Sigma\widetilde{\mathbf{G}}\Big)^{-1}\Big]\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H\Big]^{-1}\Big\}. \quad (52)$$

PROOF: See Appendix B. □

Substituting (45) and (48) back into (32), the power constraint at the relay node can be simplified to

$$Q_3(\widetilde{\mathbf{G}}) = tr\Big\{\widetilde{\mathbf{G}}(\rho^2\sigma_x^2\mathbf{\Lambda}_1 + \sigma_1^2\mathbf{I}_{N_R})\widetilde{\mathbf{G}}^H\Big\} \le p_r. \quad (53)$$

From (52) and (53), the problem of optimizing $\widetilde{\mathbf{G}}$ can be written as

$$\min_{\widetilde{\mathbf{G}}} \; T_2(\widetilde{\mathbf{G}}) \qquad s.t. \; Q_3(\widetilde{\mathbf{G}}) \le p_r. \quad (54)$$

The problem (54) does not have a closed-form solution due to the presence of $\bar{\mathbf{\Omega}}$ in the objective function. The problem (54) can be solved by resorting to numerical methods, such as the projected gradient algorithm [40].

The procedure of the iterative precoding matrices design algorithm developed in Sections III.A-III.C is summarized in Table I, where the superscript (m) denotes the variables at the mth iteration.

TABLE I
PROCEDURE OF THE PROPOSED ITERATIVE PRECODING MATRICES DESIGN ALGORITHM

1) Initialize the algorithm with $\bar{\mathbf{\Omega}}^{(0)} = \sqrt{p_s/N_S}\mathbf{I}_{N_S}$; Set $m = 0$.
2) Update $\widetilde{\mathbf{G}}^{(m)}$ by solving the problem (54) using the projected gradient algorithm [40].
3) Update $\bar{\mathbf{\Omega}}^{(m+1)}$ by (41); If $\max \text{abs}(\bar{\mathbf{\Omega}}^{(m+1)} - \bar{\mathbf{\Omega}}^{(m)}) \le \varepsilon$, then go to Step 4).
   Otherwise, let $m := m + 1$ and go to Step 2).
4) Obtain $\mathbf{F}_{opt}$ as (26), and $\mathbf{C}_{opt}$ by (21) with $\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}$ replaced by $E_{\mathbf{H}_2}\{\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\}$ in (42).

D. Simplified Precoding Matrices Design

In this subsection, we propose a precoding matrices design algorithm which has a significant computational complexity reduction compared with the iterative algorithm in Table I. In this algorithm, a lower bound of the MSE function is obtained by using the arithmetic-geometric mean inequality [37], which is given by the following lemma.

Lemma 2: For a positive semidefinite (PSD) $\mathbf{A} \in \mathbb{C}^{N \times N}$, there is

$$|\mathbf{A}|^{1/N} \le tr(\mathbf{A})/N \quad (55)$$

where the equality is achieved when $\mathbf{A}$ is a diagonal matrix with equal diagonal elements. Using Lemma 2, a lower bound of the MSE function (19) can be written as

$$\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\mathbf{F}\big|^{-1/N_S}$$
$$\le tr\big\{\mathbf{C}(\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\mathbf{F})^{-1}\mathbf{C}^H\big\}/N_S. \quad (56)$$

Since minimizing $|\mathbf{A}|^{-1}$ is equivalent to maximizing $|\mathbf{A}|$ [37], the source and relay precoding matrices design problem can be reformulated as

$$\max_{\mathbf{F},\mathbf{G}} \; X(\mathbf{F},\mathbf{G})$$
$$s.t. \; P(\mathbf{F}) \le p_s$$
$$Q_1(\mathbf{F},\mathbf{G}) \le p_r \quad (57)$$

where the objective function $X(\mathbf{F},\mathbf{G})$ can be expressed as

$$X(\mathbf{F},\mathbf{G}) = \log\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\mathbf{F}\big|$$
$$= \log\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \mathbf{F}^H\mathbf{H}_1^H\mathbf{G}^H\mathbf{H}_2^H$$
$$\times(\sigma_1^2\mathbf{H}_2\mathbf{G}\mathbf{G}^H\mathbf{H}_2^H + \sigma_2^2\mathbf{I}_{N_D})^{-1}\mathbf{H}_2\mathbf{G}\mathbf{H}_1\mathbf{F}\big|$$
$$= \log\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{F}^H\mathbf{H}_1^H\Big[\mathbf{I}_{N_R}$$
$$- \Big(\mathbf{I}_{N_R} + \frac{\sigma_1^2}{\sigma_2^2}\mathbf{G}^H\mathbf{H}_2^H\mathbf{H}_2\mathbf{G}\Big)^{-1}\Big]\mathbf{H}_1\mathbf{F}\big|. \quad (58)$$

Here the matrix inversion lemma [37] is applied to obtain the last equation. From [45], we can define that if the matrix is diagonal, then the determinant of a positive definite matrix is maximized. Hence, without loss of generality, let us express the source precoding matrix $\mathbf{F}$ in terms of the following decomposition

$$\mathbf{F} = \mathbf{V}_1\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Phi}_F \quad (59)$$

where $\mathbf{\Lambda}_F = \text{diag}\{\Lambda_{F,1}\cdots\Lambda_{F,N_S}\}$, and $\mathbf{\Phi}_F$ is a unitary matrix defined later. Let us assume that the matrix $\mathbf{G}$ which maximizes (58) can be expressed as

$$\mathbf{G} = \mathbf{V}_{\Sigma,1}\mathbf{\Lambda}_G^{\frac{1}{2}}\mathbf{U}_1^H \quad (60)$$

where $\mathbf{V}_{\Sigma,1}$ contains N_S columns of $\mathbf{V}_\Sigma$ associated with the largest N_S eigenvalues of $\mathbf{\Sigma}$, and $\mathbf{\Lambda}_G = \text{diag}\{\Lambda_{G,1},\cdots,\Lambda_{G,N_S}\}$. Substituting (45), (47), (59) and (60) in (58), we have

$$X(\mathbf{F},\mathbf{G}) = \log\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_1^{\frac{1}{2}}(\mathbf{I}_{N_S} - \mathbf{D}_5)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{\Lambda}_F^{\frac{1}{2}}\big| \quad (61)$$

where

$$\mathbf{D}_5 = \Big(\mathbf{I}_{N_S} + \frac{\sigma_1^2}{\sigma_2^2}\mathbf{\Lambda}_G^{\frac{1}{2}}\mathbf{\Lambda}_{\Sigma,1}^{\frac{1}{2}}\widetilde{\mathbf{H}}_{\omega,1}^H\widetilde{\mathbf{H}}_{\omega,1}\mathbf{\Lambda}_{\Sigma,1}^{\frac{1}{2}}\mathbf{\Lambda}_G^{\frac{1}{2}}\Big)^{-1}$$

and $\widetilde{\mathbf{H}}_{\omega,1}$ is a matrix containing the left-most N_S columns of $\widetilde{\mathbf{H}}_\omega$. It can be seen from (61) that $X(\mathbf{F},\mathbf{G})$ depends on $\widetilde{\mathbf{H}}_{\omega,1}$, which is random and unknown. In the following, we optimize $E_{\widetilde{\mathbf{H}}_{\omega,1}}\{X(\mathbf{F},\mathbf{G})\}$ given by

$$E_{\widetilde{\mathbf{H}}_{\omega,1}}\{X(\mathbf{F},\mathbf{G})\} = E_{\widetilde{\mathbf{H}}_{\omega,1}}\Big\{ \log \Big| \sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_1^{\frac{1}{2}}$$
$$\times (\mathbf{I}_{N_S} - \mathbf{D}_5)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{\Lambda}_F^{\frac{1}{2}} \Big| \Big\}. \quad (62)$$

Due to the expectation operation, maximizing (62) with respect to $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_G$ is difficult. In the following, an upper bound of $E_{\widetilde{\mathbf{H}}_{\omega,1}}\{X(\mathbf{F},\mathbf{G})\}$ is used together with the power constraints (12) and (13) to derive the suboptimal power allocation for the precoding matrices $\mathbf{F}$ and $\mathbf{G}$.

Theorem 3: The function

$$f(\mathbf{Z}) = \log\Big|\mathbf{I}_{N_S} + \sigma_x^2\sigma_1^{-2}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_1^{\frac{1}{2}}\Big[\mathbf{I}_{N_S}$$
$$-\Big(\mathbf{I}_{N_S} + \frac{\sigma_1^2}{\sigma_2^2}\mathbf{\Lambda}_G^{\frac{1}{2}}\mathbf{\Lambda}_{\Sigma,1}^{\frac{1}{2}}\mathbf{Z}\mathbf{\Lambda}_{\Sigma,1}^{\frac{1}{2}}\mathbf{\Lambda}_G^{\frac{1}{2}}\Big)^{-1}\Big]\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{\Lambda}_F^{\frac{1}{2}}\Big| \quad (63)$$

is concave with respect to a PSD $\mathbf{Z}$.

PROOF: See Appendix C. □

According to Theorem 3, $X(\mathbf{F},\mathbf{G})$ is concave in $\widetilde{\mathbf{H}}_{\omega,1}^H\widetilde{\mathbf{H}}_{\omega,1}$. Hence, $E_{\widetilde{\mathbf{H}}_{\omega,1}}\{X(\mathbf{F},\mathbf{G})\}$ has the following upper bound by using the Jensen's inequality [42]

$$X_U = \log\Big|\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_1^{\frac{1}{2}}(\mathbf{I}_{N_S}-\mathbf{D}_6)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{\Lambda}_F^{\frac{1}{2}}\Big| \quad (64)$$

where

$$\mathbf{D}_6 = \Big(\mathbf{I}_{N_S} + \frac{\sigma_1^2}{\sigma_2^2}\mathbf{\Lambda}_G^{\frac{1}{2}}\mathbf{\Lambda}_\Sigma^{\frac{1}{2}}E_{\widetilde{\mathbf{H}}_{\omega,1}}\{\widetilde{\mathbf{H}}_{\omega,1}^H\widetilde{\mathbf{H}}_{\omega,1}\}\mathbf{\Lambda}_\Sigma^{\frac{1}{2}}\mathbf{\Lambda}_G^{\frac{1}{2}}\Big)^{-1}.$$

Using the property of Gaussian random matrices with i.i.d. circularly symmetric complex entries, we have $E_{\widetilde{\mathbf{H}}_{\omega,1}}\{\widetilde{\mathbf{H}}_{\omega,1}^H\widetilde{\mathbf{H}}_{\omega,1}\} = N_D\mathbf{I}_{N_S}$, and (64) can be simplified to

$$X_U \triangleq \log\big|\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{\Lambda}_F\mathbf{\Lambda}_1(\mathbf{I}_{N_S}-\mathbf{D}_7)\big| \quad (65)$$

where

$$\mathbf{D}_7 = \Big(\mathbf{I}_{N_S} + \frac{\sigma_1^2 N_D}{\sigma_2^2}\mathbf{\Lambda}_G\mathbf{\Lambda}_\Sigma\Big)^{-1}.$$

By substituting (45), (59), and (60) back into (12) and (13), the power constraints at source and relay nodes can be simplified to

$$\sigma_x^2 tr\{\mathbf{\Lambda}_F\} \le p_s \quad (66)$$
$$tr\{(\sigma_x^2\mathbf{\Lambda}_F\mathbf{\Lambda}_1 + \sigma_1^2\mathbf{I}_{N_S})\mathbf{\Lambda}_G\} \le p_r. \quad (67)$$

Based on (65)-(67), the diagonal elements of $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_G$ can be obtained by solving the following constrained optimization problem with scalar variables

$$\max_{\{\Lambda_{F,i}\},\{\Lambda_{G,i}\}} \sum_{i=1}^{N_S} \log\Big(\sigma_x^{-2} + \frac{N_D\Lambda_{F,i}\Lambda_{1,i}\Lambda_{G,i}\Lambda_{\Sigma,i}}{\sigma_1^2 N_D\Lambda_{G,i}\Lambda_{\Sigma,i}+\sigma_2^2}\Big) \quad (68)$$

$$s.t. \sum_{i=1}^{N_S} \sigma_x^2\Lambda_{F,i} \le p_s \quad (69)$$

$$\sum_{i=1}^{N_S} (\sigma_x^2\Lambda_{F,i}\Lambda_{1,i}+\sigma_1^2)\Lambda_{G,i} \le p_r \quad (70)$$

where $\{\Lambda_{F,i}\} \triangleq \Lambda_{F,1},\cdots,\Lambda_{F,N_S}, \{\Lambda_{G,i}\} \triangleq \Lambda_{G,1},\cdots,\Lambda_{G,N_S}$. Let us introduce

$$a_i \triangleq \sigma_x^2\Lambda_{F,i}, \quad i=1,\cdots,N_S \quad (71)$$
$$b_i \triangleq (\sigma_x^2\Lambda_{F,i}\Lambda_{1,i}+\sigma_1^2)\Lambda_{G,i}, \quad i=1,\cdots,N_S. \quad (72)$$

Substituting (71) and (72) back into (68)-(70), the problem (68)-(70) can be rewritten as

$$\max_{\{a_i\},\{b_i\}} \sum_{i=1}^{N_S} \log\frac{(a_i\Lambda_{1,i}+\sigma_1^2)(N_Db_i\Lambda_{\Sigma,i}+\sigma_2^2)}{\sigma_x^2(\sigma_1^2 N_D\Lambda_{\Sigma,i}b_i + a_i\Lambda_{1,i}\sigma_2^2 + \sigma_2^2\sigma_1^2)} \quad (73)$$

$$s.t. \sum_{i=1}^{N_S} a_i \le p_s, \quad a_i \ge 0, \quad i=1,\cdots,N_S \quad (74)$$

$$\sum_{i=1}^{N_S} b_i \le p_r, \quad b_i \ge 0, \quad i=1,\cdots,N_S \quad (75)$$

where $\{a_i\} \triangleq a_1,\cdots,a_{N_S}$ and $\{b_i\} \triangleq b_1,\cdots,b_{N_S}$.

Using the KKT conditions, the solution to the problem (73)-(75) is given by

$$a_i = \varphi_{a_i}\Bigg[\sqrt{\frac{b_i^2\Lambda_{\Sigma,i}^2}{\sigma_2^4}+\frac{4b_i\Lambda_{1,i}\Lambda_{\Sigma,i}}{\mu_s N_D\sigma_1^2\sigma_2^2}} - \frac{b_i\Lambda_{\Sigma,i}}{\sigma_2^2} - \frac{2}{N_D}\Bigg]^+ \quad (76)$$

$$b_i = \varphi_{b_i}\Bigg[\sqrt{\frac{a_i^2\Lambda_{1,i}^2}{\sigma_1^4}+\frac{4N_Da_i\Lambda_{1,i}\Lambda_{\Sigma,i}}{\mu_r\sigma_1^2\sigma_2^2}} - \frac{a_i\Lambda_{1,i}}{\sigma_1^2} - 2\Bigg]^+ \quad (77)$$

$$i = 1,\cdots,N_S$$

where $[x]^+ = \max(x,0)$, μ_s and μ_r are the Lagrangian multipliers chosen to meet the power constraints (74) and (75), and

$$\varphi_{a_i} = \frac{\sigma_1^2 N_D}{2\Lambda_{1,i}}, \quad \varphi_{b_i} = \frac{\sigma_2^2}{2N_D\Lambda_{\Sigma,i}}, \quad i=1,\cdots,N_S. \quad (78)$$

The detailed derivation of (76) and (77) is shown in Appendix D.

It can be seen from (76) and (77) that $\{a_i\}$ and $\{b_i\}$ are functions of each other. Thus, directly solving (76) and (77) is difficult. To avoid this difficulty, we propose an iterative algorithm to compute $\{a_i\}$ and $\{b_i\}$. This algorithm is initialized with $a_i = \sqrt{p_s/N_S}$, $i=1,\cdots,N_S$. At each iteration, we first optimize $\{b_i\}$ according to (77) based on the initial value of $\{a_i\}$. Then we optimize $\{a_i\}$ following (76) using $\{b_i\}$. We update $\{a_i\}$ and $\{b_i\}$ iteratively until convergence. Finally, the diagonal elements of $\mathbf{\Lambda}_F$ and $\mathbf{\Lambda}_G$ can be obtained by from (71) and (72).

After obtaining the optimal source and relay precoding matrices, we now focus on determining the structure of the unitary matrix $\mathbf{\Phi}_F$ and the lower triangular matrix $\mathbf{C}$. The optimal $\mathbf{C}$ is given in (21). Substituting (45), (47), (59), and (60) back into (22), the Cholesky factorization (22) can be written as

$$\mathbf{L}\mathbf{L}^H = \mathbf{\Phi}_F^H\tilde{\mathbf{\Psi}}^{\frac{1}{2}}\tilde{\mathbf{\Psi}}^{\frac{1}{2}}\mathbf{\Phi}_F \quad (79)$$

where

$$\tilde{\mathbf{\Psi}}^{\frac{1}{2}} = \big(\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{\Lambda}_F\mathbf{\Lambda}_1(\mathbf{I}_{N_R}-\mathbf{D}_7)\big)^{-\frac{1}{2}}. \quad (80)$$

The proof of (80) can be found in [16]. Applying the GMD [39] to $\tilde{\boldsymbol{\Psi}}^{\frac{1}{2}}$, we have

$$\left(\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\boldsymbol{\Lambda}_F\boldsymbol{\Lambda}_1(\mathbf{I}_{N_R} - \mathbf{D}_7)\right)^{-\frac{1}{2}} = \mathbf{Q}_2\mathbf{R}_2\mathbf{P}_2^H. \quad (81)$$

Substituting (81) back into (79), the Cholesky factorization (79) can be written as $\mathbf{L}\mathbf{L}^H = \boldsymbol{\Phi}_F^H\mathbf{P}_2\mathbf{R}_2^H\mathbf{R}_2\mathbf{P}_2^H\boldsymbol{\Phi}_F$. Similar to (29) and (30), the unitary matrix $\boldsymbol{\Phi}_F$ can be chosen as $\boldsymbol{\Phi}_F = \mathbf{P}_2$.

IV. SIMULATION RESULTS

In this section, we investigate the performance of the proposed precoder design algorithms through numerical simulations. We simulate a two-hop nonregenerative MIMO relay system with $N_S = N_R = N_D = N$. The channel matrices $\mathbf{H}_1$ and $\mathbf{H}_\omega$ have complex Gaussian entries with zero mean and unit variance and the information-carrying symbols are generated from 16-QAM constellations.

The elements of the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{H}_2$ is generated by $[\boldsymbol{\Sigma}]_{i,j} = J_0(2\pi|i - j| \triangle d_t/\lambda_c)$ [34], where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, $\triangle$ is the angle of fading spread, λ_c is the wavelength at the center frequency, and d_t is the spacing of transmit antennas. Let us introduce $k = \lambda_c/\triangle d_t$. Unless explicitly mentioned, we set $N = 4$ and $k = 3$ in the simulations. The signal-to-noise ratios (SNRs) for the source-relay and relay-destination links are defined as $\text{SNR}_1 = \frac{\sigma_x^2}{\sigma_1^2}$ and $\text{SNR}_2 = \frac{P_r}{N_R\sigma_2^2}$, respectively.

First, we study the impact of initialization to the performance of the proposed algorithms. We tried the following three initializations for the optimal precoder design (OPT-TH-cov) algorithm in Table I: Initialization 1 is given in Table I. In Initialization 2, $\bar{\boldsymbol{\Omega}} = c_1\mathbf{D}$, where $c_1 = \sqrt{p_s/N_S}$ and $\mathbf{D}$ is a 4×4 diagonal matrix whose main diagonal elements are $[\sqrt{2}, 1, \sqrt{0.5}, \sqrt{0.5}]$. For Initialization 3, we set $\bar{\boldsymbol{\Omega}} = c_1\mathbf{U}$, where $\mathbf{U}$ is a 4×4 random Hermitian matrix. For the suboptimal precoder design (SUB-TH-cov) algorithm in Section III.D, the following two starting points are attempted: Initialization 1 as given after (78) and Initialization 2 where $a_1 = a_2 = \sqrt{2p_s/N_S}$ and $a_3 = a_4 = 0$. We observed that the proposed algorithms converge for the various initialization methods tested. Fig. 2 shows the BER performance of the two proposed algorithms using the initialization points tested. It can be seen from Fig. 2 that the system BER yielded by different starting points is quite small, and Initialization 1 has the lowest BER. Therefore, for the rest of simulations, we use Initialization 1 for both proposed algorithms.

In the following simulations, we compare the performance of two proposed algorithms with the linear transceiver-based precoding scheme such as the joint MMSE (JMMSE) scheme [7], the TH precoding based scheme with the full CSI (TH-FCSI) [28], TH-robust [32], TH-L-robust [31], and M-Schur-convex [29] schemes. Note that in contrast to other algorithms, the JMMSE and TH-FCSI schemes require the full CSI of the relay-destination link.

Fig. 3 shows the BER performance of all algorithms tested versus SNR_1 while fixing $\text{SNR}_2 = 20\text{dB}$. It can be seen from Fig. 3 that as expected, the TH-FCSI scheme has the lowest system BER. We can also observe that over the whole
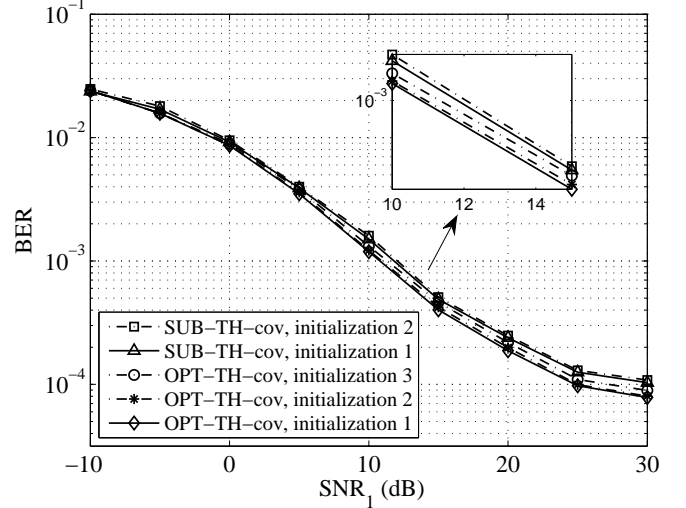

Fig. 2. BER versus SNR_1 at different number of initialization points while fixing $\text{SNR}_2 = 20\text{dB}$.
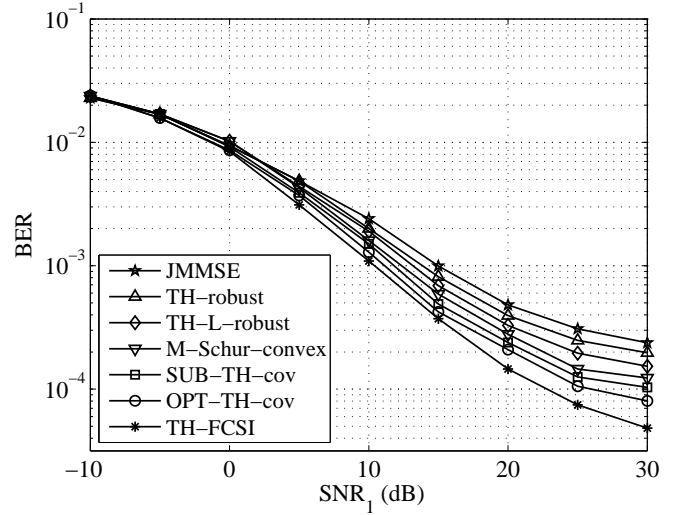

Fig. 3. BER versus SNR_1 while fixing $\text{SNR}_2 = 20\text{dB}$.

range of SNR_1, the two proposed algorithms outperform the JMMSE, TH-robust, TH-L-robust, and M-Schur-convex schemes in terms of BER. Moreover, for the whole range of SNR_1, the BER performance of the SUB-TH-cov algorithm is very close to that of the OPT-TH-cov algorithm.

In Fig. 4, we compare the performance of seven algorithms in terms of BER versus SNR_2 while fixing $\text{SNR}_1 = 20\text{dB}$. It can be noted from Fig. 4 that the proposed OPT-TH-cov and SUB-TH-cov algorithms show better BER performance over the whole range of SNR_2 than the existing schemes. Moreover, the system BER yielded by the proposed algorithms is very close to that of the system with the perfect CSI (TH-FCSI scheme).

Fig. 5 shows the BER performance comparison of the algorithms tested versus SNR_1 for $k = 3$ and $k = 10$ when SNR_2 is fixed at 20dB. It can be seen from Fig. 5 that for both value of k, the proposed OPT-TH-cov and SUB-TH-
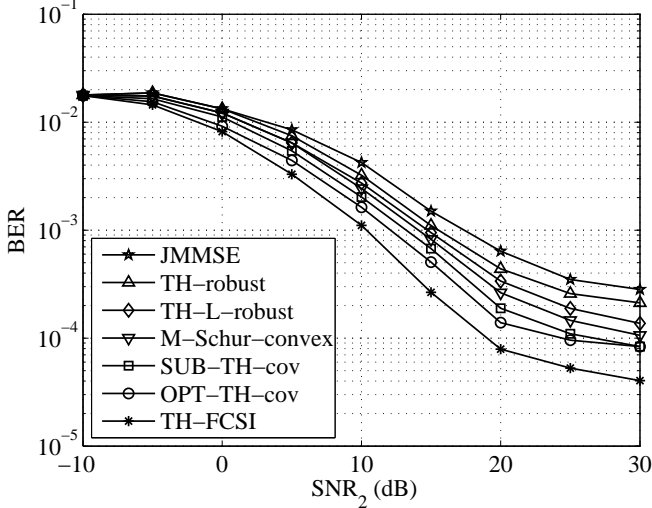
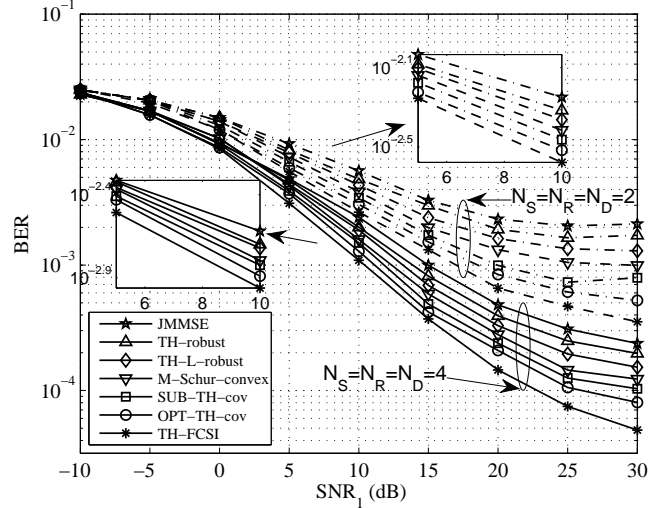Fig. 4. BER versus SNR_2 while fixing $\mathrm{SNR}_1 = 20\mathrm{dB}$.

Fig. 6. BER versus SNR_1 at different number of antennas while fixing $\mathrm{SNR}_2 = 20\mathrm{dB}$.

TABLE II
AVERAGE NUMBER OF ITERATIONS REQUIRED BY THE OPT-TH-COV
ALGORITHM TILL CONVERGENCE

$\mathrm{SNR}_1(\mathrm{dB})$	0	5	10	15	20	25	30
Number of Iterations	3	4	6	8	10	12	12

SUB-TH-cov algorithm, the complexity order of matrix inversion, matrix GMD, and matrix SVD is $\mathcal{O}(N^3)$. Since the complexity of solving the problem (73)-(75) is much lower than the matrix operations mentioned above, the complexity order of the SUB-TH-cov algorithm is $\mathcal{O}(N^3)$. In each iteration of the proposed OPT-TH-cov algorithm, the major operation is to update the relay matrix using the projected gradient method, which has a complexity order of $\mathcal{O}(I_1 N^3)$. Here I_1 is the number of projected gradient steps required to reach a stationary point. The overall computational complexity of the OPT-TH-cov algorithm also depends on the number of iterations required till convergence, which is shown in Table II. Obviously, the SUB-TH-cov algorithm has a much lower overall computational complexity than the OPT-TH-cov algorithm. Thus, the SUB-TH-cov algorithm is very attractive for practical MIMO relay communication systems.

V. CONCLUSIONS

We have addressed the challenging issue of precoding matrices optimization for a TH-based two-hop MIMO relay system where the full CSI of the source-relay link is known, while only the CCI of the relay-destination link is available at the relay node. We have derived the structure of the optimal TH precoding matrix and the source precoding matrix that minimize the MSE of the signal waveform estimation at the destination. An iterative algorithm has been developed to optimize the relay precoding matrix. We have also proposed a simplified precoding matrices design algorithm which has lower computational complexity than the iterative algorithm. Numerical results show that the proposed precoding matrices design schemes outperform existing algorithms.
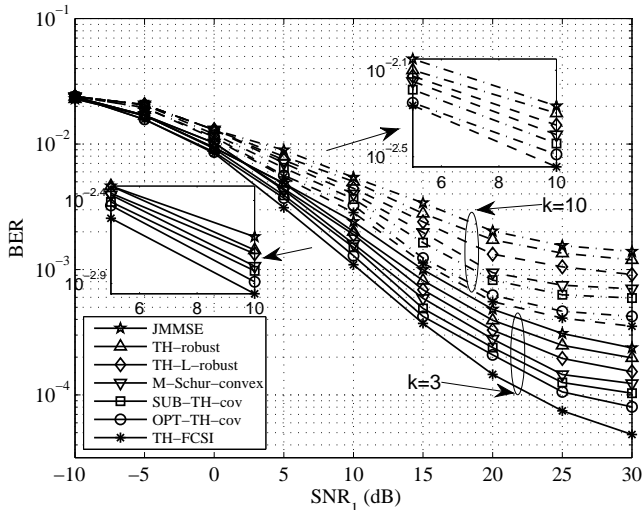

Fig. 5. BER versus SNR_1 at different correlation coefficient k while fixing $\mathrm{SNR}_2 = 20\mathrm{dB}$.

cov algorithms show better BER performance over the whole range of SNR_1 than the JMMSE, TH-robust, TH-L-robust, and M-Schur-convex schemes in terms of BER. The BER performance of the proposed OPT-TH-cov and SUB-TH-cov scheme is closer to that of the TH-FCSI scheme when k is large (i.e., the elements of $\mathbf{H}_2$ are highly correlated).

Fig. 6 shows the performance of all algorithms in terms of BER versus SNR_1 for $N = 2$ and $N = 4$, while fixing $\mathrm{SNR}_2 = 20\mathrm{dB}$. We observe that the proposed OPT-TH-cov and SUB-TH-cov algorithms outperform the JMMSE, TH-robust, TH-L-robust, and M-Schur-convex schemes over the whole range of SNR_1. It can also be seen from Fig. 6 that with increasing number of antennas at the source, relay, and destination nodes, the BER performance of all algorithms improves.

Finally, we compare the computational complexity of the proposed SUB-TH-cov and OPT-TH-cov algorithms. In the

APPENDIX A
PROOF OF THEOREM 1

The Lagrangian function associated with the problem (35) can be written as

$$\mathcal{L}_1 = -X + \mu(Q_2(\mathbf{G}) - p_r) \tag{82}$$

where μ is the Lagrangian multiplier. We denote $\nabla_G \mathcal{L}_1 = 2(\frac{\partial \mathcal{L}_1}{\partial \mathbf{G}})^*$ as the gradient of (82). The KKT conditions [40] of the problem (35) on $\mathbf{G}$ are given by

$$\nabla_G \mathcal{L}_1 = 0 \tag{83}$$

$$\mu(Q_2(\mathbf{G}) - p_r) = 0 \tag{84}$$

$$Q_2(\mathbf{G}) \leq p_r. \tag{85}$$

Using $\frac{\partial \ln|\mathbf{X}|}{\partial \mathbf{X}} = \mathbf{X}^{-T}$, $\frac{\partial tr(\mathbf{A}\mathbf{X}^{-1})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})^T$, and $\frac{\partial tr(\mathbf{X}\mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}^T$, we can write $\nabla_G \mathcal{L}_1$ as

$$\nabla_G \mathcal{L}_1 = -2\mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G} \mathbf{H}_1 \mathbf{A} \mathbf{H}_1^H$$
$$+ 2\sigma_1^2 \mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G} \mathbf{H}_1 \mathbf{A} \mathbf{H}_1^H \mathbf{G}^H \mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G}$$
$$+ 2\mu\sigma_x^2 \mathbf{G} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H. \tag{86}$$

The Lagrangian function of the problem (38)-(39) associated with $\mathbf{G}$ can be written as

$$\mathcal{L}_2 = tr\{\mathbf{\Omega}\mathbf{A}\} + \mu(Q_2(\mathbf{G}) - p_r). \tag{87}$$

The KKT conditions of the problem (38)-(39) on $\mathbf{G}$ are given by

$$\nabla_G \mathcal{L}_2 = 0 \tag{88}$$

$$\mu(Q_2(\mathbf{G}) - p_r) = 0 \tag{89}$$

$$Q_2(\mathbf{G}) \leq p_r. \tag{90}$$

The gradient $\nabla_G \mathcal{L}_2$ of the Lagrangian function (87) can be derived as

$$\nabla_G \mathcal{L}_2 = -2\mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G} \mathbf{H}_1 \mathbf{A} \mathbf{\Omega} \mathbf{A} \mathbf{H}_1^H$$
$$+ 2\sigma_1^2 \mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G} \mathbf{H}_1 \mathbf{A} \mathbf{\Omega} \mathbf{A} \mathbf{H}_1^H \mathbf{G}^H \mathbf{H}_2^H \mathbf{R}_n^{-1} \mathbf{H}_2 \mathbf{G}$$
$$+ 2\mu\sigma_x^2 \mathbf{G} \mathbf{H}_1 \mathbf{F} \mathbf{F}^H \mathbf{H}_1^H. \tag{91}$$

By comparing (83)-(86) with (88)-(91), we observe that the KKT conditions of the problem (38)-(39) are equal to those of the problem (35) when (40) holds. The derivative of (38) with respect to $\mathbf{\Omega}$ can be written as

$$\frac{\partial(tr\{\mathbf{\Omega}\mathbf{A}\} - \log|\mathbf{\Omega}|)}{\partial \mathbf{\Omega}} = \mathbf{A}^T - (\mathbf{\Omega}^{-1})^T \tag{92}$$

By equating (92) to zero, we obtain (40). Thus with given $\mathbf{G}$, the weight matrix $\mathbf{\Omega}$ minimizing (38) is given by (40).

APPENDIX B
PROOF OF THEOREM 2

Let us first introduce the following definition and lemma.

Definition 1: [42]. Let $\mathbf{\Phi}$ be a matrix-convex function. The Jensen's inequality for matrix valued functions is given by $E\{\mathbf{\Phi}(\mathbf{X})\} \geq \mathbf{\Phi}(E\{\mathbf{X}\})$, where $E\{\cdot\}$ is expectation on the random matrix $\mathbf{X}$.

Lemma 3: [43]. For positive definite Hermitian matrix $\mathbf{X}$, the matrix-valued function $\mathbf{\Phi}(\mathbf{X}) = \mathbf{X}^{-1}$ is matrix-convex. Therefore, from Definition 1, there is $E\{\mathbf{X}^{-1}\} \geq (E\{\mathbf{X}\})^{-1}$.

Now we start to prove Theorem 2. By using the matrix inversion lemma [37], (51) can be written as

$$E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\} = E_{\widetilde{\mathbf{H}}_\omega}\Big\{ tr\Big\{ \bar{\mathbf{\Omega}} \Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}\widetilde{\mathbf{G}}^H$$
$$\times \mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}} \Big(\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}} + \frac{\sigma_2^2}{\sigma_1^2}\big[\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\big]^{-1} \Big)^{-1}$$
$$\times \mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}\widetilde{\mathbf{G}}\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H \Big]^{-1} \Big\}\Big\}. \tag{93}$$

By applying Lemma 3 to (93), we have

$$E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\} \geq tr\Big\{ \bar{\mathbf{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}$$
$$\times \Big(\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}} + \frac{\sigma_2^2}{\sigma_1^2}E_{\widetilde{\mathbf{H}}_\omega}\big\{\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\big\}^{-1} \Big)^{-1}$$
$$\times \mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}\widetilde{\mathbf{G}}\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H \Big]^{-1} \Big\}$$
$$= tr\Big\{ \bar{\mathbf{\Omega}}\Big[\rho^{-2}\sigma_x^{-2}\mathbf{I}_{N_S} + \sigma_1^{-2}\mathbf{V}_1\mathbf{\Lambda}_1^{\frac{1}{2}}$$
$$\times (\mathbf{I}_{N_R} - \mathbf{D}_3)\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{V}_1^H \Big]^{-1} \Big\} \tag{94}$$

where

$$\mathbf{D}_3 = \Big(\mathbf{I}_{N_R} + \frac{\sigma_1^2}{\sigma_2^2}\widetilde{\mathbf{G}}^H\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}E_{\widetilde{\mathbf{H}}_\omega}\{\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\}\mathbf{\Lambda}_{\widetilde{\Sigma}}^{\frac{1}{2}}\widetilde{\mathbf{G}} \Big)^{-1}.$$

Using $E_{\widetilde{\mathbf{H}}_\omega}\{\widetilde{\mathbf{H}}_\omega^H\widetilde{\mathbf{H}}_\omega\} = N_D\mathbf{I}_{N_R}$, we obtain $E_{\widetilde{\mathbf{H}}_\omega}\{T_1(\widetilde{\mathbf{G}})\} \geq T_2(\widetilde{\mathbf{G}})$.

APPENDIX C
PROOF OF THEOREM 3

For the sake of notational simplicity, let us introduce $\mathbf{T}_1^{\frac{1}{2}} \triangleq \frac{\sigma_x}{\sigma_1}\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{\Lambda}_1^{\frac{1}{2}}$ and $\mathbf{T}_2^{\frac{1}{2}} \triangleq \frac{\sigma_1}{\sigma_2}\mathbf{\Lambda}_G^{\frac{1}{2}}\mathbf{\Lambda}_{\widetilde{\Sigma},1}^{\frac{1}{2}}$. Then (63) can be written as

$$f(\mathbf{Z}) = \log\Big|\mathbf{I}_{N_S} + \mathbf{T}_1^{\frac{1}{2}}\big(\mathbf{I}_{N_S} - (\mathbf{I}_{N_S} + \mathbf{T}_2^{\frac{1}{2}}\mathbf{Z}\mathbf{T}_2^{\frac{1}{2}})^{-1}\big)\mathbf{T}_1^{\frac{1}{2}}\Big|$$
$$= \log\Big|\mathbf{I}_{N_S} + \mathbf{T}_1^{\frac{1}{2}}\mathbf{T}_2^{\frac{1}{2}}(\mathbf{T}_2 + \mathbf{Z}^{-1})^{-1}\mathbf{T}_2^{\frac{1}{2}}\mathbf{T}_1^{\frac{1}{2}}\Big|$$
$$= \log|\mathbf{T}_1\mathbf{T}_2 + \mathbf{T}_2 + \mathbf{Z}^{-1}| - \log|\mathbf{T}_2 + \mathbf{Z}^{-1}|$$
$$= \log|\mathbf{I}_{N_S} + \mathbf{T}_3\mathbf{Z}| - \log|\mathbf{I}_{N_S} + \mathbf{T}_2\mathbf{Z}| \tag{95}$$

where $\mathbf{T}_3 \triangleq \mathbf{T}_1\mathbf{T}_2 + \mathbf{T}_2$. The concavity of (95) can be proven by considering an arbitrary line [44] given by $\mathbf{Z} = \mathbf{X} + t\mathbf{Y} \geq 0$. We have

$$g(t) = \log|\mathbf{I}_{N_S} + \mathbf{T}_3(\mathbf{X} + t\mathbf{Y})| - \log|\mathbf{I}_{N_S} + \mathbf{T}_2(\mathbf{X} + t\mathbf{Y})|$$
$$= \log\Big|\mathbf{I}_{N_S} + \mathbf{T}_3^{\frac{1}{2}}(\mathbf{X} + t\mathbf{Y})\mathbf{T}_3^{\frac{1}{2}}\Big|$$
$$- \log\Big|\mathbf{I}_{N_S} + \mathbf{T}_2^{\frac{1}{2}}(\mathbf{X} + t\mathbf{Y})\mathbf{T}_2^{\frac{1}{2}}\Big|$$
$$= \log|\mathbf{I}_{N_S} + t\mathbf{P}_3| - \log|\mathbf{I}_{N_S} + t\mathbf{P}_2| + \xi$$
$$= \sum_{n=1}^{N_S}(\log(1 + t\lambda_{3,n}) - \log(1 + t\lambda_{2,n})) + \xi \tag{96}$$

where $\xi \triangleq \log|\mathbf{I}_{N_S} + \mathbf{T}_3\mathbf{X}| - \log|\mathbf{I}_{N_S} + \mathbf{T}_2\mathbf{X}|$, $\lambda_{i,n}$, $i = 2, 3$, $n = 1, \cdots, N_S$, are the eigenvalues of $\mathbf{P}_i$, and for $i = 2, 3$,

$$\mathbf{P}_i = \big(\mathbf{I}_{N_S} + \mathbf{T}_i^{\frac{1}{2}}\mathbf{X}\mathbf{T}_i^{\frac{1}{2}}\big)^{-\frac{1}{2}}\mathbf{T}_i^{\frac{1}{2}}\mathbf{Y}\mathbf{T}_i^{\frac{1}{2}}\big(\mathbf{I}_{N_S} + \mathbf{T}_i^{\frac{1}{2}}\mathbf{X}\mathbf{T}_i^{\frac{1}{2}}\big)^{-\frac{1}{2}}.$$

The second-order derivative of (96) is

$$g''(t) = \sum_{n=1}^{N_S} \left(\frac{\lambda_{2,n}^2}{(1+t\lambda_{2,n})^2} - \frac{\lambda_{3,n}^2}{(1+t\lambda_{3,n})^2} \right)$$

$$= \sum_{n=1}^{N_S} \frac{(\lambda_{2,n} - \lambda_{3,n})(\lambda_{2,n} + \lambda_{3,n} + 2t\lambda_{2,n}\lambda_{2,n})}{(1+t\lambda_{2,n})^2(1+t\lambda_{3,n})^2}. \quad (97)$$

Let us introduce $\lambda(\mathbf{X})$ as the eigenvalue of $\mathbf{X}$. Using the property of $\lambda(\mathbf{AB}) = \lambda(\mathbf{BA})$, we have

$$\lambda(\mathbf{P}_i) = \lambda\left(\mathbf{T}_i^{\frac{1}{2}}\left(\mathbf{I}_{N_S} + \mathbf{T}_i^{\frac{1}{2}}\mathbf{X}\mathbf{T}_i^{\frac{1}{2}}\right)^{-1}\mathbf{T}_i^{\frac{1}{2}}\mathbf{Y}\right)$$
$$= \lambda\left(\mathbf{Y}^{\frac{1}{2}}\left(\mathbf{T}_i^{-1} + \mathbf{X}\right)^{-1}\mathbf{Y}^{\frac{1}{2}}\right), \quad i = 1, 2. \quad (98)$$

Since $\mathbf{T}_3 \geq \mathbf{T}_2$, we can see from (98) that $\lambda(\mathbf{P}_3) \geq \lambda(\mathbf{P}_2)$, i.e., $\lambda_{3,n} \geq \lambda_{2,n}$, $n = 1, \cdots, N_S$. As a result, from (97) we have $g''(t) \leq 0$. Therefore, we conclude that $f(\mathbf{Z})$ is concave.

APPENDIX D
DERIVATION OF (76) AND (77)

The Lagrangian function of (73) can be written as

$$L = -\sum_{i=1}^{N_S} \log \frac{(a_i\Lambda_{1,i} + \sigma_1^2)(N_D b_i\Lambda_{\Sigma,i} + \sigma_2^2)}{\sigma_x^2(\sigma_1^2 N_D\Lambda_{\Sigma,i}b_i + a_i\Lambda_{1,i}\sigma_2^2 + \sigma_2^2\sigma_1^2)}$$
$$+ \mu_s\left(\sum_{i=1}^{N_S} a_i - p_s\right) + \mu_r\left(\sum_{i=1}^{N_S} b_i - p_r\right) \quad (99)$$

where $\mu_s \geq 0$ and $\mu_r \geq 0$ are the Lagrangian multipliers. By using the KKT conditions, we have

$$\frac{\partial L}{\partial a_i} = -\frac{\sigma_1^2 N_D b_i\Lambda_{\Sigma,i}\Lambda_{1,i}}{(a_i\Lambda_{1,i} + \sigma_1^2)(\sigma_1^2 N_D b_i\Lambda_{\Sigma,i} + \sigma_2^2 a_i\Lambda_{1,i} + \sigma_2^2\sigma_1^2)}$$
$$+ \mu_s = 0 \quad (100)$$

$$\frac{\partial L}{\partial b_i} = -\frac{\sigma_2^2 N_D a_i\Lambda_{\Sigma,i}\Lambda_{1,i}}{(N_D b_i\Lambda_{\Sigma,i} + \sigma_2^2)(\sigma_1^2 N_D b_i\Lambda_{\Sigma,i} + \sigma_2^2 a_i\Lambda_{1,i} + \sigma_2^2\sigma_1^2)}$$
$$+ \mu_r = 0 \quad (101)$$

$$\mu_s\left(\sum_{i=1}^{N_S} a_i - p_s\right) = 0, \quad \mu_s \geq 0, \quad a_i \geq 0, i = 1, \cdots, N_S$$

$$\mu_r\left(\sum_{i=1}^{N_S} b_i - p_r\right) = 0, \quad \mu_r \geq 0, \quad b_i \geq 0, i = 1, \cdots, N_S.$$

Using (100) and (101) and after some manipulations, we obtain the optimum a_i and b_i given by (76) and (77).

ACKNOWLEDGMENT

REFERENCES

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.
[2] P. A. Angel and M. Kaveh, "On the performance of distributed space-time coding system with one and two non-regenerative relays," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 682–692, Mar. 2006.
[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062 – 3080, Dec. 2004.
[4] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 59, pp. 2381–2401, Sep. 2003.
[5] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in non-regenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, pp. 2593 – 2604, June 2007.
[6] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398–1407, Apr. 2007.
[7] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517–519, July 2008.
[8] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pp. 5062–5073, Oct. 2008.
[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, pp. 4837–4851, Dec. 2009.
[10] Y. Rong and F. Gao, "Optimal beamforming for non-regenerative MIMO relays with direct link," *IEEE Commun. Lett.*, vol. 13, pp. 926–928, Dec. 2009.
[11] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Lett.*, vol. 14, pp. 390–392, May 2010.
[12] F.-S. Tseng and W.-R. Wu, "Linear MMSE transceiver design in amplifyand-forward MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 754–765, Feb. 2010.
[13] H. W. Je, B. Lee, S. Kim, and K. B. Lee, "Design of non-regenerative MIMO-relay systems with partial channel state information," in *Proc. IEEE Int. Conf. Commun*, Beijing, 2008, pp. 4441-4445.
[14] C. Jeong and H.-M. Kim, "Precoder design of non-regenerative relays with covariance feedback," *IEEE Commun. Lett.*, vol. 13, pp. 920 – 922, Dec. 2009.
[15] M. B. Shenouda and T. N. Davidson, "Nonlinear and linear broadcasting with QoS requirements: Tractable approaches for bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, pp. 1936–1947, May 2009.
[16] L. Gopal, Y. Rong, and Z. Zang, "Joint MMSE transceiver design in non-regenerative MIMO relay systems with covariance feedback," in *Proc. IEEE 17th Asia-Pacific Conf. Commun*, Sabah, Malaysia, Oct. 2011.
[17] C. Jeong, B. Seo, S. R. Lee, H.-M. Kim, and I.-M. Kim, "Relay precoding for non-regenerative MIMO relay systems with partial CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 1698–1711, May 2012.
[18] K. Dae-Hyun and H.-M. Kim, "MMSE precoder design for a non-regenerative MIMO relay with covariance feedback," in *Proc. IEEE 21st Int. Sym. Per. Ind. Mobile Radio Commun*, Istanbul, Turkey, Aug. 2010, pp. 461-464.
[19] L. Gopal, Y. Rong, and Z. Zang, "Channel covariance information based transceiver design for AF MIMO relay systems with direct link," in *Proc. IEEE 18th Asia-Pacific Conf. Commun*, Korea, Oct. 2012.
[20] L. Gopal, Y. Rong, and Z. Zang, "MMSE based transceiver design for MIMO relay systems with mean and covariance feedback," in *Proc. 77th IEEE Veh. Technol. Conf*, Dresden, Germany, June, 2013.
[21] Y. Rong and M. R. A. Khandaker, "On uplink-downlink duality of multi-hop MIMO relay channel," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 1923–1931, June 2011.
[22] L. Sanguinetti, A. A. D. Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, pp. 1331–1346, Sep. 2012.
[23] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. H. Huber, "Space time transmission using Tomlinson-Harashima precoding," in *Proc. 4th ITC Conf. Source Channel Coding*, Berlin, Germany, 2002.
[24] A. A. D'Amico and M. Morelli, "Joint Tx-Rx MMSE design for MIMO multicarrier systems with Tomlinson-Harashima precoding," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3118–3127, Aug. 2008.
[25] L. Sanguinetti and M. Morelli, "Non-linear pre-coding for multiple-antenna multi-user downlink transmissions with different QoS requirements," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 852–856, Mar. 2007.
[26] A. A. D'Amico, "Tomlinson-Harashima precoding in MIMO systems: A unified approach to transceiver optimization based on multiplicative Schur-convexity," *IEEE Trans. Signal Process.*, vol. 56, pp. 3662–3677, Aug. 2008.

[27] L. Sun and M. Lei, "Quantized CSI-based Tomlinson-Harashima precoding in multiuser MIMO systems," *IEEE Trans. Wireless Commun., vol. 12, pp. 1118–1126, Mar. 2013.*

[28] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Joint Tomlinson-Harashima source and linear relay precoder design in amplify-and-forward MIMO relay systems via MMSE criterion," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 1687–1698, May 2011.

[29] C. Xing, M. Xia, F. Gao, and Y.-C. Wu, "Robust transceiver with Tomlinson-Harashima precoding for amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, pp. 1370–1382, Sep. 2012.

[30] A. P. Millar, S. Weiss, and R. W. Stewart, "THP transceiver design for MIMO relaying with direct link and partial CSI," *IEEE Commun. Lett.*, vol. 17, pp. 1204–1207, June 2013.

[31] F.-S. Tseng, M.-Y. Chang, and W.-R. Wu, "Robust Tomlinson-Harashima source and linear relay precoders design in amplify-and-forward MIMO relay systems," *IEEE Trans. Commun.*, vol. 60, pp. 1124–1137, Apr. 2012.

[32] A. P. Millar, S. Weiss, and R. W. Stewart, "Robust transceiver design for MIMO relay systems with Tomlinson-Harashima precoding," in *Proc. EUSIPCO 2012*, Aug. 2012, pp. 1374–1378.

[33] F. A. Dietrich, P. Breun, and W. Utschik, "Robust Tomlinson-Harashima precoding for the wireless broadcast channel," *IEEE Trans. Signal Process.*, vol. 55, pp. 631–644, Feb. 2007.

[34] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.

[35] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4791–4790, Dec. 2008.

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cilffs, NJ: Prentice Hall, 1993.

[37] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2011.

[38] M. B. Shenouda and T. N. Davidson, "A framework for designing MIMO systems with decision feedback equalization or Tomlinson-Harashima precoding," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 401–411, Feb. 2008.

[39] Y. Jiang, W. Hager, and J. Li, "The geomatric mean decomposition," *Linear Algebra Appl.*, vol. 396, pp. 373–384, Feb. 2005.

[40] E. K. P. Chong and S. H. Zak, *Introduction to Optimization*. New York, NY: Wiley-Interscience, 2001.

[41] D. P. Bertsekas, *Nonlinear Programming*. 2nd. Edition, Athena Scientific, Belmont, 1995.

[42] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York, Springer, 2011.

[43] I. Olkin and J. Pratt, "A multivariate Tchebycheff inequality," *Ann. Math. Statist.*, vol. 29, pp. 226–234, Mar. 1958.

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

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

PLACE
PHOTO
HERE

Yue Rong (S'03-M'06-SM'11) received the Ph.D. degree (summa cum laude) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently an Associate Professor. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, applications of linear algebra and optimization methods, and statistical and array signal processing. He has published over 120 journal and conference paper in these areas.

Dr. Rong was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He is an Associate Editor of theIEEE TRANSACTIONS ON SIGNAL PROCESSING. He was an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS from 2012 to 2014, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays, and was a TPC Member for the IEEE ICC, WCSP, IWCMC, and ChinaCom.

PLACE
PHOTO
HERE

Lenin Gopal (M'06-M'15) received the B.Eng., degree in Electrical and Electronics Engineering from Madurai Kamaraj University, Madurai, India, in 1996, the M.Eng. degree in Telecommunications Engineering from Multimedia University, Cyberjeya, Malaysia, in 2006, and he is currently working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, Curtin University, Australia. He joined as a principal Lecturer in the Department of Electrical and Electronics Engineering, TAFE College, Seremban, Malaysia from July 2000- 2007. In August 2007, he joined in the Department of Electronic and Communication Engineering, Bharathiyar College of Engineering and Technology, Karaikal, India as an Assistant Professor. He is currently working as a Senior Lecture with the Department of Electrical and Computer Engineering, Faculty of Engineering and Science at Curtin University, Sarawak, Malaysia. His research interests include signal processing for communications, wireless communications, FPGA Applications for Wireless Communications.