A Worst-Case Robust MMSE Transceiver Design for Nonregenerative MIMO Relaying

Hong Shen, *Student Member, IEEE,* Jiaheng Wang, *Member, IEEE,* Wei Xu, *Member, IEEE,* Yue Rong, *Senior Member, IEEE,* and Chunming Zhao, *Member, IEEE*

***Abstract*— Transceiver designs have been a key issue in guaranteeing the performance of multiple-input multiple-output (MIMO) relay systems, which are, however, often subject to imperfect channel state information (CSI). In this paper, we aim to design a robust MIMO transceiver for nonregenerative MIMO relay systems against imperfect CSI from a worst-case robust perspective. Specifically, we formulate the robust transceiver design, under the minimum mean-squared error (MMSE) criterion, as a minimax problem. Then, by decomposing the minimax problem into two subproblems with respect to the relay precoder and destination equalizer, respectively, we show that the optimal solution to each subproblem has a favorable channel-diagonalizing structure under some mild conditions. Based on this finding, we transform the two complex-matrix subproblems into their equivalent scalar forms, both of which are proven to be convex and can be efficiently solved by our proposed methods. We further propose an alternating algorithm to jointly optimize the precoder and equalizer that only requires scalar operations. Finally, the effectiveness of the proposed robust design is verified by simulation results.**

***Index Terms*— Multiple-input multiple-output (MIMO) relay systems, minimum mean-squared error (MMSE), worst-case transceiver design, imperfect channel state information (CSI).**

I. INTRODUCTION

In recent years, relaying techniques have gained a remarkable interest from both academic and industrial fields, due to the fact that relay-assisted systems can provide more reliable link quality and also expand the coverage compared to the direct transmission scheme [1], [2]. Such benefits can be greatly enhanced by incorporating multiple-input multiple-output (MIMO) techniques, leading to so-called MIMO relaying [3], [4]. In general, there are mainly two classes of relaying strategies: regenerative and nonregenerative ones, where the regenerative strategy decodes and re-encodes relayed signals, whereas the nonregenerative relaying simply performs linear processing on the received signals and retransmits them to the destination. Compared with the regenerative strategy, the nonregenerative relaying receives more attention because of its simplicity and easy implementation.

The full potential of MIMO relaying depends on proper transceiver designs exploiting available channel state information (CSI), which has been a focus of a number of recent works [5]–[11]. In particular, the optimal relay precoder that maximizes the mutual information (MI) has been investigated independently in [5] and [6], and the mean-squared error (MSE) based transceiver design has been well studied in [7] and [8]. Apart from the MI and MSE criteria, relay designs aiming to maximize the signal-to-noise-ratio (SNR) were also discussed in [8] and [9]. Recently a unified framework accommodating many commonly used design objectives was developed in [10] for nonregenerative MIMO relay systems. Revealed in most of the above works is an interesting phenomenon that the eigenmode transmission is often optimal, just as in traditional point-to-point MIMO systems, and the channel is diagonalized by the transceiver. With this important conclusion, the original matrix-variable transceiver design can be simplified into a much simpler power allocation problem.

A common assumption made in the aforementioned works is that CSI is perfectly known by the transceiver. In practical systems, due to many factors such as quantized feedback, feedback delay and channel estimation errors, one can only access imperfect CSI in general, which often results in considerable performance degradation. In order to alleviate the impact of imperfect CSI, it is necessary to design robust transceivers taking the imperfection of CSI into account. In the literature, there are generally two classes of imperfect CSI models widely used for robust designs: the stochastic and deterministic models. The stochastic model [12]–[21] assumes that the instantaneous values of CSI are unknown but its statistics, such as mean or/and covariance, are known, where the robust design usually uses the average or outage performance as the design objective. In contrast, the deterministic model [22]–[32] makes no presumption on the distribution of CSI uncertainties and assumes that the instantaneous value of CSI, although unknown, lies in a known set of possible values, e.g., a norm ball. In this case, the design goal is often to optimize the worst-case performance and achieve the so-called worst-case robustness.

Under the stochastic imperfect CSI model, the robust transceiver design for point-to-point MIMO systems has been

This work was supported by the 973 Program under 2013CB329204, 2013CB336600, National Natural Science Foundation of China under 61201174, 61101087, Natural Science Foundation of Jiangsu, China under BK2012325, the Fundamental Research Funds for the Central Universities, the Scientific Research Foundation of Graduate School of Southeast University under YBJJ1327. The work of Y. Rong was supported in part by the Australian Research Council's Discovery Projects funding scheme (project number DP110100736). The authors acknowledge research support from the Jiangsu Talent Introduction Project and Jiangsu Creativity Promotion Program.

H. Shen, J. Wang, W. Xu and C. Zhao are with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (e-mail: shenhongseu@gmail.com, jhwangee@gmail.com, wxu@seu.edu.cn, cmzhao@seu.edu.cn). H. Shen was also a visiting PhD student with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA.

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University of Technology, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

well established in [12]–[14]. The authors in [15]–[20] further developed a robust transceiver for nonregenerative MIMO relay systems with statistical CSI errors. An interesting finding made by most of these studies is that the optimal statistically robust transceiver diagonalizes the channel, as it does under the perfect CSI case. On the other hand, the worst-case robust design based on the deterministic model has been considered for a direct MIMO transmission in [22]–[24], where the authors proved that eigenmode transmission is still optimal for some specific design objectives, e.g., worst-case SNR [22] or MSE with fixed receivers or equalizers [23], [24]. In light of these existing findings, it is natural to consider whether this favorable property still holds in nonregenerative MIMO relay systems with deterministic imperfect CSI. So far as we know, this question has not been answered in the literature.

In this paper, we investigate a robust transceiver design for a two-hop nonregenerative MIMO relay system in the presence of deterministic imperfect CSI. Adopting worst-case robustness, we formulate the robust transceiver design as a minimax problem in order to achieve the minimum worst-case MSE. Our main contributions are summarized as follows:

- We first decompose the minimax problem into two subproblems with respect to the relay precoder and destination equalizer, and show that the optimal robust precoder and equalizer have a channel-diagonalizing structure under some mild conditions. This conclusion is a non-trivial generalization of the result with regard to the robust transceiver for point-to-point MIMO systems in [24]. The introduction of relay node leads to a more complicated objective function and power constraint, and thus makes it difficult to characterize the inherent structure of the precoder and equalizer.
- Built on the channel-diagonalizing structure, we further show that the robust relay precoder and destination equalizer optimization can be equivalently transformed into two convex scalar-valued problems, respectively. We then propose efficient numerical methods to find the optimal solution to both problems.
- In light of the favorable structure of robust transceiver and the related power allocation problems, we devise an alternating algorithm to jointly optimize the robust transceiver which simply requires scalar operations and converges with only a few iterations. Simulation results show that our proposed robust transceiver design outperforms the non-robust schemes by considerable gains in terms of both MSE and bit-error-rate (BER).

We would like to point out that, different from a recent work [31] where a similar topic was studied, we find the channel-diagonalizing structure of robust transceiver which provides an important insight into the robust design. In addition, based on this structural knowledge, we also propose an efficient alternating algorithm that only relies on scalar operations.

The paper is organized as follows. A system model for nonregenrative MIMO relaying is introduced in Section II. In Section III, we study the optimal structure of worst-case robust transceiver for nonregenerative MIMO relaying. We then investigate the transceiver design based on scalar optimization in Section IV. Simulation results are presented in Section V and some concluding remarks are given in Section VI.

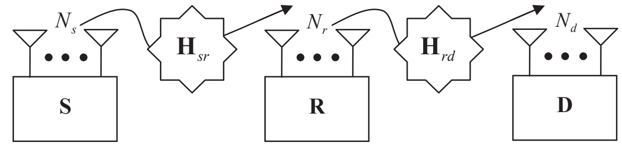


Fig. 1. A nonregenerative MIMO relay system.

Notation: We denote matrices and vectors by uppercase and lowercase boldface letters, respectively. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$ represent the transpose, conjugate transpose, inverse and pseudo inverse of matrix $\mathbf{A}$, respectively. We use vec$(\mathbf{A})$, rank$(\mathbf{A})$, tr$(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})$ to denote the vectorization, rank, trace and range space of $\mathbf{A}$, respectively. $\mathbf{A} \succeq 0$ or $\mathbf{A} \succ 0$ means the matrix $\mathbf{A}$ is positive semidefinite or definite. $(\mathbf{A})_{i,j}$ denotes the (ith, jth) element of $\mathbf{A}$. The spectral norm and Frobenius norm of $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$, respectively. The Euclidean norm of vector $\mathbf{a}$ is represented by $\|\mathbf{a}\|$. diag$\{\mathbf{a}\}$ denotes a diagonal matrix whose diagonals are the elements of $\mathbf{a}$ and blkdiag$\{\mathbf{A}, \mathbf{B}\}$ stands for a block diagonal matrix with diagonals being $\mathbf{A}$ and $\mathbf{B}$. $\mathbb{R}^{n\times m}$ and $\mathbb{C}^{n\times m}$ denote the ensemble of all $n \times m$ real and complex matrices, respectively. $\mathbf{1}$ denotes the all-ones vector. Notations $\otimes$, $\Re(\cdot)$ and $\mathbb{E}\{\cdot\}$ represent the Kronecker product, the real part of a complex number and the expectation operation, respectively.

II. System Model and Problem Formulation

We consider a dual-hop nonregenerative MIMO relay system shown in Fig. 1, where the source node has N_s antennas, the relay node is equipped with N_r antennas, and the destination node has N_d antennas. The direct link between the source and destination nodes is assumed to be sufficiently weak so that it can be ignored. At the first hop, the symbol vector $\mathbf{s} \in \mathbb{C}^{N_s}$ with $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$ is transmitted to the relay node. The received signal $\mathbf{y}_r \in \mathbb{C}^{N_r}$ at the relay can be expressed by

$$\mathbf{y}_r = \sqrt{\frac{P_s}{N_s}}\tilde{\mathbf{H}}_{sr}\mathbf{s} + \mathbf{n}_r = \mathbf{H}_{sr}\mathbf{s} + \mathbf{n}_r \tag{1}$$

where P_s is the source transmit power, $\tilde{\mathbf{H}}_{sr} \in \mathbb{C}^{N_r\times N_s}$ represents the source-relay channel, $\mathbf{H}_{sr} = \sqrt{\frac{P_s}{N_s}}\tilde{\mathbf{H}}_{sr}$ is the equivalent source-relay channel and $\mathbf{n}_r \in \mathbb{C}^{N_r}$ is the additive white Gaussian noise (AWGN) vector at the relay node with zero mean and covariance matrix $\mathbf{R}_{n_r} = \sigma_r^2\mathbf{I}$. At the second hop, the relay multiplies the received signal $\mathbf{y}_r$ by a precoding matrix $\mathbf{F}_r \in \mathbb{C}^{N_r\times N_r}$. The power constraint imposed on $\mathbf{F}_r$ is

$$\text{tr}(\mathbf{F}_r(\mathbf{H}_{sr}\mathbf{H}_{sr}^H + \sigma_r^2\mathbf{I})\mathbf{F}_r^H) \le P_r \tag{2}$$

where P_r is the total transmit power at the relay. After forwarding the signal $\mathbf{x}_r = \mathbf{F}_r\mathbf{y}_r$ from the relay to the destination, the received signal $\mathbf{y}_d \in \mathbb{C}^{N_d}$ at the destination is

$$\mathbf{y}_d = \mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr}\mathbf{s} + \mathbf{H}_{rd}\mathbf{F}_r\mathbf{n}_r + \mathbf{n}_d \tag{3}$$

where $\mathbf{H}_{rd} \in \mathbb{C}^{N_d \times N_r}$ denotes the relay-destination channel and $\mathbf{n}_d \in \mathbb{C}^{N_d}$ is the AWGN vector at the destination node with zero mean and covariance matrix $\mathbf{R}_{n_d} = \sigma_d^2 \mathbf{I}$. At the destination node, a linear equalizer $\mathbf{G} \in \mathbb{C}^{N_s \times N_d}$ is used to estimate the transmit signal $\mathbf{s}$ from $\mathbf{y}_d$, i.e. $\hat{\mathbf{s}} = \mathbf{G}\mathbf{y}_d$. To design proper relay precoder $\mathbf{F}_r$ and destination equalizer $\mathbf{G}$, we adopt MSE between $\mathbf{s}$ and $\hat{\mathbf{s}}$ as the performance metric, which is given by

$$\begin{aligned}
&\text{MSE} = \mathbb{E}\left\{\|\hat{\mathbf{s}} - \mathbf{s}\|^2\right\} \\
&= \mathbb{E}_{\mathbf{s},\mathbf{n}_r,\mathbf{n}_d}\left\{\|(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr} - \mathbf{I})\mathbf{s} + \mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\mathbf{n}_r + \mathbf{G}\mathbf{n}_d\|^2\right\} \\
&= \|\mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr} - \mathbf{I}\|_F^2 + \sigma_{n_r}^2 \|\mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\|_F^2 + \sigma_{n_d}^2 \|\mathbf{G}\|_F^2 .
\end{aligned} \tag{4}$$

In general, it is reasonable to expect that perfect CSI at the receiver (CSIR) is available with the aid of training signals, whereas accurate CSI at the transmitter (CSIT) is difficult to obtain due to, for instance, the existence of quantization and feedback errors. Therefore, for the nonregenerative relaying, we assume that the relay node knows the perfect source-relay channel $\mathbf{H}_{sr}$ but has only imperfect information about the relay-destination channel $\mathbf{H}_{rd}$. Note that this assumption has also been widely adopted in a number of recent works such as [16]–[18], [25], [30], [31], [33]. To characterize the uncertainty of the relay-destination channel, we adopt a commonly used deterministic imperfect CSI model [22]–[32] assuming that the exact channel lies in the neighborhood of the estimated or feedback channel. In accordance with this model, the actual $\mathbf{H}_{rd}$ can be expressed by

$$\mathbf{H}_{rd} = \hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd}, \;\; \mathbf{\Delta}_{rd} \in \xi_{rd} \tag{5}$$

where $\hat{\mathbf{H}}_{rd}$ represents the mismatched channel, $\mathbf{\Delta}_{rd}$ represents the channel error and $\xi_{rd} \triangleq \{\mathbf{\Delta}_{rd} : \|\mathbf{\Delta}_{rd}\|_F \le \epsilon_{rd}\}$ denotes a spherical channel uncertainty region.

To provide robustness against the deterministic channel uncertainty characterized by (5), one needs to guarantee the MSE performance for all channel realizations within the uncertainty region, which can be achieved by optimizing the worst-case MSE. Thereby, the robust design problem is formulated as

$$\begin{aligned}
\min_{\mathbf{F}_r,\mathbf{G}} \max_{\mathbf{\Delta}_{rd}} \quad & \|\mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr} - \mathbf{I}\|_F^2 \\
& + \sigma_{n_r}^2 \|\mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd})\mathbf{F}_r\|_F^2 + \sigma_{n_d}^2 \|\mathbf{G}\|_F^2 \\
\text{subject to} \quad & \text{tr}(\mathbf{F}_r(\mathbf{H}_{sr}\mathbf{H}_{sr}^H + \sigma_{n_r}^2 \mathbf{I})\mathbf{F}_r^H) \le P_r \\
& \|\mathbf{\Delta}_{rd}\|_F \le \epsilon_{rd}.
\end{aligned} \tag{6}$$

The difficulty in solving such a problem is twofold: 1) the minimax problem contains actually two problems, i.e., the inner maximization and the outer minimization; 2) the problem is inherently nonconvex in either $\mathbf{F}_r$ and $\mathbf{G}$ or $\mathbf{\Delta}_{rd}$. Furthermore, our robust design is much more difficult than that in [24] which considered only a point-to-point MIMO system. As will be shown later, the more complicated objective function and power constraint present new challenges in solving the problem.

III. Optimal Structure of Robust MMSE Transceiver

From this section, we will investigate a tractable method to handle the non-convex problem (6). As the first step, we shall study the optimal structure of relay precoder $\mathbf{F}_r$ and destination equalizer $\mathbf{G}$ in this section. This structural knowledge will be utilized to simplify the original matrix-variable problem into simpler scalar-valued problems in the next section.

To achieve this goal, we first introduce a slack variable t and separate the problem (6) into two subproblems, one for the precoder $\mathbf{F}_r$ and one for the equalizer $\mathbf{G}$, respectively. Let us first consider the subproblem optimizing $\mathbf{F}_r$ with $\mathbf{G}$ fixed

$$\begin{aligned}
\underset{\mathbf{F}_r, t}{\text{minimize}} \quad & t \\
\text{subject to} \quad & \|\mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr} - \mathbf{I}\|_F^2 \\
& + \sigma_{n_r}^2 \|\mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd})\mathbf{F}_r\|_F^2 \le t \\
& \forall \mathbf{\Delta}_{rd} : \|\mathbf{\Delta}_{rd}\|_F \le \epsilon_{rd} \\
& \text{tr}(\mathbf{F}_r(\mathbf{H}_{sr}\mathbf{H}_{sr}^H + \sigma_{n_r}^2 \mathbf{I})\mathbf{F}_r^H) \le P_r.
\end{aligned} \tag{7}$$

As can be observed, the problem (7) is still difficult to solve due to two facts: 1) the robust constraint in (7) contains in fact an infinite number of constraints; 2) the variables are complex matrices. To overcome these difficulties, in the following we show that the optimal precoder $\mathbf{F}_r$ admits a favorable channel-diagonalizing structure under some mild conditions, which paves the path to simplifying the intractable complex-matrix problem (7). Before stating our result, we would like to introduce some notations that will be used later.

Denote the singular value decompositions (SVDs) of matrices $\mathbf{H}_{sr}$ and $\hat{\mathbf{H}}_{rd}$ with $\mathbf{H}_{sr} = \mathbf{U}_{h_{sr}}\mathbf{\Sigma}_{h_{sr}}\mathbf{V}_{h_{sr}}^H$ and $\hat{\mathbf{H}}_{rd} = \mathbf{U}_{\hat{h}_{rd}}\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H$, respectively. Let $\mathbf{\Lambda}_{h_{sr}} \in \mathbb{R}^{N_s \times N_s}$ and $\mathbf{\Lambda}_{\hat{h}_{rd}} \in \mathbb{R}^{N_s \times N_s}$ be diagonal matrices whose diagonals are the largest N_s singular values of $\mathbf{H}_{sr}$ and $\hat{\mathbf{H}}_{rd}$, respectively, i.e. $\gamma_{sr,1} \ge \cdots \ge \gamma_{sr,N_s}$ and $\gamma_{rd,1} \ge \cdots \ge \gamma_{rd,N_s}$. Denote the SVDs of $\mathbf{F}_r$ and $\mathbf{G}$ with $\mathbf{F}_r = \mathbf{U}_{f_r}\mathbf{\Sigma}_{f_r}\mathbf{V}_{f_r}^H$ and $\mathbf{G} = \mathbf{U}_g\mathbf{\Sigma}_g\mathbf{V}_g^H$, respectively, where the matrices $\mathbf{\Sigma}_{f_r}$ and $\mathbf{\Sigma}_g$ can be written as

$$\mathbf{\Sigma}_{f_r} = \begin{bmatrix} \mathbf{\Lambda}_{f_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ and } \mathbf{\Sigma}_g = [\mathbf{\Lambda}_g \quad \mathbf{0}] \tag{8}$$

with $\mathbf{\Lambda}_{f_r} = \text{diag}\{[f_{r,1}, \cdots, f_{r,N_s}]^T\}$ and $\mathbf{\Lambda}_g = \text{diag}\{[g_1, \cdots, g_{N_s}]^T\}$ being real diagonal matrices.[1] The following theorem reveals the optimal structure of $\mathbf{F}_r$.

Theorem 1: Let $\mathbf{G}$ be fixed with $\mathbf{U}_g = \mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g = \mathbf{U}_{\hat{h}_{rd}}$. Then, $\mathbf{U}_{f_r} = \mathbf{V}_{\hat{h}_{rd}}$ and $\mathbf{V}_{f_r} = \mathbf{U}_{h_{sr}}$ are optimal for the subproblem (7).

Proof: See Appendix I. ■

Now we have obtained the structure of optimal $\mathbf{F}_r$ to the subproblem (7) for fixed $\mathbf{G}$. In the next, we also consider the subproblem optimizing the equalizer $\mathbf{G}$ with $\mathbf{F}_r$ fixed, which

[1] Without loss of generality, we assume $N_r \ge N_s$, $N_d \ge N_s$, rank$(\mathbf{F}_r) \le N_s$ and rank$(\mathbf{G}) \le N_s$ in our paper.

is given by

$$\begin{aligned}\underset{\mathbf{G},t}{\text{minimize}} \quad & t+\sigma_{n_d}^2\|\mathbf{G}\|_F^2\\ \text{subject to} \quad & \|\mathbf{G}(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr}-\mathbf{I}\|_F^2\\ & +\sigma_{n_r}^2\|\mathbf{G}(\hat{\mathbf{H}}_{rd}+\boldsymbol{\Delta}_{rd})\mathbf{F}_r\|_F^2\le t\\ & \forall\boldsymbol{\Delta}_{rd}:\|\boldsymbol{\Delta}_{rd}\|_F\le\epsilon_{rd}. \end{aligned}\tag{9}$$

Interestingly, the similar channel-diagonalizing structure also holds for the optimal equalizer $\mathbf{G}$, as stated in the following theorem.

Theorem 2: Let $\mathbf{F}_r$ be fixed with $\mathbf{U}_{f_r}=\mathbf{V}_{\hat{h}_{rd}}$ and $\mathbf{V}_{f_r}=\mathbf{U}_{h_{sr}}$. Then, $\mathbf{U}_g=\mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g=\mathbf{U}_{\hat{h}_{rd}}$ are optimal for the subproblem (9).

Proof: The above theorem can be proved using the similar reasoning as *Theorem* 1. Please refer to Appendix II for the detailed proof. ■

Remark 1: Theorem 1 implies that the optimal relay precoder $\mathbf{F}_r$ diagonalizes both source-relay and relay-destination channels if $\mathbf{G}$ is fixed with a specified structure. On the other hand, *Theorem* 2 shows that the optimal destination equalizer $\mathbf{G}$ also diagonalizes the relay-destination channel when $\mathbf{F}_r$ satisfies certain requirements. Note that both results in *Theorems* 1 and 2 are non-trivial generalizations of the conclusions for point-to-point MIMO systems in [24], as our work considers both the relay-destination and source-relay channels, which leads to a much more complicated objective function and power constraint, compared with the one-hop channel model in [24]. With the structural knowledge of $\mathbf{F}_r$ and $\mathbf{G}$, we can convert the subproblems (7) and (9) to equivalent power allocation problems, as will be evidenced in the next section. Moreover, we stress that the conditions in the two theorems are complementary in the sense that the conclusion in *Theorem* 1 is exactly the condition in *Theorem* 2 and vice versa. As will be shown later, this property is quite desirable to devise an alternating algorithm via scalar computation only, thus simplifying the original problem (6) to a great extent.

IV. Robust Transceiver Design Exploiting Scalar Optimization

The channel-diagonalizing structure of $\mathbf{F}_r$ and $\mathbf{G}$, revealed by *Theorems* 1 and 2, provides an important insight into the robust transceiver design. With this conclusion, the subproblems (7) and (9) can be simplified to power allocation problems that are much easier to solve.[2] We denote the ith diagonal of $\boldsymbol{\Lambda}_g$, $\boldsymbol{\Lambda}_{f_r}$, $\boldsymbol{\Lambda}_{\hat{h}_{sr}}$ and $\boldsymbol{\Lambda}_{\hat{h}_{rd}}$ with g_i, $f_{r,i}$, $\gamma_{sr,i}$ and $\gamma_{rd,i}$, respectively. The following theorem provides the equivalent scalar forms of the problems (7) and (9), respectively.

Theorem 3: Given that $\mathbf{G}$ is fixed with $\mathbf{U}_g=\mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g=\mathbf{U}_{\hat{h}_{rd}}$, the problem (7) is equivalent to the following convex form:

$$\begin{aligned}\underset{\substack{\mu\ge0\\ f_{r,i},1\le i\le N_s}}{\text{minimize}} \quad & \sum_{i=1}^{N_s}\frac{\mu((g_i'\gamma_{rd,i}f_{r,i}-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-(g_i')^2f_{r,i}^2-\sigma_r^2g_i^2f_{r,i}^2}\\ & +\mu\epsilon_{rd}^2\\ \text{subject to} \quad & f_{r,i}^2\le\mu/\tilde{g}_{m,i}^2,\ 1\le i\le N_s\\ & \sum_{i=1}^{N_s}f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2)\le P_r \end{aligned}\tag{10}$$

where $g_i'=g_i\gamma_{sr,i}$ and $\tilde{g}_{m,i}^2=(\sigma_r^2+\gamma_{sr,i}^2)\max_j\{g_j^2\},\ j=1,\cdots,N_s$.

Given that $\mathbf{F}_r$ is fixed with $\mathbf{U}_{f_r}=\mathbf{V}_{\hat{h}_{rd}}$ and $\mathbf{V}_{f_r}=\mathbf{U}_{h_{sr}}$, the problem (9) is equivalent to the following convex problem:

$$\begin{aligned}\underset{\substack{\mu\ge0\\ g_i,1\le i\le N_s}}{\text{minimize}} \quad & \sum_{i=1}^{N_s}\frac{\mu((g_i\gamma_{rd,i}f_{r,i}'-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-g_i^2(f_{r,i}')^2-\sigma_r^2g_i^2f_{r,i}^2}\\ & +\mu\epsilon_{rd}^2+\sigma_d^2\sum_{i=1}^{N_s}g_i^2\\ \text{subject to} \quad & g_i^2\le\mu/\tilde{f}_m^2,\ 1\le i\le N_s \end{aligned}\tag{11}$$

where $f_{r,i}'=f_{r,i}\gamma_{sr,i}$ and $\tilde{f}_m^2=\max_i\{\tilde{f}_i^2\}=\max_i\{(f_{r,i}')^2+\sigma_r^2f_{r,i}^2\},\ i=1,\cdots,N_s$.

Proof: See Appendix III. ■

As a consequence of *Theorem* 3, the complex-matrix subproblems (7) and (9) now can be simplified into the scalar problems (10) and (11), respectively, with $\mathbf{G}$ or $\mathbf{F}_r$ being fixed with a certain structure. Moreover, both (10) and (11) are proved to be convex problems, meaning that they can be efficiently solved via a plenty of useful tools in convex optimization. With a closer look at these two problems, we find that they both have a specific structure that can be utilized to devise efficient numerical algorithms based on, e.g., decomposition methods [34]. Such a method has the properties of easy implementation and parallel computation and thus will be our main focus in the following. Notice that the problem (11) is simpler than (10) due to the lack of sum power constraint, so we are going to solve (11) first and later come to (10).

A. Scalar Based Algorithm for Robust Equalizer Optimization

By using the primal decomposition method [34], we decompose the problem (11) into N_s convex subproblems given by

$$\begin{aligned}\underset{g_i^2\le\mu/\tilde{f}_m^2}{\text{minimize}} \quad & \psi_i(g_i)\\ & \triangleq\frac{\mu((g_i\gamma_{rd,i}f_{r,i}'-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-g_i^2(f_{r,i}')^2-\sigma_r^2g_i^2f_{r,i}^2}\\ & +\sigma_d^2g_i^2 \end{aligned}\tag{12}$$

for $i=1,\cdots,N_s$ and one master problem expressed by

$$\underset{\mu\ge0}{\text{minimize}} \quad \phi(\mu)\triangleq\sum_{l=1}^{N_s}\psi_i^*(\mu)+\epsilon_{rd}^2\mu \tag{13}$$

where $\psi_i^*(\mu)$ is the optimal value of (12).

[2]Recall that in the proof of *Theorems* 1 and 2, the subproblems (7) and (9) can be both transformed into SDP. However, the computational load of solving SDP is very high since it involves complex-matrix operations.

$$s_{\psi_i^*}(\mu) = \frac{-((g_i^*\gamma_{rd,i}f_{r,i}^{'}-1)^2+\sigma_r^2(g_i^*)^2\gamma_{rd,i}^2f_{r,i}^2)((f_{r,i}^{'})^2(g_i^*)^2+\sigma_r^2f_{r,i}^2(g_i^*)^2)+\sigma_r^2f_{r,i}^2(g_i^*)^2}{(\mu-(f_{r,i}^{'})^2(g_i^*)^2-\sigma_r^2f_{r,i}^2(g_i^*)^2)^2} \tag{16}$$

$$\begin{aligned} s_{\psi_i^*}(\mu) =& \frac{-((g_i^*\gamma_{rd,i}f_{r,i}^{'}-1)^2+\sigma_r^2(g_i^*)^2\gamma_{rd,i}^2f_{r,i}^2)((f_{r,i}^{'})^2(g_i^*)^2+\sigma_r^2f_{r,i}^2(g_i^*)^2)+\sigma_r^2f_{r,i}^2(g_i^*)^2}{(\mu-(f_{r,i}^{'})^2(g_i^*)^2-\sigma_r^2f_{r,i}^2(g_i^*)^2)^2} \\ &-\frac{\mu[g_i^*((f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2)\gamma_{rd,i}-f_{r,i}^{'}](f_{r,i}^{'}g_i^*-\mu\gamma_{rd,i})}{\tilde{f}_m^2g_i^*(\mu-(f_{r,i}^{'})^2(g_i^*)^2-\sigma_r^2f_{r,i}^2(g_i^*)^2)^2}+\frac{\sigma_d^2}{\tilde{f}_m^2} \end{aligned} \tag{18}$$

Then, our proposed numerical algorithm for solving the problem (11) consists of the following two steps:

1) Solving the N_s subproblems in (12). Due to the convexity of the subproblems in (12), the optimal g_i^* can be achieved by performing bi-section search within the interval $[0, \sqrt{\mu}/\tilde{f}_m]$. The gradient of $\psi_i(g_i)$, which is required by the bi-section method, can be obtained as

$$\begin{aligned} \psi_i^{'}(g_i) =& -2\frac{\mu(g_i((f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2)\gamma_{rd,i}-f_{r,i}^{'})(f_{r,i}^{'}g_i-\mu\gamma_{rd,i})}{(\mu-(f_{r,i}^{'})^2g_i^2-\sigma_r^2f_{r,i}^2g_i^2)^2} \\ &+2\sigma_d^2g_i. \end{aligned} \tag{14}$$

2) Solving the master problem (13). The master problem (13) can also be solved by the bi-section method via exploiting its convexity. It is easy to find that the optimal value of (11) is upper bounded by N_s, thus the initial search interval for μ is $[0, N_s/\epsilon_{rd}^2]$. Moreover, the subsequent proposition gives a subgradient of $\phi(\mu)$.

Proposition 1: A subgradient of $\phi(\mu)$ is

$$s_\phi(\mu) = \sum_{i=1}^{N_s} s_{\psi_i^*}(\mu) + \epsilon_{rd}^2 \tag{15}$$

where $s_{\psi_i^*}(\mu)$ is a subgradient of $\psi_i^*(\mu)$ and it follows

1) If $g_i^* \neq \frac{\sqrt{\mu}}{\tilde{f}_m}$ and $g_i^* \neq \frac{\sqrt{\mu}}{\tilde{f}_i}$, then $s_{\psi_i^*}(\mu)$ is calculated by (16).

2) If $g_i^* \neq \frac{\sqrt{\mu}}{\tilde{f}_m}$ and $g_i^* = \frac{\sqrt{\mu}}{\tilde{f}_i}$, then

$$s_{\psi_i^*}(\mu) = 0. \tag{17}$$

3) If $g_i^* = \frac{\sqrt{\mu}}{\tilde{f}_m}$ and $g_i^* \neq \frac{\sqrt{\mu}}{\tilde{f}_i}$, then $s_{\psi_i^*}(\mu)$ is given by (18).

4) If $g_i^* = \frac{\sqrt{\mu}}{\tilde{f}_m} = \frac{\sqrt{\mu}}{\tilde{f}_i}$, then

$$s_{\psi_i^*}(\mu) = -\frac{\gamma_{rd,i}^2}{4} + \frac{\sigma_d^2}{\tilde{f}_m^2}. \tag{19}$$

Proof: See Appendix IV. ∎

Now we summarize the detailed algorithm solving the scalar problem (11) in Algorithm 1. Note that the global optimality of this algorithm is guaranteed due to the fact that the problem (11) is convex [34]. One advantage of the proposed algorithm is its intrinsic parallel mechanism. For example, in the third step, the root of $\psi_i^{'}(g_i) = 0$ for each i can be searched independently, which can be efficiently implemented in practice.

Algorithm 1 for solving the problem (11)

1: Initialize the iteration number $n = 0$; set the maximum number of iterations $N_{\max}$ and the precision ε; $\mu_a = 0, \mu_b = N_s/\epsilon_{rd}^2, \mu = (\mu_a+\mu_b)/2$.
2: **Repeat**
3: $n = n+1$; find the root of $\psi_i^{'}(g_i) = 0$ for $i = 1, \cdots, N_s$, using bi-section method;
4: Calculate $s_\phi(\mu)$ with (15);
5: If $s_\phi(\mu) < 0$, then $\mu_a = \mu$, else $\mu_b = \mu$;
6: **Until** $n \geq N_{\max}$ or $\mu_b - \mu_a < \varepsilon$.

B. Scalar Based Algorithm for Robust Precoder Optimization

We now consider the scalar optimization of $\mathbf{F}_r$ in (10). Compared to the problem (11), the major difficulty lies in the coupling constraint $\sum_{i=1}^{N_s} f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2) \leq P_r$. To handle this, we adopt the primal-primal decomposition [34], and decompose the problem (10) into N_s subproblems, one secondary master problem, and one master problem. Specifically, we first introduce auxiliary variables $p_i,\ i = 1, \cdots, N_s$, to simplify the problem (10) as

$$\begin{aligned} \underset{\substack{\mu\geq 0, f_{r,i}, p_i \\ 1\leq i\leq N_s}}{\text{minimize}} \ & \sum_{i=1}^{N_s} \frac{\mu((g_i^{'}\gamma_{rd,i}f_{r,i}-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-(g_i^{'})^2f_{r,i}^2-\sigma_r^2g_i^2f_{r,i}^2} \\ & +\mu\epsilon_{rd}^2 \\ \text{subject to}\ & f_{r,i}^2 \leq \mu/\tilde{g}_{m,i}^2,\ f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2) \leq p_i,\ i = 1, \cdots, N_s \\ & \sum_{i=1}^{N_s} p_i \leq P_r. \end{aligned} \tag{20}$$

Then, the subproblem, each for $i = 1, \cdots, N_s$, is given by

$$\begin{aligned} \underset{f_{r,i}}{\text{minimize}} \quad & c_i(f_{r,i}) \\ & \triangleq \frac{\mu((g_i^{'}\gamma_{rd,i}f_{r,i}-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-(g_i^{'})^2f_{r,i}^2-\sigma_r^2g_i^2f_{r,i}^2} \\ \text{subject to} \quad & f_{r,i}^2 \leq \mu/\tilde{g}_{m,i}^2,\ f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2) \leq p_i \end{aligned} \tag{21}$$

which is obtained by fixing $p_i, 1 \leq i \leq N_s$ and μ. Then, the secondary master problem is

$$\underset{\mu\geq 0}{\text{minimize}} \quad d(\mathbf{p}, \mu) \triangleq \sum_{i=1}^{N_s} c_i^*(\mathbf{p}, \mu) + \mu\epsilon_{rd}^2 \tag{22}$$

where $c_i^*(\mathbf{p},\mu)$ denotes the optimal value of (21) under a given $\mathbf{p}=[p_1,\cdots,p_{N_s}]^T$. Finally, the master problem is given by

$$\begin{aligned} &\underset{\mathbf{p}}{\text{minimize}} \quad d^*(\mathbf{p}) \\ &\text{subject to} \quad \sum_{i=1}^{N_s} p_i \le P_r \end{aligned} \tag{23}$$

where $d^*(\mathbf{p})$ is the optimal value of (22). Based on the above procedure, we now show the three steps of our proposed algorithm for solving the problem (10) as follows: 1) Solving the N_s subproblems (21). Fortunately, we find each of these subproblems admits a closed-form solution, as shown in *Proposition* 2.

2) Solving the secondary master problem (22). This problem can be solved using the bi-section method. The initial interval is $\mu \in [0, N_s/\epsilon_{rd}^2]$ and a subgradient of $d(\mathbf{p},\mu)$ with respect to μ is provided in *Proposition* 2.

3) Solving the master problem (23). The subgradient projection method is ready for solving this problem. To be more specific, the solution is searched with the following expression:

$$\mathbf{p}[n+1] = (\mathbf{p}[n] - \alpha[n]\mathbf{s}_{d^*}(\mathbf{p}[n]))_{\mathcal{Q}} \tag{24}$$

where n is the iteration index, $\alpha[n]$ is the search stepsize, $\mathbf{s}_{d^*}(\mathbf{p}[n])$ is a subgradient of $d^*(\mathbf{p})$ at $\mathbf{p}[n]$ and $[\cdot]_{\mathcal{Q}}$ stands for the projection onto the set $\mathcal{Q}=\{\mathbf{p}:\mathbf{1}^T\mathbf{p}\le P_r\}$. It was pointed out in [35] that $\mathbf{p}=[\boldsymbol{\gamma}]_{\mathcal{Q}}$ can be expressed as the water-filling form $p_i=\max\{\gamma_i-\xi,0\},\forall i$, where $\xi\ge 0$ represents the water level satisfying $\mathbf{1}^T\mathbf{p}\le P_r$.

Proposition 2: The optimal solution to the problem (21) is

$$\begin{aligned} f_{r,i}^* =&\min\Big\{\sqrt{p_i/(\sigma_r^2+\gamma_{sr,i}^2)}, \sqrt{\mu}/\tilde{g}_{m,i}, \mu\gamma_{rd,i}/g_i^{'}, \\ &g_i^{'}/(((g_i^{'})^2+\sigma_r^2 g_i^2)\gamma_{rd,i})\Big\} \end{aligned} \tag{25}$$

and a subgradient of $d(\mathbf{p},\mu)$ with respect to μ is

$$s_d(\mu) = \sum_{i=1}^{N_s} s_{c_i^*}(\mu) + \epsilon_{rd}^2 \tag{26}$$

where $s_{c_i^*}(\mu)$ has the following form

1) If $f_{r,i}^* \ne \frac{\sqrt{\mu}}{\tilde{g}_{m,i}}$ and $f_{r,i}^* \ne \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then $s_{c_i^*}(\mu)$ is given by (27).

2) If $f_{r,i}^* \ne \frac{\sqrt{\mu}}{\tilde{g}_{m,i}}$ and $f_{r,i}^* = \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then

$$s_{c_i^*}(\mu) = 0. \tag{28}$$

3) If $f_{r,i}^* = \frac{\sqrt{\mu}}{\tilde{g}_{m,i}}$ and $f_{r,i}^* \ne \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then $s_{c_i^*}(\mu)$ is given by (29).

4) If $f_{r,i}^* = \frac{\sqrt{\mu}}{\tilde{g}_{m,i}}$ and $f_{r,i}^* = \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then

$$s_{c_i^*}(\mu) = -\frac{\gamma_{rd,i}^2}{4}. \tag{30}$$

The ith element of a subgradient of $d^*(\mathbf{p})$ is given as follows

1) If $f_{r,i}^* \ne \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}}$, then

$$(\mathbf{s}_{d^*}(\mathbf{p}))_i = 0 \tag{31}$$

2) If $f_{r,i}^* = \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}}$ and $f_{r,i}^* \ne \sqrt{\frac{\mu^*}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then

$$\begin{aligned} &(\mathbf{s}_{d^*}(\mathbf{p}))_i \\ &= -\frac{\mu^*[f_{r,i}^*((g_i^{'})^2+\sigma_r^2 g_i^2)\gamma_{rd,i}-g_i^{'}](g_i^{'}f_{r,i}^*-\mu^*\gamma_{rd,i})}{(\sigma_r^2+\gamma_{sr,i}^2)f_{r,i}^*(\mu^*-(g_i^{'})^2(f_{r,i}^*)^2-\sigma_r^2 g_i^2(f_{r,i}^*)^2)^2} \end{aligned} \tag{32}$$

3) If $f_{r,i}^* = \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}} = \sqrt{\frac{\mu^*}{(g_i^{'})^2+\sigma_r^2 g_i^2}}$, then

$$(\mathbf{s}_{d^*}(\mathbf{p}))_i = -\frac{\gamma_{rd,i}^2 g_i^2}{4} \tag{33}$$

where μ^* is the solution to the problem (22).

Proof: See Appendix V. ■

Based on the decomposition method we proposed above and *Proposition* 2, we make a summary on the algorithm for solving the problem (10) in Algorithm 2.

Algorithm 2 for solving the problem (10)

Initialize the outer iteration number $n_{\text{out}}=0$ and inner iteration number $n_{\text{in}}=0$; set the maximum number of outer iterations $N_{\text{max,out}}$, inner iterations $N_{\text{max,in}}$ and the precision ε; set the initial $\mathbf{p}[0]$.
Repeat
  $\mu_a=0$, $\mu_b=N_s/\epsilon_{rd}^2$, $\mu=(\mu_a+\mu_b)/2$, $n_{\text{in}}=0$;
  While $\mu_b-\mu_a>\varepsilon$ or $n_{\text{in}}<N_{\text{max,in}}$
    $n_{\text{in}}=n_{\text{in}}+1$; calculate $f_{r,i}^*$ for $i=1,\cdots,N_s$ from (25);
    Calculate $s_d(\mu)$ with (26);
    If $s_d(\mu)<0$, then $\mu_a=\mu$, else $\mu_b=\mu$;
  End while
  Calculate $\mathbf{s}_{d^*}(\mathbf{p}[n_{\text{out}}])$ according to (31)-(33);
  Update $\mathbf{p}$ using $\mathbf{p}[n_{\text{out}}+1] = (\mathbf{p}[n_{\text{out}}]-\alpha[n_{\text{out}}]\mathbf{s}_{d^*}(\mathbf{p}[n_{\text{out}}]))_{\mathcal{Q}}$; $n_{\text{out}}=n_{\text{out}}+1$;
Until $n_{\text{out}}\ge N_{\text{max,out}}$ or $|d^*(\mathbf{p}[n_{\text{out}}+1])-d^*(\mathbf{p}[n_{\text{out}}])|<\varepsilon$.

C. Iterative Algorithm for Joint Robust Transceiver Optimization

Up till now, we have addressed the matrix-valued subproblems (7) and (9) using simple scalar based algorithms. Therefore, the remaining work is to apply the alternating algorithm to deal with the original non-convex minimax problem (6), i.e, optimizing one of $\mathbf{F}_r$ and $\mathbf{G}$ with the other fixed at one time. To be more specific, we can set the initial point $\mathbf{F}_r$ with $\mathbf{V}_{\hat{h}_{rd}}\boldsymbol{\Sigma}_{f_r}\mathbf{U}_{h_{sr}}^H$, then according to *Theorem* 3, the optimization of $\mathbf{G}$ is equivalent to the scalar problem (11), which has been solved by Algorithm 1. As the resulting $\mathbf{G}$ must satisfy $\mathbf{U}_g=\mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g=\mathbf{U}_{\hat{h}_{rd}}$, it follows from *Theorem* 3 that the problem of optimizing $\mathbf{F}_r$ can be transformed into the power allocation problem (10) that can be solved with Algorithm 2. As a summary, the details of the proposed alternating algorithm are given in Algorithm 3.

Now we would like to prove that the above alternating algorithm always converges. For the i-th iteration ($i\ge 1$) of Algorithm 3, let us denote the optimized relay precoder and destination equalizer with $\mathbf{F}_r^{(i)}$ and $\mathbf{G}^{(i)}$, which are obtained

$$s_{c_i^*}(\mu) = \frac{-[(f_{r,i}^* \gamma_{rd,i} g_i^{'} - 1)^2 + \sigma_r^2 (f_{r,i}^*)^2 \gamma_{rd,i}^2 g_i^2]((g_i^{'})^2 (f_{r,i}^*)^2 + \sigma_r^2 g_i^2 (f_{r,i}^*)^2) + \sigma_r^2 g_i^2 (f_{r,i}^*)^2}{(\mu - (g_i^{'})^2 (f_{r,i}^*)^2 - \sigma_r^2 g_i^2 (f_{r,i}^*)^2)^2} \tag{27}$$

$$\begin{aligned} s_{c_i^*}(\mu) =& \frac{-((f_{r,i}^* \gamma_{rd,i} g_i^{'} - 1)^2 + \sigma_r^2 g_i^2 \gamma_{rd,i}^2 (f_{r,i}^*)^2)((f_{r,i}^*)^2 (g_i^{'})^2 + \sigma_r^2 (f_{r,i}^*)^2 g_i^2) + \sigma_r^2 (f_{r,i}^*)^2 g_i^2}{(\mu - (g_i^{'})^2 (f_{r,i}^*)^2 - \sigma_r^2 g_i^2 (f_{r,i}^*)^2)^2} \\ &- \frac{\mu [f_{r,i}^* ((g_i^{'})^2 + \sigma_r^2 g_i^2) \gamma_{rd,i} - g_i^{'}](f_{r,i}^* g_i^{'} - \mu \gamma_{rd,i})}{\tilde{g}_{m,i}^2 f_{r,i}^* (\mu - (g_i^{'})^2 (f_{r,i}^*)^2 - \sigma_r^2 g_i^2 (f_{r,i}^*)^2)^2} \end{aligned} \tag{29}$$

Algorithm 3 for solving the problem (6) **based on scalar optimization**

1: Initialize the iteration number $n = 0$; set the maximum number of iterations N_{max} and the precision ε; select $\text{MSE}_{\text{new}} \gg 0$; choose an $N_r \times N_r$ matrix $\mathbf{F}_r = \mathbf{U}_{f_r} \mathbf{\Sigma}_{f_r} \mathbf{V}_{f_r}^H$ such that $\mathbf{U}_{f_r} = \mathbf{V}_{\hat{h}_{rd}}$, $\mathbf{V}_{f_r} = \mathbf{U}_{h_{sr}}$ and $\sum_{i=1}^{N_s} f_{r,i}^2 (\sigma_r^2 + \gamma_{sr,i}^2) = P_r$.
2: **Repeat**
3: $n = n+1$; fix $\{f_{r,i}\}$ and update $\{g_i\}$ with Algorithm 1;
4: Fix $\{g_i\}$ and update $\{f_{r,i}\}$ with Algorithm 2;
5: $\text{MSE}_{\text{old}} = \text{MSE}_{\text{new}}$; update MSE_{new} with new $\{f_{r,i}\}$ and $\{g_i\}$;
6: **Until** $n \geq N_{\text{max}}$ or $\text{MSE}_{\text{old}} - \text{MSE}_{\text{new}} < \varepsilon$.
7: $\mathbf{F}_r = \mathbf{V}_{\hat{h}_{rd}} \mathbf{\Sigma}_{f_r} \mathbf{U}_{h_{sr}}^H$ and $\mathbf{G} = \mathbf{V}_{h_{sr}} \mathbf{\Sigma}_g \mathbf{U}_{\hat{h}_{rd}}^H$.

based on the optimized power allocation and the optimal channel-diagonalizing structure shown in *Theorems* 1 and 2. The MSE objective value corresponding to $(\mathbf{F}_r^{(i)}, \mathbf{G}^{(i)})$ is denoted by $\text{MSE}_{(i)}$. Then, in the first step of the $(i+1)$-th iteration, we fix $\mathbf{F}_r = \mathbf{F}_r^{(i)}$ and minimize the MSE objective, which results in an optimal solution of $\mathbf{G} = \mathbf{G}^{(i+1)}$. Denoting the MSE value corresponding to $(\mathbf{F}_r^{(i)}, \mathbf{G}^{(i+1)})$ with $\text{MSE}_{(i)}^{'}$, we must have $\text{MSE}_{(i)}^{'} \leq \text{MSE}_{(i)}$ since $\mathbf{G}^{(i+1)}$ minimizes the MSE objective function when $\mathbf{F}_r = \mathbf{F}_r^{(i)}$. In the second step of the $(i+1)$-th iteration, we fix $\mathbf{G} = \mathbf{G}^{(i+1)}$ and minimize the MSE objective to obtain an optimal solution of $\mathbf{F}_r = \mathbf{F}_r^{(i+1)}$. Letting $\text{MSE}_{(i+1)}$ be the MSE value corresponding to $(\mathbf{F}_r^{(i+1)}, \mathbf{G}^{(i+1)})$, similarly, it is easy to verify that $\text{MSE}_{(i+1)} \leq \text{MSE}_{(i)}^{'}$ and hence $\text{MSE}_{(i+1)} \leq \text{MSE}_{(i)}$. Therefore, by applying Algorithm 3, the MSE value is monotonically decreasing with each iteration. Since the MSE value is lower bounded by zero, it follows that Algorithm 3 does converge.

Remark 2: By fixing the structure of the initial relay precoder, we can alternatively update the destination equalizer and relay precoder by solving scalar-valued power allocation problems according to *Theorem* 3. In addition, the constraint on $\mathbf{\Sigma}_{f_r}$ is used to guarantee the feasibility of the initial $\mathbf{F}_r$. Although there are many possible choices for initializing power allocation matrix $\mathbf{\Sigma}_{f_r}$, we find via simulations that the converging value of Algorithm 3 is insensitive to initial values of $\mathbf{\Sigma}_{f_r}$ while the algorithm convergence speed depends on the specific initialization for $\mathbf{\Sigma}_{f_r}$. We will show detailed simulation results on the convergence issue in Section V. We also would like to note that the convergence condition $n \geq N_{\text{max}}$ could be redundant when N_{max} is set to a sufficiently large value. Nevertheless, this condition can be useful when the system designer would like to terminate the algorithm with a fixed and small number of iterations which will be convenient for practical implementation.

Remark 3: The alternating algorithm we used in this paper is a popular and efficient method to deal with difficult non-convex optimization problems with coupled variables. By alternatingly solving tractable convex subproblems, it is possible to obtain a high-quality solution to the original complicated non-convex problem. Although the solution achieved by this algorithm is generally locally optimal, it can still provide a significant gain over the non-robust scheme in the presence of norm-bounded CSI uncertainties, as verified by simulation results in Section V.

V. Simulation Results

In this section, we investigate the performance of the proposed robust transceiver design under a three-node MIMO relay system. We adopt independent and identically distributed (i.i.d.) Rayleigh fading as the channel model for both hops. The transmit power at both source and relay nodes is set to 1, i.e., $P_s = P_r = 1$. The non-robust scheme in [7] is used as a benchmark for comparison. We concern about the average worst-case MSE and BER performance, which is interpreted as the MSE/BER for a given transceiver in the worst-case channel averaged over different channel realizations, i.e., $\mathbf{H}_{sr}$ and $\hat{\mathbf{H}}_{rd}$. The worst-case channel is found by solving the following problem:

$$\begin{aligned} \underset{\|\mathbf{\Delta}_{rd}\|_F \leq \epsilon_{rd}}{\text{maxmize}} \quad & \left\| \mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd}) \mathbf{F}_r \mathbf{H}_{sr} - \mathbf{I} \right\|_F^2 \\ & + \sigma_{n_r}^2 \left\| \mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd}) \mathbf{F}_r \right\|_F^2 + \sigma_{n_d}^2 \left\| \mathbf{G} \right\|_F^2. \end{aligned} \tag{34}$$

This problem is in general non-convex, but we are still able to achieve its optimal solution, which is given in Appendix VI. We define $\epsilon_{rd}^2 = \rho \|\hat{\mathbf{H}}_{rd}\|_F^2$, where $\rho \in [0, 1)$ is a metric for evaluating the size of CSI uncertainties. Fig. 2 shows the worst-case MSE performance of the robust and non-robust transceiver designs versus different SNRs at the destination (defined by $\text{SNR}_d = P_r/\sigma_d^2$). The number of

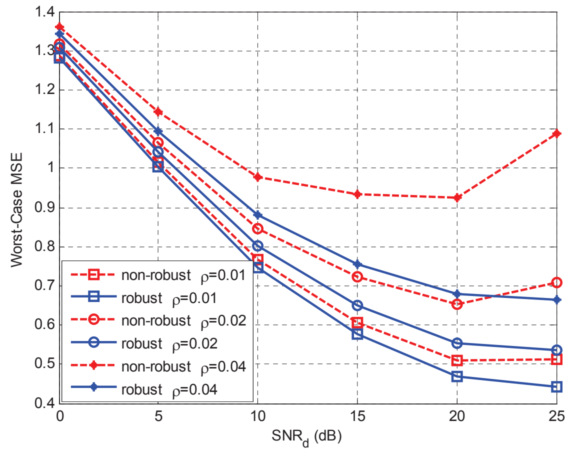


Fig. 2. Worst-case MSE versus SNR$_d$ with different ρ ($N_s = N_r = N_d = 2$, SNR$_r$ = 15 dB).

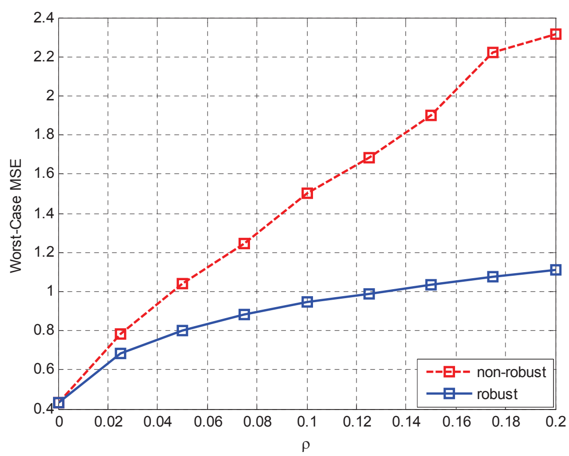


Fig. 3. Worst-case MSE versus different ρ ($N_s = N_r = N_d = 2$, SNR$_d$ = SNR$_r$ = 15 dB).

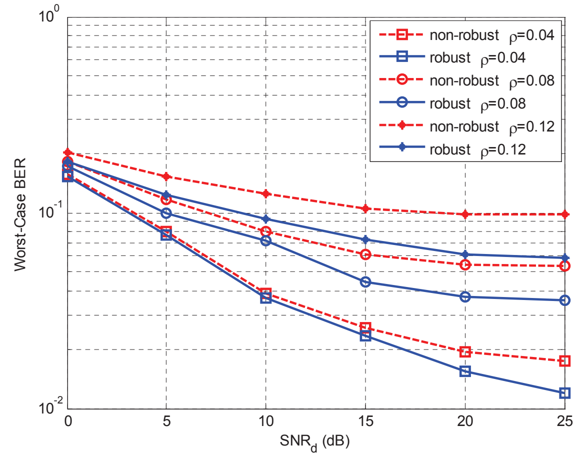


Fig. 4. Worst-case BER versus SNR$_d$ with different ρ ($N_s = N_d = 2,\ N_r = 3$, SNR$_r$ = 15 dB).

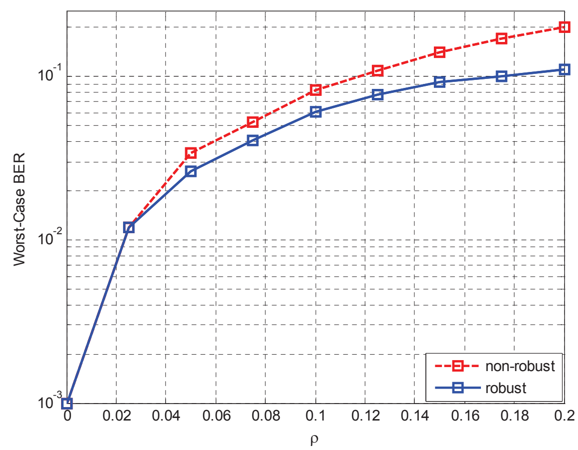


Fig. 5. Worst-case BER versus different ρ ($N_s = N_d = 2,\ N_r = 3$, SNR$_r$ = SNR$_d$ = 15 dB).

antennas of all three nodes is 2. The SNR at the relay node (defined by SNR$_r$ $= P_s/\sigma_r^2$) is set to 15 dB. It can be seen that when the size of CSI uncertainties ρ is fixed, the performance advantage of the robust scheme over the non-robust one increases gradually. On the other hand, if we fix SNR$_d$, the gap between these two schemes grows as ρ becomes larger. We can observe this phenomenon more clearly in Fig. 3, where SNR$_d$ and SNR$_r$ are set with 15 dB. From these two figures, we find that the performance gain of robust transceiver design is evident with the existence of CSI errors.

In Fig. 4, we compare the worst-case BER of robust and non-robust designs. We adopt binary phase shift keying (BPSK) modulation and fix the SNR at the relay with 15 dB. The number of antennas at the source and destination is 2 and the relay has 3 antennas. It can be found that when the size of CSI uncertainties ρ is relatively small, the robust scheme outperforms the non-robust one in both medium and high SNR regions. And the gain becomes obvious in the whole SNR region when ρ is larger. In Fig. 5, we set SNR$_d$ with 15 dB. The worst-case BER of the robust design performs better than the non-robust one with different ρ, and its superiority becomes more evident as ρ increases.

In Figs. 6 and 7, we investigate the performance of the proposed robust design under different antenna configurations. It can be observed that the advantage of robust design over the conventional non-robust one becomes more evident when the number of antennas increases.

Finally, we investigate the convergence behavior of our proposed iterative algorithm (Algorithm 3) in Figs. 8 and 9, where the notation "random" denotes that the diagonals of the initial $\mathbf{\Sigma}_{f_r}$ are randomly generated and notation "equal" means that the initial $\mathbf{\Sigma}_{f_r}$ has the same diagonals, i.e., $f_{r,i} = \sqrt{\frac{P_r}{\sum_{i=1}^{N_s}(\sigma_r^2+\gamma_{sr,i}^2)}},\ i = 1, \cdots, N_s$. From these results, we observe the following phenomena: 1) The MSE value gradually decreases with each iteration; 2) The two different initialization schemes converge to the same value; 3) The proposed alternating algorithm has a fast convergence speed, especially when we let $\mathbf{\Sigma}_{f_r}$ have equal diagonals.

VI. Conclusions

In this paper, we studied a worst-case MMSE transceiver design for nonregenerative MIMO relay systems. After de-

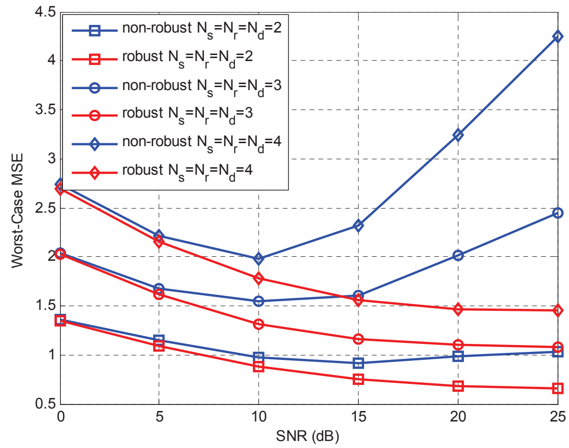


Fig. 6. Worst-case MSE performance with different antenna numbers ($\rho = 0.04$).

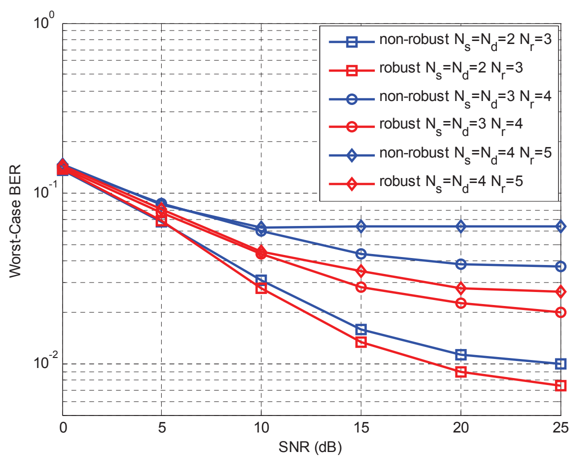


Fig. 7. Worst-case BER performance with different antenna numbers ($\rho = 0.04$).

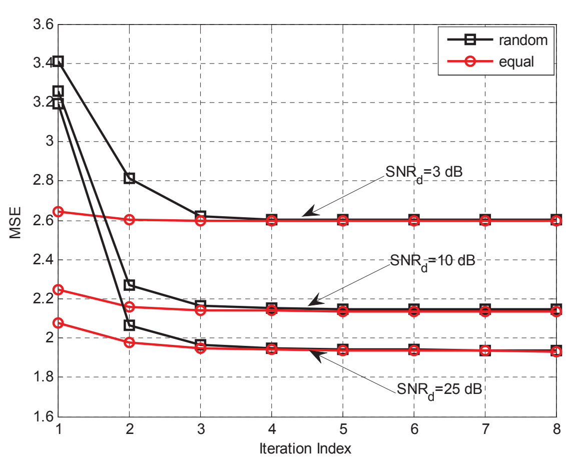


Fig. 8. Convergence behavior of Algorithm 3 with different initial $\mathbf{\Sigma}_{f_r}$ ($N_s = N_r = N_d = 4$, SNR$_r$ =15 dB, $\rho = 0.1$).

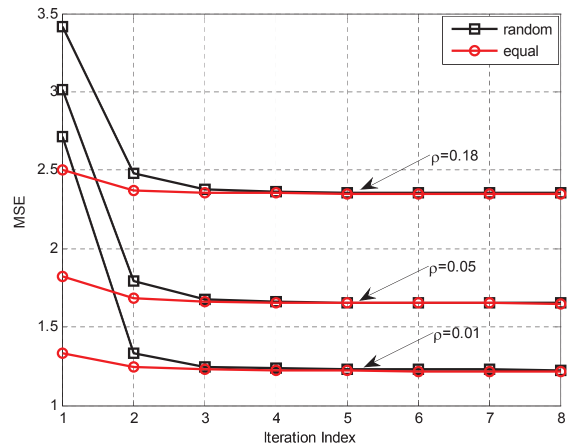


Fig. 9. Convergence behavior of Algorithm 3 with different initial $\mathbf{\Sigma}_{f_r}$ ($N_s = N_r = N_d = 4$, SNR$_r$ = SNR$_d$=15 dB).

coupling the original non-convex optimization problem into two subproblems, we proved that the optimal solution to each subproblem has an interesting channel-diagonalizing structure under some mild conditions, which is the first main result of our work. In light of this conclusion, we proposed an efficient alternating algorithm to address the worst-case robust transceiver design. The proposed robust algorithm involves simple scalar operations and has guaranteed convergence. Simulation results show that the algorithm outperforms the non-robust counterpart by a significant gain and also converges with a fast speed.

Appendix I
Proof of Theorem 1

As mentioned before, the main intricacy of the problem (7) lies in the constraint with respect to the channel uncertainty $\mathbf{\Delta}_{rd}$. Thus, we need to transform the problem (7) into an equivalent form which is irrelevant to $\mathbf{\Delta}_{rd}$ first. Concerning the left-hand side (LHS) of the first constraint, which can be expressed by

$$\begin{aligned}
&\left\|\mathbf{G}(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{H}_{sr}-\mathbf{I}\right\|_F^2+\sigma_{n_r}^2\left\|\mathbf{G}(\hat{\mathbf{H}}_{rd}+\mathbf{\Delta}_{rd})\mathbf{F}_r\right\|_F^2\\
=&\left\|\mathbf{U}_g\mathbf{\Sigma}_g\mathbf{V}_g^H(\mathbf{U}_{\hat{h}_{rd}}\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H+\mathbf{\Delta}_{rd})\mathbf{F}_r\mathbf{U}_{h_{sr}}\mathbf{\Sigma}_{h_{sr}}\mathbf{V}_{h_{sr}}^H-\mathbf{I}\right\|_F^2\\
&+\sigma_{n_r}^2\left\|\mathbf{U}_g\mathbf{\Sigma}_g\mathbf{V}_g^H(\mathbf{U}_{\hat{h}_{rd}}\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H+\mathbf{\Delta}_{rd})\mathbf{F}_r\right\|_F^2\\
\overset{(a)}{=}&\Big\|\mathbf{\Sigma}_g\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H\mathbf{F}_r\mathbf{U}_{h_{sr}}\mathbf{\Sigma}_{h_{sr}}-\mathbf{I}+\mathbf{\Sigma}_g\mathbf{U}_{\hat{h}_{rd}}^H\mathbf{\Delta}_{rd}\mathbf{V}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H\\
&\times\mathbf{F}_r\mathbf{U}_{h_{sr}}\mathbf{\Sigma}_{h_{sr}}\Big\|_F^2+\sigma_{n_r}^2\Big\|\mathbf{\Sigma}_g\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H\mathbf{F}_r\mathbf{U}_{h_{sr}}+\mathbf{\Sigma}_g\mathbf{U}_{\hat{h}_{rd}}^H\\
&\times\mathbf{\Delta}_{rd}\mathbf{V}_{\hat{h}_{rd}}\mathbf{V}_{\hat{h}_{rd}}^H\mathbf{F}_r\mathbf{U}_{h_{sr}}\Big\|_F^2\\
\overset{(b)}{=}&\left\|\begin{bmatrix}\mathrm{vec}(\mathbf{\Sigma}_g\mathbf{\Sigma}_{\hat{h}_{rd}}\hat{\mathbf{F}}_r\mathbf{\Sigma}_{h_{sr}}-\mathbf{I})\\ \sigma_{n_r}\mathrm{vec}(\mathbf{\Sigma}_g\mathbf{\Sigma}_{\hat{h}_{rd}}\hat{\mathbf{F}}_r)\end{bmatrix}+\begin{bmatrix}\mathbf{\Sigma}_{h_{sr}}^T\hat{\mathbf{F}}_r^T\otimes\mathbf{\Sigma}_g\\ \sigma_{n_r}\hat{\mathbf{F}}_r^T\otimes\mathbf{\Sigma}_g\end{bmatrix}\mathrm{vec}(\hat{\mathbf{\Delta}}_{rd})\right\|^2\\
\overset{(c)}{=}&\|\boldsymbol{\eta}+\mathbf{\Gamma}\hat{\boldsymbol{\delta}}_{rd}\|^2. \qquad (35)
\end{aligned}$$

Note that we have used the condition $\mathbf{U}_g = \mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g = \mathbf{U}_{\hat{h}_{rd}}$ in (a), while in (b), we have introduced two new matrices

$\hat{\mathbf{F}}_r = \mathbf{V}_{\hat{h}_{rd}}^H \mathbf{F}_r \mathbf{U}_{h_{sr}}$ and $\hat{\mathbf{\Delta}}_{rd} = \mathbf{U}_{\hat{h}_{rd}}^H \mathbf{\Delta}_{rd} \mathbf{V}_{\hat{h}_{rd}}$. The variables $\boldsymbol{\eta}$, $\mathbf{\Gamma}$ and $\hat{\boldsymbol{\delta}}_{rd}$ in (c) are defined by

$$\boldsymbol{\eta} = \begin{bmatrix} \mathrm{vec}(\mathbf{\Sigma}_g \mathbf{\Sigma}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r \mathbf{\Sigma}_{h_{sr}} - \mathbf{I}) \\ \sigma_{n_r} \mathrm{vec}(\mathbf{\Sigma}_g \mathbf{\Sigma}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r) \end{bmatrix}, \ \mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Sigma}_{h_{sr}}^T \hat{\mathbf{F}}_r^T \otimes \mathbf{\Sigma}_g \\ \sigma_{n_r} \hat{\mathbf{F}}_r^T \otimes \mathbf{\Sigma}_g \end{bmatrix},$$
$$\hat{\boldsymbol{\delta}}_{rd} = \mathrm{vec}(\hat{\mathbf{\Delta}}_{rd}). \tag{36}$$

Therefore, the first constraint of (7) now becomes $\|\boldsymbol{\eta} + \mathbf{\Gamma}\hat{\boldsymbol{\delta}}_{rd}\|^2 \le t$, which, by using Schur's Complement [36], amounts to

$$\begin{bmatrix} t & (\boldsymbol{\eta} + \mathbf{\Gamma}\hat{\boldsymbol{\delta}}_{rd})^H \\ \boldsymbol{\eta} + \mathbf{\Gamma}\hat{\boldsymbol{\delta}}_{rd} & \mathbf{I} \end{bmatrix} \succeq 0, \quad \forall \hat{\boldsymbol{\delta}}_{rd} : \|\hat{\boldsymbol{\delta}}_{rd}\| \le \epsilon_{rd}. \tag{37}$$

Then, by applying S-lemma [37], (37) is equivalent to the following linear matrix inequality (LMI):

$$\begin{bmatrix} t - \mu\epsilon_{rd}^2 & \boldsymbol{\eta}^H & \mathbf{0} \\ \boldsymbol{\eta} & \mathbf{I} & \mathbf{\Gamma} \\ \mathbf{0} & \mathbf{\Gamma}^H & \mu\mathbf{I} \end{bmatrix} \succeq 0. \tag{38}$$

Based on the above LMI and some matrix manipulations, we can convert the problem (7) to

$$\begin{aligned} \underset{\hat{\mathbf{F}}_r', t, \mu}{\text{minimize}} \quad & t \\ \text{subject to} \quad & \begin{bmatrix} t - \mu\epsilon_{rd}^2 & (\boldsymbol{\eta}')^H & \mathbf{0} \\ \boldsymbol{\eta}' & \mathbf{I} & \mathbf{\Gamma}' \\ \mathbf{0} & (\mathbf{\Gamma}')^H & \mu\mathbf{I} \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} P_r & \mathrm{vec}^H(\hat{\mathbf{F}}_r'\mathbf{\Lambda}_p) \\ \mathrm{vec}(\hat{\mathbf{F}}_r'\mathbf{\Lambda}_p) & \mathbf{I} \end{bmatrix} \succeq 0 \end{aligned} \tag{39}$$

where $\hat{\mathbf{F}}_r'$ is the upper left $N_s \times N_s$ submatrix of $\hat{\mathbf{F}}_r$, $\boldsymbol{\eta}'$, $\mathbf{\Gamma}'$ and $\mathbf{\Lambda}_p$ are given by

$$\boldsymbol{\eta}' = \begin{bmatrix} \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r' \mathbf{\Lambda}_{h_{sr}} - \mathbf{I}) \\ \sigma_{n_r} \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r') \end{bmatrix}, \ \mathbf{\Gamma}' = \begin{bmatrix} \mathbf{\Lambda}_{h_{sr}} (\hat{\mathbf{F}}_r')^T \otimes \mathbf{\Lambda}_g \\ \sigma_{n_r} (\hat{\mathbf{F}}_r')^T \otimes \mathbf{\Lambda}_g \end{bmatrix},$$
$$\mathbf{\Lambda}_p = (\mathbf{\Lambda}_{h_{sr}}^2 + \sigma_{n_r}^2 \mathbf{I})^{\frac{1}{2}}. \tag{40}$$

Therefore, we have equivalently transformed the original intractable problem to a semi-definite programming (SDP).

In the sequel, we will show that there exists a diagonal $\hat{\mathbf{F}}_r'$ among the solution set of the problem (39). Let $\mathbf{\Xi}_{N_s}^k \in \mathbb{R}^{N_s \times N_s}, k = 1, \cdots, 2^{N_s}$ be a diagonal matrix whose diagonals are either 1 or -1. By replacing $\hat{\mathbf{F}}_r'$ in $\mathbf{y}'$ and $\mathbf{Z}'$ with $\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k$, we have

$$\begin{aligned} \boldsymbol{\eta}_{\mathbf{\Xi}}' &= \begin{bmatrix} \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_{h_{sr}} - \mathbf{I}) \\ \sigma_{n_r} \mathrm{vec}(\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k) \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{vec}(\mathbf{\Xi}_{N_s}^k (\mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r' \mathbf{\Lambda}_{h_{sr}} - \mathbf{I}) \mathbf{\Xi}_{N_s}^k) \\ \sigma_{n_r} \mathrm{vec}(\mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_g \mathbf{\Lambda}_{\hat{h}_{rd}} \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k) \end{bmatrix} = \mathbf{T}\boldsymbol{\eta}' \\ \mathbf{\Gamma}_{\mathbf{\Xi}}' &= \begin{bmatrix} \mathbf{\Lambda}_{h_{sr}} \mathbf{\Xi}_{N_s}^k (\hat{\mathbf{F}}_r')^T \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Lambda}_g \\ \sigma_{n_r} \mathbf{\Xi}_{N_s}^k (\hat{\mathbf{F}}_r')^T \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Lambda}_g \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_{h_{sr}} (\hat{\mathbf{F}}_r')^T \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_g \mathbf{\Xi}_{N_s}^k \\ \sigma_{n_r} \mathbf{\Xi}_{N_s}^k (\hat{\mathbf{F}}_r')^T \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_g \mathbf{\Xi}_{N_s}^k \end{bmatrix} = \mathbf{T}\mathbf{\Gamma}'\mathbf{T} \\ \mathrm{vec}(\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_p) &= (\mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k) \mathrm{vec}(\hat{\mathbf{F}}_r' \mathbf{\Lambda}_p) \end{aligned} \tag{41}$$

where $\mathbf{T} = \text{blkdiag}\{\mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k, \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k\}$ and we use $(\mathbf{\Xi}_{N_s}^k)^2 = \mathbf{I}$ and $\mathbf{\Xi}_{N_s}^k \mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{\Xi}_{N_s}^k$ with $\mathbf{\Lambda}$ being diagonal. Thus, the two constraints in (39) are equivalent to

$$\begin{aligned} & \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T} \end{bmatrix} \begin{bmatrix} t - \mu\epsilon_{rd}^2 & (\boldsymbol{\eta}')^H & \mathbf{0} \\ \boldsymbol{\eta}' & \mathbf{I} & \mathbf{\Gamma}' \\ \mathbf{0} & (\mathbf{\Gamma}')^H & \mu\mathbf{I} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T} \end{bmatrix} \\ &= \begin{bmatrix} t - \mu\epsilon_{rd}^2 & (\boldsymbol{\eta}_{\mathbf{\Xi}}')^H & \mathbf{0} \\ \boldsymbol{\eta}_{\mathbf{\Xi}}' & \mathbf{I} & \mathbf{\Gamma}_{\mathbf{\Xi}}' \\ \mathbf{0} & (\mathbf{\Gamma}_{\mathbf{\Xi}}')^H & \mu\mathbf{I} \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k \end{bmatrix} \begin{bmatrix} P_r & \mathrm{vec}^H(\hat{\mathbf{F}}_r'\mathbf{\Lambda}_p) \\ \mathrm{vec}(\hat{\mathbf{F}}_r'\mathbf{\Lambda}_p) & \mathbf{I} \end{bmatrix} \\ & \times \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Xi}_{N_s}^k \otimes \mathbf{\Xi}_{N_s}^k \end{bmatrix} \\ &= \begin{bmatrix} P_r & \mathrm{vec}^H(\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_p) \\ \mathrm{vec}(\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k \mathbf{\Lambda}_p) & \mathbf{I} \end{bmatrix} \succeq 0 \end{aligned} \tag{42}$$

indicating that $\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k$ also belongs to the solution set. Since the set defined by the LMI is convex, the matrix $\mathbf{D}_{\hat{\mathbf{F}}_r'} = (1/2^{N_s}) \sum_{k=1}^{2^{N_s}} \mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k$, as a convex combination of $\mathbf{\Xi}_{N_s}^k \hat{\mathbf{F}}_r' \mathbf{\Xi}_{N_s}^k$, should also be a feasible solution. Moreover, it has been proved in [37] that the matrix $\mathbf{D}_{\hat{\mathbf{F}}_r'}$ is diagonal and satisfies $(\mathbf{D}_{\hat{\mathbf{F}}_r'})_{i,i} = (\hat{\mathbf{F}}_r')_{i,i}, \ i = 1, \cdots, N_s$. Hence, we arrive at the conclusion that there must exist a diagonal solution to (7) which is achieved by setting $\mathbf{U}_{f_r} = \mathbf{V}_{\hat{h}_{rd}}$ and $\mathbf{V}_{f_r} = \mathbf{U}_{h_{sr}}$.

Appendix II
Proof of Theorem 2

The LHS of the first constraint in the problem (9) can be expressed as

$$\begin{aligned} & \left\| \mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd}) \mathbf{F}_r \mathbf{H}_{sr} - \mathbf{I} \right\|_F^2 + \sigma_{n_r}^2 \left\| \mathbf{G}(\hat{\mathbf{H}}_{rd} + \mathbf{\Delta}_{rd}) \right. \\ & \left. \times \mathbf{F}_r \right\|_F^2 \\ & \overset{(a)}{=} \left\| \mathbf{G}(\mathbf{U}_{\hat{h}_{rd}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{V}_{\hat{h}_{rd}}^H + \mathbf{\Delta}_{rd}) \mathbf{U}_{f_r} \mathbf{\Sigma}_{f_r} \mathbf{\Sigma}_{h_{sr}} \mathbf{V}_{h_{sr}}^H - \mathbf{I} \right\|_F^2 \\ & \quad + \sigma_{n_r}^2 \left\| \mathbf{G}(\mathbf{U}_{\hat{h}_{rd}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{V}_{\hat{h}_{rd}}^H + \mathbf{\Delta}_{rd}) \mathbf{U}_{f_r} \mathbf{\Sigma}_{f_r} \mathbf{V}_{f_r}^H \right\|_F^2 \\ & \overset{(b)}{=} \left\| \mathbf{V}_{h_{sr}}^H \mathbf{G}(\mathbf{U}_{\hat{h}_{rd}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{V}_{\hat{h}_{rd}}^H + \mathbf{\Delta}_{rd}) \mathbf{U}_{f_r} \mathbf{\Sigma}_{f_r} \mathbf{\Sigma}_{h_{sr}} - \mathbf{I} \right\|_F^2 \\ & \quad + \sigma_{n_r}^2 \left\| \mathbf{V}_{h_{sr}}^H \mathbf{G}(\mathbf{U}_{\hat{h}_{rd}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{V}_{\hat{h}_{rd}}^H + \mathbf{\Delta}_{rd}) \mathbf{U}_{f_r} \mathbf{\Sigma}_{f_r} \right\|_F^2 \end{aligned} \tag{43}$$

where (a) follows the condition $\mathbf{V}_{f_r} = \mathbf{U}_{h_{sr}}$ and (b) holds since the Frobenius norm is unitary invariable. From the condition $\mathbf{U}_{f_r} = \mathbf{V}_{\hat{h}_{rd}}$, we can transform (43) into

$$\begin{aligned} & \left\| \hat{\mathbf{G}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{\Sigma}_{f_r'} - \mathbf{I} + \hat{\mathbf{G}} \hat{\mathbf{\Delta}}_{rd} \mathbf{\Sigma}_{f_r'} \right\|_F^2 + \sigma_{n_r}^2 \left\| \hat{\mathbf{G}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{\Sigma}_{f_r} \right. \\ & \left. + \hat{\mathbf{G}} \hat{\mathbf{\Delta}}_{rd} \mathbf{\Sigma}_{f_r} \right\|_F^2 \\ & = \left\| \begin{bmatrix} \mathrm{vec}(\hat{\mathbf{G}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{\Sigma}_{f_r'} - \mathbf{I}) \\ \sigma_{n_r} \mathrm{vec}(\hat{\mathbf{G}} \mathbf{\Sigma}_{\hat{h}_{rd}} \mathbf{\Sigma}_{f_r}) \end{bmatrix} + \begin{bmatrix} \mathbf{\Sigma}_{f_r'}^T \otimes \hat{\mathbf{G}} \\ \sigma_{n_r} \mathbf{\Sigma}_{f_r}^T \otimes \hat{\mathbf{G}} \end{bmatrix} \hat{\boldsymbol{\delta}}_{rd} \right\|^2 \\ & = \left\| \mathbf{x} + \mathbf{Y} \hat{\boldsymbol{\delta}}_{rd} \right\|^2 \end{aligned} \tag{44}$$

where $\mathbf{\Sigma}_{f_r'} = \mathbf{\Sigma}_{f_r}\mathbf{\Sigma}_{h_{sr}} = [\mathbf{\Lambda}_{f_r'} \quad \mathbf{0}_{N_s\times(N_r-N_s)}]^T$, $\hat{\mathbf{G}} = \mathbf{V}_{h_{sr}}^H\mathbf{G}\mathbf{U}_{\hat{h}_{rd}}$, $\hat{\mathbf{\Delta}}_{rd} = \mathbf{U}_{\hat{h}_{rd}}^H\mathbf{\Delta}_{rd}\mathbf{V}_{\hat{h}_{rd}}$, $\hat{\boldsymbol{\delta}}_{rd} = \mathrm{vec}(\hat{\mathbf{\Delta}}_{rd})$, and we define variables $\mathbf{x}$ and $\mathbf{Y}$ by

$$\mathbf{x} = \begin{bmatrix}\mathrm{vec}(\hat{\mathbf{G}}\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{\Sigma}_{f_r'} - \mathbf{I})\\ \sigma_{n_r}\mathrm{vec}(\hat{\mathbf{G}}\mathbf{\Sigma}_{\hat{h}_{rd}}\mathbf{\Sigma}_{f_r})\end{bmatrix},\ \mathbf{Y} = \begin{bmatrix}\mathbf{\Sigma}_{f_r'}^T\otimes\hat{\mathbf{G}}\\ \sigma_{n_r}\mathbf{\Sigma}_{f_r}^T\otimes\hat{\mathbf{G}}\end{bmatrix}. \quad (45)$$

Therefore, the problem (9) is equivalent to

$$\begin{aligned}&\underset{\hat{\mathbf{G}},\hat{\boldsymbol{\delta}}_{rd},t}{\mathrm{minimize}} && t + \sigma_{n_d}^2\|\hat{\mathbf{G}}\|_F^2\\ &\text{subject to} && \|\mathbf{x}+\mathbf{Y}\hat{\boldsymbol{\delta}}_{rd}\|^2 \le t, \quad \|\hat{\boldsymbol{\delta}}_{rd}\| \le \epsilon_{rd}.\end{aligned} \quad (46)$$

With similar techniques used in the proof of *Theorem* 1, it is not difficult to convert (46) to

$$\begin{aligned}&\underset{\hat{\mathbf{G}}',t,t',\mu}{\mathrm{minimize}} && t + \sigma_{n_d}^2 t'\\ &\text{subject to} && \begin{bmatrix} t-\mu\epsilon_{rd}^2 & (\mathbf{x}')^H & \mathbf{0}\\ \mathbf{x}' & \mathbf{I} & \mathbf{Y}'\\ \mathbf{0} & (\mathbf{Y}')^H & \mu\mathbf{I}\end{bmatrix}\succeq 0\\ &&& \begin{bmatrix} t' & \mathrm{vec}^H(\hat{\mathbf{G}}')\\ \mathrm{vec}(\hat{\mathbf{G}}') & \mathbf{I}\end{bmatrix}\succeq 0\end{aligned} \quad (47)$$

where $\hat{\mathbf{G}}'$ is the left $N_s\times N_s$ sub-matrix of $\hat{\mathbf{G}}$, $\mathbf{x}'$ and $\mathbf{Y}'$ are given by

$$\mathbf{x}' = \begin{bmatrix}\mathrm{vec}(\hat{\mathbf{G}}'\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r'} - \mathbf{I})\\ \sigma_{n_r}\mathrm{vec}(\hat{\mathbf{G}}'\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r})\end{bmatrix} \text{ and } \mathbf{Y}' = \begin{bmatrix}\mathbf{\Lambda}_{f_r'}\otimes\hat{\mathbf{G}}'\\ \sigma_{n_r}\mathbf{\Lambda}_{f_r}\otimes\hat{\mathbf{G}}'\end{bmatrix}. \quad (48)$$

Subsequently, we prove that there must exist a diagonal $\hat{\mathbf{G}}'$ among the optimal solution set. By replacing $\hat{\mathbf{G}}'$ in $\mathbf{x}'$, $\mathbf{Y}'$ and $\mathrm{vec}(\hat{\mathbf{G}}')$ with $\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k$, we have

$$\begin{aligned}\mathbf{x}_{\mathbf{\Xi}}' &= \begin{bmatrix}\mathrm{vec}(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r'} - \mathbf{I})\\ \sigma_{n_r}\mathrm{vec}(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r})\end{bmatrix}\\ &= \begin{bmatrix}\mathrm{vec}(\mathbf{\Xi}_{N_s}^k(\hat{\mathbf{G}}'\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r'} - \mathbf{I})\mathbf{\Xi}_{N_s}^k)\\ \sigma_{n_r}\mathrm{vec}(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r}\mathbf{\Xi}_{N_s}^k)\end{bmatrix} = \mathbf{T}\mathbf{x}'\\ \mathbf{Y}_{\mathbf{\Xi}}' &= \begin{bmatrix}\mathbf{\Lambda}_{f_r'}\otimes(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k)\\ \sigma_{n_r}\mathbf{\Lambda}_{f_r}\otimes(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k)\end{bmatrix}\\ &= \begin{bmatrix}(\mathbf{\Xi}_{N_s}^k\mathbf{\Lambda}_{f_r'}\mathbf{\Xi}_{N_s}^k)\otimes(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k)\\ \sigma_{n_r}(\mathbf{\Xi}_{N_s}^k\mathbf{\Lambda}_{f_r}\mathbf{\Xi}_{N_s}^k)\otimes(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k)\end{bmatrix} = \mathbf{T}\mathbf{Y}'\mathbf{T}\\ \mathrm{vec}(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k) &= (\mathbf{\Xi}_{N_s}^k\otimes\mathbf{\Xi}_{N_s}^k)\mathrm{vec}(\hat{\mathbf{G}}')\end{aligned} \quad (49)$$

where $\mathbf{T} = \mathrm{blkdiag}\{\mathbf{\Xi}_{N_s}^k\otimes\mathbf{\Xi}_{N_s}^k, \mathbf{\Xi}_{N_s}^k\otimes\mathbf{\Xi}_{N_s}^k\}$. Therefore, according to the LMI in (47), it immediately follows that

$$\begin{aligned}&\begin{bmatrix}1&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{T}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{T}\end{bmatrix}\begin{bmatrix} t-\mu\epsilon_{rd}^2 & (\mathbf{x}')^H & \mathbf{0}\\ \mathbf{x}' & \mathbf{I} & \mathbf{Y}'\\ \mathbf{0} & (\mathbf{Y}')^H & \mu\mathbf{I}\end{bmatrix}\begin{bmatrix}1&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{T}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{T}\end{bmatrix}\\ &= \begin{bmatrix} t-\mu\epsilon_{rd}^2 & (\mathbf{x}_{\mathbf{\Xi}}')^H & \mathbf{0}\\ \mathbf{x}_{\mathbf{\Xi}}' & \mathbf{I} & \mathbf{Y}_{\mathbf{\Xi}}'\\ \mathbf{0} & (\mathbf{Y}_{\mathbf{\Xi}}')^H & \mu\mathbf{I}\end{bmatrix}\succeq 0\\ &\begin{bmatrix}1&\mathbf{0}\\ \mathbf{0}&\mathbf{\Xi}_{N_s}^k\otimes\mathbf{\Xi}_{N_s}^k\end{bmatrix}\begin{bmatrix} t' & \mathrm{vec}^H(\hat{\mathbf{G}}')\\ \mathrm{vec}(\hat{\mathbf{G}}') & \mathbf{I}\end{bmatrix}\\ &\times\begin{bmatrix}1&\mathbf{0}\\ \mathbf{0}&\mathbf{\Xi}_{N_s}^k\otimes\mathbf{\Xi}_{N_s}^k\end{bmatrix}\\ &= \begin{bmatrix} t' & \mathrm{vec}^H(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k)\\ \mathrm{vec}(\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k) & \mathbf{I}\end{bmatrix}\succeq 0\end{aligned} \quad (50)$$

implying that $\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k$ lies in the solution set. As the LMI defined in (47) is a convex set, the linear combination $\mathbf{D}_{\hat{\mathbf{G}}'} = (1/2^{N_s})\sum_{k=1}^{2^{N_s}}\mathbf{\Xi}_{N_s}^k\hat{\mathbf{G}}'\mathbf{\Xi}_{N_s}^k$ should also belong to this set. As $\mathbf{D}_{\hat{\mathbf{G}}'}$ is diagonal matrix satisfying $(\mathbf{D}_{\hat{\mathbf{G}}'})_{i,i} = (\hat{\mathbf{G}}')_{i,i}$ [37], the optimal $\hat{\mathbf{G}}'$ can be diagonal which is achieved by setting $\mathbf{U}_g = \mathbf{V}_{h_{sr}}$ and $\mathbf{V}_g = \mathbf{U}_{\hat{h}_{rd}}$.

APPENDIX III
PROOF OF THEOREM 3

With *Theorem* 1, we obtain an equivalent form for the first constraint of the problem (39) as (51). After performing some row and column permutations, the above LMI can be transformed into $\mathrm{blkdiag}\{\mathbf{\Upsilon},\mathbf{\Theta}\}\succeq 0$, where $\mathbf{\Upsilon}$ and $\mathbf{\Theta}$ are given by

$$\mathbf{\Upsilon} = \begin{bmatrix} t-\mu\epsilon_{rd}^2 & \mathbf{0} & \boldsymbol{\zeta}^T\\ \mathbf{0} & \mu\mathbf{I} & \mathbf{\Phi}\\ \boldsymbol{\zeta} & \mathbf{\Phi}^T & \mathbf{I}\end{bmatrix}, \mathbf{\Theta} = \mathrm{blkdiag}\left\{\begin{bmatrix}\mathbf{I} & \boldsymbol{\theta}_{ij}\\ \boldsymbol{\theta}_{ij}^T & \mu\end{bmatrix}_{i\neq j}\right\} \quad (52)$$

where we let

$$\begin{aligned}\boldsymbol{\zeta} =& \big[g_1'\gamma_{rd,1}f_{r,1}-1,\cdots,g_{N_s}'\gamma_{rd,N_s}f_{r,N_s}-1,\\ &\sigma_r g_1\gamma_{rd,1}f_{r,1},\cdots,\sigma_r g_{N_s}\gamma_{rd,N_s}f_{r,N_s}\big]^T\end{aligned}$$

and

$$\begin{aligned}\mathbf{\Phi} =& \big[\mathrm{diag}\{[f_{r,1}g_1',\cdots,f_{r,N_s}g_{N_s}']^T\},\\ &\mathrm{diag}\{[\sigma_r f_{r,1}g_1,\cdots,\sigma_r f_{r,N_s}g_{N_s}]^T\}\big]\end{aligned}$$

with $g_i' = g_i\gamma_{sr,i}$ and $\boldsymbol{\theta}_{ij} = [f_{r,i}\gamma_{sr,i}g_j, \sigma_r f_{r,i}g_j]^T, 1\le i\le N_s, j\neq i$. It can be readily found that $\mathbf{\Theta}\succeq 0$ leads to $\mu\ge(f_{r,i}\gamma_{sr,i}g_j)^2+\sigma_r^2(f_{r,i}g_j)^2, i\neq j$. By using Schur's Complement [36], we find $\mathbf{\Upsilon}\succeq 0$ is equivalent to

$$\begin{bmatrix} t-\mu\epsilon_{rd}^2-\boldsymbol{\zeta}^T\boldsymbol{\zeta} & -\boldsymbol{\zeta}^T\mathbf{\Phi}^T\\ \mathbf{\Phi}\boldsymbol{\zeta} & \mu\mathbf{I}-\mathbf{\Phi}\mathbf{\Phi}^T\end{bmatrix}\succeq 0. \quad (53)$$

$$\begin{bmatrix} t-\mu\epsilon_{rd}^2 & \mathrm{vec}^T(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r}\mathbf{\Lambda}_{h_{sr}}-\mathbf{I}) & \sigma_{n_r}\mathrm{vec}^T(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r}) & \mathbf{0}\\ \mathrm{vec}(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r}\mathbf{\Lambda}_{h_{sr}}-\mathbf{I}) & \mathbf{I} & \mathbf{0} & \mathbf{\Lambda}_{f_r}\mathbf{\Lambda}_{h_{sr}}\otimes\mathbf{\Lambda}_g\\ \sigma_{n_r}\mathrm{vec}(\mathbf{\Lambda}_g\mathbf{\Lambda}_{\hat{h}_{rd}}\mathbf{\Lambda}_{f_r}) & \mathbf{0} & \mathbf{I} & \sigma_{n_r}\mathbf{\Lambda}_{f_r}\otimes\mathbf{\Lambda}_g\\ \mathbf{0} & \mathbf{\Lambda}_{f_r}\mathbf{\Lambda}_{h_{sr}}\otimes\mathbf{\Lambda}_g & \sigma_{n_r}\mathbf{\Lambda}_{f_r}\otimes\mathbf{\Lambda}_g & \mu\mathbf{I}\end{bmatrix}\succeq 0 \quad (51)$$

Then, by applying generalized Schur's Complement [38] on (53), we obtain

$$\begin{aligned}&t-\mu\epsilon_{rd}^2-\boldsymbol{\zeta}^T\boldsymbol{\zeta}-\boldsymbol{\zeta}^T\boldsymbol{\Phi}^T(\mu\mathbf{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{\dagger}\boldsymbol{\Phi}\boldsymbol{\zeta}\geq 0,\\&[\mathbf{I}-(\mu\mathbf{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}^T)(\mu\mathbf{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{\dagger}]\boldsymbol{\Phi}\boldsymbol{\zeta}=\mathbf{0},\\&\mu\mathbf{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}^T\succeq 0.\end{aligned}\tag{54}$$

In fact, $\mu\mathbf{I}-\boldsymbol{\Phi}\boldsymbol{\Phi}^T\succeq 0$ amounts to $\mu\geq(f_{r,i}g_i^{'})^2+\sigma_r^2(f_{r,i}g_i)^2,\ \forall i$. If the equality does not hold for any i, then we can convert the problem (39) to

$$\begin{aligned}\underset{\mu,f_{r,i},1\leq i\leq N_s}{\text{minimize}}\quad&\sum_{i=1}^{N_s}\varphi_i(f_{r,i},\mu)+\mu\epsilon_{rd}^2\\\text{subject to}\quad&\mu>(f_{r,i}g_i^{'})^2+\sigma_r^2(f_{r,i}g_i)^2,\ 1\leq i\leq N_s\\&\mu\geq(f_{r,i}\gamma_{sr,i}g_j)^2+\sigma_r^2(f_{r,i}g_j)^2,\ i\neq j\\&\sum_{i=1}^{N_s}f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2)\leq P_r\end{aligned}\tag{55}$$

where $\varphi_i(f_{r,i},\mu)=\frac{\mu((g_i^{'}\gamma_{rd,i}f_{r,i}-1)^2+\sigma_r^2g_i^2\gamma_{rd,i}^2f_{r,i}^2)-\sigma_r^2g_i^2f_{r,i}^2}{\mu-(g_i^{'})^2f_{r,i}^2-\sigma_r^2g_i^2f_{r,i}^2}$. Without loss of generality, we assume that $\mu=(f_{r,k}g_k^{'})^2+\sigma_r^2(f_{r,k}g_k)^2$ for a certain k. Then, based on (54), we have $(g_k^{'}\gamma_{rd,k}f_{r,k}-1)g_k^{'}f_{r,k}+\sigma_r^2g_k^2\gamma_{rd,k}f_{r,k}^2=0$ and $t-\mu\epsilon_{rd}^2-\sum_{i\neq k}^{N_s}\varphi_i(f_{r,i},\mu)-\frac{\sigma_r^2}{\sigma_r^2+\gamma_{sr,k}^2}\geq 0$. Now (39) becomes

$$\begin{aligned}\underset{\mu,f_{r,i},1\leq i\leq N_s}{\text{minimize}}\quad&\sum_{i=1,i\neq k}^{N_s}\varphi_i(f_{r,i},\mu)+\frac{\sigma_r^2}{\sigma_r^2+\gamma_{sr,k}^2}+\mu\epsilon_{rd}^2\\\text{subject to}\quad&\mu\geq(f_{r,i}\gamma_{sr,i}g_j)^2+\sigma_r^2(f_{r,i}g_j)^2,\ \forall i,j\\&\sum_{i=1}^{N_s}f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2)\leq P_r.\end{aligned}\tag{56}$$

Meanwhile, as $\mu\to(f_{r,k}g_k^{'})^2+\sigma_r^2(f_{r,k}g_k)^2$, it follows from the second equation of (54) that $f_{r,k}\to\frac{\gamma_{sr,k}}{g_k\gamma_{rd,k}(\gamma_{sr,k}^2+\sigma_r^2)}$. Therefore, the nominator and denominator of $\varphi_k(f_{r,k},\mu)$ tend to $\frac{\sigma_r^2}{\sigma_r^2+\gamma_{sr,k}^2}\left(\mu-\frac{\gamma_{sr,k}^2}{\gamma_{rd,k}^2(\gamma_{sr,k}^2+\sigma_r^2)}\right)$ and $\mu-\frac{\gamma_{sr,k}^2}{\gamma_{rd,k}^2(\gamma_{sr,k}^2+\sigma_r^2)}$, respectively, and the limit of $\varphi_k(f_{r,k},\mu)$ on the boundary of $\mu=(f_{r,k}g_k^{'})^2+\sigma_r^2(f_{r,k}g_k)^2$ is $\frac{\sigma_r^2}{\sigma_r^2+\gamma_{sr,k}^2}$. Thus, by using the limit on the boundary, we can extend (56) to

$$\begin{aligned}\underset{\mu,f_{r,i},1\leq i\leq N_s}{\text{minimize}}\quad&\sum_{i=1}^{N_s}\varphi_i(f_{r,i},\mu)+\mu\epsilon_{rd}^2\\\text{subject to}\quad&\mu\geq(f_{r,i}\gamma_{sr,i}g_j)^2+\sigma_r^2(f_{r,i}g_j)^2,\ \forall i,j\\&\sum_{i=1}^{N_s}f_{r,i}^2(\sigma_r^2+\gamma_{sr,i}^2)\leq P_r\end{aligned}\tag{57}$$

without losing any optimality.

The convexity of the objective function can be proved by constructing the following function:

$$\begin{aligned}\phi(\delta_i,f_{r,i},\mu)=&[(\gamma_{rd,i}+\delta_i)g_i^{'}f_{r,i}-1]^2+\sigma_r^2(\gamma_{rd,i}+\delta_i)^2g_i^2f_{r,i}^2\\&-\mu\delta_i^2.\end{aligned}\tag{58}$$

It can be readily shown that $\phi(\delta_i,f_{r,i},\mu)$ is convex with fixed δ_i. As $\partial^2\phi(\delta_i,f_{r,i},\mu)/\partial\delta_i^2=2((f_{r,i}g_i^{'})^2+\sigma_r^2(f_{r,i}g_i)^2-\mu)$, $\phi(\delta_i,f_{r,i},\mu)$ is concave in δ_i with fixed $(f_{r,i},\mu)$ when $\mu\geq(f_{r,i}g_i^{'})^2+\sigma_r^2(f_{r,i}g_i)^2$. With some involved calculations, we find that the maximum value of $\phi(\delta_i,f_{r,i},\mu)$ with respect to δ_i for fixed $(f_{r,i},\mu)$ equals to $\varphi_i(f_{r,i},\mu)$. As the convexity is preserved under the maximization operation [38], we conclude that $\varphi_i(f_{r,i},\mu)$ is convex. Thereby, the scalar problem (11) is convex. The second part of *Theorem* 3, involving the scalar optimization of $\mathbf{G}$, can be proved by using similar approaches. Due to the limited space, we omit the proof here.

APPENDIX IV
PROOF OF PROPOSITION 1

A subgradient of $\psi_i^*(\mu)$ with respect μ is given by [34]

$$\begin{aligned}s_{\psi_i^*}(\mu)&=\frac{\partial\psi_i(g_i^*,\mu)}{\partial\mu}+\alpha_i^*\frac{\partial((g_i^*)^2-\mu/\tilde{f}_m^2)}{\partial\mu}\\&=\frac{\partial\psi_i(g_i^*,\mu)}{\partial\mu}-\frac{\alpha_i^*}{\tilde{f}_m^2}\end{aligned}\tag{59}$$

where α_i^* is the optimal dual variable associated with the constraint $(g_i^*)^2\leq\mu/\tilde{f}_m^2$ and satisfies the Karush–Kuhn–Tucker (KKT) conditions:

$$\alpha_i^*\geq 0,\alpha_i^*((g_i^*)^2-\mu/\tilde{f}_m^2)=0\tag{60}$$

$$\partial\psi_i(g_i,\mu)/\partial g_i|_{g_i=g_i^*}+2\alpha_i^*g_i^*=0.\tag{61}$$

$$\frac{\partial\psi_i(g_i,\mu)}{\partial g_i}=\begin{cases}\frac{2\mu[g_i((f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2)\gamma_{rd,i}-f_{r,i}^{'}](\mu\gamma_{rd,i}-f_{r,i}^{'}g_i)}{(\mu-(f_{r,i}^{'})^2g_i^2-\sigma_r^2f_{r,i}^2g_i^2)^2}+2\sigma_d^2g_i, & g_i^*\neq\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\\-2\frac{\gamma_{rd,i}\gamma_{sr,i}^3f_{r,i}}{(\gamma_{sr,i}(1+g_i\gamma_{rd,i}f_{r,i}^{'})+\sigma_r^2g_i\gamma_{rd,i}f_{r,i})^2}+2\sigma_d^2g_i, & g_i^*=\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\end{cases}\tag{62}$$

$$\alpha_i^*=\begin{cases}0, & g_i^*\neq\frac{\sqrt{\mu}}{\tilde{f}_m}\\-\frac{\mu[g_i^*((f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2)\gamma_{rd,i}-f_{r,i}^{'}](\mu\gamma_{rd,i}-f_{r,i}^{'}g_i^*)}{g_i^*(\mu-(f_{r,i}^{'})^2(g_i^*)^2-\sigma_r^2f_{r,i}^2(g_i^*)^2)^2}-\sigma_d^2, & g_i^*=\frac{\sqrt{\mu}}{\tilde{f}_m}\neq\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\\\frac{\gamma_{rd,i}^2(\sigma_r^2f_{r,i}^2+(f_{r,i}^{'})^2)}{4}-\sigma_d^2, & g_i^*=\frac{\sqrt{\mu}}{\tilde{f}_m}=\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\end{cases}\tag{63}$$

$$\frac{\partial \psi_i(g_i^*,\mu)}{\partial \mu}=\begin{cases}\frac{-[(g_i^*\gamma_{rd,i}f_{r,i}^{'}-1)^2+\sigma_r^2(g_i^*)^2\gamma_{rd,i}^2f_{r,i}^2]((f_{r,i}^{'})^2(g_i^*)^2+\sigma_r^2f_{r,i}^2(g_i^*)^2)+\sigma_r^2f_{r,i}^2(g_i^*)^2}{(\mu-(f_{r,i}^{'})^2(g_i^*)^2-\sigma_r^2f_{r,i}^2(g_i^*)^2)^2}, & g_i^*\neq\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\\ 0, & g_i^*=\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}\end{cases} \tag{64}$$

$$\frac{\partial c_i(f_{r,i},\mu)}{\partial f_{r,i}}=\begin{cases}-2\frac{\mu[f_{r,i}((g_i^{'})^2+\sigma_r^2g_i^2)\gamma_{rd,i}-g_i^{'}](g_i^{'}f_{r,i}-\mu\gamma_{rd,i})}{(\mu-(g_i^{'})^2f_{r,i}^2-\sigma_r^2f_{r,i}^2g_i^2)^2}, & f_{r,i}^*\neq\sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2g_i^2}}\\ -2\frac{\gamma_{rd,i}\gamma_{sr,i}^3g_i}{(\gamma_{sr,i}(1+g_i^{'}\gamma_{rd,i}f_{r,i})+\sigma_r^2g_i\gamma_{rd,i}f_{r,i})^2}, & f_{r,i}^*=\sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2g_i^2}}\end{cases} \tag{69}$$

$$\lambda^*=\begin{cases}0, & f_{r,i}^*\neq\frac{\sqrt{\mu}}{\tilde{g}_{m,i}}\\ \frac{\mu[f_{r,i}^*((g_i^{'})^2+\sigma_r^2g_i^2)\gamma_{rd,i}-g_i^{'}](g_i^{'}f_{r,i}^*-\mu\gamma_{rd,i})}{f_{r,i}^*(\mu-(g_i^{'})^2(f_{r,i}^*)^2-\sigma_r^2(f_{r,i}^*)^2g_i^2)^2}, & f_{r,i}^*=\frac{\sqrt{\mu}}{\tilde{g}_{m,i}}\text{ and }f_{r,i}^*\neq\sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2g_i^2}}\\ \frac{\gamma_{rd,i}^2(\sigma_r^2g_i^2+(g_i^{'})^2)}{4}, & f_{r,i}^*=\frac{\sqrt{\mu}}{\tilde{g}_{m,i}}=\sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2g_i^2}}\end{cases} \tag{70}$$

Note that when the optimal $g_i^*=\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}$, we have $g_i^*=\sqrt{\frac{\mu}{(f_{r,i}^{'})^2+\sigma_r^2f_{r,i}^2}}=\frac{\gamma_{sr,i}}{f_{r,i}\gamma_{rd,i}(\gamma_{sr,i}^2+\sigma_r^2)}$ and $\psi_i(g_i,\mu)$ equals $\frac{\gamma_{sr,i}(1-g_i\gamma_{rd,i}f_{r,i}^{'})+\sigma_r^2g_i\gamma_{rd,i}f_{r,i}}{\gamma_{sr,i}(1+g_i\gamma_{rd,i}f_{r,i}^{'})+\sigma_r^2g_i\gamma_{rd,i}f_{r,i}}$. Hence, $\partial\psi_i(g_i,\mu)/\partial g_i$ is given by (62). Then, by using (60)–(62), we obtain α_i^* as (63). Since $\frac{\partial\psi_i(g_i^*,\mu)}{\partial\mu}$ can be calculated by (64), $s_{\psi_i^*}(\mu)$ is obtained immediately by substituting (63) and (64) into (59), which is readily shown to be the same as (16)-(19) in *Proposition* 1.

Appendix V
Proof of Proposition 2

It is easy to find that $c_i(f_{r,i})$ have two stationary points: $f_{r,i}=g_i^{'}/(((g_i^{'})^2+\sigma_r^2g_i^2)\gamma_{rd,i})$ and $f_{r,i}=\mu\gamma_{rd,i}/g_i^{'}$. Recall that a feasible $f_{r,i}$ should satisfy the constraints $f_{r,i}\leq\sqrt{\mu}/\tilde{g}_{m,i}$ and $f_{r,i}\leq\sqrt{p_i/(\sigma_r^2+\gamma_{sr,i}^2)}$, or equivalently, $f_{r,i}\leq\min\{\sqrt{\mu}/\tilde{g}_{m,i},\sqrt{p_i/(\sigma_r^2+\gamma_{sr,i}^2)}\}\triangleq\tau$. Thus, the optimal solution to the problem (21) must belong to the set $\{g_i^{'}/(((g_i^{'})^2+\sigma_r^2g_i^2)\gamma_{rd,i}),\mu\gamma_{rd,i}/g_i^{'},\tau\}$. Denote the minimum stationary point with $f_{r,i}^{(1)}$ and the maximum one with $f_{r,i}^{(2)}$, then from (69), we have $c_i^{'}(f_{r,i})>0$ when $f_{r,i}^{(1)}<f_{r,i}<f_{r,i}^{(2)}$, and $c_i^{'}(f_{r,i})<0$ if $f_{r,i}<f_{r,i}^{(1)}$ or $f_{r,i}>f_{r,i}^{(2)}$. Hence, if $\tau<f_{r,i}^{(1)}$, the optimal solution to the problem (21) should be τ. Otherwise, the optimal solution is $f_{r,i}^{(1)}$. This is exactly the same as (25) in *Proposition* 2.

A subgradient of $c_i^*(\mu)$ with respect to μ can be computed by

$$s_{c_i^*}(\mu)=\frac{\partial c_i(f_{r,i}^*,\mu)}{\partial\mu}+\lambda^*\frac{\partial((f_{r,i}^*)^2-\mu/\tilde{g}_{m,i}^2)}{\partial\mu} \tag{65}$$

where λ^* is the optimal Lagrange multiplier associated with the constraint $(f_{r,i}^*)^2\leq\mu/\tilde{g}_{m,i}^2$ and can be obtained by the KKT conditions as follows:

$$\lambda^*\geq 0,\lambda^*((f_{r,i}^*)^2-\mu/\tilde{g}_{m,i}^2)=0 \tag{66}$$

$$\nu^*\geq 0,\nu^*((f_{r,i}^*)^2-p_i/(\sigma_r^2+\gamma_{sr,i}^2))=0 \tag{67}$$

$$\partial c_i(f_{r,i},\mu)/\partial f_{r,i}|_{f_{r,i}=f_{r,i}^*}+2(\lambda^*+\nu^*)f_{r,i}^*=0 \tag{68}$$

where ν^* is the optimal Lagrange multiplier associated with the constraint $(f_{r,i}^*)^2\leq p_i/(\sigma_r^2+\gamma_{sr,i}^2)$. With similar techniques used in (62), we obtain (69). Then, using (66)-(69), we can calculate λ^* as (70). In addition, $\frac{\partial c_i(f_{r,i}^*,\mu)}{\partial\mu}$ can be obtained by (71). By substituting (70) and (71) into (65), we obtain (27)-(30).

To find a subgradient of $d^*(\mathbf{p})$, one requires the optimal ν^*, which can be computed by (72). Then, $(\mathbf{s}_{d^*}(\mathbf{p}))_i=-\nu^*/(\sigma_r^2+\gamma_{sr,i}^2)$ which is identical to (31)-(33).

Appendix VI
Computing Worst-case Channels for A Given Transceiver

By utilizing some matrix manipulations and noticing that optimal $\mathbf{\Delta}_{rd}$ is achieved on the boundary, the problem (34) is converted to the following trust region subproblem [39]:

$$\underset{\|\boldsymbol{\delta}_{rd}\|=\epsilon_{rd}}{\text{minimize}}\quad\boldsymbol{\delta}_{rd}^H(-\mathbf{R}^T\otimes\mathbf{S})\boldsymbol{\delta}_{rd}-2\Re\{\mathbf{d}^H\boldsymbol{\delta}_{rd}\} \tag{73}$$

where $\boldsymbol{\delta}_{rd}=\text{vec}(\mathbf{\Delta}_{rd})$, $\mathbf{R}=\mathbf{F}_r\mathbf{H}_{sr}\mathbf{H}_{sr}^H\mathbf{F}_r^H+\sigma_{n_r}^2\mathbf{F}_r\mathbf{F}_r^H$, $\mathbf{S}=\mathbf{G}^H\mathbf{G}$ and $\mathbf{d}=\text{vec}(\mathbf{G}^H(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\mathbf{H}_{sr}-\mathbf{I})\mathbf{H}_{sr}^H\mathbf{F}_r^H+\sigma_{n_r}^2\mathbf{G}^H\mathbf{G}\mathbf{H}_{rd}\mathbf{F}_r\mathbf{F}_r^H)$. It was proved in [39] that $\boldsymbol{\delta}_{rd}$ is a global optimal solution of the problem (34) if and only if the following conditions are satisfied:

$$(-\mathbf{R}^T\otimes\mathbf{S}+\omega^*\mathbf{I})\boldsymbol{\delta}_{rd}=\mathbf{d},-\mathbf{R}^T\otimes\mathbf{S}+\omega^*\mathbf{I}\succeq 0,\|\boldsymbol{\delta}_{rd}\|=\epsilon_{rd} \tag{74}$$

where ω^* is the optimal Lagrange multiplier associated with the constraint $\|\boldsymbol{\delta}_{rd}\|=\epsilon_{rd}$. Note that the dual problem of (34)

$$\frac{\partial c_i(f_{r,i}^*, \mu)}{\partial \mu} = \begin{cases} \frac{-[(f_{r,i}^*\gamma_{rd,i}g_i^{'}-1)^2+\sigma_r^2(f_{r,i}^*)^2\gamma_{rd,i}^2 g_i^2]((g_i^{'})^2(f_{r,i}^*)^2+\sigma_r^2 g_i^2(f_{r,i}^*)^2)+\sigma_r^2 g_i^2(f_{r,i}^*)^2}{(\mu-(g_i^{'})^2(f_{r,i}^*)^2-\sigma_r^2 g_i^2(f_{r,i}^*)^2)^2}, & f_{r,i}^* \neq \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}} \\ 0, & f_{r,i}^* = \sqrt{\frac{\mu}{(g_i^{'})^2+\sigma_r^2 g_i^2}} \end{cases} \tag{71}$$

$$\nu^* = \begin{cases} 0, & f_{r,i}^* \neq \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}} \\ \frac{\mu^*[f_{r,i}^*((g_i^{'})^2+\sigma_r^2 g_i^2)\gamma_{rd,i}-g_i^{'}](g_i^{'} f_{r,i}^*-\mu^*\gamma_{rd,i})}{f_{r,i}^*(\mu^*-(g_i^{'})^2(f_{r,i}^*)^2-\sigma_r^2(f_{r,i}^*)^2 g_i^2)^2}, & f_{r,i}^* = \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}} \text{ and } f_{r,i}^* \neq \sqrt{\frac{\mu^*}{(g_i^{'})^2+\sigma_r^2 g_i^2}} \\ \frac{\gamma_{rd,i}^2(\sigma_r^2 g_i^2+(g_i^{'})^2)}{4}, & f_{r,i}^* = \sqrt{\frac{p_i}{\sigma_r^2+\gamma_{sr,i}^2}} = \sqrt{\frac{\mu^*}{(g_i^{'})^2+\sigma_r^2 g_i^2}} \end{cases} \tag{72}$$

can be formulated as the following SDP form:

$$\begin{aligned} &\underset{\omega,t}{\text{maxmize}} \quad t \\ &\text{subject to} \quad \begin{bmatrix} -\mathbf{R}^T \otimes \mathbf{S} + \omega \mathbf{I} & \mathbf{d} \\ \mathbf{d}^H & -\omega\epsilon_{rd}^2 - t \end{bmatrix} \succeq 0. \end{aligned} \tag{75}$$

Hence, we can obtain the optimal ω^* with numerical tools such as SeDuMi. After obtaining ω^*, all we need to do is to solve (74) to achieve $\boldsymbol{\delta}_{rd}$. It is evident that if $-\mathbf{R}^T \otimes \mathbf{S} + \omega^* \mathbf{I} \succ 0$, $\boldsymbol{\delta}_{rd}$ has a unique solution $(-\mathbf{R}^T \otimes \mathbf{S} + \omega^* \mathbf{I})^{-1}\mathbf{d}$. On the other hand, when $-\mathbf{R}^T \otimes \mathbf{S} + \omega^* \mathbf{I} \succeq 0$, the optimal $\boldsymbol{\delta}_{rd}$ is not unique, and it can be expressed by $\boldsymbol{\delta}_{rd} = (-\mathbf{R}^T \otimes \mathbf{S} + \omega^* \mathbf{I})^{\dagger}\mathbf{d} + \beta\mathbf{f}$, where $\mathbf{f}$ is an arbitrary vector chosen from the right null space of $-\mathbf{R}^T \otimes \mathbf{S} + \omega^* \mathbf{I}$, and $\beta \in \mathbb{R}$ is selected such that the constraint $\|\boldsymbol{\delta}_{rd}\| = \epsilon_{rd}$ is met.

Acknowledgement

The authors would like to thank Prof. Zhi Ding and Prof. Bernard C. Levy for helpful discussions. The authors also thank the editor and anonymous reviewers for their valuable comments and suggestions, which have greatly improved the quality of this paper.

References

[1] 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, no. 12, pp. 3062–3080, Dec. 2004.

[2] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[3] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[4] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

[5] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

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

[7] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 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, pp. 5062–5073, Oct. 2008.

[9] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Oct. 2008.

[10] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[11] W. Xu, X. Dong, and W.-S. Lu, "Joint precoding optimization for multiuser multi-antenna relaying downlinks using quadratic programming," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1228–1235, May 2011.

[12] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[13] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 662–671, Mar. 2006.

[14] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[15] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[16] R. Mo, J. Lin, Y. H. Chew, and W. H. Chin, "Relay precoder design for non-regenerative MIMO relay networks with imperfect channel state information," in *Proc. ICC 2010*, Cape Town, South Africa, May 2010, pp. 1–5.

[17] C. Jeong, H.-M. Kim, H.-K. Song, and I.-M. Kim, "Relay precoding for non-regenerative MIMO relay systems with partial CSI in the presence of interferers," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1521–1531, Apr. 2012.

[18] 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, no. 5, pp. 1698–1711, May 2012.

[19] C. Xing, S. Ma, Z. Fei, Y.-C. Wu, and H. V. Poor, "A general robust linear transceiver design for multi-hop amplify-and-forward MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.

[20] W. Xu and X. Dong, "Optimized one-way relaying strategy with outdated CSI quantization for spatial multiplexing," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4458–4464, Aug. 2012.

[21] H. Shen, W. Xu, and C. Zhao, "Efficient joint transmit and receive optimization for multiuser MIMO systems," in *Proc. IEEE WCNC*, Paris, France, Apr. 2012, pp. 125–130.

[22] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.

[23] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known MIMO communications channels," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2918–2930, Aug. 2005.

[24] J. Wang and M. Bengtsson, "Joint optimization of the worst-case robust MMSE MIMO transceiver," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 295–298, May 2011.

[25] G. Zheng, K.-K. Wong, A. Paulraj, and B. Ottersten, "Robust collaborative-relay beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3130–3143, Aug. 2009.

[26] B. K. Chalise and L. Vandendorpe, "MIMO relay design for multipoint-to-multipoint communications with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2785–2796, Jul. 2009.

[27] B. K. Chalise and L. Vandendorpe, "Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6355–6368, Dec. 2010.

[28] D. Ponukumati, F. Gao, and L. Fan, "Robust general rank precoding design for amplify-and-forward relay network," in *Proc. Globecom 2010*, Miami, USA, Dec. 2010, pp. 1–5.

[29] P. Ubaidulla and A. Chockalingam, "Relay precoder optimization in MIMO relay networks with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5473–5484, Nov. 2011.

[30] M. Tao and R. Wang, "Robust relay beamforming for two-way relay networks," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1052–1055, Jul. 2012.

[31] Q. Sun and L. Li, "Robust transceiver optimization for non-regenerative MIMO relay systems," in *Proc. IEEE Wireless Personal Multimedia Commun. (WPMC2012)*, Taipei, Taiwan, Sep. 24–27, 2012, pp. 483–487.

[32] Y. Liu and A. Petropulu, "QoS guarantees in AF relay networks with multiple source-destination pairs in the presence of imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4225–4235, Sep. 2013.

[33] Z. Wang, W. Chen, F. Gao, and J. Li, "Capacity performance of relay beamformings for MIMO multirelay networks with imperfect R-D CSI at relays," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2608–2619, Jul. 2011.

[34] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[35] D. P. Palomar, M. Bengtsson, and B. Ottersten, "Minimum BER linear transceivers for MIMO channels via primal decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2866–2882, Aug. 2005.

[36] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[37] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

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

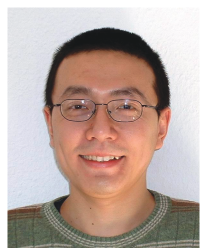
[39] D. M. Gay, "Computing optimal locally constrained steps," *SIAM J. Sci. Stat. Comput.*, vol. 2, no. 2, pp. 186–197, Jun. 1981.

Hong Shen (S'11) received the B.E. and M.S. degrees in information engineering from the Southeast University, Nanjing, China, in 2009 and 2011, respectively. He is currently working towards the Ph.D. degree in information and communication engineering with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. During his doctoral studies, he has conducted cooperative research with the Department of Electrical and Computer Engineering, University of Davis, CA, USA. His research interests include multiple-input-multiple-output systems, optimization in wireless communications, and source localization in wireless sensor networks.

Jiaheng Wang (S'08–M'10) received the Ph.D. degree in electrical engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2010. He received the B.E. and M.S. degrees from the Southeast University, Nanjing, China, in 2001 and 2006, respectively. From 2010 to 2011, he was with the Signal Processing Laboratory, ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently an Associate Professor at the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests mainly include optimization in signal processing, wireless communications and networks.

Wei Xu (S'07–M'09) received his B.Sc. degree in Electrical Engineering in 2003 and his M.S. and Ph.D. degrees in Communication and Information Engineering in 2006 and 2009, respectively, all from Southeast University, Nanjing, China. He is currently an Associate Professor at the National Mobile Communications Research Lab (NCRL), Southeast University. Between 2009 and 2010, he was a post-doctoral research fellow with the Department of Electrical and Computer Engineering, University of Victoria, Canada.

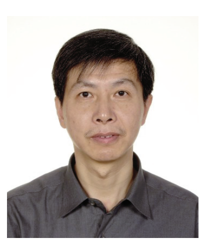
Dr. Xu is an Editor of the IEEE COMMUNICATIONS LETTERS. He has been involved in Technical Program Committees for many international conferences including *IEEE Global Communications Conference (Globecom)*, *IEEE Wireless Communications and Networking Conference (WCNC)*, *IEEE Vehicular Technology Conference (VTC)*, etc. He has coauthored more than 60 papers and had over 10 patents pending. He was a co-recipient of the best paper award of the *IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies for Wireless Communications (MAPE)*. He was elected Core Team Member of the Jiangsu Innovation Team in 2012. His research interests include communication theory, cooperative communications, and signal processing and optimization theory for wireless communications.



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.

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 Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, 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.



Chunming Zhao (M'93) received the B.S. and M.S. degrees from Nanjing Institute of Posts and Telecommunications, in 1982 and 1984, respectively. In 1993, he received his Ph.D. degree from the Department of Electrical and Electronic Engineering, University of Kaiserslautern, Germany. He has been a Postdoctoral Researcher at National Mobile Communications Research Lab., Southeast University, China, where he is currently a professor and vice director of the Lab. He has managed several key projects of Chinese Communications High Tech. Program and was awarded as "excellent researcher" from Ministry of Science and Technology, China. He also won the First Prize of National Technique Invention of China in 2011. His research interests include communication theory, coding/decoding, mobile communications, and VLSI design.