



Title	Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems
Author(s)	Xing, C; Ma, S; Wu, YC
Citation	IEEE Transactions On Signal Processing, 2010, v. 58 n. 4, p. 2273-2283
Issued Date	2010
URL	http://hdl.handle.net/10722/124720
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Robust Joint Design of Linear Relay Precoder and Destination Equalizer for Dual-Hop Amplify-and-Forward MIMO Relay Systems

Chengwen Xing, Shaodan Ma, and Yik-Chung Wu

Abstract—This paper addresses the problem of robust linear relay precoder and destination equalizer design for a dual-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay system, with Gaussian random channel uncertainties in both hops. By taking the channel uncertainties into account, two robust design algorithms are proposed to minimize the mean-square error (MSE) of the output signal at the destination. One is an iterative algorithm with its convergence proved analytically. The other is an approximated closed-form solution with much lower complexity than the iterative algorithm. Although the closed-form solution involves a minor relaxation for the general case, when the column covariance matrix of the channel estimation error at the second hop is proportional to identity matrix, no relaxation is needed and the proposed closed-form solution is the optimal solution. Simulation results show that the proposed algorithms reduce the sensitivity of the AF MIMO relay systems to channel estimation errors, and perform better than the algorithm using estimated channels only. Furthermore, the closed-form solution provides a comparable performance to that of the iterative algorithm.

Index Terms—Amplify-and-forward (AF), equalizer, minimum mean-square-error (MMSE), multiple-input multiple-output (MIMO), precoder, relay.

I. INTRODUCTION

RECENTLY, cooperative communication has gained significant research interest, due to its great potential to improve reliability, coverage and capacity of wireless links [1], [2]. Generally speaking, there are three kinds of cooperative protocols: amplify-and-forward (AF), compress-and-forward (CF), and decode-and-forward (DF). Among these three protocols, AF is a relatively simple one, in which the relay just scales the signal received from the source, and then transmits it to the destination. Because of its simplicity and low implementation complexity, AF strategy has attracted a lot of researchers' attention [3]–[6].

On the other hand, it is well known that in fully scattered environment, multiantenna systems provide substantial spatial diversity and multiplexing gains [7], [8]. In order to minimize the data estimation error from the received signal, the mean-square error (MSE) is a very important metric for transceiver design [9]–[12]. Based on implementation consideration,

linear minimum mean-square-error (LMMSE) transceiver is more preferable compared to its nonlinear counterpart which may have prohibitive complexity. Therefore, linear transceiver design for point-to-point MIMO systems has been extensively studied in various scenarios in the past decade [9]–[13].

The benefits of multiple-antenna systems can be directly introduced into cooperative communications via deployment of multiple antennas at transmitters and receivers, and such a system receives a lot of attention lately. In particular, capacity bounds for AF MIMO relay systems have been derived in [14] and [15], and optimal power allocation schemes which maximize the capacity of AF MIMO relay systems have been studied in [5], [6], and [16]–[18]. Furthermore, for practical applications (such as Winner project, LTE and IMT-Advanced [19], [20]), fixed relay stations with multiple antennas are being considered to be installed at the border of base stations' coverage area, to enhance the coverage of base stations.

In terms of transceiver design,¹ for dual-hop AF MIMO relay systems with single relay, the optimal closed-form solution for joint LMMSE transceiver design assuming perfect channel state information (CSI) has been proposed in [21]. It has also been shown in [21] that the joint design has a better performance than the various separate design schemes. In [22], the joint LMMSE transceiver design for the multiple-relay case has been discussed. This case is more general and difficult to deal with. After a relaxation on the constraint of the transmit power at the relay, a suboptimal closed-form solution has been derived in [22], and is shown to have a much better performance than the zero-forcing scheme.

Notice that the existing algorithms on LMMSE transceiver design for AF MIMO relay systems require the CSI being perfectly known. Unfortunately, in practice, CSI is generally obtained through estimation and perfect estimation is very difficult to achieve. Due to limited length of training sequences and/or time-varying nature of wireless channels, channel estimation errors inevitably exist, causing substantial system performance degradation. Robust transceiver design, which could mitigate such performance degradation by taking the channel estimation errors into account, is therefore of great importance and highly desirable for practical applications.

When channel uncertainties are considered, both min-max and stochastic (including both probability-based and Bayesian) approaches can be employed. If quality-of-service (QoS) requirement is considered (e.g., outage probability level [23], [24]), min-max or probability-based approach is preferred. On

¹In this paper, we use the term transceiver to collectively denote precoder at the relay and equalizer at destination.

Manuscript received June 02, 2009; accepted November 25, 2009. First published December 18, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Geert Leus.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: cxwing@eee.hku.hk; sdma@eee.hku.hk; ycwu@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2009.2038961

the other hand, if the goal is to minimize an average objective function over channel uncertainties, e.g., the total MSE of multiple data streams, Bayesian approach is more suitable.

In this paper, we consider an AF MIMO relay system with single relay and address the problem of robust relay precoder and destination equalizer design under imperfect CSI at both the relay and destination. With the channel estimation errors being modeled as Gaussian random variables, we employ the Bayesian approach, and robustness is incorporated into the optimization objective function in the form of MSE averaged over the channel estimation errors. Two robust design algorithms are proposed. The first one is an iterative algorithm with its convergence proved analytically. The other algorithm offers a closed-form solution with a lower complexity than the iterative algorithm. Although the closed-form solution involves a minor relaxation for the general case, when the column covariance matrix of the channel estimation error in the second hop is proportional to identity matrix, no relaxation is needed and the proposed closed-form solution is the optimal solution. Simulation results show that the two proposed robust transceivers perform better than the transceiver without taking channel estimation errors into account. Furthermore, the proposed closed-form solution has a comparable performance to the proposed iterative algorithm, but with a lower complexity.

This paper is organized as follows. System model is presented in Section II. The optimization problem for minimizing the total MSE is formulated in Section III. In Section IV, an iterative algorithm is proposed to solve the optimization problem, while an approximated closed-form solution is given in Section V. In Section VI, the iterative and closed-form solutions are further generalized to the weighted MSE criterion. Finally, simulation results are given in Section VII, and conclusions are drawn in Section VIII.

The following notations are used throughout this paper. Boldface lowercase letters denote vectors, while boldface uppercase letters denote matrices. The notations \mathbf{Z}^T , \mathbf{Z}^H , and \mathbf{Z}^* denote the transpose, Hermitian, and conjugate of the matrix \mathbf{Z} , respectively, and $\text{Tr}(\mathbf{Z})$ is the trace of the matrix \mathbf{Z} . The symbol \mathbf{I}_M denotes the $M \times M$ identity matrix, while $\mathbf{0}_{M \times N}$ denotes the $M \times N$ all zero matrix. The notation $\mathbf{Z}^{1/2}$ is the Hermitian square root of the positive semidefinite matrix \mathbf{Z} , such that $\mathbf{Z} = \mathbf{Z}^{1/2} \mathbf{Z}^{1/2}$ and $\mathbf{Z}^{1/2}$ is a Hermitian matrix. The symbol $\mathbb{E}\{\cdot\}$ represents the expectation operation. The operation $\text{vec}(\mathbf{Z})$ stacks the columns of the matrix \mathbf{Z} into a single vector and $\det(\mathbf{Z})$ denotes the determinant of \mathbf{Z} . The symbol \otimes represents the Kronecker product. The equation $\mathbf{A} \succeq \mathbf{B}$ represents that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix, while $\mathbf{A} \succ \mathbf{B}$ represents that $\mathbf{A} - \mathbf{B}$ is a positive-definite matrix.

II. SYSTEM MODEL

In this paper, a dual-hop AF cooperative communication system is considered. In the considered system, there is one source with N_S antennas, one relay with M_R receive antennas and N_R transmit antennas, and one destination with M_D antennas, as shown in Fig. 1. At the first hop, the source transmits data to the relay. The received signal \mathbf{x} at the relay is

$$\mathbf{x} = \mathbf{H}_{sr} \mathbf{s} + \mathbf{n}_1 \quad (1)$$

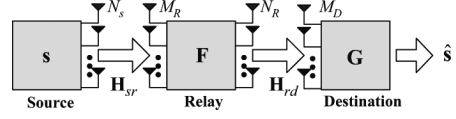


Fig. 1. Amplify-and-forward MIMO relay.

where \mathbf{s} is the $N_S \times 1$ data vector transmitted by the source with the covariance matrix $\mathbf{R}_s = \mathbb{E}\{\mathbf{s}\mathbf{s}^H\}$. The matrix \mathbf{H}_{sr} is the MIMO channel matrix between the source and the relay. Symbol \mathbf{n}_1 is the additive Gaussian noise with covariance matrix \mathbf{R}_{n_1} . At the relay, the received signal \mathbf{x} is multiplied by a precoder matrix \mathbf{F} , under a power constraint $\text{Tr}(\mathbf{F}\mathbf{R}_x\mathbf{F}^H) \leq P_r$, where $\mathbf{R}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ and P_r is the maximum transmit power. Then the resulting signal is transmitted to the destination. The received signal at the destination, \mathbf{y} , can be written as

$$\mathbf{y} = \mathbf{H}_{rd} \mathbf{F} \mathbf{H}_{sr} \mathbf{s} + \mathbf{H}_{rd} \mathbf{F} \mathbf{n}_1 + \mathbf{n}_2 \quad (2)$$

where \mathbf{H}_{rd} is the MIMO channel matrix between the relay and the destination, and \mathbf{n}_2 is the additive Gaussian noise vector at the second hop with covariance matrix \mathbf{R}_{n_2} . In order to guarantee the transmitted data \mathbf{s} can be recovered at the destination, it is assumed that M_R , N_R , and M_D are greater than or equal to N_S [22].

It is assumed that both the relay and destination have the estimated CSI. When channel estimation errors are considered, we have

$$\begin{aligned} \mathbf{H}_{sr} &= \bar{\mathbf{H}}_{sr} + \Delta \mathbf{H}_{sr} \\ \mathbf{H}_{rd} &= \bar{\mathbf{H}}_{rd} + \Delta \mathbf{H}_{rd} \end{aligned} \quad (3)$$

where $\bar{\mathbf{H}}_{sr}$ and $\bar{\mathbf{H}}_{rd}$ are the estimated CSI, while $\Delta \mathbf{H}_{sr}$ and $\Delta \mathbf{H}_{rd}$ are the corresponding channel estimation errors whose elements are zero mean Gaussian random variables. In general, the $M_R \times N_S$ matrix $\Delta \mathbf{H}_{sr}$ can be written as $\Delta \mathbf{H}_{sr} = \boldsymbol{\Sigma}_{sr}^{1/2} \mathbf{H}_{W, sr} \boldsymbol{\Psi}_{sr}^{1/2}$, where the elements of the $M_R \times N_S$ matrix $\mathbf{H}_{W, sr}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The $M_R \times M_R$ matrix $\boldsymbol{\Sigma}_{sr}$ and $N_S \times N_S$ matrix $\boldsymbol{\Psi}_{sr}^T$ are the row and column covariance matrices of $\Delta \mathbf{H}_{sr}$, respectively [7]. It is easy to see that $\text{vec}(\Delta \mathbf{H}_{sr}^T) \sim \mathcal{CN}(\mathbf{0}_{M_R N_S \times 1}, \boldsymbol{\Sigma}_{sr} \otimes \boldsymbol{\Psi}_{sr}^T)$, where $\mathcal{CN}(\mathbf{m}, \mathbf{C})$ denotes a complex Gaussian random vector with mean \mathbf{m} and covariance \mathbf{C} . Furthermore, the matrix $\Delta \mathbf{H}_{sr}$ is said to have a matrix-variate complex Gaussian distribution, which can be written as [25]

$$\Delta \mathbf{H}_{sr} \sim \mathcal{CN}_{M_R, N_S}(\mathbf{0}_{M_R \times N_S}, \boldsymbol{\Sigma}_{sr} \otimes \boldsymbol{\Psi}_{sr}^T) \quad (4)$$

with the probability density function (p.d.f.) given by [26]

$$p(\Delta \mathbf{H}_{sr}) = \frac{\exp(-\text{Tr}(\Delta \mathbf{H}_{sr}^H \boldsymbol{\Sigma}_{sr}^{-1} \Delta \mathbf{H}_{sr} \boldsymbol{\Psi}_{sr}^{-1}))}{(\pi)^{N_S M_R} [\det(\boldsymbol{\Sigma}_{sr})]^{N_S} [\det(\boldsymbol{\Psi}_{sr})]^{M_R}} \quad (5)$$

Similarly, for the estimation error in the second hop, we have

$$\Delta \mathbf{H}_{rd} \sim \mathcal{CN}_{M_D, N_R}(\mathbf{0}_{M_D \times N_R}, \boldsymbol{\Sigma}_{rd} \otimes \boldsymbol{\Psi}_{rd}^T) \quad (6)$$

where the $M_D \times M_D$ matrix $\boldsymbol{\Sigma}_{rd}$ and $N_R \times N_R$ matrix $\boldsymbol{\Psi}_{rd}^T$ are the row and column covariance matrices of $\Delta \mathbf{H}_{rd}$, respec-

tively. It is assumed that \mathbf{H}_{sr} and \mathbf{H}_{rd} are estimated independently, so the channel estimation errors, $\Delta\mathbf{H}_{sr}$ and $\Delta\mathbf{H}_{rd}$, are independent.

Remark 1: In general, the expressions of Ψ_{sr} , Σ_{sr} , Ψ_{rd} , and Σ_{rd} depend on specific channel estimation algorithms. If the channel estimation algorithm proposed in [27] is used, we have $\Psi_{sr} = \mathbf{R}_{T,sr}$, $\Sigma_{sr} = \sigma_{e,sr}^2 \mathbf{R}_{R,sr}$, $\Psi_{rd} = \mathbf{R}_{T,rd}$, and $\Sigma_{rd} = \sigma_{e,rd}^2 \mathbf{R}_{R,rd}$. The matrices $\mathbf{R}_{T,sr}$ and $\mathbf{R}_{R,sr}$ are the transmit and receive antennas correlation matrices at the source and the relay, respectively, and $\sigma_{e,sr}^2$ is the source-relay channel estimation error variance. Similarly, $\mathbf{R}_{T,rd}$, $\mathbf{R}_{R,rd}$ and $\sigma_{e,rd}^2$ are defined for the channel between the relay and the destination. On the other hand, when the channels are estimated based on the algorithm proposed in [28], we have $\Psi_{sr} = \mathbf{R}_{T,sr}$, $\Sigma_{sr} = \sigma_{e,sr}^2 (\mathbf{I}_{M_R} + \sigma_{e,sr}^2 \mathbf{R}_{R,sr}^{-1})^{-1}$, $\Psi_{rd} = \mathbf{R}_{T,rd}$, and $\Sigma_{rd} = \sigma_{e,rd}^2 (\mathbf{I}_{M_D} + \sigma_{e,rd}^2 \mathbf{R}_{R,rd}^{-1})^{-1}$. Notice that for the channel estimation algorithms in [27] and [28], Ψ_{sr} , Σ_{sr} , Ψ_{rd} , and Σ_{rd} are functions of the second-order statistics of CSI, which can be considered to change very slowly and to be known in prior [3]. However, in the following, the proposed algorithms are developed without assuming any specific form of Ψ_{sr} , Σ_{sr} , Ψ_{rd} , and Σ_{rd} .

III. PROBLEM FORMULATION

At the destination, a linear equalizer \mathbf{G} is adopted to detect the transmitted data \mathbf{s} (see Fig. 1). The problem is how to design the linear precoder matrix \mathbf{F} at the relay and the linear equalizer \mathbf{G} at the destination to minimize the MSE of the received data at the destination:

$$\text{MSE}(\mathbf{F}, \mathbf{G}) = \mathbb{E}\{\text{Tr}((\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^H)\} \quad (7)$$

where the expectation is taken with respect to \mathbf{s} , $\Delta\mathbf{H}_{sr}$, $\Delta\mathbf{H}_{rd}$, \mathbf{n}_1 , and \mathbf{n}_2 .

Since \mathbf{s} , \mathbf{n}_1 , and \mathbf{n}_2 are independent, the MSE expression (7) can be written as

$$\begin{aligned} \text{MSE}(\mathbf{F}, \mathbf{G}) &= \mathbb{E}\{||(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr} - \mathbf{I}_{N_S})\mathbf{s} + \mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{n}_1 + \mathbf{G}\mathbf{n}_2||^2\} \\ &= \mathbb{E}_{\Delta\mathbf{H}_{sr}, \Delta\mathbf{H}_{rd}}\{\text{Tr}((\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr} - \mathbf{I}_{N_S}) \\ &\quad \mathbf{R}_s(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr} - \mathbf{I}_{N_S})^H)\} \\ &\quad + \mathbb{E}_{\Delta\mathbf{H}_{rd}}\{\text{Tr}((\mathbf{G}\mathbf{H}_{rd}\mathbf{F})\mathbf{R}_{n_1}(\mathbf{G}\mathbf{H}_{rd}\mathbf{F})^H)\} \\ &\quad + \text{Tr}(\mathbf{G}\mathbf{R}_{n_2}\mathbf{G}^H) \\ &= \mathbb{E}_{\Delta\mathbf{H}_{sr}, \Delta\mathbf{H}_{rd}}\{\text{Tr}((\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr})\mathbf{R}_s(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr})^H)\} \\ &\quad + \text{Tr}(\mathbf{G}\mathbb{E}_{\Delta\mathbf{H}_{rd}}\{\mathbf{H}_{rd}\mathbf{F}\mathbf{R}_{n_1}\mathbf{F}^H\mathbf{H}_{rd}^H\}\mathbf{G}^H) \\ &\quad - \text{Tr}(\mathbf{R}_s(\mathbf{G}\bar{\mathbf{H}}_{rd}\mathbf{F}\bar{\mathbf{H}}_{sr})^H) - \text{Tr}(\mathbf{G}\bar{\mathbf{H}}_{rd}\mathbf{F}\bar{\mathbf{H}}_{sr}\mathbf{R}_s) \\ &\quad + \text{Tr}(\mathbf{R}_s) + \text{Tr}(\mathbf{G}\mathbf{R}_{n_2}\mathbf{G}^H). \end{aligned} \quad (8)$$

Because $\Delta\mathbf{H}_{sr}$ and $\Delta\mathbf{H}_{rd}$ are independent, the first term of MSE is

$$\begin{aligned} &\mathbb{E}_{\Delta\mathbf{H}_{sr}, \Delta\mathbf{H}_{rd}}\{\text{Tr}((\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr})\mathbf{R}_s(\mathbf{G}\mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr})^H)\} \\ &= \text{Tr}(\mathbf{G}\mathbb{E}_{\Delta\mathbf{H}_{rd}}\{\mathbf{H}_{rd}\mathbf{F}\mathbb{E}_{\Delta\mathbf{H}_{sr}}\{\mathbf{H}_{sr}\mathbf{R}_s\mathbf{H}_{sr}^H\}\mathbf{F}^H\mathbf{H}_{rd}^H\}\mathbf{G}^H). \end{aligned} \quad (9)$$

For the inner expectation, due to the fact that the distribution of $\Delta\mathbf{H}_{sr}$ is matrix-variate complex Gaussian with zero mean, the following equation holds [25]:

$$\begin{aligned} &\mathbb{E}_{\Delta\mathbf{H}_{sr}}\{\mathbf{H}_{sr}\mathbf{R}_s\mathbf{H}_{sr}^H\} \\ &= \mathbb{E}_{\Delta\mathbf{H}_{sr}}\{(\bar{\mathbf{H}}_{sr} + \Delta\mathbf{H}_{sr})\mathbf{R}_s(\bar{\mathbf{H}}_{sr} + \Delta\mathbf{H}_{sr})^H\} \\ &= \text{Tr}(\mathbf{R}_s\Psi_{sr})\Sigma_{sr} + \bar{\mathbf{H}}_{sr}\mathbf{R}_s\bar{\mathbf{H}}_{sr}^H \\ &\triangleq \mathbf{\Pi}_0. \end{aligned} \quad (10)$$

Applying (10) and the corresponding result for $\Delta\mathbf{H}_{rd}$ to (9), the first term of MSE becomes

$$\begin{aligned} &\text{Tr}(\mathbf{G}\mathbb{E}_{\Delta\mathbf{H}_{rd}}\{\mathbf{H}_{rd}\mathbf{F}\mathbb{E}_{\Delta\mathbf{H}_{sr}}\{\mathbf{H}_{sr}\mathbf{R}_s\mathbf{H}_{sr}^H\}\mathbf{F}^H\mathbf{H}_{rd}^H\}\mathbf{G}^H) \\ &= \text{Tr}(\mathbf{G}(\text{Tr}(\mathbf{F}\mathbf{\Pi}_0\mathbf{F}^H\Psi_{rd})\Sigma_{rd} + \bar{\mathbf{H}}_{rd}\mathbf{F}\mathbf{\Pi}_0\mathbf{F}^H\bar{\mathbf{H}}_{rd}^H)\mathbf{G}^H). \end{aligned} \quad (11)$$

Similarly, the second term of MSE in (8) can be simplified as

$$\begin{aligned} &\text{Tr}(\mathbf{G}\mathbb{E}_{\Delta\mathbf{H}_{rd}}\{\mathbf{H}_{rd}\mathbf{F}\mathbf{R}_{n_1}\mathbf{F}^H\mathbf{H}_{rd}^H\}\mathbf{G}^H) \\ &= \text{Tr}(\mathbf{G}(\text{Tr}(\mathbf{F}\mathbf{R}_{n_1}\mathbf{F}^H\Psi_{rd})\Sigma_{rd} + \bar{\mathbf{H}}_{rd}\mathbf{F}\mathbf{R}_{n_1}\mathbf{F}^H\bar{\mathbf{H}}_{rd}^H)\mathbf{G}^H). \end{aligned} \quad (12)$$

Based on (11) and (12), the MSE (8) equals to

$$\begin{aligned} \text{MSE}(\mathbf{F}, \mathbf{G}) &= \text{Tr}(\mathbf{G}(\bar{\mathbf{H}}_{rd}\mathbf{F}\mathbf{R}_s\mathbf{F}^H\bar{\mathbf{H}}_{rd}^H + \mathbf{K})\mathbf{G}^H) \\ &\quad - \text{Tr}(\mathbf{R}_s\bar{\mathbf{H}}_{sr}^H\mathbf{F}^H\bar{\mathbf{H}}_{rd}^H\mathbf{G}^H) \\ &\quad - \text{Tr}(\mathbf{R}_s\mathbf{G}\bar{\mathbf{H}}_{rd}\mathbf{F}\bar{\mathbf{H}}_{sr}) + \text{Tr}(\mathbf{R}_s) \end{aligned} \quad (13)$$

where

$$\mathbf{R}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{\Pi}_0 + \mathbf{R}_{n_1} \quad (14)$$

$$\mathbf{K} = \text{Tr}(\mathbf{F}\mathbf{R}_x\mathbf{F}^H\Psi_{rd})\Sigma_{rd} + \mathbf{R}_{n_2}. \quad (15)$$

Notice that the matrix \mathbf{R}_x is the autocorrelation matrix of the receive signal \mathbf{x} at the relay.

Subject to the transmit power constraint at the relay, the joint design of equalizer at the destination and precoder at the relay can be expressed as the following optimization problem

$$\begin{aligned} &\min_{\mathbf{F}, \mathbf{G}} \text{MSE}(\mathbf{F}, \mathbf{G}) \\ &\text{s.t.} \quad \text{Tr}(\mathbf{F}\mathbf{R}_x\mathbf{F}^H) \leq P_r. \end{aligned} \quad (16)$$

If the channels are perfectly known without estimation errors, the problem (16) has been solved in [21]. However, perfect channel estimation is difficult to achieve in practice, due to limited training and time-varying nature of wireless channels. Channel estimation errors generally exist and the formulation (13) is a very complicated function of \mathbf{F} and \mathbf{G} , making the problem difficult to solve. In this following, we propose two algorithms. One is an iterative algorithm which solves (16) without any approximation. The other is an approximated closed-form solution, which can be shown to provide a performance close to the iterative algorithm, but with a much lower complexity.

IV. THE PROPOSED ITERATIVE ALGORITHM

In this section, we derive an iterative algorithm [29], [30] to solve for \mathbf{F} and \mathbf{G} . In the following, it is shown that given \mathbf{F} , there is a closed-form solution for \mathbf{G} , and vice versa. Therefore, the proposed algorithm computes \mathbf{F} and \mathbf{G} iteratively, starting with an initial value.

A. Updating \mathbf{G} Given \mathbf{F}

Suppose the solutions for \mathbf{F} and \mathbf{G} at the i th iteration are \mathbf{F}_i and \mathbf{G}_i , respectively. First we update \mathbf{G} for given $\mathbf{F} = \mathbf{F}_i$. As the constraint in (16) is independent of \mathbf{G} , at the $(i+1)$ th iteration, the optimal \mathbf{G}_{i+1} for given \mathbf{F}_i satisfies the following condition:

$$\frac{\partial \text{MSE}(\mathbf{F}_i, \mathbf{G}_{i+1})}{\partial \mathbf{G}_{i+1}^*} = 0 \quad (17)$$

based on which we have

$$\mathbf{G}_{i+1} = \mathbf{R}_s (\bar{\mathbf{H}}_{rd} \mathbf{F}_i \bar{\mathbf{H}}_{sr})^H (\bar{\mathbf{H}}_{rd} \mathbf{F}_i \mathbf{R}_x \mathbf{F}_i^H \bar{\mathbf{H}}_{rd}^H + \mathbf{K}_i)^{-1} \quad (18)$$

where $\mathbf{K}_i = \text{Tr}(\mathbf{F}_i \mathbf{R}_x \mathbf{F}_i^H \boldsymbol{\Psi}_{rd}) \boldsymbol{\Sigma}_{rd} + \mathbf{R}_{n_2}$.

B. Updating \mathbf{F} Given \mathbf{G}

When \mathbf{G}_{i+1} is obtained as (18), we can design \mathbf{F}_{i+1} based on KKT conditions, which are obtained by differentiating the Lagrangian function of (16), $\mathcal{L}(\mathbf{F}_{i+1}, \gamma_{i+1}) = \text{MSE}(\mathbf{F}_{i+1}, \mathbf{G}_{i+1}) + \gamma_{i+1} (\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) - P_r)$, with respect to \mathbf{F}_{i+1}^* . The KKT conditions can be shown to be [see (19a)–(19c) at the bottom of the page], where γ_{i+1} is the Lagrange multiplier corresponding to \mathbf{F}_{i+1} at the $(i+1)$ th iteration. Notice that when \mathbf{G}_{i+1} is fixed, the optimization problem (16) is a convex quadratic programming problem for \mathbf{F}_{i+1} , the KKT conditions are the necessary and sufficient conditions for the optimal \mathbf{F}_{i+1} [31].

Obviously from (19a), in order to compute the optimal \mathbf{F}_{i+1} , the Lagrange multiplier γ_{i+1} should be calculated first. However, there is no closed-form solution of γ_{i+1} simultaneously satisfying (19b) and (19c) [30]. Below we propose a low complexity method to solve (19b) and (19c). First, notice that in order to have (19b) satisfied, either $\gamma_{i+1} = 0$ or $\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) = P_r$ must hold. If $\gamma_{i+1} = 0$ also makes (19c) satisfied, $\gamma_{i+1} = 0$ is a solution to (19b) and (19c). Since given \mathbf{G}_{i+1} , the optimization problem (16) is a convex quadratic programming problem of \mathbf{F}_{i+1} , which has only one solution for \mathbf{F}_{i+1} , $\gamma_{i+1} = 0$ is the only solution to (19b) and (19c) in this case. On other hand, if $\gamma_{i+1} = 0$ does not make (19c) satisfied, we have to solve $\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) = P_r$. It is proved in Appendix I that when \mathbf{G}_{i+1} is fixed, the function

$$q(\gamma_{i+1}) \triangleq \text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) \quad (20)$$

is a monotonically decreasing function of γ_{i+1} . Furthermore, it has also been proved that

$$0 \leq \gamma_{i+1} \leq \sqrt{\frac{\text{Tr}(\bar{\mathbf{H}}_{rd}^H \mathbf{G}_{i+1}^H \mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-1} \bar{\mathbf{H}}_{sr} \mathbf{R}_s \mathbf{G}_{i+1} \bar{\mathbf{H}}_{rd})}{P_r}}. \quad (21)$$

Based on these results, γ_{i+1} can be efficiently computed by a one-dimension search, such as bisection search or golden search [32]. Since $\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) = P_r$ is a stronger condition than $\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) \leq P_r$, (19c) is satisfied automatically in this case. In summary, the proposed procedure for calculating γ_{i+1} is given as follows:

$$\gamma_{i+1} = \begin{cases} 0 & \text{if } q(0) \leq P_r \\ \text{Solve } q(\gamma_{i+1}) = P_r & \text{Otherwise.} \\ \text{using bisection algorithm} & \end{cases} \quad (22)$$

C. Summary and Convergence Analysis

The proposed iterative algorithm proceeds between (18) and (19a), which can be summarized as Algorithm 1.

Algorithm 1: The Iterative Algorithm for Joint Design of \mathbf{F} and \mathbf{G}

Initialize the algorithm with \mathbf{F}_0 where the $N \times N$ principal submatrix of \mathbf{F}_0 being $\sqrt{P_r / \text{Tr}(\mathbf{R}_x)} \mathbf{I}_N$, and $N = \min\{N_R, M_R\}$, while the remaining entries of \mathbf{F}_0 being zeros.

repeat

Update \mathbf{G}_i using (18);

Compute γ_i using (22);

Update \mathbf{F}_i using (19a);

$i = i + 1$;

until $\|\mathbf{F}_i - \mathbf{F}_{i-1}\|^2$ is smaller than a threshold ϵ , where the index i denotes the i th iteration.

This iterative algorithm can be shown to converge as follows. It is obvious that when \mathbf{F}_i is given, the objective function in problem (16) is a convex quadratic function of \mathbf{G}_{i+1} . The solution given by (18) corresponds to the minimum MSE for the fixed \mathbf{F}_i . In other words, $\text{MSE}(\mathbf{F}_i, \mathbf{G}_{i+1}) \leq \text{MSE}(\mathbf{F}_i, \mathbf{G}_i)$. On the other hand, when \mathbf{G}_{i+1} is obtained, the optimization problem (16) is a convex quadratic programming problem of \mathbf{F}_{i+1} , and the KKT conditions are the necessary and sufficient condition for the optimal solution [31], so we have $\text{MSE}(\mathbf{F}_{i+1}, \mathbf{G}_{i+1}) \leq \text{MSE}(\mathbf{F}_i, \mathbf{G}_{i+1})$. It follows that each update on \mathbf{F} or \mathbf{G} will decrease the objective function and thus the iterative algorithm converges.

Although the iterative algorithm results in precoder and equalizer design that gives satisfactory performances, it requires high complexity because of iterations. Furthermore, in practice, it is not known in advance how many iterations are needed for the iterative algorithm to converge. In the next section, we propose a closed-form solution, which approximately

$$\mathbf{F}_{i+1} = (\bar{\mathbf{H}}_{rd}^H \mathbf{G}_{i+1}^H \mathbf{G}_{i+1} \bar{\mathbf{H}}_{rd} + \text{Tr}(\mathbf{G}_{i+1} \boldsymbol{\Sigma}_{rd} \mathbf{G}_{i+1}^H) \boldsymbol{\Psi}_{rd} + \gamma_{i+1} \mathbf{I})^{-1} \bar{\mathbf{H}}_{rd}^H \mathbf{G}_{i+1}^H \mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-1}, \quad (19a)$$

$$\gamma_{i+1} [\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) - P_r] = 0, \quad \gamma_{i+1} \geq 0, \quad (19b)$$

$$\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) - P_r \leq 0, \quad (19c)$$

solves (16) in the general case, but exactly when $\Psi_{rd} \propto \mathbf{I}_{N_R}$. In the simulation, we find that the closed-form solution has a comparable performance to that of the iterative algorithm.

Remark 2: As mentioned previously, when \mathbf{G}_{i+1} is given, the optimization problem (16) is a convex quadratic programming problem for \mathbf{F}_{i+1} , which can be reformulated as a semidefinite programming (SDP) problem and solved by interior-point polynomial algorithms [33]. But these algorithms have a much higher complexity compared to the proposed algorithm based on KKT conditions (19a)–(19c).

V. THE PROPOSED CLOSED-FORM SOLUTION

Since the constraint in the problem (16) does not involve the equalizer \mathbf{G} , the optimal \mathbf{G} can be directly derived as a function of \mathbf{F} , by differentiating the objective function $\text{MSE}(\mathbf{F}, \mathbf{G})$ with respect to \mathbf{G}^* and setting the result to zero. This results in

$$\hat{\mathbf{G}} = \mathbf{R}_s (\bar{\mathbf{H}}_{rd} \mathbf{F} \bar{\mathbf{H}}_{sr})^H (\bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H + \mathbf{K})^{-1} \quad (23)$$

where $\mathbf{K} = \text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H \Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2}$ was previously defined in (15). Substituting (23) into the MSE expression (13), the MSE can be written as

$$\begin{aligned} \text{MSE}(\mathbf{F}, \hat{\mathbf{G}}) &= \text{Tr}(\mathbf{R}_s) - \text{Tr}(\mathbf{R}_s \bar{\mathbf{H}}_{sr}^H [\mathbf{F}^H \bar{\mathbf{H}}_{rd}^H (\bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H + \mathbf{K})^{-1} \\ &\quad \bar{\mathbf{H}}_{rd} \mathbf{F}] \bar{\mathbf{H}}_{sr} \mathbf{R}_s) \\ &\triangleq \overline{\text{MSE}}(\mathbf{F}). \end{aligned} \quad (24)$$

Since \mathbf{K} and \mathbf{R}_x are both positive definite Hermitian matrices, exploiting the matrix inversion lemma $\mathbf{A}^H (\mathbf{A} \mathbf{A}^H + \mathbf{I})^{-1} \mathbf{A} = \mathbf{I} - (\mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1}$ [35], we have

$$\begin{aligned} &\mathbf{F}^H \bar{\mathbf{H}}_{rd}^H (\bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H + \mathbf{K})^{-1} \bar{\mathbf{H}}_{rd} \mathbf{F} \\ &= \mathbf{R}_x^{-H/2} \underbrace{\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \mathbf{K}^{-H/2}}_{\mathbf{A}^H} \\ &\quad \times \underbrace{(\mathbf{K}^{-1/2} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} \mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \mathbf{K}^{-H/2} + \mathbf{I}_{M_D})^{-1}}_{\mathbf{A}^H} \\ &\quad \times \underbrace{\mathbf{K}^{-1/2} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} \mathbf{R}_x^{-1/2}}_{\mathbf{A}} \\ &= \mathbf{R}_x^{-1} - \mathbf{R}_x^{-H/2} \\ &\quad \times (\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \mathbf{K}^{-1} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} + \mathbf{I}_{M_R})^{-1} \mathbf{R}_x^{-1/2}. \end{aligned} \quad (25)$$

Putting (25) into (24), we have (26), shown at the bottom of the page, where

$$\text{MSE}_1 \triangleq \text{Tr}(\mathbf{R}_s) - \text{Tr}(\mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-1} \bar{\mathbf{H}}_{sr} \mathbf{R}_s) \quad (27)$$

is a constant part independent of \mathbf{F} . Now, the problem becomes minimizing (26) with power constraint $\text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) \leq P_r$. Unfortunately, from (15), it is noticed that $\mathbf{K} = \text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H \Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2}$, so $\overline{\text{MSE}}(\mathbf{F})$ in (26) is a high order function of \mathbf{F} and the problem of minimizing (26) is very difficult to solve. In order to proceed, we first derive an upper bound of $\overline{\text{MSE}}(\mathbf{F})$.

For any two $M \times M$ positive-semidefinite Hermitian matrices \mathbf{A} and \mathbf{B} , the following inequalities hold [36]:

$$\text{Tr}(\mathbf{A}\mathbf{B}) \leq \sum_i \lambda_i(\mathbf{A})\lambda_i(\mathbf{B}), \quad (28)$$

where $\lambda_i(\mathbf{Z})$ denotes the i th largest eigenvalue of the matrix \mathbf{Z} . It is also obvious that

$$\begin{aligned} \sum_i \lambda_i(\mathbf{B})\lambda_i(\mathbf{A}) &\leq \left(\sum_i \lambda_i(\mathbf{A}) \right) \lambda_{\max}(\mathbf{B}) \\ &= \text{Tr}(\mathbf{A})\lambda_{\max}(\mathbf{B}) \end{aligned} \quad (29)$$

where $\lambda_{\max}(\mathbf{Z})$ represents the maximum eigenvalue of \mathbf{Z} . Combining (28) and (29), we have

$$\text{Tr}(\mathbf{A}\mathbf{B}) \leq \text{Tr}(\mathbf{A})\lambda_{\max}(\mathbf{B}). \quad (30)$$

Substituting $\mathbf{A} = \mathbf{F} \mathbf{R}_x \mathbf{F}^H$ and $\mathbf{B} = \Psi_{rd}$, we have an upper bound for \mathbf{K}

$$\begin{aligned} \mathbf{K} &= \text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H \Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2} \\ &\leq \text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) \lambda_{\max}(\Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2} \triangleq \Phi. \end{aligned} \quad (31)$$

When \mathbf{K} in (26) is replaced by its upper bound Φ , the resultant MSE is shown in (32) at the bottom of the page. It is shown in Appendix II that $\overline{\text{MSE}}_U(\mathbf{F})$ is an upper bound of $\overline{\text{MSE}}(\mathbf{F})$ (i.e., $\overline{\text{MSE}}_U(\mathbf{F}) \geq \overline{\text{MSE}}(\mathbf{F})$) and $\overline{\text{MSE}}_U(\mathbf{F}) = \overline{\text{MSE}}(\mathbf{F})$ when $\Psi_{rd} \propto \mathbf{I}_{N_R}$. By replacing the objective function $\overline{\text{MSE}}(\mathbf{F})$ with its upper bound $\overline{\text{MSE}}_U(\mathbf{F})$, the optimization problem (16) can be relaxed as

$$\begin{aligned} \min_{\mathbf{F}} &\text{Tr}(\mathbf{R}_x^{-1/2} \bar{\mathbf{H}}_{sr} \mathbf{R}_s \mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-H/2}) \\ &\quad \triangleq \text{T} \\ &\quad (\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \underbrace{\Phi^{-1} \bar{\mathbf{H}}_{rd}}_{\triangleq \Theta} \mathbf{F} \mathbf{R}_x^{1/2} + \mathbf{I}_{M_R})^{-1} + \text{MSE}_1 \\ \text{s.t.} &\text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) \leq P_r. \end{aligned} \quad (33)$$

In the following, we derive the optimal solution for the relaxed optimization problem (33). For the optimization problem (33), the optimal solution \mathbf{F} should satisfy $\text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) = P_r$. This can be proved by contradiction. Supposing that we have an optimal precoder matrix \mathbf{F} with $\text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) < P_r$, we can

$$\overline{\text{MSE}}(\mathbf{F}) = \text{Tr}(\mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-H/2} (\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \mathbf{K}^{-1} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} + \mathbf{I}_{M_R})^{-1} \mathbf{R}_x^{-1/2} \bar{\mathbf{H}}_{sr} \mathbf{R}_s) + \text{MSE}_1, \quad (26)$$

$$\overline{\text{MSE}}_U(\mathbf{F}) \triangleq \text{Tr}(\mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-H/2} (\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \Phi^{-1} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} + \mathbf{I}_{M_R})^{-1} \mathbf{R}_x^{-1/2} \bar{\mathbf{H}}_{sr} \mathbf{R}_s) + \text{MSE}_1. \quad (32)$$

find another precoder matrix $\mathbf{F}_1 = \eta \mathbf{F}$ with $\eta > 1$, such that $\text{Tr}(\mathbf{F}_1 \mathbf{R}_x \mathbf{F}_1^H) = P_r$. Furthermore, the corresponding objective function in (33) for \mathbf{F}_1 is shown in (34) at the bottom of the page, where

$$\Phi_1 = \text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) \lambda_{\max}(\Psi_{rd}) \Sigma_{rd} + \frac{\mathbf{R}_{n_2}}{\eta^2} \prec \Phi. \quad (35)$$

Since the expression of $\text{MSE}_U(\mathbf{F})$ in (32) is the same as that of $\text{MSE}_U(\mathbf{F}_1)$ in (34) except with Φ replaced by Φ_1 , with similar arguments given in Appendix II, it can be shown that $\overline{\text{MSE}}_U(\mathbf{F}_1) < \overline{\text{MSE}}_U(\mathbf{F})$, which is in contradiction with the fact that \mathbf{F} is the optimal precoder corresponding to the minimum $\overline{\text{MSE}}_U$. This result also shows that the relaxation does not change the physical meaning of traditional precoder design. Therefore, for the optimal precoder, Θ in (33) is

$$\begin{aligned} \Theta &= \bar{\mathbf{H}}_{rd}^H (\text{Tr}(\mathbf{F} \mathbf{R}_x \mathbf{F}^H) \lambda_{\max}(\Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2})^{-1} \bar{\mathbf{H}}_{rd} \\ &= \bar{\mathbf{H}}_{rd}^H (P_r \lambda_{\max}(\Psi_{rd}) \Sigma_{rd} + \mathbf{R}_{n_2})^{-1} \bar{\mathbf{H}}_{rd} \end{aligned} \quad (36)$$

which is independent of \mathbf{F} .

Based on eigendecomposition, we have

$$\mathbf{T} = \mathbf{U}_T \mathbf{\Lambda}_T \mathbf{U}_T^H \quad (37)$$

$$\Theta = \mathbf{U}_\Theta \mathbf{\Lambda}_\Theta \mathbf{U}_\Theta^H \quad (38)$$

where \mathbf{T} has been defined in (33), and Θ is given in (36). The matrices \mathbf{U}_T and \mathbf{U}_Θ consist of the eigenvectors of \mathbf{T} and Θ , respectively, while the diagonal matrices $\mathbf{\Lambda}_T$ and $\mathbf{\Lambda}_\Theta$ contains the eigenvalues of \mathbf{T} and Θ , respectively. Without loss of generality, it is assumed that the diagonal elements of $\mathbf{\Lambda}_T$ and $\mathbf{\Lambda}_\Theta$ are in decreasing order. Substituting (37) and (38) into (33) and defining

$$\tilde{\mathbf{F}} \triangleq \mathbf{U}_\Theta^H \mathbf{F} \mathbf{R}_x^{1/2} \mathbf{U}_T, \quad (39)$$

the optimization problem can be written in a compact form as

$$\begin{aligned} \min_{\tilde{\mathbf{F}}} \quad & \text{Tr}(\mathbf{\Lambda}_T (\tilde{\mathbf{F}}^H \mathbf{\Lambda}_\Theta \tilde{\mathbf{F}} + \mathbf{I}_{M_R})^{-1}) \\ \text{s.t.} \quad & \text{Tr}(\tilde{\mathbf{F}} \tilde{\mathbf{F}}^H) \leq P_r \end{aligned} \quad (40)$$

where the constant MSE_1 is neglected, which does not affect the optimization problem.

Notice that for any positive semidefinite Hermitian matrix \mathbf{B} ,

$$\text{Tr}(\mathbf{\Lambda}_T \mathbf{B}) \geq \sum_{i=1}^{M_R} \lambda_i(\mathbf{\Lambda}_T) \lambda_{M_R-i+1}(\mathbf{B}). \quad (41)$$

In (41), the equality holds when the matrix \mathbf{B} is diagonal with diagonal elements in increasing order [36, 9.H.1.h]. Therefore, applying this result to the objective function of (40), for the optimal solution, we have $(\tilde{\mathbf{F}}^H \mathbf{\Lambda}_\Theta \tilde{\mathbf{F}} + \mathbf{I}_{M_R})^{-1}$ must be diagonal

with diagonal elements in increasing order. The optimization problem (40) can be rewritten as

$$\begin{aligned} \min \quad & h(\mathbf{b}) = \mathbf{d}^T \underbrace{\{\mathbf{\Lambda}_T\} \mathbf{d} \{(\tilde{\mathbf{F}}^H \mathbf{\Lambda}_\Theta \tilde{\mathbf{F}} + \mathbf{I}_{M_R})^{-1}\}}_{\triangleq \mathbf{b}} \\ \text{s.t.} \quad & \text{Tr}(\tilde{\mathbf{F}}^H \tilde{\mathbf{F}}) \leq P_r \end{aligned} \quad (42)$$

where $\mathbf{d}\{\mathbf{Z}\}$ denotes the vector formed from the main diagonal of \mathbf{Z} . Based on the fact that the elements of \mathbf{b} are in increasing order and the diagonal elements of $\mathbf{\Lambda}_T$ are in decreasing order, the function $h(\mathbf{b})$ is a Schur-concave function of \mathbf{b} [36, 3.H.3]. Based on [10, Theorem 1], it has been proved in Appendix III that the optimal $\tilde{\mathbf{F}}$ has the following structure

$$\tilde{\mathbf{F}} = \begin{bmatrix} \text{diag}(f_1, \dots, f_N) & \mathbf{0}_{N \times (M_R - N)} \\ \mathbf{0}_{(N_R - N) \times N} & \mathbf{0}_{(N_R - N) \times (M_R - N)} \end{bmatrix} \quad (43)$$

where $N = \min(\text{Rank}(\mathbf{\Lambda}_\Theta), M_R)$, and f_i ($1 \leq i \leq N$) are unknowns to be determined. With (43), the optimization problem (42) can be rewritten as

$$\begin{aligned} \min_{f_i} \quad & \sum_{i=1}^N \frac{\lambda_i(\mathbf{\Lambda}_T)}{\lambda_i(\mathbf{\Lambda}_\Theta) f_i^2 + 1} + \sum_{i=N+1}^{M_R} \lambda_i(\mathbf{\Lambda}_T) \\ \text{s.t.} \quad & \sum_{i=1}^N f_i^2 \leq P_r. \end{aligned} \quad (44)$$

Obviously, the solution to the problem (44) is modified water-filling [37], and based on the KKT conditions of (44), we have [38]

$$f_{i,\text{opt}}^2 = \left(\sqrt{\frac{\lambda_i(\mathbf{\Lambda}_T)}{\mu \lambda_i(\mathbf{\Lambda}_\Theta)}} - \frac{1}{\lambda_i(\mathbf{\Lambda}_\Theta)} \right)^+ \quad i = 1, \dots, N \quad (45)$$

where $\mu > 0$ is the Lagrange multiplier such that $\sum_{i=1}^N f_{i,\text{opt}}^2 = P_r$ holds.

From the definition of $\tilde{\mathbf{F}}$ in (39), together with the results in (43) and (45), we can write the optimal \mathbf{F} compactly as

$$\begin{aligned} \mathbf{F}_{\text{opt}} &= \mathbf{U}_{\Theta, N} \left(\left(\frac{1}{\sqrt{\mu}} \check{\mathbf{\Lambda}}_\Theta^{-1/2} \check{\mathbf{\Lambda}}_T^{1/2} - \check{\mathbf{\Lambda}}_\Theta^{-1} \right)^+ \right)^{1/2} \\ &\quad \times \mathbf{U}_{T, N}^H \mathbf{R}_x^{-1/2} \end{aligned} \quad (46)$$

where $[(\mathbf{Z})^+]_{i,j} = \max(0, (\mathbf{Z})_{i,j})$. The matrices $\check{\mathbf{\Lambda}}_\Theta$ and $\check{\mathbf{\Lambda}}_T$ are respectively the principal submatrices of $\mathbf{\Lambda}_\Theta$ and $\mathbf{\Lambda}_T$ with dimensions of $N \times N$, while the matrices $\mathbf{U}_{\Theta, N}$ and $\mathbf{U}_{T, N}$ are the first N columns of \mathbf{U}_Θ and \mathbf{U}_T , respectively. It is obvious that the matrix \mathbf{T} only contains the CSI in the first hop, while Θ only relates to the channel in the second hop. From (46), it can be seen that the precoder at the relay in effect simultaneously diagonalizes the source-relay and relay-destination channels and pairs

$$\overline{\text{MSE}}_U(\mathbf{F}_1) = \text{Tr}(\mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-H/2} (\mathbf{R}_x^{H/2} \mathbf{F}^H \bar{\mathbf{H}}_{rd}^H \Phi_1^{-1} \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{1/2} + \mathbf{I}_{M_R})^{-1} \mathbf{R}_x^{-1/2} \bar{\mathbf{H}}_{sr} \mathbf{R}_s) + \text{MSE}_1 \quad (34)$$

the eigenchannels of the two hops based on a ‘‘best-to-best’’ criterion. Then a water-filling algorithm is used to allocate power on different eigenchannels.

Finally, applying (46) into (23), the optimal equalizer at the destination is given by

$$\mathbf{G}_{\text{opt}} = \mathbf{R}_s(\tilde{\mathbf{H}}_{rd}\mathbf{F}_{\text{opt}}\tilde{\mathbf{H}}_{sr})^H \times (\tilde{\mathbf{H}}_{rd}\mathbf{F}_{\text{opt}}\mathbf{R}_x\mathbf{F}_{\text{opt}}^H\tilde{\mathbf{H}}_{rd}^H + \mathbf{K}_{\text{opt}})^{-1} \quad (47)$$

where $\mathbf{K}_{\text{opt}} = \text{Tr}(\mathbf{F}_{\text{opt}}\mathbf{R}_x\mathbf{F}_{\text{opt}}^H\boldsymbol{\Psi}_{rd})\boldsymbol{\Sigma}_{rd} + \mathbf{R}_{n_2}$. Notice that when the source-relay link is noiseless and the first hop channel \mathbf{H}_{sr} is an identity matrix, (46) reduces to the point-to-point MIMO robust LMMSE transceiver [39]. If both source-relay and relay-destination channels are perfectly known, we have $\mathbf{K} = \mathbf{R}_{n_2}$, and after simplifying $\boldsymbol{\Theta}$ and \mathbf{T} , (46) is exactly the solution in [21].

Remark 3: If the relay is a fixed station (such as in Winner project, LTE and IMT-Advanced [19], [20]), the distance between adjacent transmit antennas at the relay is large, and it is possible that $\boldsymbol{\Psi} \propto \mathbf{I}_{N_R}$. In that case, the proposed closed-form solution is exactly the optimal solution.

VI. EXTENSION TO WEIGHTED MSE CRITERION

Previously, we focus on the unweighted MSE minimization. The proposed algorithms given in Sections IV and V can be easily extended to the weighted MSE criterion. Weighted MSE criterion is important, because it is known that based on careful design of weighting matrix, fairness between different data streams can be achieved [9].

Suppose \mathbf{W} is the weighting matrix, which is generally a positive definite Hermitian matrix [9]. The weighted MSE is

$$\begin{aligned} \text{MSE}_W(\mathbf{F}, \mathbf{G}) &= \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})^H\mathbf{W}(\mathbf{G}\mathbf{y} - \mathbf{s})\} \\ &= \mathbb{E}\{\text{Tr}((\mathbf{W}^{1/2}\mathbf{G}\mathbf{y} - \mathbf{W}^{1/2}\mathbf{s}) \\ &\quad (\mathbf{W}^{1/2}\mathbf{G}\mathbf{y} - \mathbf{W}^{1/2}\mathbf{s})^H)\} \\ &\triangleq \mathbb{E}\{\text{Tr}((\tilde{\mathbf{G}}\mathbf{y} - \tilde{\mathbf{s}})(\tilde{\mathbf{G}}\mathbf{y} - \tilde{\mathbf{s}})^H)\} \end{aligned} \quad (48)$$

where $\tilde{\mathbf{G}} = \mathbf{W}^{1/2}\mathbf{G}$ and $\tilde{\mathbf{s}} = \mathbf{W}^{1/2}\mathbf{s}$. Furthermore, the received signal \mathbf{y} at the destination in (2) can be rewritten as

$$\mathbf{y} = \mathbf{H}_{rd}\mathbf{F}\underbrace{\mathbf{H}_{sr}\mathbf{W}^{-1/2}\tilde{\mathbf{s}}}_{\triangleq \tilde{\mathbf{H}}_{sr}} + \mathbf{H}_{rd}\mathbf{F}\mathbf{n}_1 + \mathbf{n}_2. \quad (49)$$

Putting (49) into (48), $\text{MSE}_W(\mathbf{F}, \mathbf{G})$ can be written as

$$\begin{aligned} \text{MSE}_W(\mathbf{F}, \mathbf{G}) &= \mathbb{E}\{||(\tilde{\mathbf{G}}\mathbf{H}_{rd}\mathbf{F}\tilde{\mathbf{H}}_{sr} - \mathbf{I}_{N_S})\tilde{\mathbf{s}} + \tilde{\mathbf{G}}\mathbf{H}_{rd}\mathbf{F}\mathbf{n}_1 + \tilde{\mathbf{G}}\mathbf{n}_2||^2\} \end{aligned} \quad (50)$$

which has exactly the same form as the first expression in (8). Based on this fact, the solutions for unweighted MSE in Sections IV and V can be directly extended to the weighted MSE case, via the following simple substitutions

$$\mathbf{G} \rightarrow \tilde{\mathbf{G}} = \mathbf{W}^{1/2}\mathbf{G} \quad (51a)$$

$$\mathbf{R}_s \rightarrow \tilde{\mathbf{R}}_s = \mathbf{W}^{1/2}\mathbf{R}_s\mathbf{W}^{1/2} \quad (51b)$$

$$\tilde{\mathbf{H}}_{sr} \rightarrow \tilde{\tilde{\mathbf{H}}}_{sr} = \tilde{\mathbf{H}}_{sr}\mathbf{W}^{-1/2} \quad (51c)$$

$$\boldsymbol{\Psi}_{sr} \rightarrow \tilde{\boldsymbol{\Psi}}_{sr} = \mathbf{W}^{-1/2}\boldsymbol{\Psi}_{sr}\mathbf{W}^{-1/2}. \quad (51d)$$

In particular, when the source-relay link is noiseless $\mathbf{R}_{n_1} = \mathbf{0}_{N_S \times N_S}$, the source-relay channel is an identity matrix $\mathbf{H}_{sr} = \mathbf{I}_{N_S}$, and the relay-destination channel is perfectly known, after the substitutions in (51), the optimal solution given in (46) and (47) reduces to the optimal solution for point-to-point MIMO systems based on weighted MSE criterion [9].

VII. SIMULATION RESULTS AND DISCUSSIONS

A. Simulation Setup

In this section, we will investigate the performance of the proposed two algorithms: the iterative and closed-form solutions. For the purpose of comparison, the algorithm based on the estimated channel only (without taking the channel estimation errors into account) [21] is also simulated. In the following, we consider an AF MIMO relay system where the source, relay and destination are equipped with same number of antennas, i.e., $N_S = M_R = N_R = M_D = 4$. The widely used exponential model [27], [28] is chosen for the channel estimation error covariance matrices. This corresponds to the channel estimation method in [27]. More specifically, the covariance matrices are represented by

$$\begin{aligned} \boldsymbol{\Psi}_{sr} = \boldsymbol{\Psi}_{rd} &= \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix} \\ \boldsymbol{\Sigma}_{sr} = \boldsymbol{\Sigma}_{rd} &= \sigma_e^2 \begin{bmatrix} 1 & \beta & \beta^2 & \beta^3 \\ \beta & 1 & \beta & \beta^2 \\ \beta^2 & \beta & 1 & \beta \\ \beta^3 & \beta^2 & \beta & 1 \end{bmatrix} \end{aligned} \quad (52)$$

where α and β are the correlation coefficients, and σ_e^2 denotes the estimation error variance.

We define the signal-to-noise ratio for the source-relay link (SNR_{sr}) as $E_s/N_1 = \text{Tr}(\mathbf{R}_s)/\text{Tr}(\mathbf{R}_{n_1})$, and is fixed as $E_s/N_1 = 30$ dB. At the source, four independent data streams are transmitted by four antennas at the same power. For each data stream, 1000 independent QPSK symbols are transmitted and $\text{Tr}(\mathbf{R}_s)$ is normalized to 1. The signal-to-noise ratio for the relay-destination link (SNR_{rd}) is defined as $E_r/N_2 = P_r/\text{Tr}(\mathbf{R}_{n_2})$. The estimated channels, $\tilde{\mathbf{H}}_{sr}$ and $\tilde{\mathbf{H}}_{rd}$, are generated based on the following distributions:

$$\begin{aligned} \tilde{\mathbf{H}}_{sr} &\sim \mathcal{CN}_{M_R, N_S} \left(\mathbf{0}_{M_R, N_S}, \frac{(1 - \sigma_e^2)}{\sigma_e^2} \boldsymbol{\Sigma}_{sr} \otimes \boldsymbol{\Psi}_{sr}^T \right) \\ \tilde{\mathbf{H}}_{rd} &\sim \mathcal{CN}_{M_D, N_R} \left(\mathbf{0}_{M_D, N_R}, \frac{(1 - \sigma_e^2)}{\sigma_e^2} \boldsymbol{\Sigma}_{rd} \otimes \boldsymbol{\Psi}_{rd}^T \right) \end{aligned}$$

such that channel realizations $\mathbf{H}_{sr} = \tilde{\mathbf{H}}_{sr} + \Delta\mathbf{H}_{sr}$ and $\mathbf{H}_{rd} = \tilde{\mathbf{H}}_{rd} + \Delta\mathbf{H}_{rd}$ have unit variance.

In the following figures, MSE is referred to as the simulated MSE of the equalized signal at the destination. Each point in the following figures is an average of 10 000 independent channel realizations.

B. Convergence Performance of Iterative Algorithm

Fig. 2 shows the MSE performances of the iterative algorithm as a function of iteration index with $\text{SNR}_{rd} = 10$ dB and $\text{SNR}_{rd} = 30$ dB. The correlation coefficients are set as $\alpha = 0$

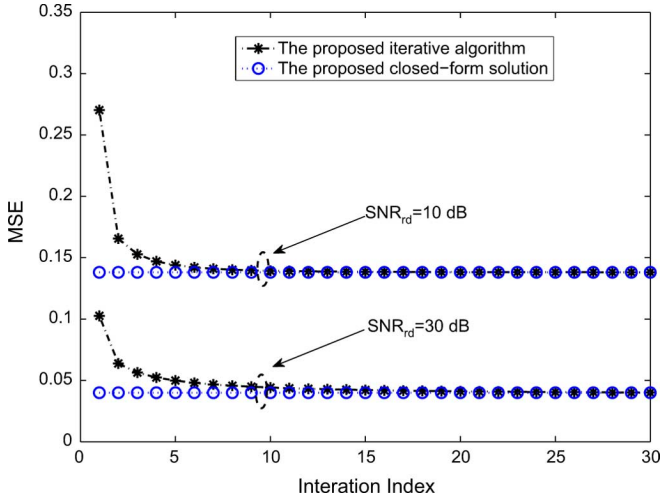


Fig. 2. Convergence performance of the iterative algorithm with $\alpha = 0$, $\beta = 0.2$, and $\sigma_e^2 = 0.002$.

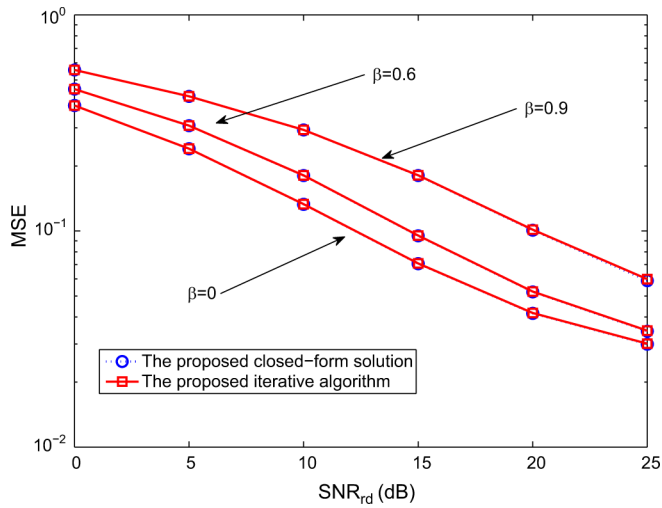


Fig. 3. The MSEs for the closed-form solution and the iterative algorithm for different β and σ_e^2 , when $\alpha = 0$ and $\sigma_e^2 = 0.002$.

(i.e., $\Psi_{rd} \propto \mathbf{I}_{N_R}$), $\beta = 0.2$ and $\sigma_e^2 = 0.002$. The threshold, ϵ , for terminating the iterative algorithm is $0.0001 \text{Tr}(\mathbf{F}_i \mathbf{F}_i^H)$. Notice that when $\Psi_{rd} \propto \mathbf{I}_{N_R}$, the proposed closed-form solution given by (46) and (47) is exactly the global optimal solution and it is also shown here as reference. From the figure, it can be seen that the iterative algorithm converges to the global optimal solution after around 15 iterations regardless of SNR_{rd} .

Fig. 3 further shows the MSE performances of the iterative algorithm and the closed-form solution as a function of SNR_{rd} with different β , when $\Psi_{rd} \propto \mathbf{I}_{N_R}$ and $\sigma_e^2 = 0.002$. In this case, the proposed closed-form solution is the global optimal solution. From Fig. 3, it can also be seen that the iterative algorithm always provides the same performance as the proposed closed-form solution, meaning that the iterative algorithm always converges to the global optimal solution, at least when $\Psi_{rd} \propto \mathbf{I}_{N_R}$.

Fig. 4 shows the effects of different initializations on convergence behaviors of the proposed iterative algorithm, when $\alpha = 0.6$, $\beta = 0.5$, and $\sigma_e^2 = 0.001$. In addition to initializing the

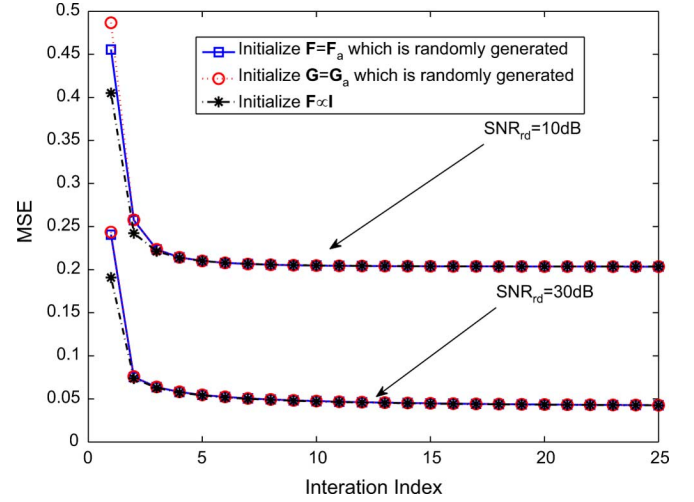


Fig. 4. Convergence behaviors of the iterative algorithm with different initializations, when $\alpha = 0.6$, $\beta = 0.5$, and $\sigma_e^2 = 0.001$.

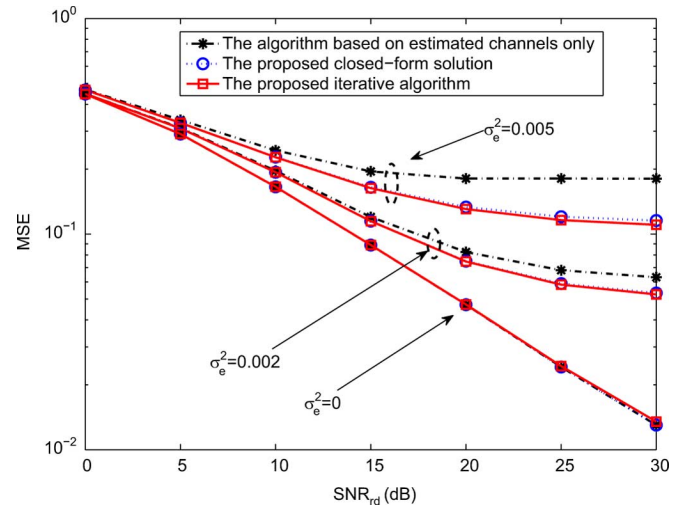


Fig. 5. The MSEs for the closed-form solution, the iterative algorithm and the algorithm based on estimated channels only for different σ_e^2 , when $\alpha = 0.6$ and $\beta = 0.45$.

iterative algorithm with $\mathbf{F} \propto \mathbf{I}$, we also simulated the case when elements of the initial \mathbf{F} are generated as independent Gaussian random variables. It is found that the convergence speed is similar to the case when initializing with $\mathbf{F} \propto \mathbf{I}$. Furthermore, when the iterative algorithm is initialized with \mathbf{G} (instead of \mathbf{F}) whose elements are generated as independent Gaussian random variables, the convergence performance is again basically the same.

C. Effect of Estimation Error σ_e^2

Fig. 5 shows the MSE performances of the different algorithms as a function of SNR_{rd} with different estimation errors σ_e^2 when $\alpha = 0.6$ and $\beta = 0.45$. It can be seen that when σ_e^2 decreases, MSE performances improve for all algorithms. Furthermore, when σ_e^2 decreases, the performance gap between the iterative algorithm and the closed-form solution becomes smaller. On the other hand, the performance of the algorithm based on estimated channels only shows degradation compared to that of the two proposed algorithms, except when $\sigma_e^2 = 0$.

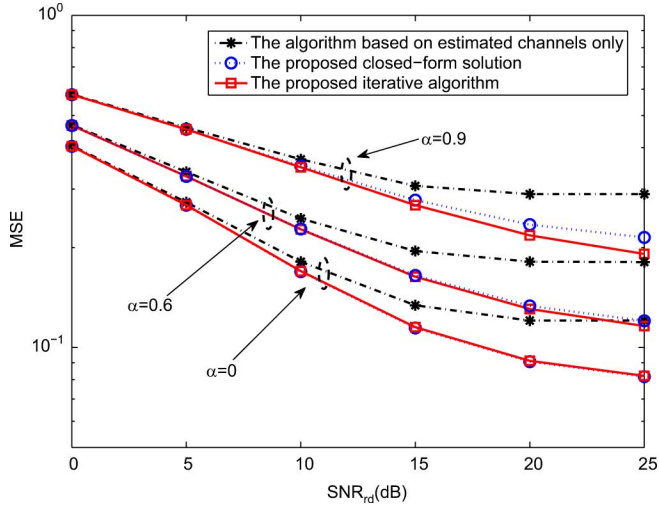


Fig. 6. The MSEs for the two proposed solutions and the algorithm based on estimated channels only for different α , when $\beta = 0.45$ and $\sigma_e^2 = 0.005$.

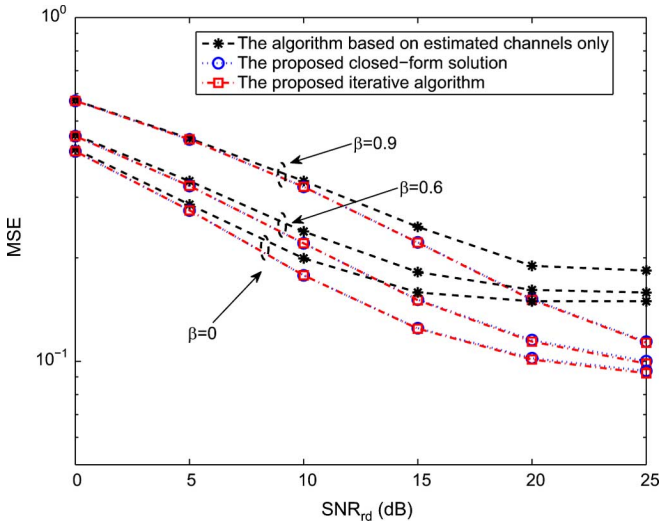


Fig. 7. The MSEs for the two proposed solutions and the algorithm based on estimated channels only for different β , when $\alpha = 0.45$ and $\sigma_e^2 = 0.005$.

D. Effect of Correlation Coefficients, α and β

Fig. 6 shows the MSE performances of the three algorithms as a function of SNR_{rd} with different α . From Fig. 6, it can be seen that when α decreases, the performances of all three algorithms improve, while the gap between the iterative algorithm and the closed-form solution decreases.

On the other hand, Fig. 7 shows the corresponding MSE performances with different β . From this figure, it is clear that when the correlation coefficient β varies, the difference between the closed-form solution and the iterative solution does not change. It is because the relaxation involved in the closed-form solution is not related to β . Furthermore, from Figs. 6 and 7, the two proposed algorithms show consistent performance improvements with respect to the algorithm based on estimated channels only.

E. BER Performance

Fig. 8 compares the bit error rates (BERs) of the two proposed algorithms and the algorithm based on estimated channel

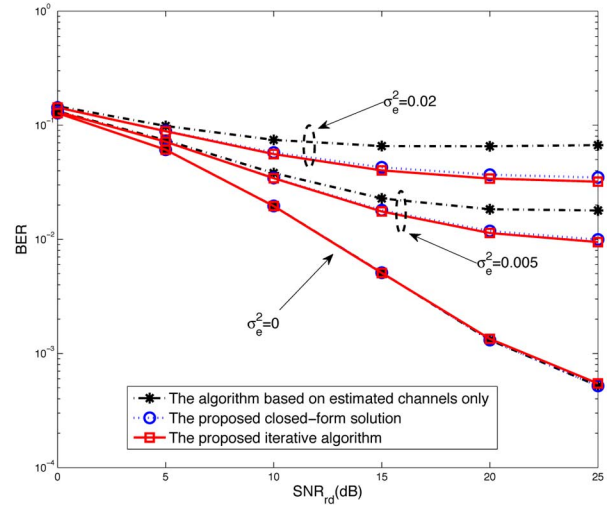


Fig. 8. The BERs for the proposed closed-form solution, iterative algorithm and the algorithm based on estimated channels only for different σ_e^2 , when $\alpha = 0.45$ and $\beta = 0.45$.

only as a function of SNR_{rd} , when $\alpha = 0.45$ and $\beta = 0.45$. It can be seen that the BER performances improve when σ_e^2 decreases, which are consistent with the MSE performances. It can also be seen that for different σ_e^2 , the proposed closed-form solution has almost the same performance as the proposed iterative algorithm, which are better than that of the algorithm based on estimated channels only.

VIII. CONCLUSION

In this paper, the joint design of linear relay precoder and destination equalizer for dual-hop AF MIMO relay systems with Gaussian random channel uncertainties in both hops was considered. The data MSE expression at the destination averaged over the random channel uncertainties was first derived. Then two robust design algorithms were proposed to minimize the average MSE. The first one is an iterative algorithm with its convergence established analytically. The second one is a closed-form solution with much lower complexity compared to the iterative algorithm. Although a mild relaxation is required for the general case, the closed-form solution was shown to be optimal when $\Psi_{rd} \propto \mathbf{I}_{N_R}$. Several existing transceiver design algorithms for point-to-point systems or relay systems under perfect CSI have been shown to be special cases of the proposed algorithms. Simulation results showed that both of the proposed algorithms perform better than the existing algorithm without taking the channel uncertainties into account. Furthermore, the performance gap between the proposed closed-form solution and iterative algorithm has been shown to be small in most cases.

APPENDIX I

PROOF OF $q(\gamma_{i+1})$ IS MONOTONICALLY DECREASING AND UPPER BOUND ON γ_{i+1}

First, we note the following two well-known results for matrix.

Property 1: For any two positive definite Hermitian matrices, \mathbf{M} and \mathbf{N} , $\mathbf{M} \succeq \mathbf{N}$ if and only if $\mathbf{M}^{-1} \preceq \mathbf{N}^{-1}$ [34, 7.7.4].

Property 2: For any two positive-semidefinite Hermitian matrices, \mathbf{M} and \mathbf{N} , $\text{Tr}(\mathbf{MN}) \geq 0$.

For notational simplicity, define

$$\mathbf{A}_{i+1} = \bar{\mathbf{H}}_{rd}^H \mathbf{G}_{i+1}^H \mathbf{G}_{i+1} \bar{\mathbf{H}}_{rd} + \text{Tr}(\mathbf{G}_{i+1} \boldsymbol{\Sigma}_{rd} \mathbf{G}_{i+1}^H) \boldsymbol{\Psi}_{rd} \quad (53)$$

$$\mathbf{B}_{i+1} = \bar{\mathbf{H}}_{rd}^H \mathbf{G}_{i+1}^H \mathbf{R}_s \bar{\mathbf{H}}_{sr}^H \mathbf{R}_x^{-1} \bar{\mathbf{H}}_{sr} \mathbf{R}_s \mathbf{G}_{i+1} \bar{\mathbf{H}}_{rd}. \quad (54)$$

Notice that \mathbf{A}_{i+1} is a positive-definite Hermitian matrix and \mathbf{B}_{i+1} is a positive-semidefinite Hermitian matrix. Putting (53) and (54) into the definition of $q(\gamma_{i+1})$ in (20), we have

$$q(\gamma_{i+1}) \triangleq \text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) = \text{Tr}((\mathbf{A}_{i+1} + \gamma_{i+1} \mathbf{I})^{-2} \mathbf{B}_{i+1}). \quad (55)$$

For any two values a and b , with $a \geq b \geq 0$,

$$q(b) - q(a) = \text{Tr}(((\mathbf{A}_{i+1} + b\mathbf{I})^{-2} - (\mathbf{A}_{i+1} + a\mathbf{I})^{-2}) \mathbf{B}_{i+1}). \quad (56)$$

Below we will show that $q(b) - q(a) \geq 0$. Because \mathbf{A}_{i+1} is a positive semidefinite Hermitian matrix and $a \geq b$, $(\mathbf{A}_{i+1} + b\mathbf{I})^2 \preceq (\mathbf{A}_{i+1} + a\mathbf{I})^2$, and with property 1, we have

$$(\mathbf{A}_{i+1} + b\mathbf{I})^{-2} - (\mathbf{A}_{i+1} + a\mathbf{I})^{-2} \succeq 0. \quad (57)$$

Together with the fact that \mathbf{B}_{i+1} is a positive semidefinite Hermitian matrix, based on property 2, we have

$$\text{Tr}(((\mathbf{A}_{i+1} + b\mathbf{I})^{-2} - (\mathbf{A}_{i+1} + a\mathbf{I})^{-2}) \mathbf{B}_{i+1}) \geq 0 \quad (58)$$

therefore,

$$q(b) - q(a) \geq 0 \quad (59)$$

and $q(\gamma_{i+1})$ is a monotonically decreasing function of γ_{i+1} .

With regard to the upper bound of γ_{i+1} , it is obvious that

$$(\mathbf{A}_{i+1} + \gamma_{i+1} \mathbf{I})^2 \succ \gamma_{i+1}^2 \mathbf{I}. \quad (60)$$

As mentioned in the proposed procedure (22) for computing γ_{i+1} , the bisection algorithm is only needed when $\gamma_{i+1} \neq 0$. So we can invert both side of (60), multiply \mathbf{B}_{i+1} from the right and take the trace of the both sides, and it follows that

$$\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) < \frac{1}{\gamma_{i+1}^2} \text{Tr}(\mathbf{B}_{i+1}). \quad (61)$$

Notice that based on (19b), if $\gamma_{i+1} \neq 0$, $\text{Tr}(\mathbf{F}_{i+1} \mathbf{R}_x \mathbf{F}_{i+1}^H) = P_r$ and then

$$\gamma_{i+1} < \sqrt{\frac{\text{Tr}(\mathbf{B}_{i+1})}{P_r}} \quad (62)$$

which is exactly the expression given in (21).

APPENDIX II

PROOF OF $\overline{\text{MSE}}_U(\mathbf{F}) \geq \overline{\text{MSE}}(\mathbf{F})$

First, we notice the following result for matrix.

Property 3: For any two Hermitian matrices, \mathbf{M} and \mathbf{N} , if $\mathbf{M} \succeq \mathbf{N}$, then $\mathbf{P}^H \mathbf{M} \mathbf{P} \succeq \mathbf{P}^H \mathbf{N} \mathbf{P}$, where \mathbf{P} is an arbitrary matrix [34, 7.7.2.a].

Defining $\mathbf{A} = \bar{\mathbf{H}}_{rd} \mathbf{F} \mathbf{R}_x^{-1/2}$ and $\mathbf{B} = \mathbf{R}_x^{-1/2} \bar{\mathbf{H}}_{sr} \mathbf{W} \mathbf{R}_s$, we have

$$\overline{\text{MSE}}(\mathbf{F}) = \text{Tr}(\mathbf{B}^H (\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} + \mathbf{I})^{-1} \mathbf{B}) + \text{MSE}_1 \quad (63)$$

$$\overline{\text{MSE}}_U(\mathbf{F}) = \text{Tr}(\mathbf{B}^H (\mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A} + \mathbf{I})^{-1} \mathbf{B}) + \text{MSE}_1. \quad (64)$$

It is straightforward to show that

$$\overline{\text{MSE}}_U(\mathbf{F}) - \overline{\text{MSE}}(\mathbf{F}) = \text{Tr}(\mathbf{B}^H ((\mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A} + \mathbf{I})^{-1} - (\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} + \mathbf{I})^{-1}) \mathbf{B}). \quad (65)$$

Below we will show that $\mathbf{B}^H ((\mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A} + \mathbf{I})^{-1} - (\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} + \mathbf{I})^{-1}) \mathbf{B} \succeq 0$. Since $\boldsymbol{\Phi}$ and \mathbf{K} are Hermitian matrices and $\boldsymbol{\Phi} \succeq \mathbf{K} \succ 0$, based on property 1 given in Appendix I, we have $\mathbf{K}^{-1} \succeq \boldsymbol{\Phi}^{-1}$. Furthermore, based on property 3, we have the following inequality:

$$\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} \succeq \mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A}. \quad (66)$$

Adding an identity matrix on both sides of (66), the inequality sign does not change. Applying the property 1 in Appendix I again, we have

$$(\mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A} + \mathbf{I})^{-1} \succeq (\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} + \mathbf{I})^{-1}. \quad (67)$$

With property 3, we have the following inequality:

$$\mathbf{B}^H (\mathbf{A}^H \boldsymbol{\Phi}^{-1} \mathbf{A} + \mathbf{I})^{-1} \mathbf{B} \succeq \mathbf{B}^H (\mathbf{A}^H \mathbf{K}^{-1} \mathbf{A} + \mathbf{I})^{-1} \mathbf{B}. \quad (68)$$

Taking the trace on both sides of (68), we finally have

$$\overline{\text{MSE}}_U(\mathbf{F}) - \overline{\text{MSE}}(\mathbf{F}) \geq 0. \quad (69)$$

Based on the definition of $\boldsymbol{\Phi}$ in (31), it is obvious that when $\boldsymbol{\Psi}_{rd} \propto \mathbf{I}_{N_R}$ is true, $\mathbf{K} = \boldsymbol{\Phi}$ and in this case, $\overline{\text{MSE}}_U(\mathbf{F}) = \overline{\text{MSE}}(\mathbf{F})$.

APPENDIX III

DERIVATION OF OPTIMAL $\tilde{\mathbf{F}}$

First, we introduce a permutation matrix with dimension of $M \times M$ as

$$\mathbf{Q}_M = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix}_M \quad (70)$$

and the optimization problem (42) can be reformulated as

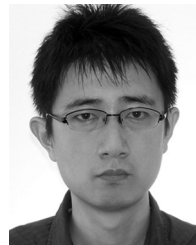
$$\begin{aligned} \min \quad & \mathbf{d}^T \{ \boldsymbol{\Lambda}'_{\mathbf{T}} \} \underbrace{\mathbf{d} \{ (\mathcal{F}^H \boldsymbol{\Lambda}'_{\mathbf{e}} \mathcal{F} + \mathbf{I}_{M_R})^{-1} \}}_{\triangleq \mathbf{b}'} \\ \text{s.t.} \quad & \text{Tr}(\mathcal{F}^H \mathcal{F}) \leq P_r \end{aligned} \quad (71)$$

where $\boldsymbol{\Lambda}'_{\mathbf{T}} = \mathbf{Q}_{M_R} \boldsymbol{\Lambda}_{\mathbf{T}} \mathbf{Q}_{M_R}$ and $\boldsymbol{\Lambda}'_{\mathbf{e}} = \mathbf{Q}_{N_R} \boldsymbol{\Lambda}_{\mathbf{e}} \mathbf{Q}_{N_R}$ are $\boldsymbol{\Lambda}_{\mathbf{T}}$ and $\boldsymbol{\Lambda}_{\mathbf{e}}$ with diagonal elements in reverse order, and $\mathcal{F} = \mathbf{Q}_{N_R} \tilde{\mathbf{F}} \mathbf{Q}_{M_R}$. Notice that because of the permutation matrices, the order of \mathbf{b}' is the reverse to that of \mathbf{b} (i.e., the elements of \mathbf{b}' are in decreasing order). Based on [10, Theorem 1], the optimal $\mathcal{F} = \mathbf{Q}_{N_R} \tilde{\mathbf{F}} \mathbf{Q}_{M_R}$ for the problem (71) has zero elements except along the rightmost main diagonal. Reversing the permutations to \mathcal{F} , we have the solution given in (43).

REFERENCES

- [1] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile Ad Hoc networks," *IEEE Signal Process. Mag.*, vol. 23, no. 5, pp. 18–29, Sep. 2006.
- [2] 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.
- [3] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.

- [4] E. Koyuncu, Y. Jing, and H. Jafarkhani, "Distributed beamforming in wireless relay networks with quantized feedback," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1429–1439, Oct. 2008.
- [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. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, pp. 2198–2206, Jun. 2007.
- [7] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [9] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [10] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2399, Sep. 2003.
- [11] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.
- [12] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.
- [13] F. A. Dietrich, *Robust Signal Processing for Wireless Communications, Foundations in Signal Processing, Communications and Networking*. New York: Springer, 2007.
- [14] 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.
- [15] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," in *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/PS_cache/arxiv/pdf/0811/0811.4565v1.pdf
- [16] H. W. Je, B. Lee, S. Kim, and K. B. Lee, "Design of non-regenerative MIMO-relay systems with partial channel state information," in *Proc. IEEE ICC 2008*, May 2008, pp. 4441–4445.
- [17] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1443, Jun. 2006.
- [18] C.-B. Chae, T. W. Tang, R. W. Heath, and S.-Y. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 727–738, Feb. 2008.
- [19] S. Stefania, I. Toufik, and M. backer, *LTE, the UMTS Long Term Evolution: From Theory to Practice*. New York: Wiley, 2009.
- [20] A. Osseiran *et al.*, "The road to IMT-Advanced communication systems: State-of-the-art and innovation areas addressed by the WINNER + project," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 38–47, Jun. 2009.
- [21] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517–519, Jul. 2008.
- [22] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pt. 2, pp. 5062–5073, Oct. 2008.
- [23] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.
- [24] 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.
- [25] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman&Hall/CRC, 2000.
- [26] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Stat.*, vol. 35, pp. 475–501, 1964.
- [27] L. Musavian, M. R. Nakhi, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 2798–2806, Sep. 2007.
- [28] 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.
- [29] S. Ulukus and A. Yener, "Iterative transmitter and receiver optimization for CDMA networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1879–1884, Nov. 2004.
- [30] J. Zhang, Y. Wu, S. Zhou, and J. Wang, "Joint linear transmitter and receiver design for the downlink of multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 9, pp. 991–993, Nov. 2005.
- [31] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. New York: Athena, 2005.
- [32] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. New York: Wiley, 1996.
- [33] A. Beck, "Quadratic matrix programming," *SIAM J. Optim.*, vol. 17, no. 4, pp. 1224–1238, 2007.
- [34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [35] S. M. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [36] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
- [37] F. Frachinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Spring Series in Operation Research*, 2003, vol. I.
- [38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [39] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceiver," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.



Chengwen Xing received the B.Eng. degree from Xidian University, Xi'an, China, in 2005. He is currently working towards the Ph.D. degree in the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong.

His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, and cooperative communication systems.



Shaodan Ma received the B.Sc. (Eng.) and the M.Eng.Sc. degrees from NanKai University, Tianjin, China, in 1999 and 2002, respectively, all in electrical engineering, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2006.

Since 2006, she has been as a Postdoctoral Fellow with the Department of Electrical and Electronic Engineering, University of Hong Kong. Her research interests include wireless communication systems, spread spectrum techniques, MIMO systems, OFDM

technique, and digital signal processing.



Yik-Chung Wu received the B.Eng. (EEE) and M.Phil. degrees from the University of Hong Kong (HKU) in 1998 and 2001, respectively, and the Ph.D. degree from the Texas A&M University, College Station, in 2005, where his studies were fully supported by the prestigious Croucher Foundation scholarship.

After receiving the Master's degree, he became a Research Assistant at HKU. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with the

University of Hong Kong as an Assistant Professor. His research interests are in the general area of signal processing and communication systems, and in particular receiver algorithm design, synchronization techniques, channel estimation, and equalization.

Dr. Wu was a TPC member for the IEEE Vehicular Technology Conference Fall 2005, IEEE GLOBECOM 2006 and 2008, and the International Conference on Communications (ICC) 2007 and 2008. He is currently serving as an Associate Editor for the IEEE COMMUNICATIONS LETTERS.