Joint source and relay design for MIMO multi-relay systems using projected gradient approach

RESEARCH   **Open Access**

Joint source and relay design for MIMO multi-relay systems using projected gradient approach

Apriana Toding[1], Muhammad RA Khandaker[2] and Yue Rong[3*]

Abstract

In this paper, we develop the optimal source precoding matrix and relay amplifying matrices for non-regenerative multiple-input multiple-output (MIMO) relay communication systems with parallel relay nodes using the projected gradient (PG) approach. We show that the optimal relay amplifying matrices have a beamforming structure. Exploiting the structure of relay matrices, an iterative joint source and relay matrices optimization algorithm is developed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination using the PG approach. The performance of the proposed algorithm is demonstrated through numerical simulations.

Keywords: MIMO relay; Parallel relay network; Beamforming; Non-regenerative relay; Projected gradient

1 Introduction

Recently, multiple-input multiple-output (MIMO) relay communication systems have attracted much research interest and provided significant improvement in terms of both spectral efficiency and link reliability [1-17]. Many works have studied the optimal relay amplifying matrix for the source-relay-destination channel. In [2,3], the optimal relay amplifying matrix maximizing the mutual information (MI) between the source and destination nodes was derived, assuming that the source covariance matrix is an identity matrix. In [4-6], the optimal relay amplifying matrix was designed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination.

A few research has studied the joint optimization of the source precoding matrix and the relay amplifying matrix for the source-relay-destination channel. In [7], both the source and relay matrices were jointly designed to maximize the source-destination MI. In [8,9], source and relay matrices were developed to jointly optimize a broad class of objective functions. The author of [10] investigated the

joint source and relay optimization for two-way MIMO relay systems using the projected gradient (PG) approach. The source and relay optimization for multi-user MIMO relay systems with single relay node has been investigated in [11-14].

All the works in [1-14] considered a single relay node at each hop. In general, joint source and relay precoding matrices design for MIMO relay systems with multiple relay nodes is more challenging than that for single-relay systems. The authors of [15] developed the optimal relay amplifying matrices with multiple relay nodes. A matrix-form conjugate gradient algorithm has been proposed in [16] to optimize the source and relay matrices. In [17], the authors proposed a suboptimal source and relay matrices design for parallel MIMO relay systems by first relaxing the power constraint at each relay node to the sum relay power constraints at the output of the second-hop channel and then scaling the relay matrices to satisfy the individual relay power constraints.

In this paper, we propose a jointly optimal source precoding matrix and relay amplifying matrices design for a two-hop non-regenerative MIMO relay network with multiple relay nodes using the PG approach. We show that the optimal relay amplifying matrices have a beamforming structure. This new result is not available in [16]. It generalizes the optimal source and relay matrices design

*Correspondence: y.rong@curtin.edu.au
[3]Department of Electrical and Computer Engineering, Curtin University of Technology, Bentley, WA 6102, Australia
Full list of author information is available at the end of the article

from a single relay node case [8] to multiple parallel relay nodes scenarios. Exploiting the structure of relay matrices, an iterative joint source and relay matrices optimization algorithm is developed to minimize the MSE of the signal waveform estimation. Different to [17], in this paper, we develop the optimal source and relay matrices by directly considering the transmission power constraint at each relay node. Simulation results demonstrate the effectiveness of the proposed iterative joint source and relay matrices design algorithm with multiple parallel relay nodes using the PG approach.

The rest of this paper is organized as follows. In Section 2, we introduce the model of a non-regenerative MIMO relay communication system with parallel relay nodes. The joint source and relay matrices design algorithm is developed in Section 3. In Section 4, we show some numerical simulations. Conclusions are drawn in Section 5.

2 System model

In this section, we introduce the model of a two-hop MIMO relay communication system consisting of one source node, K parallel relay nodes, and one destination node as shown in Figure 1. We assume that the source and destination nodes have N_s and N_d antennas, respectively, and each relay node has N_r antennas. The generalization to systems with different number of antennas at each relay node is straightforward. Due to its merit of simplicity, a linear non-regenerative strategy is applied at each relay node. The communication process between the source and destination nodes is completed in two time slots. In the first time slot, the $N_b \times 1$ ($N_b \leq N_s$) modulated source symbol vector $\mathbf{s}$ is linearly precoded as

$$\mathbf{x} = \mathbf{B}\,\mathbf{s}, \tag{1}$$

where $\mathbf{B}$ is an $N_s \times N_b$ source precoding matrix. We assume that the source signal vector satisfies $\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{I}_{N_b}$, where $\mathbf{I}_n$ stands for an $n \times n$ identity matrix, $(\cdot)^H$ is the matrix (vector) Hermitian transpose, and $\mathrm{E}[\cdot]$ denotes statistical expectation. The precoded vector $\mathbf{x}$ is transmitted to K parallel relay nodes. The $N_r \times 1$ received signal vector at the ith relay node can be written as

$$\mathbf{y}_{\mathrm{r},i} = \mathbf{H}_{\mathrm{sr},i}\mathbf{x} + \mathbf{v}_{\mathrm{r},i}, \qquad i = 1, \cdots, K, \tag{2}$$

where $\mathbf{H}_{\mathrm{sr},i}$ is the $N_r \times N_s$ MIMO channel matrix between the source and the ith relay nodes and $\mathbf{v}_{\mathrm{r},i}$ is the additive Gaussian noise vector at the ith relay node.

In the second time slot, the source node is silent, while each relay node transmits the linearly amplified signal vector to the destination node as

$$\mathbf{x}_{\mathrm{r},i} = \mathbf{F}_i\,\mathbf{y}_{\mathrm{r},i}, \qquad i = 1, \cdots, K, \tag{3}$$

where $\mathbf{F}_i$ is the $N_r \times N_r$ amplifying matrix at the ith relay node. The received signal vector at the destination node can be written as

$$\mathbf{y}_{\mathrm{d}} = \sum_{i=1}^{K} \mathbf{H}_{\mathrm{rd},i}\mathbf{x}_{\mathrm{r},i} + \mathbf{v}_{\mathrm{d}}, \tag{4}$$

where $\mathbf{H}_{\mathrm{rd},i}$ is the $N_d \times N_r$ MIMO channel matrix between the ith relay and the destination nodes, and $\mathbf{v}_{\mathrm{d}}$ is the additive Gaussian noise vector at the destination node.

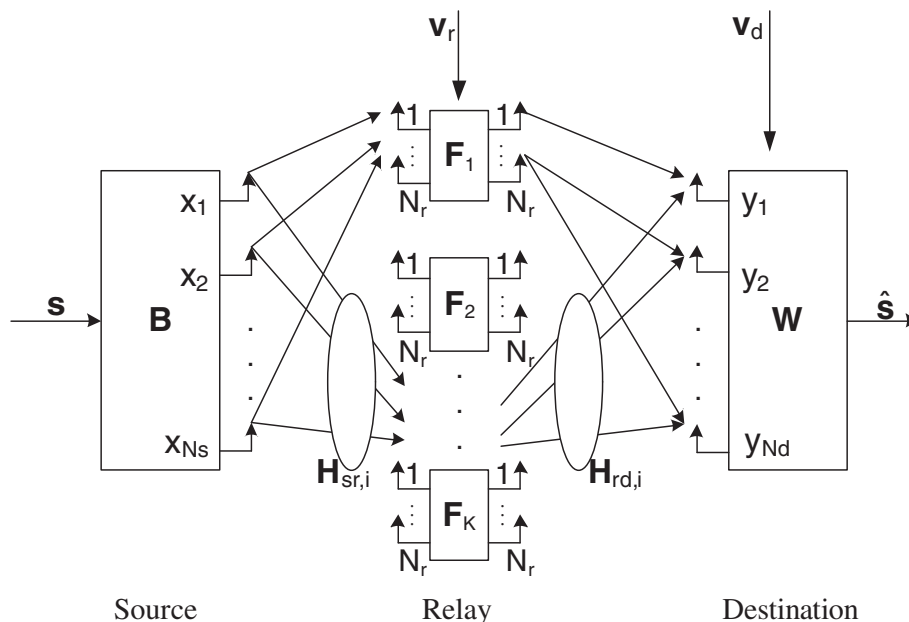**Figure 1 Block diagram of a parallel MIMO relay communication system.**

Substituting (1) to (3) into (4), we have

$$\mathbf{y}_\mathrm{d} = \sum_{i=1}^{K} (\mathbf{H}_{\mathrm{rd},i}\mathbf{F}_i\mathbf{H}_{\mathrm{sr},i}\mathbf{B}\mathbf{s} + \mathbf{H}_{\mathrm{rd},i}\mathbf{F}_i\mathbf{v}_{\mathrm{r},i}) + \mathbf{v}_\mathrm{d}$$

$$= \mathbf{H}_{\mathrm{rd}}\mathbf{F}\mathbf{H}_{\mathrm{sr}}\mathbf{B}\mathbf{s} + \mathbf{H}_{\mathrm{rd}}\mathbf{F}\mathbf{v}_\mathrm{r} + \mathbf{v}_\mathrm{d} \triangleq \tilde{\mathbf{H}}\mathbf{s} + \tilde{\mathbf{v}}, \qquad (5)$$

where $\mathbf{H}_\mathrm{sr} \triangleq \left[\mathbf{H}_{\mathrm{sr},1}^T, \mathbf{H}_{\mathrm{sr},2}^T, \cdots, \mathbf{H}_{\mathrm{sr},K}^T\right]^T$ is a $KN_\mathrm{r} \times N_\mathrm{s}$ channel matrix between the source node and all relay nodes, $\mathbf{H}_\mathrm{rd} \triangleq \left[\mathbf{H}_{\mathrm{rd},1}, \mathbf{H}_{\mathrm{rd},2}, \cdots, \mathbf{H}_{\mathrm{rd},K}\right]$ is an $N_\mathrm{d} \times KN_\mathrm{r}$ channel matrix between all relay nodes and the destination node, $\mathbf{F} \triangleq \mathrm{bd}[\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_K]$ is the $KN_\mathrm{r} \times KN_\mathrm{r}$ block diagonal equivalent relay matrix, $\mathbf{v}_\mathrm{r} \triangleq \left[\mathbf{v}_{\mathrm{r},1}^T, \mathbf{v}_{\mathrm{r},2}^T, \cdots, \mathbf{v}_{\mathrm{r},K}^T\right]^T$ is obtained by stacking the noise vectors at all the relays, $\tilde{\mathbf{H}} \triangleq \mathbf{H}_\mathrm{rd}\mathbf{F}\mathbf{H}_\mathrm{sr}\mathbf{B}$ is the effective MIMO channel matrix of the source-relay-destination link, and $\tilde{\mathbf{v}} \triangleq \mathbf{H}_\mathrm{rd}\mathbf{F}\mathbf{v}_\mathrm{r} + \mathbf{v}_\mathrm{d}$ is the equivalent noise vector. Here, $(\cdot)^T$ denotes the matrix (vector) transpose, $\mathrm{bd}[\cdot]$ constructs a block-diagonal matrix. We assume that all noises are independent and identically distributed (i.i.d.) Gaussian noise with zero mean and unit variance. The transmission power consumed by each relay node (3) can be expressed as

$$\mathrm{E}\left[\mathrm{tr}\left(\mathbf{x}_{\mathrm{r},i}\mathbf{x}_{\mathrm{r},i}^H\right)\right] = \mathrm{tr}\left(\mathbf{F}_i\left[\mathbf{H}_{\mathrm{sr},i}\mathbf{B}\mathbf{B}^H\mathbf{H}_{\mathrm{sr},i}^H + \mathbf{I}_{N_\mathrm{r}}\right]\mathbf{F}_i^H\right),$$

$$i = 1, \cdots, K, \qquad (6)$$

where $\mathrm{tr}(\cdot)$ stands for the matrix trace.

Using a linear receiver, the estimated signal waveform vector at the destination node is given by $\hat{\mathbf{s}} = \mathbf{W}^H\mathbf{y}_\mathrm{d}$, where $\mathbf{W}$ is an $N_\mathrm{d} \times N_\mathrm{b}$ weight matrix. The MSE of the signal waveform estimation is given by

$$\mathrm{MSE} = \mathrm{tr}\left(\mathrm{E}\left[\left(\hat{\mathbf{s}} - \mathbf{s}\right)\left(\hat{\mathbf{s}} - \mathbf{s}\right)^H\right]\right)$$

$$= \mathrm{tr}\left(\left(\mathbf{W}^H\tilde{\mathbf{H}} - \mathbf{I}_{N_\mathrm{b}}\right)\left(\mathbf{W}^H\tilde{\mathbf{H}} - \mathbf{I}_{N_\mathrm{b}}\right)^H + \mathbf{W}^H\tilde{\mathbf{C}}\mathbf{W}\right), \qquad (7)$$

where $\tilde{\mathbf{C}}$ is the equivalent noise covariance matrix given by $\tilde{\mathbf{C}} = \mathrm{E}\left[\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\right] = \mathbf{H}_\mathrm{rd}\mathbf{F}\mathbf{F}^H\mathbf{H}_\mathrm{rd}^H + \mathbf{I}_{N_\mathrm{d}}$. The weight matrix $\mathbf{W}$ which minimizes (7) is the Wiener filter and can be written as

$$\mathbf{W} = (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \tilde{\mathbf{C}})^{-1}\tilde{\mathbf{H}}, \qquad (8)$$

where $(\cdot)^{-1}$ denotes the matrix inversion. Substituting (8) back into (7), it can be seen that the MSE is a function of $\mathbf{F}$ and $\mathbf{B}$ and can be written as

$$\mathrm{MSE} = \mathrm{tr}\left(\left[\mathbf{I}_{N_\mathrm{b}} + \tilde{\mathbf{H}}^H\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{H}}\right]^{-1}\right). \qquad (9)$$

3 Joint source and relay matrix optimization
In this section, we address the joint source and relay matrix optimization problem for MIMO multi-relay systems with a linear minimum mean-squared error (MMSE) receiver at the destination node. In particular, we show that optimal relay matrices have a general beamforming

structure. Based on (6) and (9), the joint source and relay matrices optimization problem can be formulated as

$$\min_{\{\mathbf{F}_i\},\mathbf{B}} \mathrm{tr}\left(\left[\mathbf{I}_{N_\mathrm{b}} + \tilde{\mathbf{H}}^H\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{H}}\right]^{-1}\right) \qquad (10)$$

$$\mathrm{s.t.}\ \mathrm{tr}\left(\mathbf{B}\mathbf{B}^H\right) \le P_\mathrm{s} \qquad (11)$$

$$\mathrm{tr}\left(\mathbf{F}_i\left[\mathbf{H}_{\mathrm{sr},i}\mathbf{B}\mathbf{B}^H\mathbf{H}_{\mathrm{sr},i}^H + \mathbf{I}_{N_\mathrm{r}}\right]\mathbf{F}_i^H\right)$$

$$\le P_{\mathrm{r},i}, \quad i = 1, \cdots, K, \qquad (12)$$

where $\{\mathbf{F}_i\} \triangleq \{\mathbf{F}_i, i = 1, \cdots, K\}$, (11) is the transmit power constraint at the source node, while (12) is the power constraint at each relay node. Here, $P_\mathrm{s} > 0$ and $P_{\mathrm{r},i} > 0, i = 1, \cdots, K$, are the corresponding power budget. Obviously, to avoid any loss of transmission power in the relay system when a linear receiver is used, there should be $N_\mathrm{b} \le \min(KN_\mathrm{r}, N_\mathrm{d})$. The problem (10)-(12) is non-convex, and a globally optimal solution of $\mathbf{B}$ and $\{\mathbf{F}_i\}$ is difficult to obtain with a reasonable computational complexity. In this paper, we develop an iterative algorithm to optimize $\mathbf{B}$ and $\{\mathbf{F}_i\}$. First, we show the optimal structure of $\{\mathbf{F}_i\}$.

3.1 Optimal structure of relay amplifying matrices
For given source matrix $\mathbf{B}$ satisfying (11), the relay matrices $\{\mathbf{F}_i\}$ are optimized by solving the following problem:

$$\min_{\{\mathbf{F}_i\}} \mathrm{tr}\left(\left[\mathbf{I}_{N_\mathrm{b}} + \tilde{\mathbf{H}}^H\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{H}}\right]^{-1}\right) \qquad (13)$$

$$\mathrm{s.t.}\ \mathrm{tr}\left(\mathbf{F}_i\left[\mathbf{H}_{\mathrm{sr},i}\mathbf{B}\mathbf{B}^H\mathbf{H}_{\mathrm{sr},i}^H + \mathbf{I}_{N_\mathrm{r}}\right]\mathbf{F}_i^H\right) \le P_{\mathrm{r},i}, \qquad i = 1, \cdots, K. \qquad (14)$$

Let us introduce the following singular value decompositions (SVDs):

$$\mathbf{H}_{\mathrm{sr},i}\mathbf{B} = \mathbf{U}_{\mathrm{s},i}\Lambda_{\mathrm{s},i}\mathbf{V}_{\mathrm{s},i}^H, \quad \mathbf{H}_{\mathrm{rd},i} = \mathbf{U}_{\mathrm{r},i}\Lambda_{\mathrm{r},i}\mathbf{V}_{\mathrm{r},i}^H, \qquad i = 1, \cdots, K, \qquad (15)$$

where $\Lambda_{\mathrm{s},i}$ and $\Lambda_{\mathrm{r},i}$ are $R_{\mathrm{s},i} \times R_{\mathrm{s},i}$ and $R_{\mathrm{r},i} \times R_{\mathrm{r},i}$ diagonal matrices, respectively. Here, $R_{\mathrm{s},i} \triangleq \mathrm{rank}(\mathbf{H}_{\mathrm{sr},i}\mathbf{B})$, $R_{\mathrm{r},i} \triangleq \mathrm{rank}(\mathbf{H}_{\mathrm{rd},i}), i = 1, \cdots, K$, and $\mathrm{rank}(\cdot)$ denotes the rank of a matrix. Based on the definition of matrix rank, $R_{\mathrm{s},i} \le \min(N_\mathrm{r}, N_\mathrm{b})$ and $R_{\mathrm{r},i} \le \min(N_\mathrm{r}, N_\mathrm{d})$. The following theorem states the structure of the optimal $\{\mathbf{F}_i\}$.

Theorem 1. *Using the SVDs of (15), the optimal structure of $\mathbf{F}_i$ as the solution to the problem (13)-(14) is given by*

$$\mathbf{F}_i = \mathbf{V}_{\mathrm{r},i}\mathbf{A}_i\mathbf{U}_{\mathrm{s},i}^H, \quad i = 1, \cdots, K, \qquad (16)$$

where $\mathbf{A}_i$ is an $R_{\mathrm{r},i} \times R_{\mathrm{s},i}$ matrix, $i = 1, \cdots, K$.

Proof. See Appendix 1. □

The remaining task is to find the optimal $\mathbf{A}_i$, $i = 1, \cdots, K$. From (31) and (32) in Appendix 1, we can equivalently rewrite the optimization problem (13)-(14) as

$$
\min_{\{\mathbf{A}_i\}} \mathrm{tr}\left(\left[\mathbf{I}_{N_\mathrm{b}} + \sum_{i=1}^{K} \mathbf{V}_{\mathrm{s},i}\Lambda_{\mathrm{s},i}\mathbf{A}_i^H \Lambda_{\mathrm{r},i}\mathbf{U}_{\mathrm{r},i}^H \right.\right.
$$
$$
\times \left(\sum_{i=1}^{K} \mathbf{U}_{\mathrm{r},i}\Lambda_{\mathrm{r},i}\mathbf{A}_i\mathbf{A}_i^H \Lambda_{\mathrm{r},i}\mathbf{U}_{\mathrm{r},i}^H + \mathbf{I}_{N_\mathrm{d}}\right)^{-1}
$$
$$
\left.\left.\times \sum_{i=1}^{K} \mathbf{U}_{\mathrm{r},i}\Lambda_{\mathrm{r},i}\mathbf{A}_i\Lambda_{\mathrm{s},i}\mathbf{V}_{\mathrm{s},i}^H\right]^{-1}\right) \qquad (17)
$$

s.t. $\mathrm{tr}\left(\mathbf{A}_i(\Lambda_{\mathrm{s},i}^2 + \mathbf{I}_{R_{\mathrm{s},i}})\mathbf{A}_i^H\right) \leq P_{\mathrm{r},i}, \qquad i = 1, \cdots, K.$ (18)

Both the problem (13)-(14) and the problem (17)-(18) have matrix optimization variables. However, in the former problem, the optimization variable $\mathbf{F}_i$ is an $N_\mathrm{r} \times N_\mathrm{r}$ matrix, while the dimension of $\mathbf{A}_i$ is $R_{\mathrm{r},i} \times R_{\mathrm{s},i}$, which may be smaller than that of $\mathbf{F}_i$. Thus, solving the problem (17)-(18) has a smaller computational complexity than solving the problem (13)-(14). In general, the problem (17)-(18) is non-convex, and a globally optimal solution is difficult to obtain with a reasonable computational complexity. Fortunately, we can resort to numerical methods, such as the projected gradient algorithm [18] to find (at least) a locally optimal solution of (17)-(18).

Theorem 2. *Let us define the objective function in (17) as* $f(\mathbf{A}_i)$. *Its gradient* $\nabla f(\mathbf{A}_i)$ *with respect to* $\mathbf{A}_i$ *can be calculated by using results on derivatives of matrices in [19] as*

$$
\nabla f(\mathbf{A}_i) = 2\left(\mathbf{R}_i^H\mathbf{M}_i^H\left(\mathbf{E}_i\mathbf{S}_i^H + \mathbf{D}_i^H\right) - \mathbf{R}_i^H\mathbf{G}_i^{-H}\mathbf{E}_i\mathbf{S}_i^H\right),
$$
$$
i = 1, \cdots, K, \qquad (19)
$$

where $\mathbf{M}_i$, $\mathbf{R}_i$, $\mathbf{S}_i$, $\mathbf{D}_i$, $\mathbf{E}_i$, *and* $\mathbf{G}_i$ *are defined in Appendix 2.*

Proof. See Appendix 2. □

In each iteration of the PG algorithm, we first obtain $\tilde{\mathbf{A}}_i = \mathbf{A}_i - s_n\nabla f(\mathbf{A}_i)$ by moving $\mathbf{A}_i$ one step towards the negative gradient direction of $f(\mathbf{A}_i)$, where $s_n > 0$ is the step size. Since $\tilde{\mathbf{A}}_i$ might not satisfy the constraint (18), we need to project it onto the set given by (18). The projected matrix $\bar{\mathbf{A}}_i$ is obtained by minimizing the Frobenius norm of $\bar{\mathbf{A}}_i - \tilde{\mathbf{A}}_i$ (according to [18]) subjecting to (18), which can be formulated as the following optimization problem:

$$
\min_{\bar{\mathbf{A}}_i} \mathrm{tr}\left((\bar{\mathbf{A}}_i - \tilde{\mathbf{A}}_i)(\bar{\mathbf{A}}_i - \tilde{\mathbf{A}}_i)^H\right) \qquad (20)
$$

s.t. $\mathrm{tr}\left(\bar{\mathbf{A}}_i\left(\Lambda_{\mathrm{s},i}^2 + \mathbf{I}_{R_{\mathrm{s},i}}\right)\bar{\mathbf{A}}_i^H\right) \leq P_{\mathrm{r},i}.$ (21)

Obviously, if $\mathrm{tr}\left(\tilde{\mathbf{A}}_i\left(\Lambda_{\mathrm{s},i}^2 + \mathbf{I}_{R_{\mathrm{s},i}}\right)\tilde{\mathbf{A}}_i^H\right) \leq P_{\mathrm{r},i}$, then $\bar{\mathbf{A}}_i = \tilde{\mathbf{A}}_i$. Otherwise, the solution to the problem (20)-(21) can be obtained by using the Lagrange multiplier method, and the solution is given by

$$
\bar{\mathbf{A}}_i = \tilde{\mathbf{A}}_i\left[(\lambda + 1)\mathbf{I}_{R_{\mathrm{s},i}} + \lambda\Lambda_{\mathrm{s},i}^2\right]^{-1},
$$

where $\lambda > 0$ is the solution to the non-linear equation of

$$
\mathrm{tr}\left(\tilde{\mathbf{A}}_i\left[(\lambda + 1)\mathbf{I}_{R_{\mathrm{s},i}} + \lambda\Lambda_{\mathrm{s},i}^2\right]^{-1}\left(\Lambda_{\mathrm{s},i}^2 + \mathbf{I}_{R_{\mathrm{s},i}}\right)\right.
$$
$$
\left.\times \left[(\lambda + 1)\mathbf{I}_{R_{\mathrm{s},i}} + \lambda\Lambda_{\mathrm{s},i}^2\right]^{-1}\tilde{\mathbf{A}}_i^H\right) = P_{\mathrm{r},i}. \qquad (22)
$$

Equation (22) can be efficiently solved by the bisection method [18].

The procedure of the PG algorithm is listed in Algorithm 1, where $(\cdot)^{(n)}$ denotes the variable at the nth iteration, δ_n and s_n are the step size parameters at the nth iteration, $\|\cdot\|$ denotes the maximum among the absolute value of all elements in the matrix, and ε is a positive constant close to 0. The step size parameters δ_n and s_n are determined by the Armijo rule [18], i.e., $s_n = s$ is a constant through all iterations, while at the nth iteration, δ_n is set to be γ^{m_n}. Here, m_n is the minimal non-negative integer that satisfies the following inequality $f\left(\mathbf{A}_i^{(n+1)}\right) - f\left(\mathbf{A}_i^{(n)}\right) \leq \alpha\gamma^{m_n}\mathrm{tr}\left(\left(\nabla f\left(\mathbf{A}_i^{(n)}\right)\right)^H\left(\bar{\mathbf{A}}_i^{(n)} - \mathbf{A}_i^{(n)}\right)\right)$, where α and γ are constants. According to [18], usually α is chosen close to 0, for example, $\alpha \in [10^{-5}, 10^{-1}]$, while a proper choice of γ is normally from 0.1 to 0.5.

3.2 Optimal source precoding matrix
With fixed $\{\mathbf{F}_i\}$, the source precoding matrix $\mathbf{B}$ is optimized by solving the following problem:

Algorithm 1 Procedure of applying the projected gradient algorithm to solve the problem (17)-(18)

---

1. Initialize the algorithm at a feasible $\mathbf{A}_i^{(0)}$ for $i = 1, \cdots, K$; set $n = 0$.
2. For $i = 1, \cdots, K$,
   Compute the gradient of (17) $\nabla f\left(\mathbf{A}_i^{(n)}\right)$;
   Project $\tilde{\mathbf{A}}_i^{(n)} = \mathbf{A}_i^{(n)} - s_n\nabla f\left(\mathbf{A}_i^{(n)}\right)$ to obtain $\bar{\mathbf{A}}_i^{(n)}$;
   Update $\mathbf{A}_i$ with $\mathbf{A}_i^{(n+1)} = \mathbf{A}_i^{(n)} + \delta_n\left(\bar{\mathbf{A}}_i^{(n)} - \mathbf{A}_i^{(n)}\right)$.
3. If $\max\|\mathbf{A}_i^{(n+1)} - \mathbf{A}_i^{(n)}\| \leq \varepsilon$, then end. Otherwise, let $n := n + 1$ and go to step 2.

$$
\min_{\mathbf{B}} \mathrm{tr}\left(\left[\mathbf{I}_{N_\mathrm{b}} + \mathbf{B}^H\Psi\mathbf{B}\right]^{-1}\right) \qquad (23)
$$

s.t. $\mathrm{tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_\mathrm{s},$ (24)

$\mathrm{tr}\left(\mathbf{F}_i\mathbf{H}_{\mathrm{sr},i}\mathbf{B}\mathbf{B}^H\mathbf{H}_{\mathrm{sr},i}^H\mathbf{F}_i^H\right) \leq \breve{P}_{\mathrm{r},i}, \qquad i = 1, \cdots, K,$ (25)

where $\Psi \triangleq \mathbf{H}_{\mathrm{sr}}^H \mathbf{F}^H \mathbf{H}_{\mathrm{rd}}^H \left(\mathbf{H}_{\mathrm{rd}} \mathbf{F} \mathbf{F}^H \mathbf{H}_{\mathrm{rd}}^H + \mathbf{I}_{N_\mathrm{d}} \right)^{-1} \mathbf{H}_{\mathrm{rd}} \mathbf{F} \mathbf{H}_{\mathrm{sr}}$, and $\check{P}_{\mathrm{r},i} \triangleq P_{\mathrm{r},i} - \mathrm{tr}\left(\mathbf{F}_i \mathbf{F}_i^H \right)$, $i = 1, \cdots, K$. Let us introduce $\Omega \triangleq \mathbf{B}\mathbf{B}^H$, and a positive semi-definite (PSD) matrix $\mathbf{X}$ with $\mathbf{X} \succeq \left(\mathbf{I}_{N_\mathrm{s}} + \Psi^{\frac{1}{2}} \Omega \Psi^{\frac{1}{2}} \right)^{-1}$, where $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a PSD matrix. By using the Schur complement [20], the problem (23)-(25) can be equivalently converted to the following problem:

$$\min_{\mathbf{X},\Omega} \mathrm{tr}\,(\mathbf{X}) - N_\mathrm{s} + N_\mathrm{b} \tag{26}$$

$$\text{s.t.} \begin{pmatrix} \mathbf{X} & \mathbf{I}_{N_\mathrm{s}} \\ \mathbf{I}_{N_\mathrm{s}} & \mathbf{I}_{N_\mathrm{s}} + \Psi^{\frac{1}{2}} \Omega \Psi^{\frac{1}{2}} \end{pmatrix} \succeq 0, \tag{27}$$

$$\mathrm{tr}\,(\Omega) \leq P_\mathrm{s}, \quad \Omega \succeq 0, \tag{28}$$

$$\mathrm{tr}\left(\mathbf{F}_i \mathbf{H}_{\mathrm{sr},i} \Omega \mathbf{H}_{\mathrm{sr},i}^H \mathbf{F}_i^H \right) \leq \check{P}_{\mathrm{r},i}, \qquad i = 1, \cdots, K. \tag{29}$$

The problem (26)-(29) is a convex semi-definite programming (SDP) problem which can be efficiently solved by the interior point method [20]. Let us introduce the eigenvalue decomposition (EVD) of $\Omega = \mathbf{U}_\Omega \Lambda_\Omega \mathbf{U}_\Omega^H$, where Λ_Ω is a $R_\Omega \times R_\Omega$ eigenvalue matrix with $R_\Omega = \mathrm{rank}(\Omega)$. If $R_\Omega = N_\mathrm{b}$, then from $\Omega = \mathbf{B}\mathbf{B}^H$, we have $\mathbf{B} = \mathbf{U}_\Omega \Lambda_\Omega^{\frac{1}{2}}$. If $R_\Omega > N_\mathrm{b}$, the randomization technique [21] can be applied to obtain a possibly suboptimal solution of $\mathbf{B}$ with rank N_b. If $R_\Omega < N_\mathrm{b}$, it indicates that the system (channel) cannot support N_b independent data streams, and thus, in this case, a smaller N_b should be chosen in the system design.

Now, the original joint source and relay optimization problem (10)-(12) can be solved by an iterative algorithm as shown in Algorithm 2, where $(\cdot)^{(m)}$ denotes the variable at the mth iteration. This algorithm is first initialized at a random feasible $\mathbf{B}$ satisfying (11). At each iteration, we first update $\{\mathbf{F}_i\}$ with fixed $\mathbf{B}$ and then update $\mathbf{B}$ with fixed $\{\mathbf{F}_i\}$. Note that the conditional updates of each matrix may either decrease or maintain but cannot increase the objective function (10). Monotonic convergence of $\{\mathbf{F}_i\}$ and $\mathbf{B}$ towards (at least) a locally optimal solution follows directly from this observation. Note that in each iteration of this algorithm, we need to update the relay amplifying matrices according to the procedure listed in Algorithm 1 at a complexity order of $\mathcal{O}\left(K \left(N_\mathrm{d}^3 + N_\mathrm{r}^3 + N_\mathrm{b}^3 \right) \right)$ and update the source precoding matrix through solving the SDP problem (26)-(29) at a complexity cost that is at most $\mathcal{O}\left(\left(N_\mathrm{s}^2 + K + 1 \right)^{3.5} \right)$ using interior point methods [22]. Therefore, the per-iteration computational complexity order of the proposed algorithm is $\mathcal{O}\left(K \left(N_\mathrm{d}^3 + N_\mathrm{r}^3 + N_\mathrm{b}^3 \right) + \left(N_\mathrm{s}^2 + K + 1 \right)^{3.5} \right)$. The overall complexity of this algorithm depends on the number of iterations until convergence, which will be studied in the next section.

Algorithm 2 Procedure of solving the problem (10)-(12)

1. Initialize the algorithm at a feasible $\mathbf{B}^{(0)}$ satisfying constraint (11); set $m = 0$.
2. For fixed $\mathbf{B}^{(m)}$, obtain $\{\mathbf{F}_i\}^{(m)}$ by solving the problem (17)-(18) using the PG algorithm.
3. Update $\mathbf{B}^{(m+1)}$ by solving the problem (26)-(29) with known $\{\mathbf{F}_i\}^{(m)}$.
4. If $\|\mathbf{B}^{(m+1)} - \mathbf{B}^{(m)}\| \leq \varepsilon$, then end. Otherwise, let $m := m + 1$ and go to step 2.

---

4 Simulations

In this section, we study the performance of the proposed jointly optimal source and relay matrix design for MIMO multi-relay systems with linear MMSE receiver. All simulations are conducted in a flat Rayleigh fading environment where the channel matrices have zero-mean entries with variances $\sigma_\mathrm{s}^2/N_\mathrm{s}$ and $\sigma_\mathrm{r}^2/(KN_\mathrm{r})$ for $\mathbf{H}_{\mathrm{sr}}$ and $\mathbf{H}_{\mathrm{rd}}$, respectively. For the sake of simplicity, we assume $P_{\mathrm{r},i} = P_\mathrm{r}$, $i = 1, \cdots, K$. The BPSK constellations are used to modulate the source symbols, and all noises are i.i.d. Gaussian with zero mean and unit variance. We define $\mathrm{SNR}_\mathrm{s} = \sigma_\mathrm{s}^2 P_\mathrm{s} K N_\mathrm{r}/N_\mathrm{s}$ and $\mathrm{SNR}_\mathrm{r} = \sigma_\mathrm{r}^2 P_\mathrm{r} N_\mathrm{d}/(KN_\mathrm{r})$ as the signal-to-noise ratio (SNR) for the source-relay link and the relay-destination link, respectively. We transmit $1000N_\mathrm{s}$ randomly generated bits in each channel realization, and all simulation results are averaged over 200 channel realizations. In all simulations, we set $N_\mathrm{b} = N_\mathrm{s} = N_\mathrm{r} = N_\mathrm{d} = 3$, and the MMSE linear receiver in (8) is employed at the destination for symbol detection.

In the first example, a MIMO relay system with $K = 3$ relay nodes is simulated. We compare the normalized MSE performance of the proposed joint source and relay optimization algorithm using the projected gradient (JSR-PG) algorithm in Algorithm 2, the optimal relay-only algorithm using the projected gradient (ORO-PG) algorithm in Algorithm 1 with $\mathbf{B} = \sqrt{P_\mathrm{s}/N_\mathrm{s}}\mathbf{I}_{N_\mathrm{s}}$, and the naive amplify-and-forward (NAF) algorithm. Figure 2 shows the normalized MSE of all algorithms versus SNR_s for $\mathrm{SNR}_\mathrm{r} = 20$ dB. While Figure 3 demonstrates the normalized MSE of all algorithms versus SNR_r for SNR_s fixed at 20 dB. It can be seen from Figures 2 and 3 that the JSR-PG and ORO-PG algorithms have a better performance than the NAF algorithm over the whole SNR_s and SNR_r range. Moreover, the proposed JSR-PG algorithm yields the lowest MSE among all three algorithms.

The number of iterations required for the JSR-PG algorithm to converge to $\varepsilon = 10^{-3}$ in a typical channel realization are listed in Table 1, where we set $K = 3$ and $\mathrm{SNR}_\mathrm{r} = 20$ dB. It can be seen that the JSR-PG algorithm converges within several iterations, and thus, it is realizable with the advancement of modern chip design.
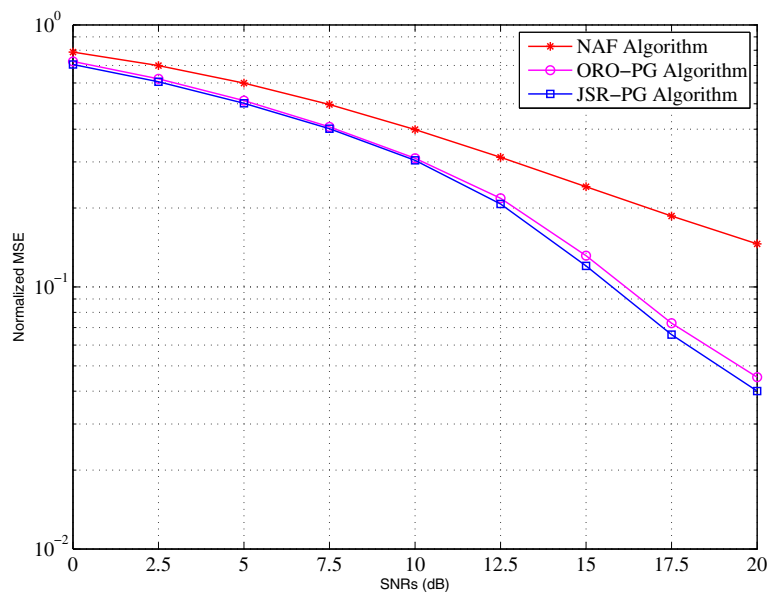
Figure 2 Example 1. Normalized MSE versus SNR_s with $K = 3$, $SNR_r = 20$ dB.

In the second example, we compare the bit error rate (BER) performance of the proposed JSR-PG algorithm in Algorithm 2, the ORO-PG algorithm in Algorithm 1, the suboptimal source and relay matrix design in [17], the one-way relay version of the conjugate gradient-based source and relay algorithm in [16], and the NAF algorithm. Figure 4 displays the system BER versus SNR_s for a MIMO relay system with $K = 3$ relay nodes and fixed SNR_r at 20 dB. It can be seen from Figure 4 that the proposed JSR-PG algorithm has a better BER performance than the existing algorithms over the whole SNR_s range.

In the third example, we study the effect of the number of relay nodes to the system BER performance using the JSR-PG and ORO-PG algorithms. Figure 5 displays the system BER versus SNR_s with $K = 2$, 3, and 5 for fixed SNR_r at 20 dB. It can be seen that at BER $= 10^{-2}$, for both the ORO-PG algorithm and JSR-PG algorithm, we can achieve approximately 3-dB gain by increasing from $K = 2$ to $K = 5$. It can also be seen that the performance gain of the JSR-PG algorithm over the ORO-PG algorithm increases with the increasing number of relay nodes.
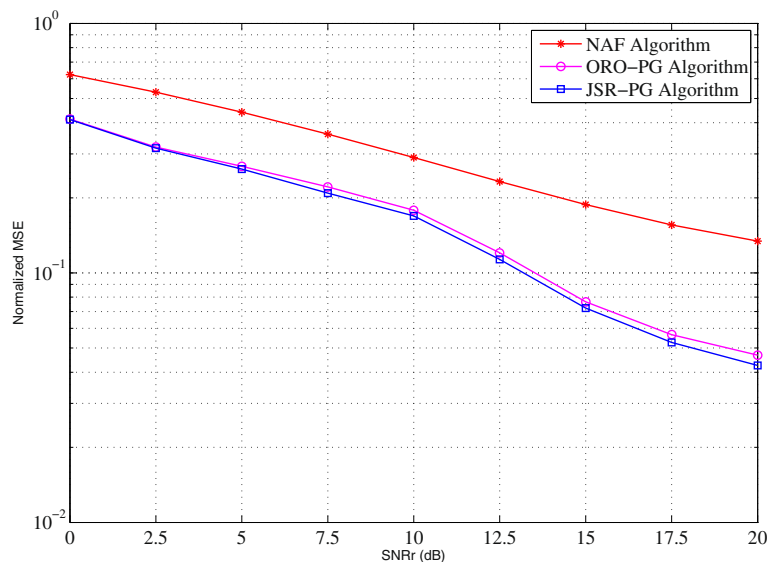

Figure 3 Example 1. Normalized MSE versus SNR_r with $K = 3$, $SNR_s = 20$ dB.

Table 1 Iterations required until convergence in the JSR-PG algorithm

SNR$_s$ (dB)	Iterations
0	3
2.5	3
5	3
7.5	4
10	4
12.5	5
15	5
17.5	5
20	6

5 Conclusions

In this paper, we have derived the general structure of the optimal relay amplifying matrices for linear non-regenerative MIMO relay communication systems with multiple relay nodes using the projected gradient approach. The proposed source and relay matrices minimize the MSE of the signal waveform estimation. The simulation results demonstrate that the proposed algorithm has improved the MSE and BER performance compared with existing techniques.

Appendices

Appendix 1

Proof of Theorem 1

Without loss of generality, $\mathbf{F}_i$ can be written as

$$
\mathbf{F}_i = \begin{bmatrix} \mathbf{V}_{r,i} & \mathbf{V}_{r,i}^\perp \end{bmatrix} \begin{bmatrix} \mathbf{A}_i & \mathbf{X}_i \\ \mathbf{Y}_i & \mathbf{Z}_i \end{bmatrix} \begin{bmatrix} \mathbf{U}_{s,i}^H \\ (\mathbf{U}_{s,i}^\perp)^H \end{bmatrix}, \qquad i = 1, \cdots, K,
$$

(30)

where $\mathbf{V}_{r,i}^\perp (\mathbf{V}_{r,i}^\perp)^H = \mathbf{I}_{N_r} - \mathbf{V}_{r,i}\mathbf{V}_{r,i}^H$, $\mathbf{U}_{s,i}^\perp (\mathbf{U}_{s,i}^\perp)^H = \mathbf{I}_{N_r} - \mathbf{U}_{s,i}\mathbf{U}_{s,i}^H$, such that $\bar{\mathbf{V}}_{r,i} \triangleq \begin{bmatrix} \mathbf{V}_{r,i}, \mathbf{V}_{r,i}^\perp \end{bmatrix}$ and $\bar{\mathbf{U}}_{s,i} \triangleq \begin{bmatrix} \mathbf{U}_{s,i}, \mathbf{U}_{s,i}^\perp \end{bmatrix}$ are $N_r \times N_r$ unitary matrices. Matrices $\mathbf{A}_i, \mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i$ are arbitrary matrices with dimensions of $R_{r,i} \times R_{s,i}$, $R_{r,i} \times (N_r - R_{s,i})$, $(N_r - R_{r,i}) \times R_{s,i}$, $(N_r - R_{r,i}) \times (N_r - R_{s,i})$, respectively. Substituting (15) and (30) back into (13), we obtain that $\mathbf{H}_{rd,i}\mathbf{F}_i\mathbf{H}_{sr,i}\mathbf{B} = \mathbf{U}_{r,i}\Lambda_{r,i}\mathbf{A}_i\Lambda_{s,i}\mathbf{V}_{s,i}^H$ and $\mathbf{H}_{rd,i}\mathbf{F}_i\mathbf{F}_i^H\mathbf{H}_{rd,i}^H = \mathbf{U}_{r,i}\Lambda_{r,i}(\mathbf{A}_i\mathbf{A}_i^H + \mathbf{X}_i\mathbf{X}_i^H)\Lambda_{r,i}\mathbf{U}_{r,i}^H$. Thus, we can rewrite (13) as

$$
\text{MSE tr} \left(\left[\mathbf{I}_{N_b} + \sum_{i=1}^{K} \mathbf{V}_{s,i}\Lambda_{s,i}\mathbf{A}_i^H\Lambda_{r,i}\mathbf{U}_{r,i}^H \right. \right.
$$

$$
\times \left(\sum_{i=1}^{K} \mathbf{U}_{r,i}\Lambda_{r,i}(\mathbf{A}_i\mathbf{A}_i^H + \mathbf{X}_i\mathbf{X}_i^H)\Lambda_{r,i}\mathbf{U}_{r,i}^H + \mathbf{I}_{N_d} \right)^{-1}
$$

$$
\left. \left. \times \sum_{i=1}^{K} \mathbf{U}_{r,i}\Lambda_{r,i}\mathbf{A}_i\Lambda_{s,i}\mathbf{V}_{s,i}^H \right]^{-1} \right).
$$

(31)

It can be seen that (31) is minimized by $\mathbf{X}_i = \mathbf{0}_{R_{r,i} \times (N_r - R_{s,i})}$, $i = 1, \cdots, K$.

Substituting (15) and (30) back into the left-hand side of the transmission power constraint (14), we have

$$
\text{tr} \left(\mathbf{F}_i [\mathbf{H}_{sr,i}\mathbf{B}\mathbf{B}^H\mathbf{H}_{sr,i}^H + \mathbf{I}_{N_r}] \mathbf{F}_i^H \right)
$$
$$
= \text{tr} \left(\mathbf{A}_i (\Lambda_{s,i}^2 + \mathbf{I}_{R_{s,i}}) \mathbf{A}_i^H + \mathbf{Y}_i (\Lambda_{s,i}^2 + \mathbf{I}_{R_{s,i}}) \mathbf{Y}_i^H \right.
$$
$$
\left. + \mathbf{X}_i\mathbf{X}_i^H + \mathbf{Z}_i\mathbf{Z}_i^H \right), \quad i = 1, \cdots, K.
$$

(32)

From (32), we find that $\mathbf{X}_i = \mathbf{0}_{R_{r,i} \times (N_r - R_{s,i})}$, $\mathbf{Y}_i = \mathbf{0}_{(N_r - R_{r,i}) \times R_{s,i}}$, and $\mathbf{Z}_i = \mathbf{0}_{(N_r - R_{r,i}) \times (N_r - R_{s,i})}$ minimize the power consumption at each relay node. Thus, we have $\mathbf{F}_i = \mathbf{V}_{r,i}\mathbf{A}_i\mathbf{U}_{s,i}^H$, $i = 1, \cdots, K$.
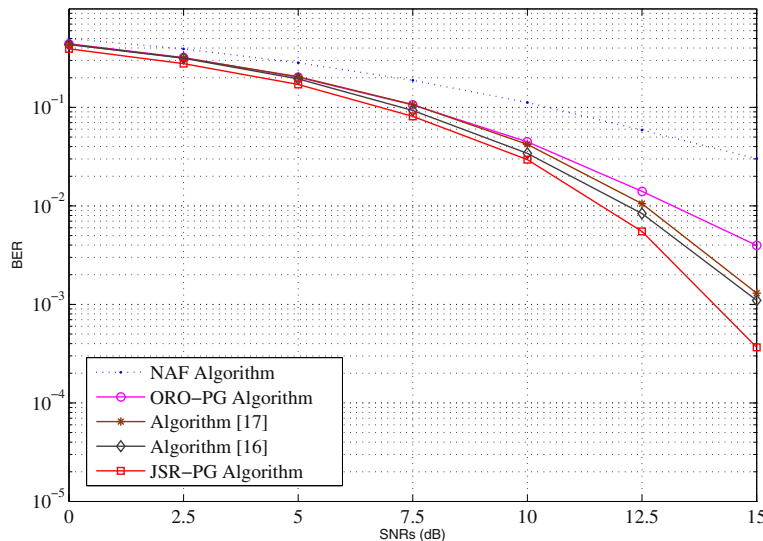

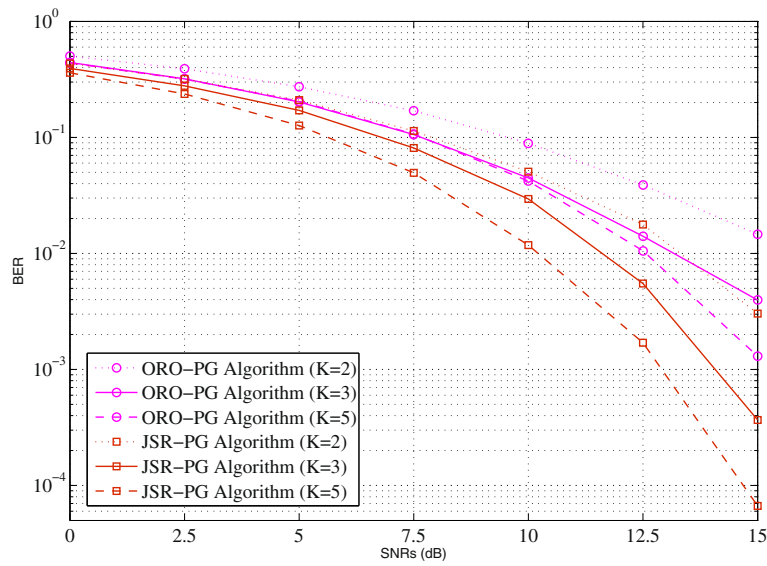**Figure 4 Example 2.** BER versus SNR$_s$ with $K = 3$, SNR$_r = 20$ dB.

Figure 5 Example 3. BER versus SNR$_s$ for different K, SNR$_r$ = 20 dB.

Appendix 2
Proof of Theorem 2

Let us define $\mathbf{Z}_i \triangleq \sum_{j=1, j \neq i}^{K} \mathbf{U}_{\mathrm{r},j} \Lambda_{\mathrm{r},j} \mathbf{A}_j \Lambda_{\mathrm{s},j} \mathbf{V}_{\mathrm{s},j}^H$ and $\mathbf{Y}_i \triangleq \sum_{j=1, j \neq i}^{K} \mathbf{U}_{\mathrm{r},j} \Lambda_{\mathrm{r},j} \mathbf{A}_j \mathbf{A}_j^H \Lambda_{\mathrm{r},j} \mathbf{U}_{\mathrm{r},j}^H + \mathbf{I}_{N_{\mathrm{d}}}$. Then, $f(\mathbf{A}_i)$ can be written as

$$
\begin{aligned}
f(\mathbf{A}_i) = \mathrm{tr} \Big(& \big[\mathbf{I}_{N_{\mathrm{b}}} + \big(\mathbf{Z}_i^H + \mathbf{V}_{\mathrm{s},i} \Lambda_{\mathrm{s},i} \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big) \\
& \times \big(\mathbf{Y}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big)^{-1} \\
& \times \big(\mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big) \big]^{-1} \Big).
\end{aligned}
\tag{33}
$$

Applying $\big(\mathbf{I}_{N_{\mathrm{b}}} + \mathbf{A}^H \mathbf{C}^{-1} \mathbf{A} \big)^{-1} = \mathbf{I}_{N_{\mathrm{b}}} - \mathbf{A}^H (\mathbf{A}\mathbf{A}^H + \mathbf{C})^{-1} \mathbf{A}$, (33) can be written as

$$
\begin{aligned}
f(\mathbf{A}_i) = \mathrm{tr} \Big(& \mathbf{I}_{N_{\mathrm{b}}} - \big(\mathbf{Z}_i^H + \mathbf{V}_{\mathrm{s},i} \Lambda_{\mathrm{s},i} \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big) \\
& \times \big[\big(\mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big) \\
& \times \big(\mathbf{Z}_i^H + \mathbf{V}_{\mathrm{s},i} \Lambda_{\mathrm{s},i} \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big) \\
& + \mathbf{Y}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big]^{-1} \\
& \times \big(\mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big) \Big).
\end{aligned}
\tag{34}
$$

Let us now define $\mathbf{E}_i \triangleq \mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H$, $\mathbf{K}_i \triangleq \mathbf{Y}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H$, and $\mathbf{G}_i \triangleq \mathbf{E}_i \mathbf{E}_i^H + \mathbf{K}_i$. We can rewrite (34) as

$$
f(\mathbf{A}_i) = \mathrm{tr} \Big(\mathbf{I}_{N_{\mathrm{b}}} - \mathbf{E}_i^H \mathbf{G}_i^{-1} \mathbf{E}_i \Big) = \mathrm{tr} \Big(\mathbf{I}_{N_{\mathrm{b}}} - \mathbf{E}_i \mathbf{E}_i^H \mathbf{G}_i^{-1} \Big).
\tag{35}
$$

The derivative of $f(\mathbf{A}_i)$ with respect to $\mathbf{A}_i$ is given by

$$
\begin{aligned}
\frac{\partial f(\mathbf{A}_i)}{\partial \mathbf{A}_i} = & - \frac{\partial}{\partial \mathbf{A}_i} \mathrm{tr} \Big(\mathbf{E}_i \mathbf{E}_i^H \mathbf{G}_i^{-1} \Big) \\
= & \frac{\partial}{\partial \mathbf{A}_i} \mathrm{tr} \Big(\mathbf{G}_i^{-1} \mathbf{E}_i \mathbf{E}_i^H \mathbf{G}_i^{-1} \big(\big(\mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big) \mathbf{E}_i^H \\
& + \mathbf{Y}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H \big) \Big) \\
& - \frac{\partial}{\partial \mathbf{A}_i} \mathrm{tr} \Big(\mathbf{E}_i^H \mathbf{G}_i^{-1} \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \Big).
\end{aligned}
\tag{36}
$$

Defining $\mathbf{M}_i \triangleq \mathbf{G}_i^{-1} \mathbf{E}_i \mathbf{E}_i^H \mathbf{G}_i^{-1}$, $\mathbf{R}_i \triangleq \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i}$, $\mathbf{S}_i \triangleq \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H$, and $\mathbf{D}_i \triangleq \mathbf{A}_i^H \Lambda_{\mathrm{r},i} \mathbf{U}_{\mathrm{r},i}^H$, we can rewrite (36) as

$$
\begin{aligned}
\frac{\partial f(\mathbf{A}_i)}{\partial \mathbf{A}_i} = & \frac{\partial}{\partial \mathbf{A}_i} \mathrm{tr} \big(\mathbf{M}_i \big(\mathbf{Z}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big) \mathbf{E}_i^H \\
& + \mathbf{M}_i \big(\mathbf{Y}_i + \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \mathbf{A}_i \mathbf{D}_i \big) \big) \\
& - \big(\mathbf{E}_i^H \mathbf{G}_i^{-1} \mathbf{U}_{\mathrm{r},i} \Lambda_{\mathrm{r},i} \big)^T \big(\Lambda_{\mathrm{s},i} \mathbf{V}_{\mathrm{s},i}^H \big)^T \\
= & \frac{\partial}{\partial \mathbf{A}_i} \mathrm{tr} \big(\mathbf{M}_i \mathbf{R}_i \mathbf{A}_i \mathbf{S}_i \mathbf{E}_i^H + \mathbf{M}_i \mathbf{R}_i \mathbf{A}_i \mathbf{D}_i \big) \\
& - \big(\mathbf{E}_i^H \mathbf{G}_i^{-1} \mathbf{R}_i \big)^T \mathbf{S}_i^T.
\end{aligned}
\tag{37} \tag{38}
$$

Finally, the gradient of $f(\mathbf{A}_i)$ is given by

$$
\begin{aligned}
\nabla f(\mathbf{A}_i) & = 2 \left(\frac{\partial f(\mathbf{A}_i)}{\partial \mathbf{A}_i} \right)^* \\
& = 2 \left((\mathbf{M}_i \mathbf{R}_i)^T \big(\mathbf{S}_i \mathbf{E}_i^H \big)^T + (\mathbf{M}_i \mathbf{R}_i)^T \mathbf{D}_i^T - \big(\mathbf{E}_i^H \mathbf{G}_i^{-1} \mathbf{R}_i \big)^T \mathbf{S}_i^T \right)^* \\
& = 2 \left(\mathbf{R}_i^H \mathbf{M}_i^H \big(\mathbf{E}_i \mathbf{S}_i^H + \mathbf{D}_i^H \big) - \mathbf{R}_i^H \mathbf{G}_i^{-H} \mathbf{E}_i \mathbf{S}_i^H \right)
\end{aligned}
\tag{39}
$$

where $(\cdot)^*$ stands for complex conjugate.

Author details
[1] Department of Electrical Engineering, Universitas Kristen Indonesia Paulus, Jln. Perintis Kemerdekaan No 28 Daya, Makassar 90243, Indonesia. [2] Department of Electronic and Electrical Engineering, University College London, Gower Street, London WC1E 7JE, UK. [3] Department of Electrical and Computer Engineering, Curtin University of Technology, Bentley, WA 6102, Australia.

References
1. B Wang, J Zhang, A Høst-Madsen, On the capacity of MIMO relay channels. IEEE Trans. Inf. Theory. **51**, 29–43 (2005)
2. X Tang, Y Hua, Optimal design of non-regenerative MIMO wireless relays. IEEE Trans. Wireless Commun. **6**, 1398–1407 (2007)
3. O Muñoz-Medina, J Vidal, A Agustín, Linear transceiver design in nonregenerative relays with channel state information. IEEE Trans. Signal Process. **55**, 2593–2604 (2007)
4. W Guan, H Luo, Joint MMSE transceiver design in non-regenerative MIMO relay systems. IEEE Commun. Lett. **12**, 517–519 (2008)
5. G Li, Y Wang, T Wu, J Huang, Joint linear filter design in multi-user cooperative non-regenerative MIMO relay systems. EURASIP J. Wireless Commun. Netw. **2009**, 670265 (2009)
6. Y Rong, Linear non-regenerative multicarrier MIMO relay communications based on MMSE criterion. IEEE Trans. Commun. **58**, 1918–1923 (2010)
7. Z Fang, Y Hua, JC Koshy, Joint source and relay optimization for a non-regenerative MIMO relay, in *Proc. IEEE Workshop Sensor Array Multi-Channel Signal Process* (Waltham, WA, USA, 12–14 July 2006, pp. 239–243
8. Y Rong, X Tang, Y Hua, A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems. IEEE Trans. Signal Process. **57**, 4837–4851 (2009)
9. Y Rong, Y Hua, Optimality of diagonalization of multi-hop MIMO relays. IEEE Trans. Wireless Commun. **8**, 6068–6077 (2009)
10. Y Rong, Joint source and relay optimization for two-way linear non-regenerative MIMO relay communications. IEEE Trans. Signal Process. **60**, 6533–6546 (2012)
11. MRA Khandaker, Y Rong, Joint transceiver optimization for multiuser MIMO relay communication systems. IEEE Trans. Signal Process. **60**, 5997–5986 (2012)
12. MRA Khandaker, Y Rong, Joint source and relay optimization for multiuser MIMO relay communication systems, in *Proceedings of the 4th International Conference on Signal Processing and Communication Systems* Gold Coast, Australia, 13–15 Dec 2010
13. H Wan, W Chen, Joint source and relay design for multiuser MIMO nonregenerative relay networks with direct links. IEEE Trans. Veh. Technol. **61**, 2871–2876 (2012)
14. J Zeng, Z Chen, L Li, Iterative joint source and relay optimization for multiuser MIMO relay systems, in *Proceedings of the IEEE Vehicular Technology Conference* Quebec City, QC, Canada, 3–6 Sept 2012
15. AS Behbahani, R Merched, AM Eltawil, Optimizations of a MIMO relay network. IEEE Trans. Signal Process. **56**, 5062–5073 (2008)
16. C-C Hu, Y-F Chou, Precoding design of MIMO AF two-way multiple-relay systems. IEEE Signal Process. Lett. **20**, 623–626 (2013)
17. A Toding, MRA Khandaker, Y Rong, Joint source and relay optimization for parallel MIMO relay networks. EURASIP J. Adv. Signal Process. **2012**, 174 (2012)
18. DP Bertsekas, *Nonlinear Programming*, 2nd edn, Athena Scientific, Belmont,1999)
19. KB Petersen, MS Petersen, The Matrix Cookbook. http://www2.imm.dtu.dk/pubdb/p.php?3274. Accessed 9 Sept 2014
20. S Boyd, L Vandenberghe, *Convex Optimization*. (Cambridge University Press, Cambridge, 2004)
21. P Tseng, Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation. SIAM J. Optim. **14**(1), 268–283 (2003)
22. Y Nesterov, A Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*, (SIAM, Philadelphia, 1994)