

On Linear Precoding of Nonregenerative MIMO Relay Networks for Finite-Alphabet Source

Xiao Liang¹, Member, IEEE, Zhi Ding, Fellow, IEEE, and Chengshan Xiao, Fellow, IEEE

Abstract—Multiple input and multiple output (MIMO) relay could provide broader wireless coverage, better diversity, and higher throughput. Most existing precoder designs for either source or relay node are based on the assumption of Gaussian input signals. However, recent works have revealed possible performance loss of MIMO systems originally optimized for Gaussian source signals when applied to practical finite-alphabet source signals. In this work, we investigate the design problem of joint MIMO precoding for wireless two-hop nonregenerative cooperative relay networks under finite-alphabet source signals. We identify several structural properties of optimal precoders. Specifically, we provided the optimal left singular vectors of the relay precoder, and proved the convexity of mutual information with respect to the square of relay precoder singular value. These results generalize the two-hop relay networks in Gaussian input assumption to the cooperative relay networks in arbitrary finite-alphabet input signals. Furthermore, we propose gradient-based numerical iterative optimization algorithms not only for arbitrary finite-alphabet source signal precoding but also for cooperative relay networks which may or may not have a direct source to destination link. Our results demonstrate substantial performance improvement over existing precoder designed traditionally under Gaussian input assumption, which indicates that the waterfilling based precoding strategy is not suitable for finite-alphabet constellation source inputs.

Index Terms—Amplify-and-forward, mutual information, minimum mean-square error, convex.

I. INTRODUCTION

RELAY stations have already been widely deployed in wireless networks, such as 4G long-term evolution (LTE) and wireless local area networks [1], [2]. Specifically, when the des-

tinuation is far from the source (e.g. at the edge of cell), a relay station activated between the source and destination can substantially improve source-to-destination transport. Compared with point-to-point transmissions, such a three-node cooperative relay networks [3], [4] is a promising approach to offer higher power efficiency and throughput, deliver better reliability through diversity, reduce potential interference among users and provide broader wireless coverage. Generally, there exist three different types of relays: amplify-and-forward (AF) [5], decode-and-forward (DF) [6] and compress-and-forward (CF) retransmissions [7]. AF protocol is usually the simplest realization with shorter processing delay and potential lower cost.

Therefore, the problem of optimum cooperative AF relay networks design has attracted substantial research attention in recent years as a simple form of cooperative wireless networking topology [8]. Many researches have attempted to jointly optimize source and relay transmission, especially in MIMO scenarios. For instance, Behbahani and coauthors [9] investigated MIMO relay network design to minimize MMSE and maximize signal-noise ratio (SNR). Tang, *et al.* in [10] studied relay precoder design for maximizing channel capacity by providing a closed form solution without direct source to destination link. Source precoding was also jointly optimized with relay precoder in [11]. Rong, *et al.* in [12] further extended this three-node relay prototype with a unified work for MMSE and capacity criterion. [13] presented semi-analytical derivations of the achievable rate of relay networks. To accommodate a generalized configuration with multiple sources, multiple relays, and multiple destinations, a unified approach to optimal transceiver design for AF MIMO relaying has been investigated in [14]. Meanwhile, without dropping the source-destination link, [15] further presented a semi-closed form solution for AF MIMO relay. By considering the impact of imperfect channel state information (CSI), [16] studied the precoder design and the achieved information rate for AF MIMO relay network.

It should be noted that the aforementioned works were implicitly based on the well known assumption of Gaussian source signals. However, practical communication modulate source signals use a finite constellation alphabet (e.g. QAM). In such case, [17] proposed a mercury/waterfilling (MWF) power allocation algorithm for parallel channels, and illustrated severe performance loss when using traditional waterfilling algorithm. [18] and [19] further investigated the optimal precoder design for point-to-point MIMO channel. To reduce the computational burden of precoder design, [20] derived a lower bound of mutual information. Recently, the precoder design with finite constellation alphabet constraints has been studied for scenarios of the generalized spatial modulation [21], MIMO H-ARQ channel [22], multiple access channel [23], relay networks [24], cognitive radio networks [25], interference channel [26], multi-input

Manuscript received March 27, 2017; accepted June 9, 2017. Date of publication June 21, 2017; date of current version November 10, 2017. This work was supported in part by 973 Program under Grant 2013CB329204, in part by the National Natural Science Foundation of China under Grant 61521061, Grant 61571107, Grant 61601115, and Grant 61461136003, and in part by the National Science Foundation of Jiangsu Province under Grant BK20160069. The work of Z. Ding was supported by the National Science Foundation under Grant CNS-1702752 and Grant CCF-1443870. The work of C. Xiao was supported in part by USA National Science Foundation under Grant ECCS-1231848 and Grant ECCS-1539316. The review of this paper was coordinated by Dr. H. Zhu. This work was presented in part at the IEEE International Conference on Communications, Ottawa, ON, Canada, June 2012. (*Corresponding author: Xiao Liang.*)

X. Liang is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (w-mail: xiaoliang@seu.edu.cn).

Z. Ding is with Southeast University, Nanjing 210096, China. He is now with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: zding@ucdavis.edu).

C. Xiao is with the Department of Electrical & Computer Engineering, Missouri University of Science & Technology, Rolla, MO 65409 USA (e-mail: xiaoc@mst.edu).

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/TVT.2017.2717925

single-output downlink multicasting channel [27], and cooperative multi-cell MIMO downlink channel [28], among others. All results have exhibited the potential performance loss in terms of achievable mutual information between source and destination if designs optimized for Gaussian sources are naively applied to practical systems driven by finite-alphabet input signals.

In this work, we consider the well known two-hop three-node relay network of multi-antenna transceivers under finite-alphabet inputs. The source signal are first precoded before transmission. Then, an AF node relays its received signals after linear precoding to destination. Both source and relay nodes are under the sum power constraints, respectively. We study optimal designs of source-relay precoders that maximize the system mutual information.

Unlike the case involving Gaussian inputs, the literature does not contain much work on rate maximization for relay networks with finite-alphabet inputs. Since the mutual information of relay networks is generally non-concave with respect to either source precoder or relay precoder, this maximization of mutual information for rate increase is a more challenging problem. Moreover, the mutual information generally lacks closed-form expression, and it also involves higher computational complexity. In this work, we present several important insights and propose a gradient based iterative numerical optimization algorithm for a heuristic precoder structures for source and relay precoders with and without source-to-destination channel. More specifically,

- 1) We prove that the left singular vectors of the relay precoder coincide with the right singular vectors of the relay-to-destination MIMO channel.
- 2) Extending results of [19] for point-to-point MIMO transceiver, we further prove that the mutual information is a concave function with respect to a quadratic function of the singular values of the relay precoder.
- 3) By adopting the optimal singular vectors generated according to Gaussian input assumption, we prove that the mutual information is also a concave function of source precoder power. Thus, we extend results in [12] to the same system under finite-alphabet constellations.

Our contributions differ from two published works on relay networks with finite-alphabet constraints. In comparison with the works in [24], we consider the more general MIMO relay channels whereas [24] considers single antenna transceivers. Second, the problem in [24] can be completely transformed into a point-to-point MIMO problem whereas our focus is on a more general relay problem which poses practical challenges and cannot be similarly mapped. Our results also extend the work of [29], which only targets the design of relay precoder but not at the source. Furthermore, due to the space limitation of conference paper, the proof of those conclusions are absent. In this work, we provide a full proof for those conclusions. Moreover, we present a joint source-relay precoders design that can achieve substantial performance improvement.

We organize this manuscript as follows. Section II describes the three-node relay networks model and presents the basic problem formulation and preliminary information. Section III presents our main results on the convexity of some singular-value decomposition (SVD) components of precoders and results on the derivatives of mutual information with respect to some precoder parameters. In Section IV, we propose a special precoder structure and utilize our main results to design gradient based optimization algorithms. We present the simulation results in Section V before concluding with Section VI.

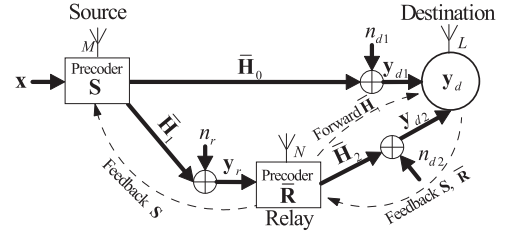


Fig. 1. A two-hop three-node orthogonal MIMO relay model.

II. SYSTEM MODEL AND PRELIMINARIES

A. Notations

Normal fonts denote scalars. Bold lower and bold upper case symbols denote vectors and matrices, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^\dagger$ denote conjugate, transpose and conjugate transpose operation, respectively. Vector $\mathbf{e}_i$ is the standard column vector whose elements are zero except for its i -th unit element. $\text{Tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. $\|\mathbf{A}\| = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^\dagger)}$ is the Frobenius norm of $\mathbf{A}$. $\mathbb{E}\{\cdot\}$ denotes statistical expectation. $(b)_+ = \max(b, 0)$. $\text{vec}(\mathbf{A})$ returns a column vector from matrix $\mathbf{A}$ by stacking its column vectors. $\otimes$ and $\odot$ are the Kronecker product and Hadamard product, respectively. $\oplus$ is the Kronecker sum operator. For $\mathbf{A}$ and $\mathbf{B}$ as m -by- m and n -by- n matrices respectively, $\mathbf{A} \oplus \mathbf{B}$ is an mn -by- mn matrix $(\mathbf{A} \otimes \mathbf{I}_n) + (\mathbf{I}_m \otimes \mathbf{B})$. $[\mathbf{A}]_i$, with one subscript, extracts the i -th column from $\mathbf{A}$ whereas $[\mathbf{A}]_{ij}$ denotes the (i, j) -th element of $\mathbf{A}$. Suppose $\mathbf{A}$ is positive semi-definite, and the corresponding eigen-decomposition of $\mathbf{A}$ is $\mathbf{U}_\mathbf{A} \Sigma_\mathbf{A} \mathbf{U}_\mathbf{A}^H$. Then the square-root of $\mathbf{A}$ is defined as $\mathbf{A}^{\frac{1}{2}} \triangleq \mathbf{U}_\mathbf{A} \Sigma_\mathbf{A}^{\frac{1}{2}} \mathbf{U}_\mathbf{A}^H$. Furthermore, when $\mathbf{A}$ is strict positive definite, $\mathbf{A}^{-\frac{1}{2}} \triangleq \mathbf{U}_\mathbf{A} \Sigma_\mathbf{A}^{-\frac{1}{2}} \mathbf{U}_\mathbf{A}^H$.

B. System Model

Consider a two-hop three-node MIMO relay network shown in Fig. 1. Assume the source-to-destination channel $\bar{\mathbf{H}}_0$ and relay-to-destination channel $\bar{\mathbf{H}}_2$ are orthogonal, either in time or frequency. Source node broadcasts signals sequentially. When the source-to-destination channel is unreliable, relay may assist destination by amplifying and forwarding its received signals. Based on both sets of signal received from the source and the relay, the destination can jointly detect the source signals. Our focus is the design of linear MIMO precoders $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}$, respectively, for relay and source, under individual power constraints.

In our design, we assume that the precoder optimization is carried out at the destination. Before the finite-precision effect evaluations on CSI and precoders in Section V-E, we suppose that the destination feedbacks the accurate source and relay precoders to relay node, and the relay node forwards the accurate source-to-relay CSI and feedbacks the accurate source precoder, respectively, to the destination and source through a dedicated, band unlimited and error free channel, respectively.

We denote the numbers of antennas at source, destination and relay to be M , L and N , respectively. For reasons stated in [12], we consider square precoders $\bar{\mathbf{R}} \in \mathcal{C}^{N \times N}$ and $\bar{\mathbf{S}} \in \mathcal{C}^{M \times M}$. Our diagram denotes the source-to-destination channel, the source-to-relay channel and the relay-to-destination channel as $\bar{\mathbf{H}}_0 \in \mathcal{C}^{L \times M}$, $\bar{\mathbf{H}}_1 \in \mathcal{C}^{N \times M}$ and $\bar{\mathbf{H}}_2 \in \mathcal{C}^{L \times N}$, respectively.

Let $\mathbf{x}$ be the source signal vector whose elements are independent, zero mean with unit power from finite alphabets. Let $\mathbf{n}_r \in \mathcal{CN}(\mathbf{0}, \sigma_r^2 \mathbf{I})$ be the relay channel noise vector. The

received signal vector at relay is

$$\mathbf{y}_r = \sqrt{\frac{P_s}{M}} \bar{\mathbf{H}}_1 \mathbf{S} \mathbf{x} + \mathbf{n}_r \quad (1)$$

where $\frac{P_s}{M}$ is the normalized average power on each transmit antenna such that P_s is the total source transmission power.

At the destination, the received vectors from source and AF relay, respectively, are

$$\mathbf{y}_{d1} = \sqrt{\frac{P_s}{M}} \bar{\mathbf{H}}_0 \mathbf{S} \mathbf{x} + \mathbf{n}_{d1} \quad (2)$$

$$\mathbf{y}_{d2} = \bar{\mathbf{H}}_2 \bar{\mathbf{R}} \mathbf{y}_r + \mathbf{n}_{d2} \quad (3)$$

where $\mathbf{n}_{d1} \in \mathcal{CN}(\mathbf{0}, \sigma_{d1}^2 \mathbf{I})$ and $\mathbf{n}_{d2} \in \mathcal{CN}(\mathbf{0}, \sigma_{d2}^2 \mathbf{I})$ are the noise vectors for the orthogonal source-destination channel and relay-destination channel, respectively. We can first whiten the noise in $\mathbf{y}_{d2}$ by applying an invertible noise whitening filter

$$\mathbf{F} \triangleq \left(\sigma_{d2}^2 \mathbf{I} + \sigma_r^2 \bar{\mathbf{H}}_2 \bar{\mathbf{R}} \bar{\mathbf{R}}^\dagger \bar{\mathbf{H}}_2^\dagger \right)^{-1/2}$$

without loss of any signal information.

Thus, the overall received vector $\bar{\mathbf{y}}$ at destination is obtained by applying the noise whitening filter $\mathbf{F}$ into (3) and stacking received vectors of (2) and (3),

$$\bar{\mathbf{y}} = \sqrt{\frac{P_s}{M}} \cdot \begin{bmatrix} \bar{\mathbf{H}}_0 \\ \mathbf{F} \bar{\mathbf{H}}_2 \bar{\mathbf{R}} \bar{\mathbf{H}}_1 \end{bmatrix} \mathbf{S} \mathbf{x} + \bar{\mathbf{n}} \quad (4)$$

where $\bar{\mathbf{n}}$, the equivalent $2L \times 1$ noise vector at the destination, follows $\mathcal{CN}(\mathbf{0}, \text{diag}(\sigma_{d1}^2 \mathbf{I}, \mathbf{I}))$.

Our goal is to design the source precoder $\mathbf{S}$ and the relay precoder $\bar{\mathbf{R}}$ under individual power constraints to maximize the mutual information between source and destination:

$$\max_{\mathbf{S}, \bar{\mathbf{R}}} \mathcal{I}(\mathbf{x}; \bar{\mathbf{y}}) \quad (5a)$$

$$\text{s.t.} \quad \mathbb{E}(\|\mathbf{S} \mathbf{x}\|^2) \leq M \quad (5b)$$

$$\mathbb{E}(\|\bar{\mathbf{R}} \mathbf{y}_r\|^2) \leq P_r. \quad (5c)$$

Obviously, the relay power constraint (5c) can be written as

$$\text{Tr} \left[\bar{\mathbf{R}} (\sigma_r^2 \mathbf{I} + \frac{P_s}{M} \bar{\mathbf{H}}_1 \mathbf{S} \mathbf{S}^\dagger \bar{\mathbf{H}}_1^\dagger) \bar{\mathbf{R}}^\dagger \right] \leq P_r.$$

To simplify the constraint, we define a new precoder matrix $\mathbf{R}$,

$$\mathbf{R} \triangleq \sqrt{\frac{\sigma_r^2 \cdot N}{P_r}} \bar{\mathbf{R}} (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger)^{\frac{1}{2}} \quad (6)$$

where we define the equivalent channel $\mathbf{H}_1 \triangleq \sqrt{\frac{P_s}{\sigma_r^2 M}} \bar{\mathbf{H}}_1$.

Hence, the constraint (5c) is converted into

$$\mathbb{E}(\|\bar{\mathbf{R}} \mathbf{y}_r\|^2) \leq P_r \Rightarrow \|\mathbf{R}\|^2 \leq N. \quad (7)$$

From (6), we can easily compute $\bar{\mathbf{R}}$ once $\mathbf{R}$ is found. The relationship among $\bar{\mathbf{R}}$, $\mathbf{R}$ and relay 'whitening' filter, which whitens the received signal $\mathbf{y}_r$, could be shown in Fig. 2.

Henceforth, unless explicitly specified otherwise, we would focus on the new defined relay precoder $\mathbf{R}$ and the associated power constraint (7).

To further simplify our formulation, we also define $\mathbf{H}_0 \triangleq \sqrt{\frac{P_s}{\sigma_{d1}^2 M}} \bar{\mathbf{H}}_0$ and $\mathbf{H}_2 \triangleq \sqrt{\frac{P_r}{\sigma_{d2}^2 N}} \bar{\mathbf{H}}_2$. Consequently, the receive

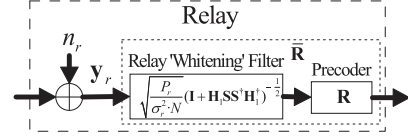


Fig. 2. Relay precoder concatenated with the received vector $\mathbf{y}_r$ whitening pre-filter.

vector $\bar{\mathbf{y}}$ is equivalently reformulated by substituting definitions $\mathbf{F}$, $\mathbf{R}$, $\mathbf{H}_0$, $\mathbf{H}_1$ and $\mathbf{H}_2$ into (4) as follows

$$\mathbf{y} = \mathbf{W} \mathbf{x} + \mathbf{n} \quad (8)$$

where $\mathbf{n} \in \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is normalized with unit variance, and we define

$$\mathbf{W} \triangleq \begin{bmatrix} \mathbf{H}_0 \\ (\mathbf{I} + \mathbf{T} \mathbf{T}^\dagger)^{-\frac{1}{2}} \mathbf{T} \mathbf{H}_1 \end{bmatrix} \mathbf{S} \quad (9)$$

$$\mathbf{T} \triangleq \mathbf{H}_2 \mathbf{R} (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger)^{-\frac{1}{2}}.$$

Our problem of (5a)–(5c) is equivalently

$$\begin{aligned} \max_{\mathbf{S}, \mathbf{R}} \quad & \mathcal{I}(\mathbf{x}; \mathbf{y}) \\ \text{s.t.} \quad & \|\mathbf{S}\|^2 \leq M \\ & \|\mathbf{R}\|^2 \leq N. \end{aligned} \quad (10)$$

From now on, unless explicitly specified otherwise, we shall work on the equivalent model defined in (8), (9) and (10).

C. Preliminaries

Before we present our main results, it is necessary to first introduce some basic facts. Consider a complex-field vector channel model in (11), where $\mathbf{A}$ is the equivalent channel matrix, $\mathbf{x}$ is the source vector with arbitrary distributions, the vector $\mathbf{n}$ is $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ and observed vector $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{n}. \quad (11)$$

The MMSE estimation of $\mathbf{x}$ based on the observation of output $\mathbf{y}$ is the conditional mean $\hat{\mathbf{x}}(\mathbf{y}) = \mathbb{E}(\mathbf{x} | \mathbf{y})$ and the corresponding MMSE matrix and companion matrix conditioned on a specific observation $\mathbf{y}$ [19], respectively, are defined as

$$\Phi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y}) \triangleq \mathbb{E}_{\mathbf{x}} \left\{ [\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})] [\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})]^\dagger | \mathbf{y} \right\} \quad (12)$$

$$\Psi_{\mathbf{x}\mathbf{x}^T}(\mathbf{y}) \triangleq \mathbb{E}_{\mathbf{x}} \left\{ [\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})] [\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})]^T | \mathbf{y} \right\}. \quad (13)$$

Furthermore, we can obtain the MMSE matrix and companion matrix by averaging over the observation $\mathbf{y}$,

$$\Phi_{\mathbf{x}\mathbf{x}^\dagger} = \mathbb{E}_{\mathbf{y}} [\Phi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y})]$$

$$\Psi_{\mathbf{x}\mathbf{x}^T} = \mathbb{E}_{\mathbf{y}} [\Psi_{\mathbf{x}\mathbf{x}^T}(\mathbf{y})].$$

Note that $p(\mathbf{y} | \mathbf{x}) \propto \exp(-\|\mathbf{y} - \mathbf{A} \mathbf{x}\|^2)$. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = \mathbb{E}_{\mathbf{x}, \mathbf{y}} \left[\log \frac{p(\mathbf{y} | \mathbf{x})}{p(\mathbf{y})} \right]. \quad (14)$$

For Gaussian input, a unitary transformation $\mathbf{U}$ on the input and a unitary transformation $\mathbf{V}$ on the output does not change the

mutual information between channel input and output signals. However, for non-Gaussian input, unitary $\mathbf{U}$ no longer preserves mutual information, i.e.,

$$\mathcal{I}(\mathbf{x}; \mathbf{V}\mathbf{y}) = \mathcal{I}(\mathbf{x}; \mathbf{y}) \neq \mathcal{I}(\mathbf{U}\mathbf{x}; \mathbf{y}) .$$

III. MAIN RESULTS

In this section, we will present some properties for the components of SVD of source and relay precoders. Before giving a review of our results, we first take precoders and channels apart by SVD, respectively, as

$$\begin{aligned} \mathbf{S} &\stackrel{\text{SVD}}{=} \mathbf{U}_s \Sigma_s \mathbf{V}_s^\dagger & \mathbf{R} &\stackrel{\text{SVD}}{=} \mathbf{U}_r \Sigma_r \mathbf{V}_r^\dagger \\ \mathbf{H}_1 &\stackrel{\text{SVD}}{=} \mathbf{U}_{H_1} \Sigma_{H_1} \mathbf{V}_{H_1}^\dagger & \mathbf{H}_2 &\stackrel{\text{SVD}}{=} \mathbf{U}_{H_2} \Sigma_{H_2} \mathbf{V}_{H_2}^\dagger . \end{aligned}$$

As a preview, in this section, Proposition 1 characterizes the optimal left singular vectors for relay precoder $\mathbf{R}$. Theorem 2 highlights the convexity of the mutual information with respect to the power allocations for relay precoder. To develop algorithms of gradient search, Subsection III-B presents the derivatives of the mutual information with respect to the right singular vectors of $\mathbf{S}$, the precoder $\mathbf{S}$ and $\mathbf{R}$, respectively. These properties are useful in the next section when searching for optimal precoders for the relay networks.

A. Structural Properties of Optimal Precoders

To begin, we recite a useful theorem discovered by Xiao, *et al.* in [19]. In particular, we simply let $\mathbf{H} = \mathbf{I}$ and replace $\mathbf{G}$ with $\mathbf{A}$ in Theorem 1 of [19].

Theorem 1: For the general model of (11), the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ depends on $\mathbf{A}$ only through $\mathbf{Q} = \mathbf{A}^\dagger \mathbf{A}$. The gradient of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with respect to $\mathbf{Q}$ is

$$\frac{\partial}{\partial \mathbf{Q}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \Phi_{\mathbf{x}\mathbf{x}^\dagger} \quad (15)$$

where $\Phi_{\mathbf{x}\mathbf{x}^\dagger}$ is the MMSE matrix. Furthermore, $\mathcal{I}(\mathbf{x}; \mathbf{y})$ is a concave function of $\mathbf{Q}$.

Note that, Theorem 1 states that the derivative of mutual information with respect to $\mathbf{Q}$ is always the MMSE matrix $\Phi_{\mathbf{x}\mathbf{x}^\dagger}$. It is indeed the general (vector) form of the well-known relationship between the mutual information and MMSE of x in the following scalar channel [30]

$$y = \sqrt{\rho}x + n$$

by replacing $\mathbf{Q}$ and $\Phi_{\mathbf{x}\mathbf{x}^\dagger}$ with ρ and $\text{mmse}(x)$, respectively, in (15).

This theorem is very useful. If $\mathbf{A} = \mathbf{H}\mathbf{G}$, it captures the point-to-point MIMO precoding problem in [19]. If $\mathbf{A} = [(\mathbf{H}_1 \mathbf{S}_1)^\dagger \cdots (\mathbf{H}_K \mathbf{S}_K)^\dagger]^\dagger$, it can be applied for an MIMO H-ARQ problem in [22]. In our formulation, defining

$$\mathbf{B} \triangleq \mathbf{W}^\dagger \mathbf{W} \quad (16)$$

we conclude that

$$\frac{\partial}{\partial \mathbf{B}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \Phi_{\mathbf{x}\mathbf{x}^\dagger} .$$

Furthermore, $\mathcal{I}(\mathbf{x}; \mathbf{y})$ is a concave function of $\mathbf{B}$.

Proposition 1: For the optimization problem in (10), the left singular vectors of the optimal precoder $\mathbf{R}$, can be always chosen to coincide with the right singular vectors of channel $\mathbf{H}_2$, i.e., $\mathbf{U}_r = \mathbf{V}_{H_2}$.

Proof: See Appendix A. ■

Theorem 2: When $\mathbf{U}_r = \mathbf{V}_{H_2}$, the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ of (8) is a concave function of Σ_r^2 . Furthermore, the gradient of the mutual information with respect to Σ_r^2 is given by

$$\frac{\partial}{\partial \Sigma_r^2} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \text{diag} (\Sigma_{H_2} \mathbf{V}_r^\dagger \mathbf{K}^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger} \mathbf{K} \mathbf{V}_r \Sigma_{H_2}) \quad (17)$$

in which

$$\begin{aligned} \mathbf{K} &\triangleq \mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger)^{\frac{1}{2}} \\ &\quad \times (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger + \mathbf{R}^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{R})^{-1} . \end{aligned} \quad (18)$$

Proof: See Appendix B. ■

Furthermore, consider the scenario without direct link. Suppose the singular vectors of source and relay precoders are set with the value shown in (20). We could obtain the follows by substituting the values in (20) into (28) (defined in Appendix A),

$$\mathbf{B} =$$

$$\mathbf{V}_s \Sigma_s^2 \Sigma_r^2 \Sigma_{H_1}^2 \Sigma_{H_2}^2 (\mathbf{I} + \Sigma_{H_1}^2 \Sigma_s^2 + \Sigma_{H_2}^2 \Sigma_r^2)^{-1} \mathbf{V}_s^\dagger . \quad (19)$$

Obviously, the singular vectors of source and relay precoders are structural symmetric. Combining the result of Theorem 2, we conclude that the mutual information is also a concave function of Σ_s^2 . This result is summarized in the following corollary.

Corollary 1: Consider the scenario without direct link, i.e., $\mathbf{H}_0 = \mathbf{0}$. If source and relay precoders select what are optimal for Gaussian source signals [12], i.e.,

$$\mathbf{U}_s = \mathbf{V}_{H_1}, \quad \mathbf{V}_r = \mathbf{U}_{H_1}, \quad \mathbf{U}_r = \mathbf{V}_{H_2} \quad (20)$$

then the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ of (8) is a concave function of Σ_s^2 . Moreover, the partial derivative of the mutual information with respect to Σ_s^2 is given by

$$\begin{aligned} \frac{\partial}{\partial \Sigma_s^2} \mathcal{I}(\mathbf{x}; \mathbf{y}) &= \text{diag} \{ \mathbf{V}_s^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger} \mathbf{V}_s \Sigma_{H_1}^2 \Sigma_{H_2}^2 \Sigma_r^2 (\mathbf{I} + \Sigma_{H_2}^2 \Sigma_r^2) \\ &\quad \times (\mathbf{I} + \Sigma_{H_2}^2 \Sigma_r^2 + \Sigma_{H_1}^2 \Sigma_s^2)^{-2} \} . \end{aligned} \quad (21)$$

Remark 1: It should be emphasized that Corollary 1 strongly depends on the choice of singular vector of source and relay precoders. Once the choices in (20) is violated, the convexity of mutual information with respect to Σ_s^2 could not be preserved. Note that our convexity results on power allocation for source and relay are more general than that in [12] because

- 1) Both Theorem 2 and Corollary 1 state that the individual power allocation for source precoder and relay precoder are convex problems under certain conditions.
- 2) For relay power allocation, we obtain a looser condition of $\mathbf{U}_r = \mathbf{V}_{H_2}$ versus the condition in [12] which requires diagonalizing the relay networks without direct link, and further requires the Gaussian source assumption.
- 3) For the source power allocation, we obtain the same conclusion as in [12] but without requiring the Gaussian signal assumption.

B. Gradient Properties of the Mutual Information

Thus far, we have presented the convex properties of separate power allocations for source and relay. However, the joint power allocation problem for source and relay is generally not convex, as shown in [12]. As a result, we would like to impose some

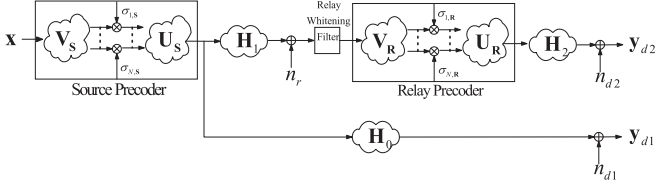


Fig. 3. General joint source and relay precoders design.

structural constraints on the precoders and propose specially designed numerical algorithms to search for good power allocation under the precoder structural constraints. We will present our gradient based numerical algorithms in the next section. For gradient based methods, we present the following results that can be helpful. Details are in Appendix C.

- 1) The gradient of the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with respect to $\mathbf{V}_s$ of the source precoder $\mathbf{S}$ is

$$\frac{\partial}{\partial \mathbf{V}_s} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \mathbf{V}_s \mathbf{B} \Phi_{\mathbf{x}\mathbf{x}^\dagger}. \quad (22)$$

- 2) The gradient of the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with respect to source precoder $\mathbf{R}$ is

$$\frac{\partial}{\partial \mathbf{R}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{R} \mathbf{K}^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger} \mathbf{K}. \quad (23)$$

- 3) The gradient of the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with respect to the source precoder $\mathbf{S}$ is

$$\frac{\partial}{\partial \mathbf{S}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \left[\mathbf{H}_0^\dagger \mathbf{H}_0 + \mathbf{H}_1^\dagger \mathbf{T}^\dagger \mathbf{T} (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{H}_1 \right] \mathbf{S} \Phi_{\mathbf{x}\mathbf{x}^\dagger} - \mathbf{C} \quad (24)$$

in which the element of matrix $\mathbf{C}$ can be obtained via

$$\mathbf{C}_{ij} = [\text{vec}(\Phi_{\mathbf{x}\mathbf{x}^\dagger})]^T (\mathbf{D}^* \otimes \mathbf{D}) (\mathbf{L}^* \oplus \mathbf{L}) \times [(\mathbf{O}^* \oplus \mathbf{O}) (\mathbf{O}^* \otimes \mathbf{O})]^{-1} ([\mathbf{H}_1^\dagger]_i \otimes [\mathbf{H}_1 \mathbf{S}]_j)$$

and matrices $\mathbf{D}$, $\mathbf{L}$ and $\mathbf{O}$ are defined as follows to simplify $\mathbf{C}$

$$\mathbf{D} \triangleq \mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \quad (25)$$

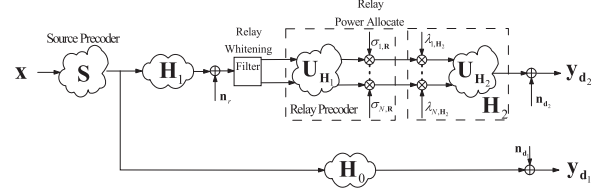
$$\mathbf{L} \triangleq \mathbf{T}^\dagger \mathbf{H}_2 \mathbf{R} \quad (26)$$

$$\mathbf{O} \triangleq \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{\frac{1}{2}}. \quad (27)$$

IV. ITERATIVE PRECODER DESIGNS

The joint source-relay precoder design has been shown in Fig. 3. Both precoders have been decomposed by SVD. Based on Proposition 1 and Theorem 2, we can always choose the optimal $\mathbf{U}_r = \mathbf{V}_{H_2}$, which converts the power allocation for relay precoder into a convex problem. However, the remaining three unitary transformations $\mathbf{V}_s$, $\mathbf{U}_s$ and $\mathbf{V}_r$ have no known simple relationships with respect to the mutual information. They pose challenges to the precoder optimization problem that we seek to tackle. We now address them differently.

We have learned from existing works [17], [19] that, in point-to-point MIMO communications, the linear transformation by $\mathbf{V}_s$ at the source before power allocation is highly critical when non-Gaussian source signals and mutual information are considered. It is therefore intuitive to conclude that $\mathbf{V}_s$ is equally vital in our relay network and must be handled with extra caution. In particular, $\mathbf{V}_s$ should be optimized without imposing structural assumptions.

Fig. 4. The precoders design with direct link $\mathbf{H}_0$.

Observe from the diagram of Fig. 3, that the source signal vector $\mathbf{x}$ is processed sequentially by these three unitary transformations. Based on the Central Limit Theorem, linear transformation of even the non-Gaussian source signals leads to random variables that are more Gaussian-like. For this reason, later transformations in this diagram are likely to be more tolerant of the Gaussian assumption without causing substantial performance loss. Therefore, to simplify our relay design, we let $\mathbf{V}_r = \mathbf{U}_{H_1}$, which is the optimal selection under Gaussian source signals [12]. Under this selection, we reduce the relay precoder design to the power allocation problem.

Next is the choice of $\mathbf{U}_s$. It turns out that the two cases arising from relay networks with and without direct links need to be considered separately.

- 1) When there is no direct link, based on Corollary 1, we assume that the transformation by $\mathbf{V}_s$ has been optimized so that the transformed signal arriving at $\mathbf{U}_s$ is closer to Gaussian. Thus, we set $\mathbf{U}_s = \mathbf{V}_{H_1}$, which is optimal with Gaussian source inputs. With this special precoder constraint, the power allocation problem of source precoder has been proven convex and can be optimized. Therefore, without direct link, we will optimize $\mathbf{V}_s$ by constraining the selection of $\mathbf{U}_s$ and $\mathbf{V}_r$ to take the forms optimal under Gaussian.
- 2) When there exists the direct link $\mathbf{H}_0$, the design of $\mathbf{S}$ is more complex as it depends on two channels. We should note that even for the well known Gaussian source assumption, the optimal solution remains elusive. For this reason, we do not further assume any extra structural constraint on $\mathbf{S}$. Instead, we will optimize $\mathbf{S}$ by applying gradient search algorithms based on the gradient properties in Section III-B.

Before we discuss our precoder designs for these two cases in more detail, it should be noted that, our next proposed alternative optimization benefits from the new defined relay precoder in (6), which decouples the original coupled source-relay power constraint.

A. Relay Networks With Direct Link $\mathbf{H}_0$

Our proposed design of relay network precoders with direct link $\mathbf{H}_0$ is based on the choices of $\mathbf{U}_r$ and $\mathbf{V}_r$ as discussed earlier. The structure of the relay network precoding is illustrated in Fig. 4. More specifically, for relay precoder, we reduce the problem into the optimal power allocation $\Sigma_{\mathbf{R}}^2$. For source precoder $\mathbf{S}$, we apply the gradient descending method to search for the optimal $\mathbf{S}$.

Although the joint problem $\max_{\mathbf{S}, \Sigma_{\mathbf{R}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y})$ is not convex, we still could optimize these two unknown matrices iteratively. Hence, we have two optimization sub-problems as follows.

$$\begin{aligned} \max_{\mathbf{S}} \quad & \mathcal{I}(\mathbf{x}; \mathbf{y} | \Sigma_{\mathbf{R}}^2) \quad \max_{\Sigma_{\mathbf{R}}^2} \quad \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{S}) \\ \text{s.t.} \quad & \mathbb{E}(\|\mathbf{S}\|^2) \leq M \quad \text{s.t.} \quad \text{Tr}(\Sigma_{\mathbf{R}}^2) \leq N \end{aligned}$$

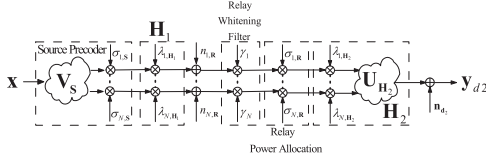


Fig. 5. Proposed joint source-relay precoder design without direct link $\mathbf{H}_0$.

Similar to the method proposed in [12], we can obtain good results by optimizing these two sub-problems iteratively.

The classic text [31] established that a gradient descending algorithm for constrained optimization problem can consist of three steps: the gradient calculation, the projection of variables on constraint surface, and setting the stepsize. Based on Theorem 2 and Appendix C, the gradient of these two sub-problems are

$$\nabla_{\mathbf{S}} = \frac{\partial}{\partial \mathbf{S}^*} \mathcal{I}(\mathbf{x}; \mathbf{y})$$

$$\nabla_{\Sigma_{\mathbf{R}}^2} = \frac{\partial}{\partial \Sigma_{\mathbf{R}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y}) - \frac{1}{N} \text{Tr} \left(\frac{\partial}{\partial \Sigma_{\mathbf{R}}^2} \mathcal{I} \right) \mathbf{I}.$$

The projections for $\mathbf{S}$ and $\Sigma_{\mathbf{R}}^2$ are

$$\pi_{\mathbf{S}} = \sqrt{\frac{M}{\text{Tr}(\mathbf{S}\mathbf{S}^\dagger)}} \cdot \mathbf{S}, \quad \pi_{\Sigma_{\mathbf{R}}^2} = \frac{N}{\text{Tr}[(\Sigma_{\mathbf{R}}^2)_+]} (\Sigma_{\mathbf{R}}^2)_+$$

respectively. We acquire a proper gradient descending stepsize according to the Armijo scheme [31]. Our iterative algorithm begins with randomly generated $\mathbf{S}$ and $\Sigma_{\mathbf{R}}^2$. Because this iterative process promises improvement after each optimization step with the power constraint, it ensures local convergence.

B. Relay Networks Without Direct Link $\mathbf{H}_0$

Although we may treat the case of three-node relay network without direct link as a special case of $\mathbf{H}_0 = 0$, we can in fact do better by utilizing its special properties and structure. More specifically, we choose the singular vectors of source and relay precoders to match the singular vector of corresponding channels. Therefore, the overall channel has been diagonalized as shown in Fig. 5.

To optimize, we have three unknown matrices to acquire: $\Sigma_{\mathbf{S}}^2$, $\Sigma_{\mathbf{R}}^2$ and $\mathbf{V}_s$. Again, since the joint problem

$$\max_{\mathbf{V}_s, \Sigma_{\mathbf{S}}^2, \Sigma_{\mathbf{R}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y})$$

is not convex, we may optimize them one-by-one in an iterative fashion by freezing the remaining ones when optimizing one matrix.

Applying Theorem 2 and Corollary 1, we find that the sub-problem of optimizing $\Sigma_{\mathbf{R}}^2$ and the sub-problem of optimizing $\Sigma_{\mathbf{S}}^2$ while freezing the remaining two unknown matrices are both convex. This convex property makes the respective sub-problem iteration easier. However, the optimization of $\mathbf{V}_s$ itself is not convex. Similar to the method for the first case with direct link $\mathbf{H}_0$, we divide the design problem into three sub-problems to solve in Fig. 6.

Because the gradient and the corresponding projection for $\Sigma_{\mathbf{R}}^2$ have been given in the previous subsection for networks with direct link, we will only provide the gradients for $\Sigma_{\mathbf{S}}^2$

$\max_{\Sigma_{\mathbf{S}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y} \mathbf{V}_s, \Sigma_{\mathbf{R}}^2)$ $\text{s.t.} \quad \text{Tr}(\Sigma_{\mathbf{S}}^2) \leq M$	$\max_{\Sigma_{\mathbf{R}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y} \mathbf{V}_s, \Sigma_{\mathbf{S}}^2)$ $\text{s.t.} \quad \text{Tr}(\Sigma_{\mathbf{R}}^2) \leq N$	$\max_{\mathbf{V}_s} \mathcal{I}(\mathbf{x}; \mathbf{y} \Sigma_{\mathbf{S}}^2, \Sigma_{\mathbf{R}}^2)$ $\text{s.t.} \quad \mathbf{V}_s \mathbf{V}_s^\dagger = \mathbf{I}$
--	--	---

Fig. 6. Three sub-optimization problems.



Fig. 7. Iterative optimization for the case without direct link $\mathbf{H}_0$.

and $\mathbf{V}_s$,

$$\nabla_{\Sigma_{\mathbf{S}}^2} = \frac{\partial}{\partial \Sigma_{\mathbf{S}}^2} \mathcal{I}(\mathbf{x}; \mathbf{y}) - \frac{1}{M} \text{Tr} \left(\frac{\partial}{\partial \Sigma_{\mathbf{S}}^2} \mathcal{I} \right) \mathbf{I}$$

$$\nabla_{\mathbf{V}_s} = \mathbf{V}_s \left(\frac{\partial}{\partial \mathbf{V}_s^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) \right)^\dagger \mathbf{V}_s - \frac{\partial}{\partial \mathbf{V}_s^*} \mathcal{I}.$$

Accordingly, the constraint projections are

$$\pi_{\Sigma_{\mathbf{S}}^2} = \frac{M}{\text{Tr}[(\Sigma_{\mathbf{S}}^2)_+]} (\Sigma_{\mathbf{S}}^2)_+$$

$$\pi_{\mathbf{V}_s} = \arg \min_{\mathbf{P} \mathbf{P}^\dagger = \mathbf{I}} \|\mathbf{V}_s - \mathbf{P}\|^2.$$

Noting that the optimization for $\mathbf{V}_s$ is defined on the Stiefel manifold, we adopted the gradient and projection for $\mathbf{V}_s$ from [32]. We also adjust the stepsizes for the three matrices by applying the Armijo algorithm. To summarize, these three unknown matrices are iteratively and sequentially updated in the manner shown in Fig. 7.

It should be noted that, despite in the proof of our Theorem 2 that the mutual information of cooperative system is convex with respect to the power vector of relay precoder, the overall joint source-relay precoder design problem is still non-convex. As our special case, the problem with Gaussian input [12], [33] has already reported this non-convexity. Consequently, all our proposed algorithms optimizing parameters alternatively are inherent sub-optimal. Nevertheless, we could pick the best precoders as final solution by applying optimizing under sufficiently many random initializations.

V. SIMULATION RESULTS

A. Test Setup and Comparisons

In our simulations, we adopt the definition of [10] and define the signal-noise ratio (SNR) for the source and the relay as $\rho_s = P_s / \sigma_r^2 M$ and $\rho_r = P_r / \sigma_d^2 N$, respectively. Essentially the SNR is the ratio of average power on each transmit antenna to average additive Gaussian noise power on each receiver antenna. Without loss of generality, we set all noise variances to one and examine the effect on the mutual information between source and destination with respect to different ρ_s and ρ_r designs. As described in Section II, the elements of three MIMO channel, source-destination channel $\bar{\mathbf{H}}_0$, source-relay channel $\bar{\mathbf{H}}_1$ and relay-destination channel $\bar{\mathbf{H}}_2$, are independent and identically distribution (i.i.d) zero-mean and unit variance Gaussian entries following $\mathcal{CN}(0, 1)$. They are also assumed to be block independent and will not change within one relay phase. We measure the ergodic mutual information by averaging over more than 200 independent random channels.

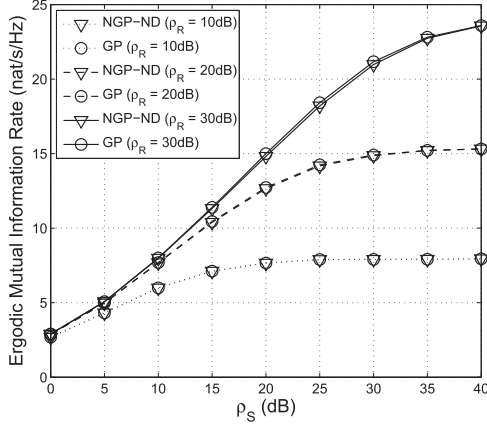


Fig. 8. Capacity performance of none direct-link relay networks with Gaussian source signals.

For comparison, we will evaluate our results against the algorithm by Rong in [12]. Under the Gaussian source assumption, Rong's algorithm is designed for MIMO relay systems without direct link. We directly apply this precoder design to the case of finite-alphabet signals, and label the results by GP (Gaussian source Precoder). The results of our proposed algorithms for networks with and without direct link are denoted by NGP-DL and NGP-ND, respectively.

In addition, we also compare the result of NGP-DL with those obtained by a pure gradient (PG) method (labeled by PG-NGP), which directly applies the gradient results in Section III-B to evaluate the gradient of mutual information with respect to $\mathbf{S}$ and $\mathbf{R}$. The PG algorithm iteratively optimizes $\mathbf{S}$ and $\mathbf{R}$ by applying gradient algorithm search in two iterative steps to find the optimum precoders in a ping-pong fashion.

B. Gaussian Source Test

Firstly, consider the networks without direct link. We set the antenna number of each node to 4 and evaluate the performances of NGP-ND and GP algorithms. Applying our relay precoder defined in (6), which moves the whitening filter outside relay precoder, the capacity optimization problem in (10) for Gaussian signals could be written as follows,

$$\begin{aligned} \max_{\sigma_{i,S}^2, \sigma_{i,R}^2} \quad & \sum_{i=1}^M \ln \\ & \times \left(\frac{1 + \lambda_{i,H_1}^2 \sigma_{i,S}^2 + \lambda_{i,H_2}^2 \sigma_{i,R}^2}{1 + \lambda_{i,H_1}^2 \sigma_{i,S}^2 + \lambda_{i,H_2}^2 \sigma_{i,R}^2 + \lambda_{i,H_1}^2 \sigma_{i,S}^2 \lambda_{i,H_2}^2 \sigma_{i,R}^2} \right) \\ \text{s.t.} \quad & \sum_{i=1}^M \sigma_{i,S}^2 \leq M, \quad \sum_{i=1}^M \sigma_{i,R}^2 \leq N. \end{aligned}$$

We observe that, along with our reformulated systems, source precoder $\mathbf{S}$ and relay precoder $\mathbf{R}$ actually play symmetric roles in the term of ergodic capacity because the random source-relay channel $\mathbf{H}_1$ and relay-destination channel $\mathbf{H}_2$ share the same distribution. Consequently, we will only need to illustrate the resulting network capacity with a fixed relay precoder power.

In Fig. 8, we illustrate the performance of relay networks with the fixed relay precoder power at 10 dB, 20 dB, and 30 dB, respectively. We observe that although the system capacity is an increasing function of the source and relay powers,

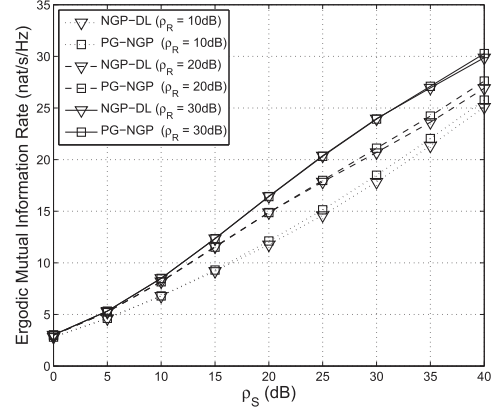


Fig. 9. Capacity performance under Gaussian source signals with 10 dB attenuated direct link and fixed relay power.

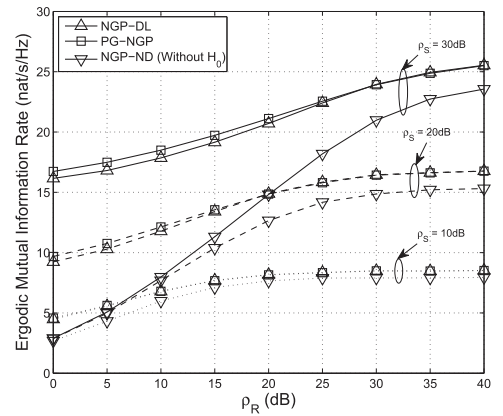


Fig. 10. Capacity performance of relay networks with Gaussian source signals and 10 dB attenuated direct link for fixed source power.

the overall capacity remains bounded once the power of source precoder is fixed. Simulations show that our generic NGP-ND precoder achieves nearly the same performance as Rong's algorithm specifically designed for Gaussian input. When ρ_R is set to 30 dB, a noticeable small gap is merely a result of numerical choices in optimization termination and gradient step selection.

Next, we consider networks with a direct link and carry out the evaluation of NGP-DL and PG-NGP algorithms. A fixed 10 dB attenuation is added to the source-destination channel to model the path loss. We first test networks with fixed relay power $\rho_R = 10$ dB, 20 dB and 30 dB, respectively.

As Fig. 9 shows, with the aid of direct link, the mutual information now continues to grow as the source increases. Our results illustrate that the convergence speed of NGP-DL is much faster than that of PG-NGP. Our simulation also demonstrates that NGP-DL has almost the same performance as PG-NGP when the effective relay link (source-relay-destination) channel dominates the direct source-destination link channel. On the other hand, we recognize that by fixing the right singular vectors of $\mathbf{R}$, NGP-DL has a stronger structure constraint than pure-gradient search of PG-NGP. Indeed, NGP-DL suffers up to 1 dB loss if the source-destination channel is dominant. However, such case is rare in reality as relays are typically activated and utilized only when the direct link is nominally weak.

In Fig. 10, we freeze instead the source power ρ_S to 10 dB, 20 dB and 30 dB, respectively. The networks capacity is shown for different relay power ρ_R . When the source power is fixed,

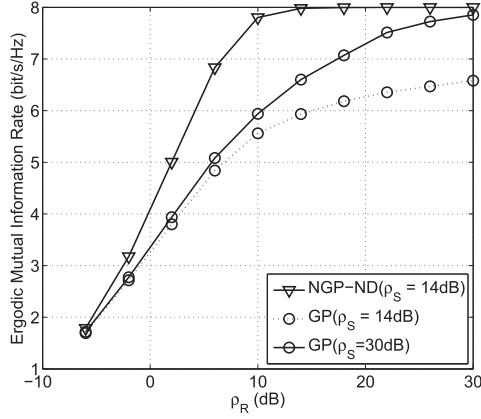


Fig. 11. QPSK with fixed source power and without direct link.

the network capacity increases with growing relay power but becomes bounded instead. As the relay power increases, the network changes from source-destination channel dominant to relay dominant. As such, we see that the gain achieved by the direct link $\mathbf{H}_0$ also narrows with increasing relay power. Similar as the case when relay power fixed, when the overall channel is relay channel dominant, our proposed NGP-DL and PG-NGP achieve nearly the same results.

For comparison, the performance without direct link are also plotted. Obviously, with the help of direct link, the achieved capacity is better than those without direct link.

Despite the structural constraints we introduced, our test results on Gaussian input signals show that, the proposed NGP-DL and NGP-ND algorithms can achieve close to optimized performance by precoders specifically designed for Gaussian inputs (using pure gradient search). The newly proposed precoders not only converges faster, they are also not sensitive to the type of input signals to the relay network.

Next we move to test the case involving finite-alphabet signals.

C. Simulations With Non-Gaussian Sources

First, consider the case that without direct link. Based on our choice for the singular vectors in NGP-ND, the $\mathbf{B}$ matrix could be written as (19). Similar to the case for Gaussian source signals, the singular value for source and relay precoders play symmetric roles. Combining this observation with Theorem 1, we conclude that the ergodic capacity for source precoder with fixed power of relay is same as that for relay precoder with fixed power of source. Therefore, without direct link, we only need to show results with fixed source power.

The results in Fig. 11 are driven by QPSK source. The antenna number of each node is 4. Four independent streams are sent. Without direct link $\mathbf{H}_0$, source power are fixed at $\rho_S = 14$ dB and $\rho_S = 30$ dB, respectively.

We observe from Fig. 11 that NGP out-performs GP precoder. Indeed, the GP precoder exhibits a greater performance loss, especially under medium-to-high relay SNR. It should be noted that as long as the source power is high enough, the ceiling of achieved mutual information is limited by the constellation size of source signal; therefore, the performance loss between NGP and GP is reduced to zero, as shown with $\rho_S = 30$ dB. On the other hand, at $\rho_S = 14$ dB, regardless of high relay power, the achieved mutual information of GP saturates at 6.7 bit since

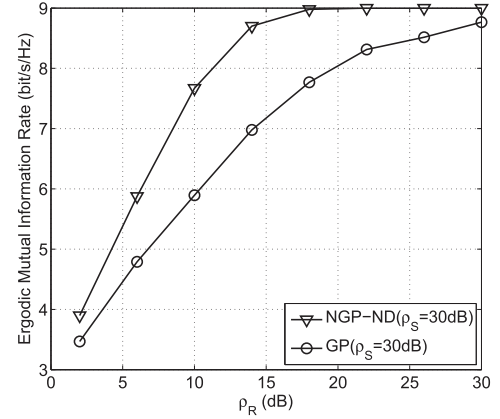


Fig. 12. 8PSK with fixed 30 dB source power and without direct link.

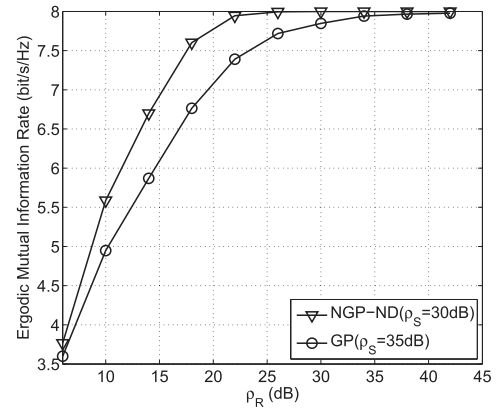


Fig. 13. 16QAM with fixed 35 dB source power and without direct link.

the mutual information is mainly limited by the source-relay channel $\mathbf{H}_1$ with the waterfilling power allocation. In fact, the GP precoder requires much higher value of ρ_S (as high as 30 dB) in order to close the transmission rate gap.

In Fig. 12, we change the input signals to 8PSK without direct link. To save computation time, we consider three antennas at each node to carry three independent data symbol streams. Setting source node power to 30 dB, we find that the GP precoder requires more than 5 dB excess relay power in order to achieve the same mutual information rate (≥ 7 bit/s/Hz) of NGP-ND.

The performance degradation of the GP precoder is primarily due to the difference of power allocation for Gaussian signals versus for finite-alphabet signals. With Gaussian source, it is always advantageous to allocate more power (as in waterfilling) to the stronger channels and less power to the weaker channels to achieve better channel capacity. However, when transmitting non-Gaussian source signals, due to limited constellation size, larger power allocation does not necessarily lead to larger mutual information gain. Consequently, Gaussian power allocation policy for some signals of finite alphabet may be worse. A second reason is that, for Gaussian sources, the choice of $\mathbf{V}_S$ does not affect the final result. However, for finite-alphabet signals, this unitary transformation is critical [18] [19].

Next, we change the input signals to 16-QAM for the relay network without direct link. In Fig. 13, we test the case when each node has 2 antennas to support two parallel data streams. We find that NGP-ND precoder not only requires 5 dB lower source node power than GP precoder, but also exhibits a more

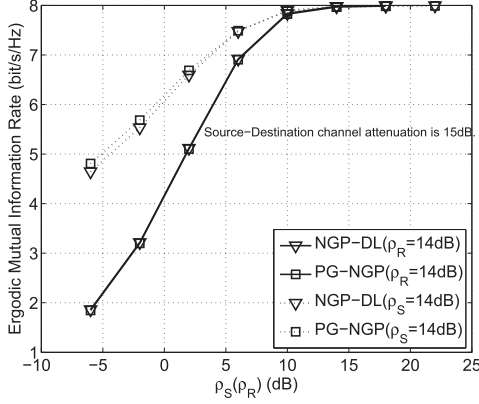


Fig. 14. The performance of QPSK inputs with direct link networks.

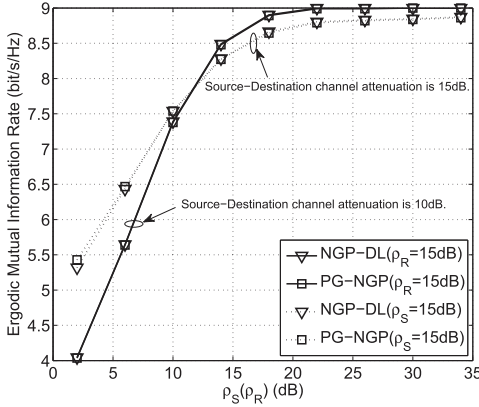


Fig. 15. The performance of 8PSK inputs with direct link networks.

than 4 dB performance gain over GP precoder when the mutual information rate is greater than 7 bit/s/Hz.

Finally, we also consider the effect of source-destination link $\mathbf{H}_0$. With non-negligible $\mathbf{H}_0$, the symmetric ergodic mutual information performances between source and relay precoders no longer holds. So we will individually evaluate the performances with fixed ρ_S and different ρ_R and the performances with fixed ρ_R and different ρ_S . Our proposed NGP-DL algorithm and the pure gradient algorithm PG-NGP algorithms will be tested.

In Fig. 14, the results are driven by QPSK source. Each node has 4 antennas to send four independent streams. The attenuation of source-destination channel is 15 dB. Solid curves are for fixed relay transmit power ($\rho_R = 14$ dB), and dotted curves are for fixed source transmit power ($\rho_S = 14$ dB). We observe that due to finite-alphabet source, the mutual information could not go unbounded even driven with infinite source and relay power. Moreover, the performances under fixed source power and under fixed relay power are no longer same (symmetric). We also observe a similar performance relationship between NGP-DL and PG-NGP curves, which has been shown in Gaussian signals. In particular, our NGP-DL algorithm achieves almost the same performance with PG-NGP algorithm when the effective relay link channel dominates the direct source-destination link channel. On the other hand, NGP-DL suffers around 0.5 dB performance loss due to our fixed right singular vectors of relay precoder.

We also present 8PSK and 16-QAM performances in Figs. 15 and 16, respectively. For the case of 8PSK signal, each node has three antennas to carry three independent streams. While evaluating mutual information with different relay transmission

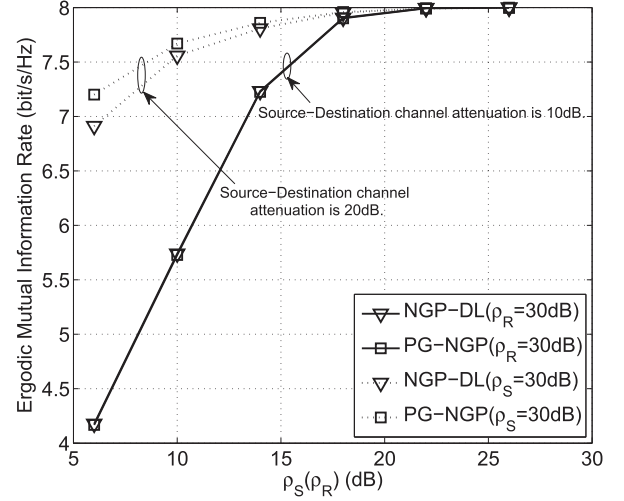


Fig. 16. The performance of 16-QAM inputs with direct link networks.

power, the power of source precoder is 15 dB and the attenuation of source-destination channel is set to 15 dB. On the other hand, while evaluating performance with different source transmission power, the power of relay precoder is 15 dB and the attenuation of source-destination channel is set to 10 dB. Similarly, for the case of 16-QAM signal, to save computation time each node is equipped with two antennas to support two independent streams. To obtain the performances with different relay power, $\rho_S = 30$ dB and the attenuation of source-destination channel is set to 20 dB. On the other hand, to get the performances with different source power, $\rho_R = 30$ dB and the attenuation of source-destination channel is set to 10 dB. We observe that both Figs. 15 and 16 exhibit similar performances as those for QPSK.

Finally, we conclude that our proposed NGP-ND and NGP-DL algorithm are efficient. Despite we have imposed some structural constraints on our precoders, they achieve a fairly close performance by precoders designed for Gaussian input signals. With finite-alphabet input signals, they illustrate a substantial performance improvement over the existing precoder designed with Gaussian signals assumption. Even when there exists direct link between source and destination, the performance of our proposed algorithm is still comparable to that of with full gradient search results, especially when the effective relay link channel is dominant.

D. The Convergence of Proposed Algorithms

Next, we will plot the mutual information evolution procedure with random initializations for NGP-ND in Fig. 17. The pure gradient-based alternative optimization method (PG-ND algorithm), is also considered as a benchmark.

Since there are multiple variables to be optimized, our proposed algorithm works by optimizing one while freezing the others in a round-robin fashion. Therefore, in each turn we would not obtain a worse solution. In addition, since the power of two precoders are limited, our proposed algorithms promise to be converge.

It should be noted that since the original problem is non-concave, a gradient-based optimization may be trapped into a local maximum. Therefore, it is necessary to run optimization enough times with random initialization and select the best one. In Fig. 18, a histogram of optimized mutual information with

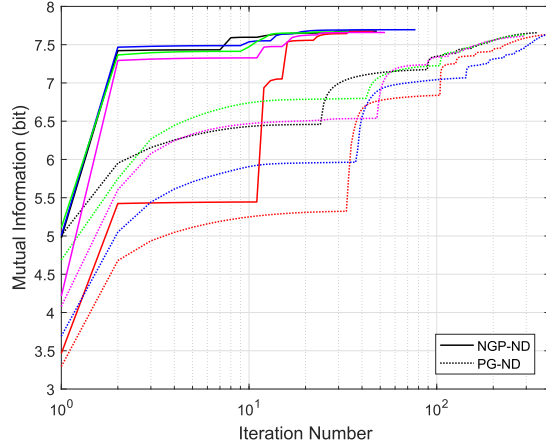


Fig. 17. The convergence of five random trails are demonstrated and marked with black, red, green, blue and magenta color, respectively. For NGP-ND, three sub-optimizations, $\mathbf{V}_S$, Σ_S^2 and Σ_R^2 , are invoked sequentially and alternatively. For PG-ND, two sub-optimizations, $\mathbf{S}$ and $\mathbf{R}$, are invoked sequentially and alternatively.

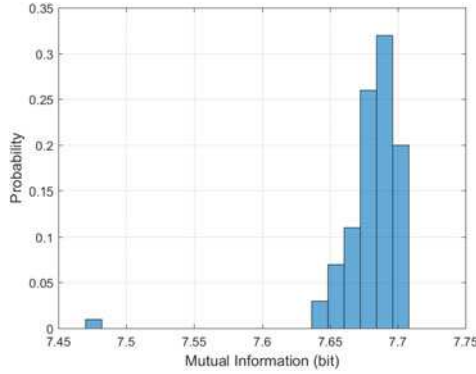


Fig. 18. A histogram of optimized mutual information with 100 random trials. There is no source-destination link. The antenna number is four for each node, and the modulation is QPSK. The power of source node is 14 dB and the power of relay node is 10 dB.

100 random initializations is depicted for a scenario of 4 antenna relay system without source-destination link at 14 dB source signal power and 10 dB relay power. We observe that all results except for one fall in a small range of 7.65–7.7 bit/Hz/s. It shows that our proposed algorithm is not sensitive to the initial points. The probability of the optimized mutual information falls in the maximum range is around 20%.

Clearly, because our proposed NGP-ND algorithm utilizes the special precoder structure, it converges much faster than those pure gradient-based algorithm. Usually, the NGP-ND algorithm require 20–40 iterations to reach the maximum, whereas the PG-ND algorithm demands more than 200 iterations to approach a similar mutual information result.

E. Frame Error Rate With Non-Gaussian Sources

Finally, we evaluate and compare the frame error rate (FER) of our proposed precoders with those designed under the assumption of Gaussian inputs. For channel error correction, a LDPC code with length 2304 and rate $R = 5/6$ is employed. Before carrying out source precoding, the coded bits are Gray-mapped to symbol (e.g. QPSK, 8PSK or 16QAM). In receiver, a maximum a posteriori (MAP) detector provides soft information to LDPC decoder. The decoder utilizes the standard sum-product

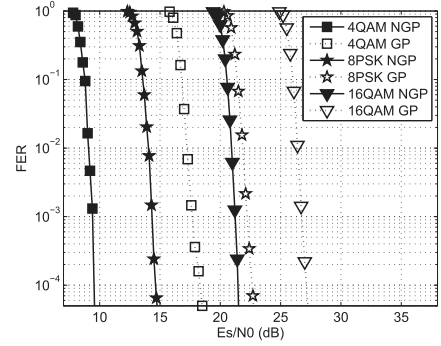


Fig. 19. The FER comparison with Gray-mapped 4QAM, 8PSK and 16QAM modulations.

algorithm (SPA) [34] in which the maximum iteration number of decoding is fixed to 50. To ultimately exploit the extra information provided by source and relay precoders, LDPC decoder also feedbacks soft information to MAP detector to form a joint detection and decoding. Such iteration number in joint detection and decoding is fixed to 10 in our test.

In Fig. 19, the case that without direct source-destination link is considered. Same as Figs. 11–13, the antenna number of each node are 4, 3 and 2, respectively, for 4QAM, 8PSK and 16QAM sources. Consider the case of fixed source power with 30 dB. We observe the FER performance with respect to the variation of relay node power. The ergodic FER performance is obtained by averaging over 200 independent channel samples.

The simulation clearly demonstrates the FER performance gain achieved by our proposed NGP precoders. The gaps are 10 dB for QPSK, 8 dB for 8PSK and 5 dB for 16QAM, respectively, which also slightly greater than the corresponding gap of mutual information shown in Figs. 11–13, respectively.

F. The Finite-Precision Effects on CSI $\mathbf{H}_1$ and Precoders

Our previous performances are obtained under the assumptions of accurate CSI at destination and accurate source and relay precoders known by source and relay, respectively. Generally, such assumptions may not be fully practical. In this section, we will demonstrate the finite-precision effects on both CSI and precoders.

In our setup, a joint source-relay precoder optimization is carried out at the destination node. Hence, the relay should forward a quantized source-relay CSI $\mathbf{H}_1$ to the destination. After optimization, the destination should feedback the quantized source and relay precoders to relay, and relay sends the quantized source precoder to source. We further assume that the relay-destination CSI forward channel, the destination-relay precoder feedback channel, and the relay-source precoder feedback channel are error free without severe delay.

In Fig. 20, we evaluate the finite-precision effects on CSI $\mathbf{H}_1$ for 4QAM, 8PSK and 16QAM constellations, respectively. The real and image parts of each element of $\mathbf{H}_1$ are individually quantized. Therefore, considering an N -bit quantization, the total bit number for M -by- M $\mathbf{H}_1$ is $2M^2N$. The quantizer is following the fixed-point definition in Matlab, e.g. (n,m) means that the total signed word is n bit and the fraction part is m bit.

Our simulations show that, a 4-bit quantization is much better than a 2-bit quantization. For the 2-bit quantization, the mutual information degradation appears large in the middle range of

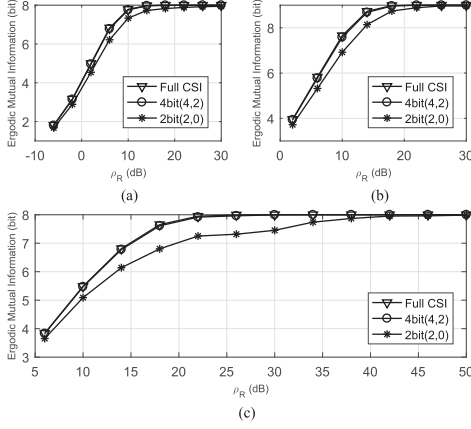


Fig. 20. Finite-precision effects on the source-relay channel state information ($\mathbf{H}_1$) feed-forwards from relay node to destination node. In each sub-figure, the source-relay CSI is quantized to 4-bit length (4,2) and 2-bit length (2,0), respectively. There is no source-destination link. In sub-figure (a), (b) and (c), the modulation are 4QAM, 8PSK and 16QAM, respectively, and the antenna number is 4, 3 and 2, respectively. In each sub-figure, the antenna number of source, relay and destination are same. The power levels of source node are 14 dB, 30 dB and 30 dB for sub-figure (a), (b) and (c), respectively.

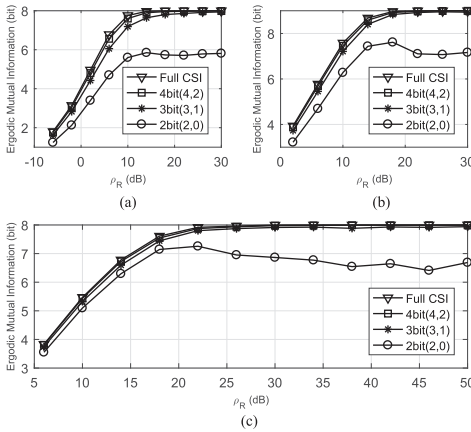


Fig. 21. Finite-precision effects on source and relay precoder feedbacks from destination node to relay and source node. For one curve, source and relay precoders are quantized with same bit length. There is no source-destination link. In sub-figure (a), (b) and (c), the modulation is 4QAM, 8PSK and 16QAM, respectively, and the antenna number is 4, 3 and 2, respectively. In each sub-figure, the antenna number of source, relay and destination are same. The power of source node is 14 dB, 30 dB and 30 dB for sub-figure (a), (b) and (c), respectively.

tested SNR. Also as the constellation size grows, such quantization degradation also becomes greater. Finally, our simulations show that a 4-bit (4,2) quantization is good enough for $\mathbf{H}_1$ with the constellations of 4QAM, 8PSK and 16QAM. And the corresponding forward bit number are 128, 72 and 32, respectively, for 4QAM (4 antennas), 8PSK (3 antennas) and 16QAM (2 antennas).

Next, since the source precoder and the relay precoder should also be feedback to source and relay nodes, respectively. In Fig. 21, we evaluate the finite-precision effects on precoders. At the same time, $\mathbf{H}_1$ is quantized with (4,2) quantizer. Similar with the previous quantization processes, we illustrate the mutual information performances with (4,2), (3,1) and (2,0) quantizer, respectively, at different modulations. Simulations show that the (2,0) quantization has the worst performance. The

degradation becomes more serious with the increase of relay SNR. It should be noted that, based on our current setup, the amount of CSI feedback through the destination-relay feedback channel is two times larger than those through the relay-source feedback channel.

VI. CONCLUSION

This work has studied the joint source-relay linear precoding method for a three-node two-hop MIMO cooperative relay networks. Dropping the unrealistic the Gaussian source assumption, we incorporated the more general and practical finite-alphabet signals such as M-QAM and M-PSK. Our goal is to maximize the achievable rate by maximizing mutual information between source signal and the destination output. We derived the optimal left singular vector of relay precoder and proved the concavity of mutual information with respect to the power vector of relay precoder. Utilizing such results, we developed effective numerical optimization algorithms to iteratively search for the optimal solution. Monte Carlo simulation results under finite-alphabet input have demonstrated significant performance gain achieved by the proposed methods in terms of the ergodic mutual information and the coded FER over designs optimized for Gaussian input signals.

APPENDIX

A. Proof of Proposition 1

From Theorem 1, we conclude that the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ only depends on $\mathbf{B}$ of (16). This proposition further shows that $\mathbf{B}$ does not depend on the left singular vectors of $\mathbf{R}$.

First, definition (9) states that

$$\mathbf{B} = \mathbf{S}^\dagger \left[\mathbf{H}_0^\dagger \mathbf{H}_0 + \mathbf{H}_1^\dagger \mathbf{T}^\dagger (\mathbf{I} + \mathbf{T} \mathbf{T}^\dagger)^{-1} \mathbf{T} \mathbf{H}_1 \right] \mathbf{S}. \quad (28)$$

Applying the matrix inversion lemma [35], $(\mathbf{I} + \mathbf{T} \mathbf{T}^\dagger)^{-1} = \mathbf{I} - \mathbf{T} (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{T}^\dagger$ to (28) yields

$$\begin{aligned} \mathbf{B} &= \mathbf{S}^\dagger \left(\mathbf{H}_0^\dagger \mathbf{H}_0 \right. \\ &\quad \left. + \mathbf{H}_1^\dagger \left[\mathbf{T}^\dagger \mathbf{T} - \mathbf{T}^\dagger \mathbf{T} (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{T}^\dagger \mathbf{T} \right] \mathbf{H}_1 \right) \mathbf{S} \\ &= \mathbf{S}^\dagger \left(\mathbf{H}_0^\dagger \mathbf{H}_0 + \mathbf{H}_1^\dagger \mathbf{T}^\dagger \mathbf{T} (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{H}_1 \right) \mathbf{S}. \end{aligned} \quad (29)$$

Here we can see that $\mathbf{B}$ depends on the basic element $\mathbf{T}^\dagger \mathbf{T}$,

$$\begin{aligned} \mathbf{T}^\dagger \mathbf{T} &= \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}} \mathbf{R}^\dagger \mathbf{H}_2^\dagger \\ &\quad \times \mathbf{H}_2 \mathbf{R} \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}}. \end{aligned} \quad (30)$$

Borrowing a result from [36] (Proposition 1), given any $\mathbf{R}_1$, we can always find $\mathbf{R}_2$ with the left singular vectors coinciding with the right singular vectors of channel $\mathbf{H}_2$ such that

$$\mathbf{R}_1^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{R}_1 = \mathbf{R}_2^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{R}_2.$$

Consequently, we can always choose the right singular vectors of $\mathbf{H}_2$ as the left singular vectors of $\mathbf{R}$. ■

B. Proof of Theorem 2

First, we will restate the definitions in [37] useful to obtain our proof. Let $\mathbf{F}(\mathbf{Z}, \mathbf{Z}^*)$ be a complex-valued matrix function of $\mathbf{Z}$ and $\mathbf{Z}^*$. The Jacobian matrices of $\mathbf{F}$ with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$ are respectively given

$$\mathcal{D}_{\mathbf{Z}}(\mathbf{F}) \triangleq \frac{\partial \text{vec}(\mathbf{F})}{\partial \text{vec}^T(\mathbf{Z})} \quad \text{and} \quad \mathcal{D}_{\mathbf{Z}^*}(\mathbf{F}) \triangleq \frac{\partial \text{vec}(\mathbf{F})}{\partial \text{vec}^T(\mathbf{Z}^*)}.$$

Let $\mathbf{Z}_0$ and $\mathbf{Z}_1$ be two complex-valued matrices, and let f be a real-valued scalar function of $\mathbf{Z}_0$ and $\mathbf{Z}_1$. Then, the complex Hessian of f with respect to $\mathbf{Z}_0$ and $\mathbf{Z}_1$ is defined as

$$\mathcal{H}_{\mathbf{Z}_0, \mathbf{Z}_1} f \triangleq \frac{\partial}{\partial \text{vec}^T(\mathbf{Z}_0)} \left[\frac{\partial f}{\partial \text{vec}^T(\mathbf{Z}_1)} \right]^T.$$

We now introduce two Lemmas as intermediate results.

Lemma 1: The following results on matrix calculus hold:

$$(1\text{-a}) : d\text{vec}(\mathbf{Z}\mathbf{Z}^\dagger) = (\mathbf{Z}^* \otimes \mathbf{I}_N) d\text{vec}(\mathbf{Z}) + \mathbf{K}_{N,N} (\mathbf{Z} \otimes \mathbf{I}_N) d\text{vec}(\mathbf{Z}^*)$$

where $\mathbf{K}_{N,N}$ is the commutation matrix satisfying

$$\mathbf{K}_{N,N} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$$

$$(1\text{-b}) : d\text{vec}(\mathbf{Z}^{-1}) = -((\mathbf{Z}^T)^{-1} \otimes \mathbf{Z}^{-1}) d\text{vec}(\mathbf{Z})$$

$$(1\text{-c}) : d\text{vec}(\mathbf{Z}^{-\frac{1}{2}}) = -\left(\mathbf{Z}^{-\frac{T}{2}} \otimes \mathbf{Z}^{-\frac{1}{2}}\right) \times \left(\mathbf{Z}^{\frac{T}{2}} \oplus \mathbf{Z}^{\frac{1}{2}}\right)^{-1} d\text{vec}(\mathbf{Z}).$$

Proof: Equalities (1-a) and (1-b) are directly from [37]. We prove equality (1-c) here. Since $\mathbf{Z} = \mathbf{Z}^{\frac{1}{2}} \mathbf{Z}^{\frac{1}{2}}$, we have

$$d\text{vec}(\mathbf{Z}) = (\mathbf{Z}^{\frac{T}{2}} \oplus \mathbf{Z}^{\frac{1}{2}}) d\text{vec}(\mathbf{Z}^{\frac{1}{2}})$$

or, equivalently,

$$d\text{vec}(\mathbf{Z}^{\frac{1}{2}}) = \left(\mathbf{Z}^{\frac{T}{2}} \oplus \mathbf{Z}^{\frac{1}{2}}\right)^{-1} d\text{vec}(\mathbf{Z}).$$

Consequently,

$$\begin{aligned} d\text{vec}(\mathbf{Z}^{-\frac{1}{2}}) &= -\left(\mathbf{Z}^{-\frac{T}{2}} \otimes \mathbf{Z}^{-\frac{1}{2}}\right) d\text{vec}(\mathbf{Z}^{\frac{1}{2}}) \\ &= -\left(\mathbf{Z}^{-\frac{T}{2}} \otimes \mathbf{Z}^{-\frac{1}{2}}\right) \left(\mathbf{Z}^{\frac{T}{2}} \oplus \mathbf{Z}^{\frac{1}{2}}\right)^{-1} d\text{vec}(\mathbf{Z}). \end{aligned}$$

Directly applying Lemmas 1–7 of [19], we have

Lemma 2: The following second-order derivatives of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ are given by

$$\mathcal{D}_{\mathbf{B}}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}) = \mathcal{H}_{\mathbf{B}, \mathbf{B}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = -\mathbb{E}_{\mathbf{y}} \{ \Phi_{\mathbf{x}\mathbf{x}^\dagger}^*(\mathbf{y}) \otimes \Phi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y}) \}$$

$$\mathcal{D}_{\mathbf{B}^*}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}) = \mathcal{H}_{\mathbf{B}^*, \mathbf{B}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) = -\mathbb{E}_{\mathbf{y}} \{ \Psi_{\mathbf{x}\mathbf{x}^\dagger}^*(\mathbf{y}) \otimes \Psi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y}) \}.$$

Now we are ready to start the proof of Theorem 2. Following the definition of first derivation in [38]–[40], since

$$\begin{aligned} \frac{\partial}{\partial [\Sigma_{\mathbf{R}}^2]_{kk}} \mathcal{I}(\mathbf{x}; \mathbf{y}) &= \text{Tr} \left\{ \left(\frac{\partial}{\partial \mathbf{B}} \mathcal{I}(\mathbf{x}; \mathbf{y}) \right)^T \frac{\partial}{\partial [\Sigma_{\mathbf{R}}^2]_{kk}} \mathbf{B} \right\} \\ &= (\text{vec}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}^T))^T \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\mathbf{B}). \end{aligned} \quad (31)$$

according to [19], we need to evaluate $\mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\mathbf{B})$.

Obviously, $\mathbf{B}$ in (29) is required to be rewritten as a function of $\Sigma_{\mathbf{R}}^2$. From Proposition 1, we have $\mathbf{U}_{\mathbf{R}} = \mathbf{V}_{\mathbf{H}_2}$. Then $\mathbf{B}$ could be given as

$$\mathbf{B} = \mathbf{S}^\dagger \mathbf{H}_0^\dagger \mathbf{H}_0 \mathbf{S} + \mathbf{L} \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2 (\mathbf{J} + \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2)^{-1} \mathbf{J} \mathbf{L}^\dagger$$

where the new notations $\mathbf{L}$ and $\mathbf{J}$ are defined below for brevity of expressions,

$$\mathbf{L} \triangleq \mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger)^{-\frac{1}{2}} \mathbf{V}_{\mathbf{R}} \quad (32)$$

$$\mathbf{J} \triangleq \mathbf{V}_{\mathbf{R}}^\dagger (\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger) \mathbf{V}_{\mathbf{R}}. \quad (33)$$

Based on the chain rule provided by Theorem 1 of [37], we have

$$\begin{aligned} &\mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\mathbf{B}) \\ &= \left[\left((\mathbf{J} + \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2)^{-1} \mathbf{J} \mathbf{L}^\dagger \right)^T \otimes \mathbf{L} \right] \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2) \\ &\quad + \left((\mathbf{J} \mathbf{L}^\dagger)^T \otimes \mathbf{L} \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2 \right) \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}} \left[(\mathbf{J} + \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2)^{-1} \right] \\ &= \left\{ \left[(\mathbf{J} + \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2)^{-1} \mathbf{J} \mathbf{L}^\dagger \right]^T \right. \\ &\quad \left. \otimes \mathbf{L} \mathbf{J} (\mathbf{J} + \Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2)^{-1} \right\} \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\Sigma_{\mathbf{H}_2}^2 \Sigma_{\mathbf{R}}^2) \\ &= \text{vec} \left(\mathbf{K} \mathbf{V}_{\mathbf{R}} \Sigma_{\mathbf{H}_2} \mathbf{e}_k \mathbf{e}_k^\dagger \Sigma_{\mathbf{H}_2} \mathbf{V}_{\mathbf{R}}^\dagger \mathbf{K}^\dagger \right). \end{aligned} \quad (34)$$

Note that the second step follows the results of Lemma 1; the last step is from the definition of $\mathbf{K}$ in (18). Substituting (34) into (31), we have the final result (17).

Next, we will investigate the concavity of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with respect to $\Sigma_{\mathbf{R}}^2$ by fixing $\mathbf{V}_{\mathbf{R}}$ and $\mathbf{S}$. The Hessian of mutual information with respect to $\Sigma_{\mathbf{R}}^2$ will be evaluated. First, by the chain rule, the element-wise second order derivative is

$$\begin{aligned} \frac{\partial}{\partial [\Sigma_{\mathbf{R}}^2]_{jj}} \left(\frac{\partial \mathcal{I}}{\partial [\Sigma_{\mathbf{R}}^2]_{kk}} \right) &= \underbrace{\left(\mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\mathbf{B}) \right)^T}_{c_1} \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{jj}}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}) \\ &\quad + \underbrace{(\text{vec}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}^T))^T}_{c_2} \mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{jj}} \left(\mathcal{D}_{[\Sigma_{\mathbf{R}}^2]_{kk}}(\mathbf{B}) \right). \end{aligned} \quad (35)$$

Based on the chain rule and Lemma 2,

$$c_1 = \underbrace{\left(\mathcal{D}_{[\Sigma_R^2]_{kk}}(\mathbf{B}) \right)^T \mathcal{H}_{\mathbf{B}, \mathbf{B}^*} \mathcal{I}(\mathbf{x}; \mathbf{y}) \mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{B})}_{c_{11}} + \underbrace{\left(\mathcal{D}_{[\Sigma_R^2]_{kk}}(\mathbf{B}) \right)^T \mathcal{H}_{\mathbf{B}^*, \mathbf{B}} \mathcal{I}(\mathbf{x}; \mathbf{y}) \mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{B}^*)}_{c_{12}}. \quad (36)$$

Let us define

$$\mathbf{g} \triangleq \Sigma_{H_2} \mathbf{V}_R^\dagger \mathbf{K}^\dagger \mathbf{x}. \quad (37)$$

Applying Lemma 2 and (34), we obtain the next equation after some manipulations.

$$\begin{aligned} c_{11} &= -\mathbb{E}_{\mathbf{y}} \left[\mathbf{e}_k^\dagger \Sigma_{H_2} \mathbf{V}_R^\dagger \mathbf{K}^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y}) \mathbf{K} \mathbf{V}_R \Sigma_{H_2} \mathbf{e}_j \right. \\ &\quad \left. \times \mathbf{e}_j^\dagger \Sigma_{H_2} \mathbf{V}_R^\dagger \mathbf{K}^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger}(\mathbf{y}) \mathbf{K} \mathbf{V}_R \Sigma_{H_2} \mathbf{e}_k \right] \\ &= -\mathbb{E}_{\mathbf{y}} \left[\mathbf{e}_k^\dagger \Phi_{\mathbf{g}\mathbf{g}^\dagger}(\mathbf{y}) \mathbf{e}_j \mathbf{e}_j^\dagger \Phi_{\mathbf{g}\mathbf{g}^\dagger}(\mathbf{y}) \mathbf{e}_k \right] \\ &= -\left[\mathbb{E}_{\mathbf{y}} \left\{ \Phi_{\mathbf{g}\mathbf{g}^\dagger}(\mathbf{y}) \odot \Phi_{\mathbf{g}\mathbf{g}^\dagger}^*(\mathbf{y}) \right\} \right]_{kj} \end{aligned} \quad (38)$$

where $\Phi_{\mathbf{g}\mathbf{g}^\dagger}(\mathbf{y})$ in the second step follows the definition in (12).

Similarly, with definitions (37) and (13), we could express

$$c_{12} = -\left[\mathbb{E}_{\mathbf{y}} \left\{ \Psi_{\mathbf{g}\mathbf{g}^T}(\mathbf{y}) \odot \Psi_{\mathbf{g}\mathbf{g}^T}^*(\mathbf{y}) \right\} \right]_{kj}. \quad (39)$$

Furthermore, utilizing the identity $\mathbf{r}\mathbf{s}^\dagger \odot (\mathbf{r}\mathbf{s}^\dagger)^* = [\mathbf{r} \odot \mathbf{r}^*][\mathbf{s} \odot \mathbf{s}^*]^\dagger$ with $\mathbf{s}$ and $\mathbf{r}$ being complex-valued column vectors, and defining vector $\mathbf{z}_{\mathbf{y}}$ as

$$\mathbf{z}_{\mathbf{y}} \triangleq \mathbb{E}_{\mathbf{g}} \{ [\mathbf{g} - \hat{\mathbf{g}}(\mathbf{y})] \odot [\mathbf{g} - \hat{\mathbf{g}}(\mathbf{y})]^* | \mathbf{y} \} \quad (40)$$

we have the following conclusions,

$$\begin{aligned} \Phi_{\mathbf{g}\mathbf{g}^\dagger}(\mathbf{y}) \odot \Phi_{\mathbf{g}\mathbf{g}^\dagger}^*(\mathbf{y}) &= \mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^\dagger \succeq 0 \\ \Psi_{\mathbf{g}\mathbf{g}^T}(\mathbf{y}) \odot \Psi_{\mathbf{g}\mathbf{g}^T}^*(\mathbf{y}) &= \mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^T \succeq 0 \end{aligned}$$

where the second positive-semidefinite property is from $\mathbf{z}_{\mathbf{y}} \in \mathcal{R}^L$. Hence, we have obtained the first part of the Hessian of mutual information with respect to Σ_R^2 , which is

$$c_1 = -\left[\mathbb{E}_{\mathbf{y}} \left\{ \mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^\dagger + \mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^T \right\} \right]_{k,j}.$$

Next, we will evaluate the Jacobian $\mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathcal{D}_{[\Sigma_R^2]_{kk}}(\mathbf{B}))$ in c_2 . For conciseness, we define vector $\mathbf{c}$, where $\mathbf{c}\mathbf{c}^\dagger$ is exactly $\mathcal{D}_{[\Sigma_R^2]_{kk}}(\mathbf{B})$

$$\mathbf{c} \triangleq \mathbf{L}\mathbf{J}(\mathbf{J} + \Sigma_{H_2}^2 \Sigma_R^2)^{-1} \Sigma_{H_2} \mathbf{e}_k. \quad (41)$$

If we apply (41) and Lemma 1, we can write

$$c_2 = \underbrace{\left(\text{vec}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}^T) \right)^T (\mathbf{c}^* \otimes \mathbf{I}_N) \mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{c})}_{c_{21}} + \underbrace{\left(\text{vec}(\Phi_{\mathbf{x}\mathbf{x}^\dagger}^T) \right)^T \mathbf{K}_{M,M}(\mathbf{c} \otimes \mathbf{I}_M) \mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{c}^*)}_{c_{22}}. \quad (42)$$

Utilizing (41), $\mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{c})$ could be easily obtained as follows,

$$\mathcal{D}_{[\Sigma_R^2]_{jj}}(\mathbf{c}) = -\text{vec} \left\{ \mathbf{K} \mathbf{V}_R \Sigma_{H_2} \mathbf{e}_j \mathbf{e}_j^\dagger \Sigma_{H_2} \times (\mathbf{J} + \Sigma_{H_2}^2 \Sigma_R^2)^{-1} \Sigma_{H_2} \mathbf{e}_k \right\}. \quad (43)$$

Let us define

$$\mathbf{M} \triangleq \Sigma_{H_2} \mathbf{V}_R^\dagger \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger + \mathbf{R}^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{R} \right)^{-1} \mathbf{V}_R \Sigma_{H_2}.$$

Applying (43) and some simplification, we can write

$$\begin{aligned} c_{21} &= -\mathbf{e}_k^\dagger \Sigma_{H_2} \mathbf{V}_R^\dagger \mathbf{K}^\dagger \Phi_{\mathbf{x}\mathbf{x}^\dagger} \mathbf{K} \mathbf{V}_R \Sigma_{H_2} \mathbf{e}_j \\ &\quad \times \left[\mathbf{e}_k^\dagger \Sigma_{H_2} (\mathbf{J} + \Sigma_{H_2}^2 \Sigma_R^2)^{-1} \Sigma_{H_2} \mathbf{e}_j \right]^* \\ &= -[\Phi_{\mathbf{g}\mathbf{g}^\dagger} \odot \mathbf{M}^*]_{k,j}. \end{aligned} \quad (44)$$

Since $\Phi_{\mathbf{g}\mathbf{g}^\dagger}$ and $\mathbf{M}$ are positive-semidefinite matrices, according to the Schur Product Theorem in [41], we conclude that

$$\Phi_{\mathbf{g}\mathbf{g}^\dagger} \odot \mathbf{M}^* \succeq 0.$$

Similarly, we can obtain

$$c_{22} = -[\Phi_{\mathbf{g}\mathbf{g}^\dagger}^* \odot \mathbf{M}]_{k,j}. \quad (45)$$

The Hessian of mutual information is eventually obtained as follows by combining c_{11} , c_{12} , c_{21} , and c_{22}

$$\begin{aligned} \mathcal{H}_{\Sigma_R^2, \Sigma_R^2} \mathcal{I}(\mathbf{x}, \mathbf{y}) &= -\mathbb{E}_{\mathbf{y}} \left(\mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^\dagger + \mathbf{z}_{\mathbf{y}} \mathbf{z}_{\mathbf{y}}^T \right) \\ &\quad - \left(\Phi_{\mathbf{g}\mathbf{g}^\dagger}^* \odot \mathbf{M} + \Phi_{\mathbf{g}\mathbf{g}^\dagger} \odot \mathbf{M}^* \right) \preceq 0 \end{aligned}$$

which has been shown to be negative-semidefinite. Therefore, the mutual information is a concave function of Σ_R^2 , which conclude the proof of Theorem 2. $\blacksquare$

C. The Derivation of Gradients in III-B

Observing the definition in (9), the matrix $\mathbf{T}$ in fact is irrelevant to $\mathbf{V}_s$. Therefore, applying the result of (22) in [42], we could directly obtain the first derivative in III-B, the gradient of mutual information with respect to the right vector of source precoder.

According to (31) in the proof of Theorem 2, the second derivative, the gradient of mutual information with respect to relay precoder, easily follows from the chain rule and the fol-

lowing derivatives: (46) and (47),

$$\mathcal{D}_{[\mathbf{R}]_{kj}^*}(\mathbf{B}) = \left([(\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{H}_1 \mathbf{S}]^T \right. \\ \left. \otimes \mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \right) \mathcal{D}_{[\mathbf{R}]_{kj}^*}(\mathbf{T}^\dagger \mathbf{T}) \quad (46)$$

$$\mathcal{D}_{[\mathbf{S}]_{kj}^*}(\mathbf{T}^\dagger \mathbf{T}) = \text{vec} \left\{ \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}} \mathbf{e}_j \mathbf{e}_k^\dagger \mathbf{H}_2^\dagger \right. \\ \left. \times \mathbf{H}_2 \mathbf{R} \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}} \right\}. \quad (47)$$

To derive the third derivative, the gradient of mutual information with respect to relay precoder, we need to evaluate $\mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{B})$. According to the chain rule, we have

$$\mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{B}) = \text{vec} \left[\mathbf{e}_j \mathbf{e}_i^\dagger \left(\mathbf{H}_0^\dagger \mathbf{H}_0 + \mathbf{H}_1^\dagger \mathbf{T}^\dagger \mathbf{T} (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \mathbf{H}_1 \right) \mathbf{S} \right] \\ + \left[\mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \right]^* \\ \otimes \left[\mathbf{S}^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{T}^\dagger \mathbf{T})^{-1} \right] \mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{T}^\dagger \mathbf{T}). \quad (48)$$

The derivative term in (48) is

$$\mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{T}^\dagger \mathbf{T}) = [(\mathbf{T}^\dagger \mathbf{H}_2 \mathbf{R})^* \oplus (\mathbf{T}^\dagger \mathbf{H}_2 \mathbf{R})] \\ \times \mathcal{D}_{[\mathbf{S}]_{ij}^*} \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}}. \quad (49)$$

Applying Lemma 1, the derivative term in (49) is expanded into

$$\mathcal{D}_{[\mathbf{S}]_{ij}^*} \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}} \\ = - \left[\left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{T}{2}} \otimes \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{-\frac{1}{2}} \right] \\ \times \left[\left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{\frac{T}{2}} \oplus \left(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger \right)^{\frac{1}{2}} \right] \\ \times \mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger). \quad (50)$$

We easily obtain that

$$\mathcal{D}_{[\mathbf{S}]_{ij}^*}(\mathbf{I} + \mathbf{H}_1 \mathbf{S} \mathbf{S}^\dagger \mathbf{H}_1^\dagger) = \text{vec}(\mathbf{H}_1 \mathbf{S} \mathbf{e}_j^\dagger \mathbf{e}_i \mathbf{H}_1^\dagger) \\ = [\mathbf{H}_1]_i^* \otimes [\mathbf{H}_1 \mathbf{S}]_j. \quad (51)$$

Substitute (51), (50), and (49) into (48). Following the steps in Theorem 2 and by applying the definitions (25), (26), and (27), we hence obtain the final result in (24).

REFERENCES

- [1] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, "Relay architectures for 3GPP LTE-Advanced," *EURASIP J. Wireless Commun. Net.*, vol. 2009, no. 1, pp. 618–787, 2009.
- [2] K. Loa *et al.*, "IMT-advanced relay standards [WiMAX/LTE Update]," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 40–48, Aug. 2010.
- [3] 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.
- [4] R. Zhang, C. C. Chai, and Y.-C. Liang, "Joint beamforming and power control for multiantenna relay broad channel with QoS constraints," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 726–737, Feb. 2009.
- [5] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.
- [6] T. Q. Duong, V. N. Q. Bao, and H. j. Zepernick, "On the performance of selection decode-and-forward relay networks over Nakagami-m fading channels," *IEEE Commun. Lett.*, vol. 13, no. 3, pp. 172–174, Mar. 2009.
- [7] X. Wu and L. L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, May 2013.
- [8] 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.
- [9] A. Behbahani, R. Merched, and A. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.
- [10] 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.
- [11] Z. Fang, Y. Hua, and J. C. Koshy, "Joint source and relay optimization for a non-regenerative MIMO relay," in *Proc. 4th IEEE Workshop Sens. Array Multichannel Process.*, Jul. 2006, pp. 239–243.
- [12] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.
- [13] S. Zhou, G. Alfano, C. F. Chiasserini, and A. Nordin, "MIMO relay network with precoding," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 316–319, Feb. 2016.
- [14] C. Zhao and B. Champagne, "A unified approach to optimal transceiver design for nonregenerative MIMO relaying," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 2938–2951, Jul. 2015.
- [15] H. Shen, W. Xu, and C. Zhao, "A semi-closed form solution to MIMO relaying optimization with source-destination link," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 247–251, Feb. 2016.
- [16] R. Mo, Y. H. Chew, and C. Yuen, "Information rate and relay precoder design for amplify-and-forward MIMO relay networks with imperfect channel state information," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 3958–3968, Nov. 2012.
- [17] A. Lozano, A. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.
- [18] F. Perez-Cruz, M. Rodrigues, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.
- [19] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.
- [20] W. Zeng, C. Xiao, and J. Lu, "A low-complexity design of linear precoding for MIMO channels with finite-alphabet inputs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 38–41, Feb. 2012.
- [21] S. R. Jin, W. C. Choi, J. H. Park, and D. J. Park, "Linear precoding design for mutual information maximization in generalized spatial modulation with finite alphabet inputs," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1323–1326, Aug. 2015.
- [22] X. Liang, C. Zhao, and Z. Ding, "Sequential linear MIMO precoder optimization for hybrid ARQ retransmission of QAM signals," *IEEE Commun. Lett.*, vol. 15, no. 9, pp. 913–915, Sep. 2011.
- [23] Y. Wu, C. K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear precoding for the MIMO multiple access channel with finite alphabet inputs and statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 983–997, Feb. 2015.
- [24] W. Zeng, Y. R. Zheng, M. Wang, and J. Lu, "Linear precoding for relay networks: A perspective on finite-alphabet inputs," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1146–1157, Mar. 2012.
- [25] W. Zeng, C. Xiao, J. Lu, and K. B. Letaief, "Globally optimal precoder design with finite-alphabet inputs for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1861–1874, Nov. 2012.
- [26] Y. Wu, C. Xiao, X. Gao, J. D. Matyjas, and Z. Ding, "Linear precoder design for MIMO interference channels with finite-alphabet signaling," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3766–3780, Sep. 2013.
- [27] S. X. Wu, Q. Li, A. M. C. So, and W. K. Ma, "Rank-two beamforming and stochastic beamforming for MISO physical-layer multicasting with finite-alphabet inputs," *Signal Process. Lett.*, vol. 22, no. 10, pp. 1614–1618, Oct. 2015.
- [28] W. Wu, K. Wang, W. Zeng, Z. Ding, and C. Xiao, "Cooperative multi-cell MIMO downlink precoding with finite-alphabet inputs," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 766–779, Mar. 2015.
- [29] X. Liang, Z. Ding, and C. Xiao, "On linear precoding of non-regenerative MIMO relays for QAM inputs," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 1–6.

- [30] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [31] D. G. Luenberger, *Linear and Non-Linear Programming*. New York, NY, USA: Springer, 2008.
- [32] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
- [33] Z. Fang, Y. Hua, and J. C. Koshy, "Joint source and relay optimization for a non-regenerative MIMO relay," in *Proc. 4th IEEE Workshop Sensor Array Multichannel Process.*, Jul. 2006, pp. 239–243.
- [34] D. MacKay and R. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 33, no. 6, pp. 457–458, Mar. 1997.
- [35] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge U.K.: Cambridge Univ. Press, Feb. 1990.
- [36] M. Payaro and D. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2009, pp. 1085–1089.
- [37] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.
- [38] M. R. Bhatnagar and A. Hjørungnes, "Linear precoding of STBC over correlated Ricean MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1832–1836, Jun. 2010.
- [39] M. R. Bhatnagar, A. Hjørungnes, and L. Song, "Precoded differential orthogonal space-time modulation over correlated Ricean MIMO channels," *IEEE J. Sel. Signal Process.*, vol. 2, no. 2, pp. 124–134, Apr. 2008.
- [40] M. R. Bhatnagar, A. Hjørungnes, and L. Song, "Differential coding for non-orthogonal space-time block codes with non-unitary constellations over arbitrarily correlated Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3985–3995, Aug. 2009.
- [41] R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications*. Cambridge U.K.: Cambridge Univ. Press, 1997.
- [42] D. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.



Xiao Liang (S'12–M'13) received the B.Sc., M.S., and Ph.D. degrees in communication and information engineering in 2000, 2005, and 2013, respectively, all from Southeast University, Nanjing, China.

He is currently a Lecturer with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests include signal processing for communications and wireless networking.



Zhi Ding (S'88–M'90–SM'95–F'03) received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, USA, in 1990. From 1990 to 2000, he was a Faculty Member at Auburn University and later, University of Iowa. He has held visiting positions in Australian National University, Hong Kong University of Science and Technology, NASA Lewis Research Center, and USAF Wright Laboratory. He has active collaboration with researchers from several countries including Australia, China, Japan, Canada, Taiwan, Korea, Singapore, and Hong

Kong. He is currently a Professor of electrical and computer engineering at the University of California, Davis, CA, USA. He is a coauthor of the text: *Modern Digital and Analog Communication Systems*, 4th edition (Oxford University Press, 2009).

Dr. Ding has been an active member of IEEE, serving on technical programs of several workshops and conferences. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994–1997, 2001–2004, and an Associate Editor of IEEE SIGNAL PROCESSING LETTERS 2002–2005. He was a member of technical committee on Statistical Signal and Array Processing and member of technical committee on Signal Processing for Communications (1994–2003). He was the General Chair of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing and the Technical Program Chair of the 2006 IEEE Globecom. He was also an IEEE Distinguished Lecturer (Circuits and Systems Society, 2004–2006, Communications Society, 2008–2009). He served on as IEEE Transactions on Wireless Communications Steering Committee Member (2007–2009) and its Chair (2009–2010).



Chengshan Xiao (M'99–SM'02–F'10) received the B.Sc. degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, the M.Sc. degree in electronic engineering from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree in electrical engineering from the University of Sydney, Sydney, NSW, Australia, in 1997.

Dr. Xiao is a Professor in the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA, (formerly the University of Missouri—Rolla). He is currently serving as a Program Director at USA National Science Foundation through intergovernmental personnel act assignment. Previously, he was a senior member of scientific staff with Nortel Networks, Ottawa, Canada, a faculty member at Tsinghua University, Beijing, China, the University of Alberta, Edmonton, Canada, and the University of Missouri, Columbia, MO, USA. He also held Visiting Professor positions in Germany and Hong Kong. His research interests include wireless communications, signal processing, and underwater acoustic communications. He is the holder of three U.S. patents. His invented algorithms have been implemented into Nortel's base station radio products after successful technical field trials and network integration.

Dr. Xiao served as an elected member of Board of Governors, a member of Fellow Evaluation Committee, Director of Conference Publications, Distinguished Lecturer of the IEEE Communications Society, and Distinguished Lecturer of the IEEE VEHICULAR TECHNOLOGY. He also served as an Editor, an Area Editor, and the Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I. He was the Technical Program Chair of the 2010 IEEE International Conference on Communications, Cape Town, South Africa. He served as the founding Chair of the IEEE Wireless Communications Technical Committee. He is a Fellow of Canadian Academy of Engineering. He received several distinguished awards including the 2014 Humboldt Research Award, the 2014 IEEE Communications Society Joseph LoCicero Award, and the 2015 IEEE Wireless Communications Technical Committee Recognition Award.