Linear Precoding for Relay Networks: A Perspective on Finite-Alphabet Inputs

Weiliang Zeng, *Student Member, IEEE*, Yahong Rosa Zheng, *Senior Member, IEEE*
Mingxi Wang, *Student Member, IEEE*, and Jianhua Lu, *Senior Member, IEEE*

***Abstract*—This paper considers the precoder design for dual-hop amplify-and-forward relay networks and formulates the design from the standpoint of finite-alphabet inputs. In particular, the mutual information is employed as the utility function, which, however, results in a nonlinear and nonconcave problem. This paper exploits the structure of the optimal precoder that maximizes the mutual information and develops a two-step algorithm based on convex optimization and optimization on the Stiefel manifold. By doing so, the proposed algorithm is insensitive to initial point selection and able to achieve a near global optimal precoder solution. Besides, it converges fast and offers high mutual information gain. These advantages are verified by numerical examples, which also show the large performance gain in mutual information also represents the large gain in the coded bit-error rate.**

***Index Terms*—Relay networks, amplify-and-forward, precoder design, finite-alphabet input.**

I. INTRODUCTION

RELAYING technology is promising to provide reliable communication, high throughput, and broad coverage for wireless networks. These benefits can be achieved with judicious designs that exploit network configurations and/or relaying strategies (see [1]–[9] and references therein). The existing design methods for improving the end-to-end transmission rate and reliability of the relay networks may be categorized into two groups: diversity oriented designs and data-rate oriented designs. The diversity oriented designs try to obtain the highest diversity order, by reaching the steepest asymptotic slope of either the outage capacity or uncoded bit error rate (BER) versus signal-to-noise ratio (SNR) curve, but are not necessary to achieve the highest coding gain. Examples of such designs include the distributed space-time coding in [1], [2] and the precoder design for maximal diversity in [3], [4]. The data-rate oriented designs try to obtain the highest data rate of each end-to-end source-destination pair by maximizing the output SNR [5], [6] or maximizing the network capacity [5], [7]–[9] with Gaussian input assumption.

This paper considers two-hop relay networks employing simple relays that are equipped with single antenna and adopt the amplify-and-forward (AF) strategy [1]. It focuses on the design of precoder and the selection of relay node that maximize the mutual information of one source-destination pair. The channel model of such a relay network becomes very similar to that of a standard multiple-input multiple-output (MIMO) channel and the existing precoding methods developed for MIMO settings are applicable to the relay networks. However, most of the existing works in precoder design use extensively the Gaussian input assumption so that the mutual information between the transmit and receive signals is a simple and elegant function of the precoder and channel matrix and the optimization problem becomes easier to solve than that of finite-alphabet inputs. A practical system usually utilizes finite-alphabet constellations, such as pulse amplitude modulation (PAM), phase shift keying (PSK) modulation, or quadrature amplitude modulation (QAM). The mutual information of the discrete-input systems depart significantly from that of the Gaussian inputs [10]–[13]. Therefore, applying the precoder/relaying schemes designed for Gaussian inputs to discrete-input systems results in a significant performance loss from those designed directly for finite-alphabet inputs [14].

Several recent works target on the maximization of mutual information with finite-alphabet inputs and complex precoding matrices. Lozano *et. al.* [10] constrains the channel and precoding matrices to be diagonal such that the mutual information becomes a concave function of the squared precoder and the global maximal can be solved. The work in [15] discovers the linear relationship between the gradients of mutual information and the minimum mean squared error (MMSE) versus SNR curves and proposes a gradient descent (GD) method to solve for the precoder iteratively. The GD method is then further explored in [16]. However, as will be shown later in this paper, the GD method is very sensitive to initial point selection and can easily get stuck in a local maximal because the mutual information is a non-concave function of the precoder. More recently, the work in [13] derives the asymptotic mutual information expressions for relay networks in large-system regime with several parameters approaching infinity and provides necessary conditions that the optimal precoder satisfies. Hessian and concavity results are developed in [17] for real-valued channels, but with no

Manuscript received April 21, 2011; revised October 11, 2011; accepted December 8, 2011. The associate editor coordinating the review of this paper and approving it for publication was Y. Zhang.

The material in this paper has been presented in part at the IEEE International Communications Conference (ICC), Kyoto, Japan, 2011. This work was supported in part by the National Basic Research Program of China (2007CB310601), China National Science Foundation under Grant 61101071, and the US National Science Foundation under Grant CCF-0915846.

W. Zeng and J. Lu are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: zengwl07@mails.tsinghua.edu.cn, lhh-dee@mail.tsinghua.edu.cn). This work has been carried out while W. Zeng was a visiting scholar at Missouri University of Science and Technology.

Y. R. Zheng and M. Wang are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA (e-mail: zhengyr@mst.edu, mw9zd@mail.mst.edu).

Digital Object Identifier 10.1109/TWC.2012.012412.110747

precoder design. A possible design is presented in [18] that uses the analysis with real-valued channel assumption, without rigorous proof, in the complex-valued channel case. The work in [19] proposes a two-layer iterative algorithm that finds the precoding matrix by iterations between the bottom layer (*i.e.*, the precoding matrix) and top layer (*i.e.*, the compound channel-precoding matrix) using the concavity of the mutual information on the compound channel-precoding matrix. The global optimal solutions can be achieved for some combinations of discrete constellations and MIMO configurations.

This paper proposes a novel two-step iterative algorithm and a new framework for linear precoder design with finite-alphabet inputs by exploring the structure of the precoder that maximizes the mutual information. The proposed method first separates precoder and channel matrices, by the singular value decomposition (SVD), into product of the left singular vectors, diagonal power allocation matrix, and right singular vectors. Then making use of the results that the left singular vectors of the optimal precoder coincide with the right singular vectors of the effective channel matrix [20, Appendix 3.B] and that the mutual information is a concave function on the squared singular values of the linear precoder, the proposed algorithm maximizes the mutual information by first designing the power allocation matrix with a given set of the right singular vectors then optimizing the right singular vectors with the obtained power allocation matrix. The algorithm iterates through the two steps until the mutual information is maximized.

The success of the two-step iteration is based on the result that the mutual information is a concave function on the squared singular values of the linear precoder for a complex-valued channel. Although optimizing the right singular vectors is extremely difficult even for real-valued channel and precoder matrices [21], we reformulate the complex-valued problem on the complex Stiefel manifold and solve it by the gradient method with projection. The proposed two-step algorithm is applied to a two-hop relay network and the maximization of the end-to-end mutual information with multiple relay nodes is also considered by relay selection. Our simulation examples achieve 68% and 38% of mutual information improvement over no precoding for Binary PSK (BPSK) and Quadrature PSK (QPSK) systems, respectively, and similar performance gain is expected to be achievable when applied to standard MIMO channels.

The proposed precoder design algorithm has several advantages over the exiting works. First, the proposed algorithm is applicable to general AF relay networks and complex-valued MIMO systems with arbitrary combinations of constellations and antenna configurations. It contains the *unitary* matrix based precoder design method [3], [4]) as its special cases. It can also handle the special combinations of the discrete constellations and MIMO configurations addressed in [19] and achieve almost the same performance. Second, the proposed algorithm is insensitive to initial point selection and it converges much faster than the GD method. Third, the proposed algorithm, although suboptimal, can reach most of the maximal capacity predicted by Gaussian inputs at high probability when using finite-alphabet inputs and random initial points.

The remainder of this paper is organized as follows: Section II introduces the system model for the two-hop relaying scheme. The properties of mutual information are addressed in Section III. Section IV proposes the precoding scheme to maximize the mutual information, which includes the design of the left singular vectors, the power allocation matrix by convex optimization, and the right singular vectors using optimization on the complex Stiefel manifold. Section V presents several numerical examples to demonstrate the performance gain of this scheme over the existing ones. Finally, Section VI offers conclusions.

Notation: Real and complex spaces are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. The superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^+$ stand for transpose, complex conjugate, Hermitian, and Moore-Penrose pseudoinverse operations, respectively. The scalar with subscript c_i denotes the i-th element of vector $\mathbf{c}$, whereas $[\mathbf{A}]_{i,j}$ and $[\mathbf{A}]_{:,j}$ denote the (i,j)-th element and j-th column of matrix $\mathbf{A}$, respectively; $\mathbf{Diag}(\mathbf{a})$ represent a diagonal matrix whose nonzero elements are given by the elements of vector $\mathbf{a}$; $\mathbf{vec}(\mathbf{A})$ represents the vector obtained by stacking the columns of $\mathbf{A}$; $\mathbf{I}$ and $\mathbf{0}$ represent identity matrix and zero matrix of appropriate dimensions, respectively. The Kronecker matrix product is represented by $\mathbf{A} \otimes \mathbf{B}$; $\mathrm{Tr}(\mathbf{A})$ denotes the trace operation; $\mathbb{E}$ denotes statistical expectation; $\log$ and $\ln$ are used, respectively, for the base two logarithm and natural logarithm.

II. System Model

Consider a relay network with one transmit-and-receive pair, where the source node communicates to the destination node with the assistance of k relays, $r_1, r_2, \cdots, r_k$. All nodes are equipped with a single antenna and operate in the half-duplex mode. The relaying transmission system is assumed to be flat fading. The channel gain from the source to the destination is denoted by h_0, and those from the source to the i-th relay and from the i-th relay to the destination are denoted as h_i and g_i, respectively. These channel gains are assumed to remain unchanged during a period of observation. The relay system adopts the two-hop AF protocols [7] combined with single-relay selection [22]. The signals are transmitted in blocks with block length $2L$, where $L \geq 1$. The selected relay node receives in the first period of length L and transmits in the second period of length L.

The original signal at the source node is denoted by

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a^T & \mathbf{x}_b^T \end{bmatrix}^T$$

where $\mathbf{x}_a = [x_1, \cdots, x_L]^T$ and $\mathbf{x}_b = [x_{L+1}, \cdots, x_{2L}]^T$ with x_l being the symbol at the l-th time slot for $l = 1, \cdots, 2L$. It is assumed to be equally probable from a discrete constellation set, such as PSK, PAM, or QAM, with a unit covariance matrix, *i.e.*, $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}$.

The original signal is processed by a precoding matrix before being transmitted from the source node. The precoded data $\mathbf{s} = \begin{bmatrix} \mathbf{s}_a^T & \mathbf{s}_b^T \end{bmatrix}^T$ can be written as

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_a \\ \mathbf{s}_b \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix}$$

where $\mathbf{P} \in \mathbb{C}^{2L\times 2L}$ is a matrix to be designed to improve the end-to-end performance.

The source node transmits the signal $\sqrt{P_s}\mathbf{s}_a$ with power P_s during the first L time slots. Let $\mathbf{y}_i$ and $\mathbf{y}_a$ be the received signals at the i-th relay node and the destination, respectively. We have

$$\mathbf{y}_i = \sqrt{P_s} h_i \mathbf{s}_a + \mathbf{n}_i$$
$$\mathbf{y}_a = \sqrt{P_s} h_0 \mathbf{s}_a + \mathbf{n}_a$$

where $\mathbf{n}_i$ and $\mathbf{n}_a$ denote the independent and identically distributed (i.i.d.) zero-mean circularly Gaussian noise vector with unit variance at the i-th relay and the destination, respectively.

Assume that the i-th relay node is selected for the information forwarding in the second time slot and that the relay knows only the second-order statistics of h_i. The selected relay node scales the received signal by a factor b, which guarantees that the average transmit power of the i-th relay $b^2 \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{y}_i\mathbf{y}_i^H\right]/L\right)$ is less than or equal to the power constraint P_r. If the channel gains are assumed to have unit variance, then b can be chosen as $\sqrt{P_r/(1+2P_s)}$.

At the same time, the source node sends the signal $\sqrt{P_s}\mathbf{s}_b$. Hence, the destination node receives the superposition of the relay transmission and the source transmission during the second time slot:

$$\begin{aligned}\mathbf{y}_b &= b g_i \mathbf{y}_i + \sqrt{P_s} h_0 \mathbf{s}_b + \mathbf{n}_b \\ &= \sqrt{P_s} b h_i g_i \mathbf{s}_a + \sqrt{P_s} h_0 \mathbf{s}_b + \mathbf{n}_e\end{aligned}$$

where $\mathbf{n}_b$ denotes the noise vector of the destination at the second time slot, and $\mathbf{n}_e$ denotes the effective end-to-end noise with complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, N_d\mathbf{I})$ and $N_d = 1 + b^2|g_i|^2$.

For convenience of presentation, $\mathbf{y}_b$ is normalized by a factor $w = 1/\sqrt{N_d}$, and the received signal vector for the two time slots is denoted as $\mathbf{y} = \left[\mathbf{y}_a^T \ \ w\mathbf{y}_b^T\right]^T$. Thus, the effective input-output relationship for the two-hop transmission with precoding is summarized as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} = \mathbf{H}\mathbf{P}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ is the original transmitted signal vector; $\mathbf{n} = \left[\mathbf{n}_a^T \ \ w\mathbf{n}_e^T\right]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$; $\mathbf{H}$ is the effective channel matrix of the two-hop relay channel,

$$\mathbf{H} = \sqrt{P_s}\begin{bmatrix} h_0\mathbf{I} & \mathbf{0} \\ w b h_i g_i \mathbf{I} & w h_0 \mathbf{I}\end{bmatrix} \tag{2}$$

which is full rank for any nonzero channel gain h_0.

The precoding matrix $\mathbf{P}$ is thus designed to maximize the mutual information with finite-alphabet inputs. Note that to effectively implement the proposed algorithm in practice, a low-rate feedback shall be allowed from the destination to the relays and to the source, respectively, for delivering information of node selection and precoding. The feedback will be discussed in Section IV-D.

III. Mutual Information for Relay Networks

Consider conventional equiprobable discrete constellations such as M-ary PSK, PAM, or QAM, where M is the number of points in the constellation set. The mutual information between the input $\mathbf{x}$ and the output $\mathbf{y}$, with the equivalent channel matrix $\mathbf{H}$ and the precoding matrix $\mathbf{P}$ known at the receiver, is $\mathcal{I}(\mathbf{x};\mathbf{y})$ given by [12]

$$\mathcal{I}(\mathbf{x};\mathbf{y}) = \log M - \frac{1}{2LM^{2L}} \sum_{m=1}^{M^{2L}} \mathbb{E}_{\mathbf{n}} \log \sum_{k=1}^{M^{2L}} \exp(-d_{mk}) \tag{3}$$

where d_{mk} is $\|\mathbf{HP}(\mathbf{x}_m - \mathbf{x}_k) + \mathbf{n}\|^2 - \|\mathbf{n}\|^2$ and $\|\cdot\|$ denotes the Euclidean norm of a vector. Both $\mathbf{x}_m$ and $\mathbf{x}_k$ contain $2L$ symbols, taken independently from the M-ary signal constellation.

The mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ is fully determined by the distribution of $\|\mathbf{HP}(\mathbf{x}_m - \mathbf{x}_k) + \mathbf{n}\|^2$ for $m, k \in \{1, \cdots, M^{2L}\}$, which remain unchanged when a unitary transform $\mathbf{U}$ is applied on the output signal $\mathbf{y}$ because a unitary matrix is an *isometry* for the Euclidean norm. That is,

$$\mathcal{I}(\mathbf{x};\mathbf{y}) = \mathcal{I}(\mathbf{x};\mathbf{U}\mathbf{y}). \tag{4}$$

However, if the linear transform is applied on the input signal, then $\mathcal{I}(\mathbf{U}\mathbf{x};\mathbf{y})$ may change to a different value, even though the transmit power is not altered. That is,

$$\mathcal{I}(\mathbf{U}\mathbf{x};\mathbf{y}) \neq \mathcal{I}(\mathbf{x};\mathbf{y}). \tag{5}$$

Note that the property of mutual information for the discrete input vector is different from the case of Gaussian inputs. For Gaussian inputs, the mutual information $\mathcal{I}^{\mathcal{G}}(\mathbf{x};\mathbf{y})$ is unchanged when either the transmitted signal $\mathbf{x}$ or the received signal $\mathbf{y}$ is rotated by a unitary matrix:

$$\mathcal{I}^{\mathcal{G}}(\mathbf{x};\mathbf{y}) = \mathcal{I}^{\mathcal{G}}(\mathbf{U}\mathbf{x};\mathbf{y}) = \mathcal{I}^{\mathcal{G}}(\mathbf{x};\mathbf{U}\mathbf{y}). \tag{6}$$

The case of finite inputs does not follow the same rule; thus, a new opportunity is available here to improve system performance by transforming the input signal.

The $2L\times 2L$ complex precoding matrix $\mathbf{P}$ implements linear transform on $\mathbf{x}$, thus can exploit property (5) to maximize the mutual information of relay systems with finite-alphabet inputs. The optimization problem is formulated as:

$$\begin{aligned}&\text{maximize} && \mathcal{I}(\mathbf{x};\mathbf{y}) \\ &\text{subject to} && \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right]\right\} = \mathrm{Tr}\left(\mathbf{P}\mathbf{P}^H\right) \leq 2L\end{aligned} \tag{7}$$

which is difficult to solve because it is nonconcave over $\mathbf{P}$ [21]. The next section provides a new algorithm for this problem.

IV. Precoder Design to Maximize Mutual Information

This section examines the properties of the precoding matrix $\mathbf{P}$ under finite-alphabet inputs and presents several new results for complex-valued relay channels. These results are the foundation for the development of the proposed algorithm, which maximizes the mutual information.

A. Optimal Left Singular Vectors

The first step is to characterize the dependence of mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ on the precoding matrix $\mathbf{P}$. Given the signal constellation and the SNR, $\mathcal{I}(\mathbf{x};\mathbf{y})$ is a function of the variable $\|\mathbf{HP}(\mathbf{x}_m-\mathbf{x}_k)+\mathbf{n}\|^2$

$$\|\mathbf{HP}(\mathbf{x}_m-\mathbf{x}_k)+\mathbf{n}\|^2=\mathrm{Tr}\Big[\hat{\mathbf{e}}_{mk}\hat{\mathbf{e}}_{mk}^H\mathbf{P}^H\mathbf{H}^H\mathbf{HP} \\ +2\Re\left(\hat{\mathbf{e}}_{mk}^H\mathbf{P}^H\mathbf{H}^H\mathbf{n}\right)+\mathbf{nn}^H\Big]$$

where $\hat{\mathbf{e}}_{mk}=\mathbf{x}_m-\mathbf{x}_k$, and $\Re$ denotes the real part of a complex number; $\mathcal{I}(\mathbf{x};\mathbf{y})$ changes based on the distribution of $\|\mathbf{HP}(\mathbf{x}_m-\mathbf{x}_k)+\mathbf{n}\|^2$, which depends on $\mathbf{P}$ through $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$: the first term of the right-hand side depends on $\mathbf{P}$ through $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$; the second term $\hat{\mathbf{e}}_{mk}^H\mathbf{P}^H\mathbf{H}^H\mathbf{n}$ is a Gaussian random variable determined by its zero mean and its variance $\hat{\mathbf{e}}_{mk}^H\mathbf{P}^H\mathbf{H}^H\mathbf{HP}\hat{\mathbf{e}}_{mk}$, which also depends on $\mathbf{P}$ through $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$; the last term is independent of the precoding matrix. Therefore, $\mathcal{I}(\mathbf{x};\mathbf{y})$ depends on $\mathbf{P}$ through $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$, which is called the compound channel-precoding matrix.

Consider the SVD of the $2L\times 2L$ channel matrix $\mathbf{H}=\mathbf{U_H}\mathbf{Diag}(\boldsymbol{\sigma})\mathbf{V}_\mathbf{H}^H$, where $\mathbf{U_H}$ and $\mathbf{V_H}$ are unitary matrices, and the vector $\boldsymbol{\sigma}$ contains nonnegative entries in decreasing order. We also decompose the precoding matrix by SVD as $\mathbf{P}=\mathbf{U_P}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V}_\mathbf{P}^H$, where the vector $\boldsymbol{\lambda}$ is nonnegative constrained by the transmit power. To simplify notation, we define $\mathbf{U}=\mathbf{U_P}$ and $\mathbf{V}=\mathbf{V}_\mathbf{P}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are called left and right singular vectors of $\mathbf{P}$, respectively.

For a given matrix $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$ and a specific mutual information value, the structure of the precoding matrix can be used to minimize the transmit power $\mathrm{Tr}(\mathbf{PP}^H)$, as shown in Appendix 3.B of [20], by letting the left singular vectors of $\mathbf{P}$ coincide with the right singular vectors of $\mathbf{H}$, that is, $\mathbf{P}=\mathbf{V_H}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V}$. Similar results based on real-valued channels are reported in [18], [21]. In other words, for a given matrix $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$ and a specific power constraint, letting the left singular vectors of $\mathbf{P}$ equal to the right singular vectors of $\mathbf{H}$ maximizes the mutual information for general channel conditions and arbitrary inputs.

Adopting the optimal left singular vectors $\mathbf{U}=\mathbf{V_H}$, the channel matrix (1) can be further simplified by (4) to

$$\mathbf{y}=\mathbf{Diag}(\boldsymbol{\sigma})\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{Vx}+\mathbf{n}. \tag{8}$$

It is clear from (8) that the mutual information is now dependent only on the squared singular values of the precoder $\boldsymbol{\lambda}$ and the right singular vectors $\mathbf{V}$. We use the notations $\mathcal{I}(\boldsymbol{\lambda})$ and $\mathcal{I}(\mathbf{V})$ to develop a two-step optimization algorithm that first maximizes the mutual information via optimal power allocation $\boldsymbol{\lambda}$, and then via optimal right singular vectors $\mathbf{V}$.

B. Optimal Power Allocation

Given the right singular vectors of the precoder, the optimization problem (7) is addressed over $\boldsymbol{\lambda}$:

$$\begin{aligned}\text{maximize}\quad & \mathcal{I}(\boldsymbol{\lambda})\\ \text{subject to}\quad & \mathrm{Tr}\left(\mathbf{PP}^H\right)=\mathbf{1}^T\boldsymbol{\lambda}\le 2L\\ & \boldsymbol{\lambda}\succeq\mathbf{0}\end{aligned} \tag{9}$$

where $\mathbf{1}$ and $\mathbf{0}$ denote the column vector with all entries being one and zero, respectively.

We extend the Hessian and concavity results of real-valued case in [17, Theorem 5] to a complex-valued channel which are as follows:

Proposition 1: Based on the simplified channel model in (8), the mutual information is a concave function of $\boldsymbol{\lambda}$; that is, the Hessian of mutual information satisfies $\mathcal{H}_{\boldsymbol{\lambda}}\mathcal{I}(\boldsymbol{\lambda})\preceq\mathbf{0}$. Moreover, the gradient and Hessian of the mutual information are given, respectively, as

$$\nabla_{\boldsymbol{\lambda}}\mathcal{I}(\boldsymbol{\lambda})=\mathbf{R}\cdot\mathbf{vec}\left(\mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{VEV}^H\right) \tag{10}$$

and

$$\begin{aligned}\mathcal{H}_{\boldsymbol{\lambda}}\mathcal{I}(\boldsymbol{\lambda})=&-\frac{1}{2}\mathbf{R}\left[\mathbf{I}\otimes\mathbf{Diag}^2(\boldsymbol{\sigma})\right]\mathbb{E}\left\{\tilde{\boldsymbol{\Phi}}(\mathbf{y})^*\otimes\tilde{\boldsymbol{\Phi}}(\mathbf{y})\right\}\\ &\left[\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{Diag}^2(\boldsymbol{\sigma})\otimes\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\right]\\ &\mathbf{R}^T\mathbf{Diag}^{-1}(\boldsymbol{\lambda})\\ &-\frac{1}{2}\mathbf{R}\left[\mathbf{I}\otimes\mathbf{Diag}^2(\boldsymbol{\sigma})\right]\mathbb{E}\left\{\tilde{\boldsymbol{\Psi}}(\mathbf{y})^*\otimes\tilde{\boldsymbol{\Psi}}(\mathbf{y})\right\}\\ &\left[\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\otimes\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{Diag}^2(\boldsymbol{\sigma})\right]\\ &\mathbf{KR}^T\mathbf{Diag}^{-1}(\boldsymbol{\lambda})\end{aligned} \tag{11}$$

where $\mathbf{E}$ is the minimum mean square error (MMSE) matrix, defined as

$$\mathbf{E}\triangleq\mathbb{E}\left\{[\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y})][\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y})]^H\right\};$$

and $\tilde{\boldsymbol{\Phi}}(\mathbf{y})=\mathbf{V}\boldsymbol{\Phi}(\mathbf{y})\mathbf{V}^H$ and $\tilde{\boldsymbol{\Psi}}(\mathbf{y})=\mathbf{V}\boldsymbol{\Psi}(\mathbf{y})\mathbf{V}^T$ with $\boldsymbol{\Phi}(\mathbf{y})$ and $\boldsymbol{\Psi}(\mathbf{y})$ being the MMSE matrix and companion MMSE matrix conditioned on a realization of the received signal $\mathbf{y}$, defined by

$$\boldsymbol{\Phi}(\mathbf{y})\triangleq\mathbb{E}\left\{[\mathbf{x}-\mathbb{E}\{\mathbf{x}|\mathbf{y}\}][\mathbf{x}-\mathbb{E}\{\mathbf{x}|\mathbf{y}\}]^H|\mathbf{y}\right\}$$

and

$$\boldsymbol{\Psi}(\mathbf{y})\triangleq\mathbb{E}\left\{[\mathbf{x}-\mathbb{E}\{\mathbf{x}|\mathbf{y}\}][\mathbf{x}-\mathbb{E}\{\mathbf{x}|\mathbf{y}\}]^T|\mathbf{y}\right\}$$

respectively. And $\mathbf{R}\in\mathbb{R}^{2L\times 4L^2}$ is a reduction matrix with entries given by $[\mathbf{R}]_{i,2L(j-1)+k}=\delta_{ijk}$.

Proof: See Appendix A. □

The concavity property in Proposition 1 ensures that a global optimal power allocation vector can be found given right singular vectors $\mathbf{V}$. In addition, the gradient and Hessian results in (10) and (11) permit either the steepest descent or Newton-type algorithms to solve for the global optimal power allocation vector. However, the existing general-purpose solvers for convex problem (*e.g.*, CVX [23]) fail to address this problem because of the complexity of the objective function $\mathcal{I}(\mathbf{x};\mathbf{y})$. Therefore, a specialized interior-point method is developed here.

The first step is to re-write problem (9), making the inequality constraints implicit in the objective function:

$$\text{minimize}\quad f(\boldsymbol{\lambda})=-\mathcal{I}(\boldsymbol{\lambda})+\sum_{i=1}^{2L}\phi(-\lambda_i)+\phi(\mathbf{1}^T\boldsymbol{\lambda}-2L) \tag{12}$$

where $\phi(u)$ is the logarithmic barrier function approximating an indicator as whether constraints are violated:

$$\phi(u) = \begin{cases} -(1/t)\ln(-u), & u < 0 \\ +\infty, & u \geq 0 \end{cases}$$

in which the parameter $t > 0$ sets the accuracy of the approximation [24].

Based on Proposition 1, the gradient of the objective function (12) is written as

$$\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda}) = -\mathbf{R} \cdot \mathbf{vec}\left(\mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{V}\mathbf{E}\mathbf{V}^H\right) - \frac{1}{t}\left(\mathbf{q} - \frac{1}{2L - \mathbf{1}^T\boldsymbol{\lambda}}\right) \quad (13)$$

where $q_i = 1/\lambda_i$ is the i-th element of vector $\mathbf{q}$. Thus the steepest descent direction is chosen as

$$\Delta\boldsymbol{\lambda} = -\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda}).$$

Combining this search direction with the backtracking line search conditions [24], we establish Algorithm 1 for the optimal power allocation vector, which ensures the convergence because of the concavity.

Algorithm 1: Maximize Mutual Information Over the Power Allocation Vector:

1. Given a feasible vector $\boldsymbol{\lambda}$, $t := t^{(0)} > 0$, $\alpha > 1$, tolerance $\epsilon > 0$.
2. Compute the gradient of f at $\boldsymbol{\lambda}$, $\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$, as (13) and the descent direction $\Delta\boldsymbol{\lambda} = -\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$.
3. Evaluate $\|\Delta\boldsymbol{\lambda}\|^2$. If it is sufficiently small, then go to Step 6; else go to Step 4.
4. Choose step size γ so that $f(\boldsymbol{\lambda} + \gamma\Delta\boldsymbol{\lambda}) < f(\boldsymbol{\lambda})$ by backtracking line search.
5. Set $\boldsymbol{\lambda} := \boldsymbol{\lambda} + \gamma\Delta\boldsymbol{\lambda}$. Go to Step 2.
6. Stop if $1/t < \epsilon$, else $t := \alpha t$, and go to step 2.

C. Optimization Over Right Singular Vectors

Now we consider maximizing the mutual information over the right singular vectors $\mathbf{V}$ for a given $\boldsymbol{\lambda}$:

$$\begin{array}{ll} \text{maximize} & \mathcal{I}(\mathbf{V}) \\ \text{subject to} & \mathbf{V}^H\mathbf{V} = \mathbf{V}\,\mathbf{V}^H = \mathbf{I}. \end{array} \quad (14)$$

This unitary-matrix constrained problem can be formulated as an unconstrained optimization in a constrained search space:

$$\text{minimize} \quad g(\mathbf{V})$$

with domain restricted to the Stiefel manifold $\mathrm{St}(n)$ [25]:

$$\mathrm{dom}\, g = \{\mathbf{V} \in \mathrm{St}(n)\}$$

and

$$\mathrm{St}(n) = \left\{\mathbf{V} \in \mathbb{C}^{n\times n} | \mathbf{V}^H\mathbf{V} = \mathbf{I}\right\}$$

where the function $g(\mathbf{V})$ is defined as $-\mathcal{I}(\mathbf{V})$. For each point $\mathbf{V} \in \mathrm{St}(n)$, the search direction $\Delta\mathbf{V}$ on the tangent space has been suggested in [26] to minimize the objective function,

$$\Delta\mathbf{V} = -\nabla_{\mathbf{V}} g(\mathbf{V}) = \nabla_{\mathbf{V}}\mathcal{I}(\mathbf{V}) - \mathbf{V}(\nabla_{\mathbf{V}}\mathcal{I}(\mathbf{V}))^H\mathbf{V} \quad (15)$$

where $\nabla_{\mathbf{V}}\mathcal{I}(\mathbf{V})$ is the gradient of mutual information with respect to $\mathbf{V}$, given by $\mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{Diag}(\boldsymbol{\lambda})\mathbf{V}\mathbf{E}$.

Note that moving along the descent direction on the tangent space may cause the unitary property being lost. Therefore, it needs to be restored in each step via projection. For an arbitrary matrix $\mathbf{W} \in \mathbb{C}^{n\times n}$, its projection $\pi(\mathbf{W})$ on the Stiefel manifold is defined as the point closest to $\mathbf{W}$ in the Euclidean norm:

$$\pi(\mathbf{W}) = \arg\min_{\mathbf{Q}\in\mathrm{St}(n)} \|\mathbf{W} - \mathbf{Q}\|^2.$$

If the SVD of $\mathbf{W}$ is $\mathbf{W} = \mathbf{U_W}\boldsymbol{\Sigma}\mathbf{V_W}$, the projection can be expressed by $\mathbf{U_W}\mathbf{V_W}$ [27, Sec. 7.4.8].

Combining the search direction and the projection with the backtracking line search condition, we develop Algorithm 2 to maximize the mutual information over the right singular vectors $\mathbf{V}$.

Algorithm 2: Maximize the Mutual Information on Complex Stiefel Manifold:

1. Given a feasible $\mathbf{V} \in \mathbb{C}^{n\times n}$ such that $\mathbf{V}^H\mathbf{V} = \mathbf{I}$.
2. Compute the descent direction $\Delta\mathbf{V}$ as (15). Set the step size $\gamma := 1$.
3. Evaluate $\|\Delta\mathbf{V}\|^2 = \mathrm{Tr}\{(\Delta\mathbf{V})^H\Delta\mathbf{V}\}$. If it is sufficiently small, then stop; else go to Step 4.
4. Choose step size γ so that $g(\pi(\mathbf{V} + \gamma\Delta\mathbf{V})) < g(\mathbf{V})$ by backtracking line search.
5. Set $\mathbf{V} := \pi(\mathbf{V} + \gamma\Delta\mathbf{V})$. Go to Step 2.

D. Two-Step Approach to Optimize Precoder

A complete two-step approach can now be developed to maximize the mutual information over a generalized precoding matrix $\mathbf{P}$ by combining Algorithm 1 and Algorithm 2.

Algorithm 3: Two-Step Algorithm to Maximize the Mutual Information Over a Generalized Precoder:

1. *Initialization*. Set the left singular vectors of the precoder $\mathbf{U} := \mathbf{V_H}$. Specify a feasible $\boldsymbol{\lambda}$ and $\mathbf{V}$.
2. *Update power allocation vector*: Run Algorithm 1 given $\mathbf{V}$.
3. *Update right singular vectors*: Run Algorithm 2 given the obtained $\boldsymbol{\lambda}$ in Step 2.
4. Repeat Step 2 and Step 3 until convergence.

The proposed algorithm converges to the globally optimal solution in the low SNR region because mutual information is maximized by optimizing $\boldsymbol{\lambda}$ [16]. For medium to high SNR, the proposed method, theoretically, converges to a local maximum. However, extensive numerical examples show that different initial points have limited effect on the algorithm (see Sec. V); that is, the two-step method achieves near global optimal performance.

The destination node applies Algorithm 3 to each relay node and calculates the corresponding achievable mutual information and precoder. Denote the mutual information of the i-th

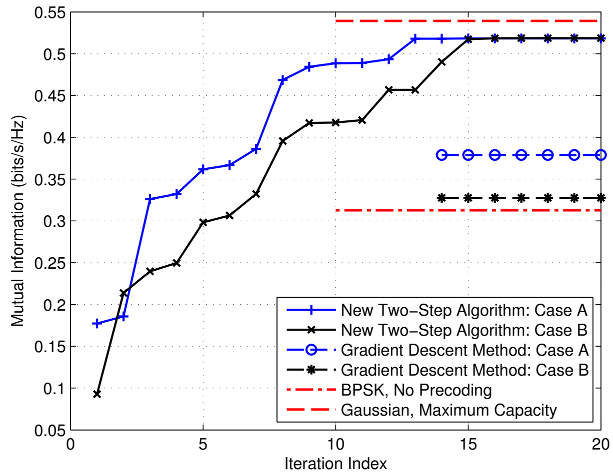

Fig. 1. Evolution of the mutual information as the linear precoder is iteratively optimized with the two-step algorithm when input signal is drawn from BPSK and SNR is 0 dB.

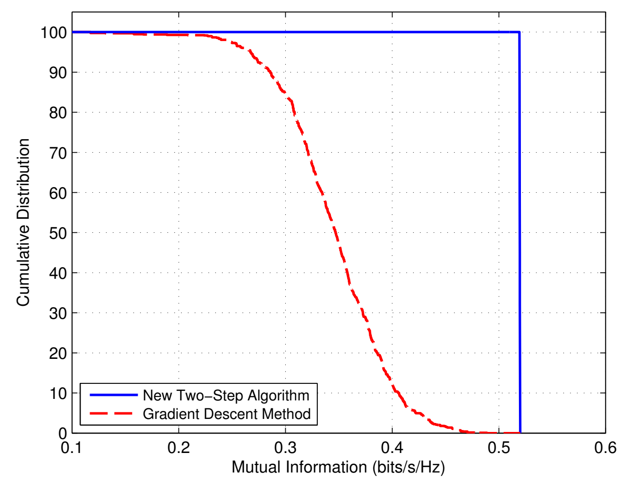

Fig. 2. Cumulative distribution of the optimized mutual information from various initial points for the proposed two-step algorithm and the GD method. The input signal is drawn from BPSK and SNR is 0 dB.

relay node as $\mathcal{I}_i$. The best relay node is then selected by

$$R_s = \arg \max_{i=1,\cdots,k} \mathcal{I}_i.$$

Next, the index of the selected node and the corresponding precoder are transmitted via a feedback channel from the destination to the relay and to the source node, respectively.

V. Simulation Results

This section examines the efficacy of the linear precoder on relay networks by several examples. These examples consider a single-relay network with a block length $L = 1$ and channel coefficients $h_0 = 0.4$, $h_1 = 1.2$, and $g_1 = -0.9\jmath$. The same transmit power is assumed at the source and relay node (*i.e.*, $P_s = P_r = P$).

The proposed two-step iterative algorithm is first tested from different feasible starting points. A general 2×2 unitary matrix group can be expressed as [28]

$$\mathbf{V} = \underbrace{\begin{bmatrix} e^{\jmath\alpha_1} & 0 \\ 0 & e^{\jmath\alpha_2} \end{bmatrix}}_{\mathbf{V}_1} \cdot \underbrace{\begin{bmatrix} \cos\psi & e^{-\jmath\theta}\sin\psi \\ -e^{\jmath\theta}\sin\psi & \cos\psi \end{bmatrix}}_{\mathbf{V}_2}.$$

For the simplified channel model in (8), the mutual information remains unchanged under the rotation of the diagonal unitary matrix $\mathbf{V}_1$. Thus, only the structure of $\mathbf{V}_2$ is used to generate feasible right singular vectors $\mathbf{V}$.

A. Mutual Information Performance

We first consider two different starting points: Case A: $\boldsymbol{\lambda} = [0.5; 0.5]$, $\psi = \pi/6$, and $\theta = \pi/4$; Case B: $\boldsymbol{\lambda} = [0.2; 0.8]$, $\psi = \pi/10$, and $\theta = \pi/10$. Figure 1 illustrates the convergence for BPSK inputs when the SNR is 0 dB. For comparison, the figure also shows the mutual information corresponding to the cases of no precoding, optimal precoder with Gaussian inputs [5], [8], and optimal precoding with the GD method [15], [16]. The GD method is influenced by its initial point selection and converges to different mutual information levels with different initial points. The proposed two-step iteration of Algorithm 3, however, is insensitive to the initial points and converges to almost the same value, which increases the mutual information to 0.52 bps/Hz or 68% over that of no precoding. It also approaches the upper bound that is achieved by Gaussian inputs. Moreover, the progress of the proposed method exhibits a staircase shape, where each stair is associated with either the shift between the optimizations of the power allocation vector and the right singular vector or the iteration for the parameter t within Algorithm 1.

The cumulative distribution of the optimized mutual information from different initial points for the proposed two-step algorithm and the GD method are depicted in Fig. 2, which is obtained by maximizing the mutual information via 5,000 uniformly random initial points that are feasible to the considered problem. From Fig. 2, the two-step algorithm has a narrow and sharp curve, and the difference between the highest and lowest optimized mutual information is less than 0.003 bps/Hz; that is, the iterative algorithm is able to obtain a near global optimal solution even if the problem is nonconcave. In contrast, the GD method depends highly on the initial point selection, and the difference between the highest and lowest optimized mutual information is as wide as 0.3 bps/Hz. It means the GD method may provide a solution even much worse than that of no precoding if the initial point is not chosen carefully.

Similar performance is also observed for QPSK with SNR given by 5 dB, as shown in Fig. 3 and 4. The convergence of the proposed two-step algorithm is achieved in 15 iterations, similar to the case of BPSK. It also achieves 38% improvement of mutual information over no precoding. From Fig. 4, the proposed algorithm obtains mutual information within 98% of the maximum capacity of Gaussian inputs with 69% of the initial point selections, while the GD method reaches within 70% of the maximum capacity of Gaussian inputs with only 5% of the initial point selections.

Figures 5 and 6 show the mutual information of the proposed algorithm versus SNR for BPSK and QPSK inputs, respectively, in comparison with other schemes such as no precoding, maximum diversity design in [3], maximum cod-

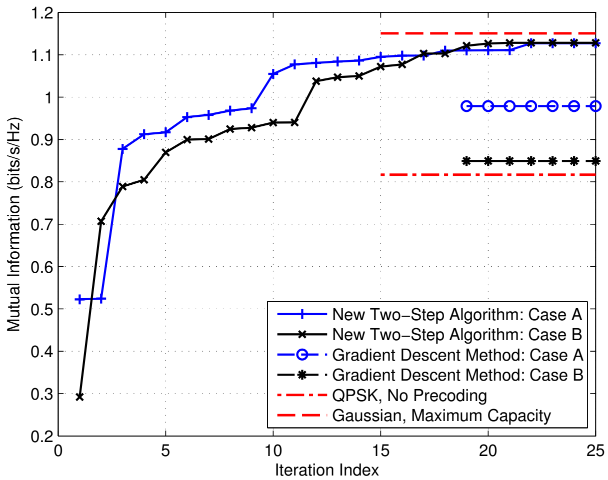

Fig. 3. Evolution of the mutual information as the linear precoder is iteratively optimized with the two-step algorithm when input signal is drawn from QPSK and SNR is 5 dB.

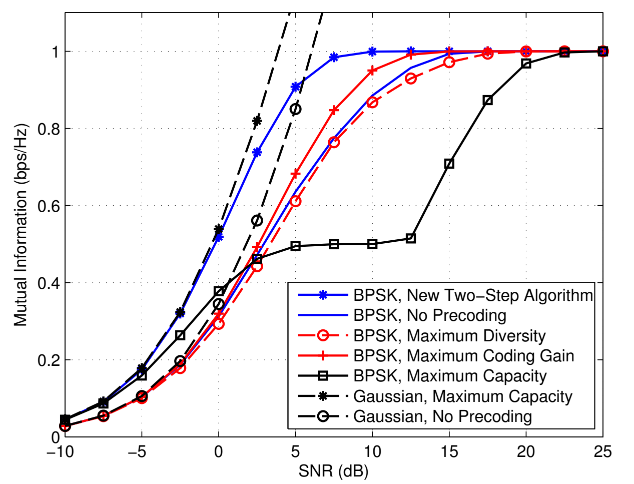

Fig. 5. Mutual information versus the SNR for different relaying strategies. The block length $L = 1$, and the transmit data signals are drawn from BPSK.

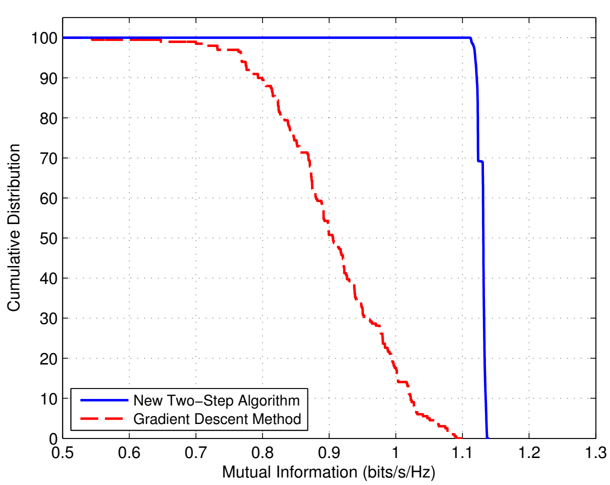

Fig. 4. Cumulative distribution of mutual information from various initial points. The input signal is drawn from QPSK and SNR is 5 dB.

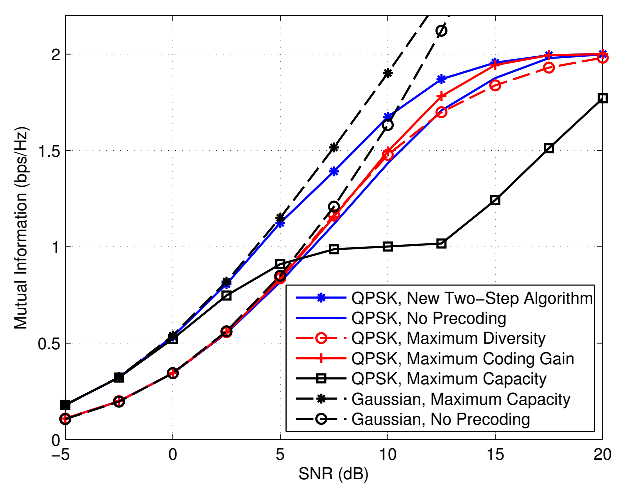

Fig. 6. Mutual information versus the SNR for different relaying strategies. The block length $L = 1$, and the transmit data signals are drawn from QPSK.

ing gain design in [3], [4], and maximum capacity design assuming Gaussian inputs in [5], [8]. Figures 5 and 6 suggest the following observations:

First, when the elements of the transmitted signal $\mathbf{x}$ are drawn from BPSK or QPSK, the mutual information for relay networks is bounded by 1 bps/Hz and 2 bps/Hz, respectively, which are achieved by all precoding schemes when SNR is high.

Second, the precoder design based on maximizing capacity with the assumption of Gaussian inputs may result in a significant loss for systems employing discrete inputs. This loss comes from differences in designing the power allocation vector and the right singular vectors. For Gaussian inputs, allocating more power to the stronger subchannels and less to the weaker subchannels is helpful to maximize capacity. This strategy, however, does not work for finite-alphabet inputs, because the mutual information with finite inputs is bounded, therefore little incentive can be gained by allocating more power to the subchannels already close to saturation. Moreover, the right singular vectors for Gaussian-input systems is an arbitrary unitary matrix [see eq. (6)]. Systems with finite inputs, however, have to carefully select the right singular vectors to maximize the mutual information.

Third, although the maximum coding gain method [3] performs better than the maximum diversity and no precoding methods, it is valid only for the case of block length $L = 1$ and QPSK inputs or the case of $L = 1$ and M-QAM inputs [4]. In comparison, the proposed two-step method can be used for an arbitrary block length L and input type. Our algorithm exploits the degrees of freedom in the optimal left singular vectors, the optimal power allocation vector, and the local optimal right singular vectors simultaneously. Therefore, providing significant gains of mutual information in a wide range of SNR. For example, with input BPSK and 3/4 channel coding rate, the performance of the proposed method is about 4 dB, 5.5 dB, and 6 dB better than those of the maximum coding gain, no precoding, and maximum capacity methods, respectively.

Last, the proposed method achieves mutual information very close to maximum capacity with Gaussian inputs when

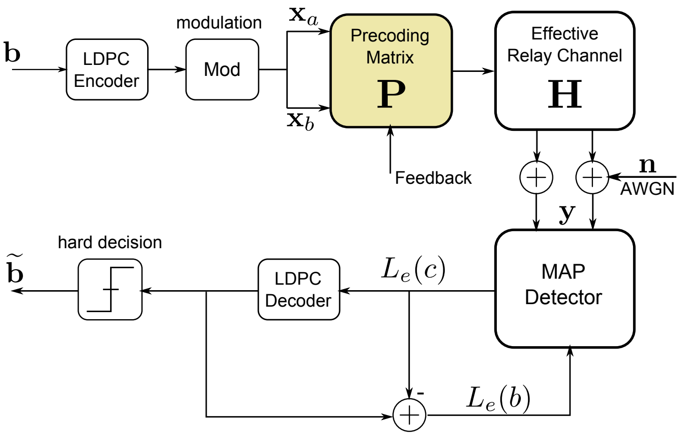

Fig. 7. Block diagram of the relay network with precoding at the source node and iterative detection and decoding at the destination node.

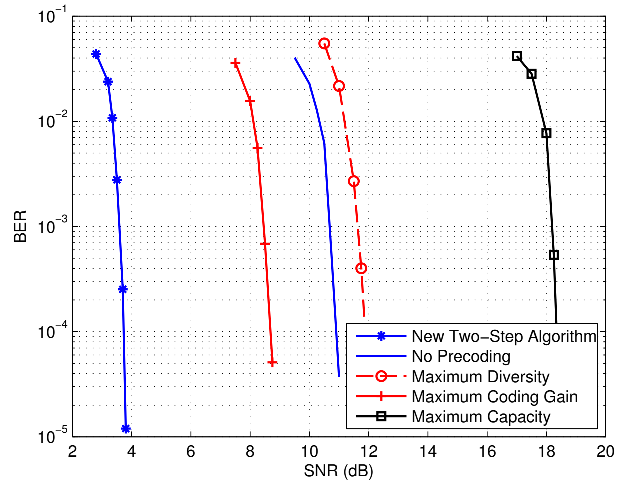

Fig. 8. Bit error performance comparison for the proposed two-step algorithm and other methods. The block length is 1, the data signals are from BPSK inputs, the channel coding rate is 3/4, coding length is 2400, and the iteration between the MAP detector and the LDPC decoder is 5.

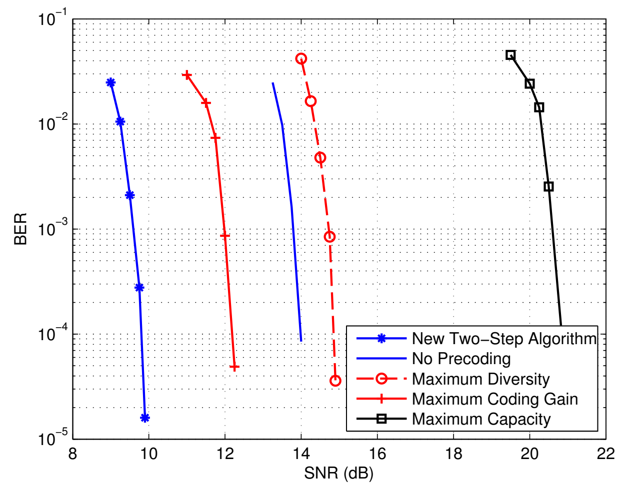

Fig. 9. Bit error performance comparison for the proposed two-step algorithm and other methods. The block length is 1, the data signals are from QPSK inputs, the channel coding rate is 2/3, coding length is 2400, and the iteration between the MAP detector and the LDPC decoder is 5.

the channel coding rate is below 0.6 for both BPSK and QPSK; it also outperforms the case of Gaussian inputs with no precoding when the channel coding rate is below 0.9.

B. Coded BER Performance

To evaluate the coded BER of the proposed method, channel coding is used at the source node and turbo principle [29], [30] is used at the destination node, as illustrated in Fig. 7. Note that the interleaver is not shown in the block diagram because of the usage of the low-density parity-check, or LDPC, codes [31]. The signal sequence $\mathbf{b}$ is encoded by the LDPC encoder and mapped according to the conventional equiprobable discrete signaling constellations. It is then divided into $\mathbf{x}_a$ and $\mathbf{x}_b$ and transmitted at the two time slots, respectively; the selected relay node simply amplifies and forwards the signal $\mathbf{x}_a$ in the second time slot.

At the receiver of the destination node, the maximum *a posteriori* (MAP) detector takes channel observations $\mathbf{y}$ and the *a priori* information $L_e(b)$ from the decoder and computes the extrinsic information $L_e(c)$ for each coded bit. Thus, the extrinsic information between the MAP detector and the decoder is exchanged iteratively until the desired performance is achieved [29].

In our simulation of the end-to-end system of Fig. 7, we use the block length for relaying transmission $L = 1$. The LDPC encoder and decoder modules are derived from [32] with coding rate 1/2. The channel coding length 2400, the coding rate 3/4 and 2/3 for BPSK and QPSK, respectively, and five iterations between the MAP detector and the LDPC decoder. Figures 8 and 9 show that the designed precoder that maximizes the mutual information provides large performance gains in the coded BER over other schemes including on precoding, the maximum coding gain, and maximum capacity methods. Note that the precoder designed by maximum capacity with Gaussian input offers worse BER performance than no precoding in practical BPSK and QPSK systems.

VI. Conclusion

This paper has proposed a new two-step iterative algorithm for linear precoder design that maximizes mutual information with finite-alphabet inputs. By setting the left singular vectors of the precoding matrix equal to the right singular vectors of the channel matrix, the mutual information has proved to be a concave function on the squared singular values of the linear precoder for a complex-valued channel. Consequently, the proposed algorithm first optimizes the mutual information via the power allocation matrix or singular values of the linear precoder with a given set of right singular vectors, then solves the second maximization problem for right singular vectors with the obtained power allocation matrix, and iterates through the two steps until convergence. The second optimization problem is difficult and we have reformulated it on the complex Stiefel manifold and have solved it by the gradient method with projection. The proposed two-step linear precoder design algorithm has several advantages: being able to handle general complex-valued channels and system configurations, insensitive to initial point selection, fast convergence, and high mutual information gain. A numerical example of a two-hop relay network with amplify-and-forward strategy and

$$\begin{aligned}\mathbf{E} &\triangleq \mathbb{E}\left\{[\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y})][\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y})]^H\right\} \\ &= \mathbf{I} - \frac{1}{(\pi M)^{2L}} \int_{\mathbf{y}} \frac{\left[\sum_{l=1}^{M^{2L}} \mathbf{x}_l \exp\left(-\|\mathbf{y}-\mathbf{HPx}_l\|^2\right)\right]\left[\sum_{k=1}^{M^{2L}} \mathbf{x}_k^H \exp\left(-\|\mathbf{y}-\mathbf{HPx}_k\|^2\right)\right]}{\sum_{m=1}^{M^{2L}} \exp\left(-\|\mathbf{y}-\mathbf{HPx}_m\|^2\right)} d\mathbf{y}. \end{aligned} \tag{17}$$

Binary/Quadrature Phase Shift Keying (BPSK/QPSK) constellations is presented to demonstrate the performance gains of the proposed linear precoder algorithm.

Appendix A: Proof of Proposition 1

To prove Proposition 1, we first introduce the definitions of gradient, Jacobian, and Hessian of a complex matrix [33].

1) Definition: The gradient matrix with respect to a complex-valued matrix $\mathbf{Z}$ is defined as

$$\nabla_{\mathbf{Z}} f \triangleq \frac{\partial f}{\partial \mathbf{Z}^*}$$

where the (i,j)-th element of the gradient matrix is $[\nabla_{\mathbf{Z}} f]_{ij} = \partial f / \partial [\mathbf{Z}^*]_{ij}$.

Let $\mathbf{F}(\mathbf{Z},\mathbf{Z}^*)$ be a complex matrix function of $\mathbf{Z}$ and $\mathbf{Z}^*$; the Jacobian matrices of $\mathbf{F}$ with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$ are then given by:

$$\mathcal{D}_{\mathbf{Z}}\mathbf{F} \triangleq \frac{\partial \mathbf{vec}(\mathbf{F})}{\partial \mathbf{vec}^T(\mathbf{Z})} \qquad \text{and} \qquad \mathcal{D}_{\mathbf{Z}^*}\mathbf{F} \triangleq \frac{\partial \mathbf{vec}(\mathbf{F})}{\partial \mathbf{vec}^T(\mathbf{Z}^*)}. \tag{16}$$

Let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be two complex-valued matrices, and let f be a real-valued scalar function of $\mathbf{Z}_1$ and $\mathbf{Z}_2$. The complex Hessian matrix of f with respect to $\mathbf{Z}_1$ and $\mathbf{Z}_2$ is defined by:

$$\mathcal{H}_{\mathbf{Z}_1,\mathbf{Z}_2} f \triangleq \mathcal{D}_{\mathbf{Z}_1}\left(\mathcal{D}_{\mathbf{Z}_2} f\right) = \frac{\partial}{\partial \mathbf{vec}^T(\mathbf{Z}_1)}\left[\frac{\partial f}{\partial \mathbf{vec}^T(\mathbf{Z}_2)}\right]^T. \tag{17}$$

2) Jacobian of the MMSE Matrix: The gradient of mutual information as the first-order characteristic has been derived in [15]:

$$\nabla_{\mathbf{P}} \mathcal{I}(\mathbf{x};\mathbf{y}) = \mathbf{H}^H\mathbf{HPE}$$

where $\mathbf{E}$ is the MMSE matrix given by (17).

The MMSE estimate of $\mathbf{x}$ (the conditional mean) is

$$\mathbb{E}(\mathbf{x}|\mathbf{y}) = \sum_{l=1}^{M^{2L}} \mathbf{x}_l p(\mathbf{x}_l|\mathbf{y}) = \mathbb{E}_{\mathbf{x}}\left[\mathbf{x}\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right]$$

and the conditional probability density function of the received signal $\mathbf{y}$ is calculated as

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^{2L}} \exp\left(-\|\mathbf{y}-\mathbf{HPx}\|^2\right).$$

The Jacobian of the MMSE matrix $\mathcal{D}_{\mathbf{P}^*}\mathbf{E}$ is now introduced as a Lemma for deriving the second-order characteristic function of the mutual information. Note that Lemma 1 hold for the general complex-valued channel case, which agrees with [17, Theorem 3] for real-valued signal model.

Lemma 1: The Jacobian of the MMSE matrix $\mathbf{E}$ with respect to $\mathbf{P}^*$ is given by

$$\begin{aligned}\mathcal{D}_{\mathbf{P}^*}\mathbf{E} = &-\mathbb{E}\left\{\mathbf{\Phi}^*(\mathbf{y})\otimes\mathbf{\Phi}(\mathbf{y})\right\}\mathbf{K}\left[\mathbf{I}\otimes\left(\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right)\right] \\ &-\mathbb{E}\left\{\mathbf{\Psi}^*(\mathbf{y})\otimes\mathbf{\Psi}(\mathbf{y})\right\}\left[\mathbf{I}\otimes\left(\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right)\right]\end{aligned} \tag{18}$$

where $\mathbf{K}$ is a unique $4L^2 \times 4L^2$ permutation matrix with $\mathbf{vec}\left(\mathbf{\Phi}^T(\mathbf{y})\right) = \mathbf{K}\cdot\mathbf{vec}\left(\mathbf{\Phi}(\mathbf{y})\right)$.

Proof: The (i,j)-th element of the MMSE matrix $\mathbf{E}$ is given by $[\mathbf{E}]_{ij} = \mathbb{E}\left\{x_i x_j^*\right\} - \mathbb{E}\{\mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j|\mathbf{y}\}\}$, from which it follows that

$$\begin{aligned}\frac{\partial[\mathbf{E}]_{ij}}{\partial[\mathbf{P}]_{kl}^*} = &-\int \frac{\partial p(\mathbf{y})}{\partial[\mathbf{P}]_{kl}^*}\mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}d\mathbf{y} \\ &-\int p(\mathbf{y})\frac{\partial\mathbb{E}\{x_i|\mathbf{y}\}}{\partial[\mathbf{P}]_{kl}^*}\mathbb{E}\{x_j^*|\mathbf{y}\}d\mathbf{y} \\ &-\int p(\mathbf{y})\mathbb{E}\{x_i|\mathbf{y}\}\frac{\partial\mathbb{E}\{x_j^*|\mathbf{y}\}}{\partial[\mathbf{P}]_{kl}^*}d\mathbf{y}.\end{aligned} \tag{19}$$

Employing $\mathcal{D}_{\mathbf{y}}\|\mathbf{y}-\mathbf{HPx}\|^2 = (\mathbf{y}-\mathbf{HPx})^H$ yields

$$\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) = -p(\mathbf{y}|\mathbf{x})(\mathbf{y}-\mathbf{HPx})^H \tag{20}$$

$$\mathcal{D}_{\mathbf{y}} p(\mathbf{y}) = \mathbb{E}\{\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\} = -\mathbb{E}\left\{p(\mathbf{y}|\mathbf{x})(\mathbf{y}-\mathbf{HPx})^H\right\}.$$

The gradient of probability density function $\nabla_{\mathbf{P}} p(\mathbf{y})$ can then be written as

$$\begin{aligned}\nabla_{\mathbf{P}} p(\mathbf{y}) &= \mathbb{E}\left[\nabla_{\mathbf{P}} p(\mathbf{y}|\mathbf{x})\right] \\ &= \mathbb{E}\left[p(\mathbf{y}|\mathbf{x})\mathbf{H}^H(\mathbf{y}-\mathbf{HPx})\mathbf{x}^H\right] \\ &= -\mathbf{H}^H\mathbb{E}\left\{[\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})]^H\mathbf{x}^H\right\}\end{aligned} \tag{21}$$

and the gradient of conditional expectation $\nabla_{\mathbf{P}}\mathbb{E}\{x_i|\mathbf{y}\}$ becomes

$$\begin{aligned}\nabla_{\mathbf{P}}\mathbb{E}\{x_i|\mathbf{y}\} &= \nabla_{\mathbf{P}}\mathbb{E}\left\{x_i\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right\} \\ &= -\mathbb{E}\left\{x_i\frac{\mathbf{H}^H[\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})]^H\mathbf{x}^H}{p(\mathbf{y})}\right\} \\ &\quad + \mathbb{E}\left\{x_i\frac{p(\mathbf{y}|\mathbf{x})\mathbf{H}^H\mathbb{E}\left\{[\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})]^H\mathbf{x}^H\right\}}{[p(\mathbf{y})]^2}\right\} \\ &= \frac{1}{p(\mathbf{y})}\mathbf{H}^H\Big[-\mathbb{E}\left\{x_i[\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})]^H\mathbf{x}^H\right\} \\ &\quad +\mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\left\{[\mathcal{D}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})]^H\mathbf{x}^H\right\}\Big].\end{aligned} \tag{22}$$

The Jacobian $\mathcal{D}_{\mathbf{y}}\mathbb{E}\{\mathbf{x}|\mathbf{y}\}$ is derived as follows:

$$\begin{aligned}
\mathcal{D}_{\mathbf{y}}\mathbb{E}\{\mathbf{x}|\mathbf{y}\} &= \mathcal{D}_{\mathbf{y}}\mathbb{E}\left\{\mathbf{x}\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right\} \\
&= \mathbb{E}\left\{\mathbf{x}\frac{\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})\cdot p(\mathbf{y}) - p(\mathbf{y}|\mathbf{x})\cdot \mathcal{D}_{\mathbf{y}}p(\mathbf{y})}{[p(\mathbf{y})]^2}\right\} \\
&= \mathbb{E}\left\{\mathbf{x}\frac{-p(\mathbf{y}|\mathbf{x})\cdot(\mathbf{y}-\mathbf{HPx})^H}{p(\mathbf{y})}\right\} \\
&\quad +\mathbb{E}\left\{\mathbf{x}\frac{p(\mathbf{y}|\mathbf{x})\cdot(\mathbf{y}-\mathbf{HP}\mathbb{E}\{\mathbf{x}|\mathbf{y}\})^H}{p(\mathbf{y})}\right\} \\
&= \mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\mathbf{P}^H\mathbf{H}^H\right\} \\
&\quad -\mathbb{E}\left\{\mathbf{x}\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\mathbb{E}\{\mathbf{x}|\mathbf{y}\}^H\mathbf{P}^H\mathbf{H}^H\right\} \\
&= \left[\mathbb{E}\left\{\mathbf{x}\mathbf{x}^H|\mathbf{y}\right\} - \mathbb{E}\{\mathbf{x}|\mathbf{y}\}\mathbb{E}\{\mathbf{x}|\mathbf{y}\}^H\right]\mathbf{P}^H\mathbf{H}^H \\
&= \mathbf{\Phi}(\mathbf{y})\mathbf{P}^H\mathbf{H}^H. \qquad (23)
\end{aligned}$$

Following the similar steps in (23), the Jacobian can also be obtained:

$$\mathcal{D}_{\mathbf{y}^*}\mathbb{E}\{\mathbf{x}|\mathbf{y}\} = \mathbf{\Psi}(\mathbf{y})\mathbf{P}^H\mathbf{H}^H. \qquad (24)$$

Substituting (21) and (22) into (19) yields

$$\begin{aligned}
&\frac{\partial[\mathbf{E}]_{ij}}{\partial[\mathbf{P}]^*_{kl}} \\
&= -\int \mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}d\mathbf{y} \\
&\quad + \int \mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{x_i[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}\mathbb{E}\{x_j^*|\mathbf{y}\}d\mathbf{y} \\
&\quad + \int \mathbb{E}\{x_i|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{x_j^*[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}d\mathbf{y} \qquad (25)
\end{aligned}$$

where $\mathbf{e}_k$ and $\mathbf{e}_l$ are, respectively, the k-th and l-th columns of an identity matrix with appropriate dimensions.

The first term in (25) can be rewritten as

$$\begin{aligned}
&-\int \mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}d\mathbf{y} \\
&\stackrel{(a)}{=} -\int \mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\frac{\partial p(\mathbf{y})\mathbb{E}\{x_l^*|\mathbf{y}\}}{\partial\mathbf{vec}(\mathbf{y}^*)}d\mathbf{y} \\
&\stackrel{(b)}{=} \int p(\mathbf{y})\mathbb{E}\{x_l^*|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\frac{\partial\mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}}{\partial\mathbf{vec}(\mathbf{y}^*)}d\mathbf{y} \\
&\stackrel{(c)}{=} \int p(\mathbf{y})\mathbb{E}\{x_l^*|\mathbf{y}\}\mathbb{E}\{x_j^*|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Psi}(\mathbf{y})\mathbf{e}_i d\mathbf{y} \\
&\quad + \int p(\mathbf{y})\mathbb{E}\{x_l^*|\mathbf{y}\}\mathbb{E}\{x_i|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Phi}(\mathbf{y})\mathbf{e}_j d\mathbf{y} \qquad (26)
\end{aligned}$$

where (a) follows from the definition of Jacobian matrix (16) and the fact $\mathbb{E}\{x_l^*|\mathbf{y}\} = \mathbb{E}\left\{x_l^* p(\mathbf{y}|\mathbf{x})/p(\mathbf{y})\right\}$; (b) is the result of integration by parts; (c) is from the result of (23) and (24).

Similarly, the second and third terms in (25) can be re-expressed, respectively, as

$$\begin{aligned}
&\int \mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{x_i[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}\mathbb{E}\{x_j^*|\mathbf{y}\}d\mathbf{y} \\
&\quad = -\int \mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Phi}(\mathbf{y})\mathbf{e}_j\mathbb{E}\{x_i x_l^*|\mathbf{y}\}p(\mathbf{y})d\mathbf{y} \qquad (27)
\end{aligned}$$

and

$$\begin{aligned}
&\int \mathbb{E}\{x_i|\mathbf{y}\}\mathbf{e}_k^T\mathbf{H}^H\mathbb{E}\left\{x_j^*[\mathcal{D}_{\mathbf{y}}p(\mathbf{y}|\mathbf{x})]^H x_l^*\right\}d\mathbf{y} \\
&\quad = -\int \mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Psi}(\mathbf{y})\mathbf{e}_j\mathbb{E}\{x_j^* x_l^*|\mathbf{y}\}p(\mathbf{y})d\mathbf{y}. \qquad (28)
\end{aligned}$$

Substituting (26), (28), and (27) into (25) yields

$$\begin{aligned}
\frac{\partial[\mathbf{E}]_{ij}}{\partial[\mathbf{P}]^*_{kl}} &= -\int \mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Phi}(\mathbf{y})\mathbf{e}_j p(\mathbf{y})\big[\mathbb{E}\{x_i x_l^*|\mathbf{y}\} \\
&\qquad -\mathbb{E}\{x_i|\mathbf{y}\}\mathbb{E}\{x_l^*|\mathbf{y}\}\big]d\mathbf{y} \\
&\quad -\int \mathbf{e}_k^T\mathbf{H}^H\mathbf{HP\Psi}(\mathbf{y})\mathbf{e}_j p(\mathbf{y})\big[\mathbb{E}\{x_j^* x_l^*|\mathbf{y}\} \\
&\qquad -\mathbb{E}\{x_j^*|\mathbf{y}\}\mathbb{E}\{x_l^*|\mathbf{y}\}\big]d\mathbf{y} \\
&= -\mathbb{E}_{\mathbf{y}}\left\{\mathbf{e}_i^T\mathbf{\Phi}(\mathbf{y})\mathbf{e}_l\mathbf{e}_j^T\mathbf{\Phi}^T(\mathbf{y})\mathbf{G}^T\mathbf{H}^T\mathbf{H}^*\mathbf{e}_k\right\} \\
&\qquad -\mathbb{E}_{\mathbf{y}}\left\{\mathbf{e}_j^T\mathbf{\Psi}^*(\mathbf{y})\mathbf{e}_l\mathbf{e}_i^T\mathbf{\Psi}(\mathbf{y})\mathbf{G}^T\mathbf{H}^T\mathbf{H}^*\mathbf{e}_k\right\} \\
&= -\mathbb{E}_{\mathbf{y}}\Big\{\Big[\mathbf{K}\big(\mathbf{\Phi}(\mathbf{y})\otimes \\
&\qquad (\mathbf{\Phi}^T(\mathbf{y})\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*)\big)\Big]_{i+(j-1)2L,k+(l-1)2L}\Big\} \\
&\quad -\mathbb{E}_{\mathbf{y}}\Big\{\Big[\mathbf{\Psi}^*(\mathbf{y})\otimes \\
&\qquad \Big(\mathbf{\Psi}(\mathbf{y})\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\Big)\Big]_{i+(j-1)2L,k+(l-1)2L}\Big\}
\end{aligned}$$

where $\mathbf{K}\in\mathbb{R}^{4L^2\times 4L^2}$ is a unique permutation matrix that satisfies [34]:

$$\mathbf{vec}\left(\mathbf{\Phi}^T(\mathbf{y})\right) = \mathbf{K}\cdot\mathbf{vec}\left(\mathbf{\Phi}(\mathbf{y})\right).$$

From $\partial[\mathbf{E}]_{ij}/\partial[\mathbf{P}]^*_{kl} = [\mathcal{D}_{\mathbf{P}^*}\mathbf{E}]_{i+(j-1)2L,k+(l-1)2L}$, we have

$$\begin{aligned}
\mathcal{D}_{\mathbf{P}^*}\mathbf{E} &= -\mathbf{K}\mathbb{E}_{\mathbf{y}}\left\{\mathbf{\Phi}(\mathbf{y})\otimes\left[\mathbf{\Phi}^T(\mathbf{y})\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right]\right\} \\
&\quad -\mathbb{E}_{\mathbf{y}}\left\{\mathbf{\Psi}^*(\mathbf{y})\otimes\left[\mathbf{\Psi}(\mathbf{y})\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right]\right\} \\
&= -\mathbf{K}\mathbb{E}_{\mathbf{y}}\left\{\mathbf{\Phi}(\mathbf{y})\otimes\mathbf{\Phi}^*(\mathbf{y})\right\}\left[\mathbf{I}\otimes\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right] \\
&\quad -\mathbb{E}_{\mathbf{y}}\left\{\mathbf{\Psi}^*(\mathbf{y})\otimes\mathbf{\Psi}(\mathbf{y})\right\}\left[\mathbf{I}\otimes\mathbf{P}^T\mathbf{H}^T\mathbf{H}^*\right]. \qquad (29)
\end{aligned}$$

Applying of the property of permutation matrix $\mathbf{K}$ [34], $\mathbf{K}\left(\mathbf{A}\otimes\mathbf{B}\right) = \left(\mathbf{B}\otimes\mathbf{A}\right)\mathbf{K}$, the proof is complete. □

3) Proof of Proposition 1: Let us denote the remaining part of the precoder, $\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V}$, as $\mathbf{G}$, which allows the power allocation vector $\boldsymbol{\lambda}$ to be rewritten as

$$\boldsymbol{\lambda} = \mathbf{R}\cdot\mathbf{vec}(\mathbf{G}\mathbf{G}^H).$$

According to (17), the Hessian of the mutual information can be obtained from

$$\mathcal{H}_{\boldsymbol{\lambda}}\mathcal{I}\left(\boldsymbol{\lambda}\right) = \mathcal{D}_{\boldsymbol{\lambda}}\left[\mathcal{D}_{\boldsymbol{\lambda}}\mathcal{I}\left(\boldsymbol{\lambda}\right)\right]. \qquad (30)$$

Now $\mathcal{D}_{\boldsymbol{\lambda}}\mathcal{I}\left(\boldsymbol{\lambda}\right)$ can be calculated as

$$\begin{aligned}
\mathcal{D}_{\boldsymbol{\lambda}}\mathcal{I}(\boldsymbol{\lambda}) &= \mathcal{D}_{\mathbf{G}\mathbf{G}^H}\mathcal{I}(\boldsymbol{\lambda})\cdot\mathbf{R}^T \\
&\stackrel{(a)}{=} \mathbf{vec}^T\left(\mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{VEV}^H\right)\mathbf{R}^T \\
&\stackrel{(b)}{=} \mathbf{vec}^T\left(\mathbf{E}\right)\left(\mathbf{V}^H\otimes\mathbf{V}^T\mathbf{Diag}^2(\boldsymbol{\sigma})\right)\mathbf{R}^T
\end{aligned}$$

where (a) follows from the results in [15, Theorem 2], and (b) follows from the property [34]:

$$\mathbf{vec}\left(\mathbf{ATB}\right) = \left(\mathbf{B}^T\otimes\mathbf{A}\right)\mathbf{vec}(\mathbf{T})$$

in which $\mathbf{A}$, $\mathbf{T}$, and $\mathbf{B}$ are matrices with appropriate dimensions. The result in (30) can then be written as

$$\begin{aligned}\mathcal{H}_{\boldsymbol{\lambda}}\mathcal{I}(\boldsymbol{\lambda}) &= \mathcal{D}_{\boldsymbol{\lambda}}\left[\mathbf{R}\left(\mathbf{V}^* \otimes \mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{V}\right)\mathbf{vec}\left(\mathbf{E}\right)\right] \\ &= \mathbf{R}\left(\mathbf{V}^* \otimes \mathbf{Diag}^2(\boldsymbol{\sigma})\mathbf{V}\right)\mathcal{D}_{\boldsymbol{\lambda}}\mathbf{E}. \end{aligned} \tag{31}$$

From the chain rule of the Jacobian [33], it yields

$$\begin{aligned}\mathcal{D}_{\mathbf{G}^*}\mathbf{E} &= \mathcal{D}_{\boldsymbol{\lambda}}\mathbf{E}\cdot\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda} + \mathcal{D}_{\boldsymbol{\lambda}^*}\mathbf{E}\cdot\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda}^* \\ &= 2\mathcal{D}_{\boldsymbol{\lambda}}\mathbf{E}\cdot\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda}\end{aligned} \tag{32}$$

where $\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda}$ can be derived from the definition of Jacobian (16):

$$\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda} = \mathcal{D}_{\mathbf{G}^*}\left[\mathbf{R}\cdot\mathbf{vec}\left(\mathbf{G}\mathbf{G}^H\right)\right] = \mathbf{R}\left(\mathbf{I}\otimes\mathbf{G}\right)\mathbf{K}.$$

Because $\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda}$ has full column rank, it is possible to invert the transformation in (32) to obtain

$$\begin{aligned}\mathcal{D}_{\boldsymbol{\lambda}}\mathbf{E} &= \frac{1}{2}\left(\mathcal{D}_{\mathbf{G}^*}\mathbf{E}\right)\left(\mathcal{D}_{\mathbf{G}^*}\boldsymbol{\lambda}\right)^+ \\ &= \frac{1}{2}\left(\mathcal{D}_{\mathbf{G}^*}\mathbf{E}\right)\cdot\left(\mathbf{G}^H\otimes\mathbf{I}\right)\mathbf{K}^T\mathbf{R}^T\mathbf{Diag}^{-1}\left(\boldsymbol{\lambda}\right).\end{aligned} \tag{33}$$

Plugging (18) and (33) into (31) yields the Hessian result provided in (11), which can be readily identified as negative semi-definite. □

Acknowledgment

The authors would like to thank Prof. Chengshan Xiao for invaluable discussions on this topic. They would also like to thank the editor and reviewers for helpful comments and suggestions that greatly improved the quality of the paper.

References

[1] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[2] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524–3536, Dec. 2006.

[3] Y. Ding, J. Zhang, and K. Wong, "The amplify-and-forward half-duplex cooperative system: pairwise error probability and precoder design," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 605–617, Feb. 2007.

[4] ——, "Optimal precoder for amplify-and-forward half-duplex relay system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 2890–2895, Aug. 2008.

[5] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.

[6] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming: from receive to transmit and network designs," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, Mar. 2010.

[7] R. Nabar, H. Bolcskei, and F. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[8] R. Mo and Y. Chew, "Precoder design for non-regenerative MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5041–5049, Oct. 2009.

[9] W. Zeng, C. Xiao, Y. Wang, and J. Lu, "Opportunistic cooperation for multi-antenna multi-relay networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3189–3199, Oct. 2010.

[10] A. Lozano, A. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[11] G. Caire and K. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[12] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. 2008 IEEE Globecom*.

[13] C. Wen, K. Wong, and C. Ng, "On the asymptotic properties of amplify-and-forward MIMO relay channels," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 590–602, Feb. 2009.

[14] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for relay networks with finite-alphabet constraints," in *2011 IEEE Int. Conf. Commun.*

[15] 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.

[16] F. Pérez-Cruz, M. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[17] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.

[18] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in *Proc. 2009 Int. Symp. on Wireless Commun. Syst.*, pp. 26–30.

[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, July 2011.

[20] D. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory*. Now Publishers Inc., 2006.

[21] M. Payaró and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, pp. 1085–1089.

[22] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[23] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," retrieved Apr. 2011. [Online]. Available: http://cvxr.com/cvx

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[25] P. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[26] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

[27] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[28] F. Murnaghan, *Lectures on Applied Mathematics: The Unitary and Rotation Groups, Vol. 3*. Spartan Books, 1962.

[29] B. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[30] R. Koetter, A. Singer, and M. Tuchler, "Turbo equalization," *IEEE Signal Process Mag.*, vol. 21, no. 1, pp. 67–80, Jan. 2004.

[31] B. Lu, X. Wang, and K. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: performance analysis and receiver design," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 74–88, Jan. 2002.

[32] M. Valenti, "Iterative solutions coded modulation library (ISCML)." Available: http://www.iterativesolutions.com/Matlab.htm

[33] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, 2011.

[34] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edition. John Wiley & Sons, 2007.

Weiliang Zeng received the B.S. degree in electronic engineering from the University of Electronic Sciences and Technology of China (UESTC), Chengdu, China, in July 2007 with the highest honors. Since September 2007, he has been a Ph.D. candidate at the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing, China. During his doctoral studies, he has conducted cooperative research at the Department of Electronic Engineering, Missouri University of Science and Technology, USA. His research interests include communications and information theory with special emphasis on wireless communications and signal processing. He is the recipient of first-class scholarship for graduate students at Tsinghua University.

Yahong Rosa Zheng received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, and the M.S. degree from Tsinghua University, Beijing, China, in 1989, both in electrical engineering. She received the Ph.D. degree from the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, in 2002. She was an NSERC Postdoctoral Fellow from Jan. 2003 to April 2005 at University of Missouri-Columbia. Since fall 2005, she has been an assistant then associate professor with the Department of Electrical and Computer Engineering at the Missouri University of Science and Technology. Her research interests include array signal processing, wireless communications, and wireless sensor networks. She has been a senior member of the IEEE since 2007. She is the recipient of an NSF CAREER award in 2009. She served as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS for 2006-2008. She is currently associate editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Mingxi Wang received his B.S degree in telecommunication engineering from Beihang University (formerly: Beijing University of Aeronautics and Astronautics), China, in 2006, and M.S degree in signal and information processing from the same university in 2009. He is currently a Ph.D. candidate of department of electrical and computer engineering at Missouri University of Science and Technology (formerly: University of Missouri-Rolla), USA. His research interests include wireless communications, signal processing, MIMO, OFDM, linear precoding and iterative receiver.

Jianhua Lu received B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and his Ph.D. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon. Since 1989 he has been with the Department of Electronic Engineering, Tsinghua University, where he is currently a professor. His research interests include broad- band wireless communication, multimedia signal processing, satellite communication, and wireless networking. He has published more than 180 technical papers in international journals and conference proceedings. He has been an active member of professional societies. He was one of the recipients of best paper awards at the International Conference on Communications, Circuits and Systems 2002 and ChinaCom 2006, and was awarded the National Distinguished Young Scholar Fund by the NSF committee of China in 2006. He has served in numerous IEEE conferences as a member of Technical Program Committees and served as Lead Chair of the General Symposium of IEEE ICC 2008, as well as a Program Committee Co-Chair of the 9th IEEE International Conference on Cognitive Informatics. He is a Senior Member of the IEEE Signal Processing Society.