

# Linear Precoding for Relay Networks with Finite-Alphabet Constraints

Weiliang Zeng\*, Chengshan Xiao<sup>†</sup>, Mingxi Wang<sup>†</sup>, and Jianhua Lu\*

\*State Key Laboratory on Microwave and Digital Communications  
Tsinghua National Laboratory for Information Science and Technology  
Dept. of Electronic Engineering, Tsinghua University, Beijing, China

<sup>†</sup>Dept. of Electrical & Computer Engineering, Missouri University of Science and Technology, MO, USA  
Email: {zengwl, lujh}@wmc.ee.tsinghua.edu.cn, {xiaoc, mw9zd}@mst.edu

**Abstract**—In this paper, we investigate the optimal precoding scheme for relay networks with finite-alphabet constraints. We show that the previous work utilizing various design criteria to maximize either the diversity order or the transmission rate with the Gaussian-input assumption may lead to significant loss for a practical system with finite constellation set constraint. A linear precoding scheme is proposed to maximize the mutual information for relay networks. We exploit the structure of the optimal precoding matrix and develop a unified two-step iterative algorithm utilizing the theory of convex optimization and optimization on the complex Stiefel manifold. Numerical examples show that this novel iterative algorithm achieves significant gains compared to its conventional counterpart.

## I. INTRODUCTION

Cooperative relaying is an emerging technology, which provides reliable high data rate transmission for wireless networks without the need of multiple antennas at each node. These benefits can be further exploited by utilizing judicious cooperative design (see [1]–[5] and the references therein).

Most of the existing designs optimize the performance of relay networks with Gaussian-input assumptions, for example, maximizing output signal-to-noise (SNR) [1], [2], minimizing mean square error (MSE) [1], [3] and maximizing channel capacity [1], [4], [5]. Despite the information theoretic optimality of Gaussian inputs, they can never be realized in practice. Rather, the inputs must be drawn from a finite constellation set, such as pulse amplitude modulation (PAM), phase shift keying (PSK) modulation and quadrature amplitude modulation (QAM), in a practical communication system. These kinds of discrete constellations depart significantly from the Gaussian idealization [6], [7]. Therefore, there exhibits a big performance gap between the scheme designed with the Gaussian-input assumption and the scheme designed from the standing point of finite-alphabet constraint [8].

In this paper, we consider the two-hop relay networks with finite-input constraint and utilize linear precoder to improve the maximal possible transmission rate of networks. We exploit the optimal structure of the precoding matrix under

finite-alphabet constraint and develop a unified framework to solve this nonconvex optimization problem. We prove that the left singular matrix of the precoder coincides with the right singular matrix of the effective channel matrix; the mutual information is a concave function of the power allocation vector of the precoder; the optimization of the right singular matrix with unitary constraint can be formulated as an unconstrained one on the complex Stiefel manifold. Once these important results are provided, the optimal precoder is solved with a complete two-step iterative algorithm utilizing the theory of convex optimization and optimization on the manifold. We show that this novel iterative algorithm achieves significant gains compared to its conventional counterpart.

*Notation:* Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. The superscripts  $(\cdot)^T$  and  $(\cdot)^H$  stand for transpose and Hermitian operations, respectively. The subscripts  $c_i$  and  $[\mathbf{A}]_{ij}$  denote the  $i$ -th element of vector  $\mathbf{c}$  and the  $(i, j)$ -th element of matrix  $\mathbf{A}$ , respectively. The operator  $\text{Diag}(\mathbf{a})$  represents a diagonal matrix whose nonzero elements are given by the elements of vector  $\mathbf{a}$ . Furthermore,  $\text{vec}(\mathbf{A})$  represents the vector obtained by stacking the columns of  $\mathbf{A}$ ;  $\mathbf{I}$  and  $\mathbf{0}$  represents an identity matrix and a zero matrix of appropriate dimensions, respectively;  $\text{Tr}(\mathbf{A})$  denotes the trace operation. Besides, all logarithms are base 2.

## II. SYSTEM MODEL

Consider a relay network with one transmit-and-receive pair, where the source node attempts to communicate to the destination node with the assistance of  $k$  relays ( $r_1, r_2, \dots, r_k$ ). All nodes are equipped with a single antenna and operated in half-duplex mode. We consider a flat fading cooperative transmission system, in which the channel gain from the source to the destination is denoted by  $h_0$ , whereas 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.

We focus on the amplify-and-forward protocols combined with single relay selection [5]. The signal transmission is carried out by blocks with block length  $2L$ ,  $L \geq 1$ . For the selected relay node, there is a data receiving period of length  $L$  before a data transmitting period of length  $L$ .

The original data at the source node is denoted by

$$\mathbf{x} = [\mathbf{x}_a^T \quad \mathbf{x}_b^T]^T,$$

This work was supported in part by the National Basic Research Program of China (2007CB310601) and the U.S. National Science Foundation under Grant CCF-0915846.

This work has been carried out while W. Zeng is a visiting scholar at Missouri University of Science and Technology.

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = \log M - \frac{1}{2LM^{2L}} \sum_{m=1}^{M^{2L}} \mathbb{E}_{\mathbf{n}} \left\{ \log \sum_{k=1}^{M^{2L}} \exp \left[ -\|\mathbf{H}\mathbf{P}(\mathbf{x}_m - \mathbf{x}_k) + \mathbf{n}\|^2 + \|\mathbf{n}\|^2 \right] \right\}. \quad (3)$$

$$\mathbf{E} = \mathbf{I} - \frac{1}{(\pi M)^{2L}} \int_{\mathbf{y}} \frac{\left[ \sum_{l=1}^{M^{2L}} \mathbf{x}_l \exp(-\|\mathbf{y} - \mathbf{H}\mathbf{P}\mathbf{x}_l\|^2) \right] \left[ \sum_{k=1}^{M^{2L}} \mathbf{x}_k^H \exp(-\|\mathbf{y} - \mathbf{H}\mathbf{P}\mathbf{x}_k\|^2) \right]}{\sum_{m=1}^{M^{2L}} \exp(-\|\mathbf{y} - \mathbf{H}\mathbf{P}\mathbf{x}_m\|^2)} d\mathbf{y}. \quad (9)$$

where  $\mathbf{x}_a = [x_1, \dots, x_L]^T$ ,  $\mathbf{x}_b = [x_{L+1}, \dots, x_{2L}]^T$  with  $x_l$  being the data symbol at the  $l$ -th time slot,  $l = 1, \dots, 2L$ . We assume that the original information signals are equally probable from discrete signaling constellations such as PSK, PAM, or QAM with unit covariance matrix, *i.e.*,  $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}$ .

The original data is processed by a precoding matrix before being transmitted from the source node. The precoded data  $\mathbf{s} = [\mathbf{s}_a^T \ \mathbf{s}_b^T]^T$  is given by  $\mathbf{s} = \mathbf{P}\mathbf{x}$ , where  $\mathbf{P}$  is a *generalized* complex precoding matrix.

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

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

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

Assuming the  $i$ -th relay node is selected at the second  $L$ -time slots, it scales the received signal by a factor  $b$  (so that its average transmit power is  $P_r$ ) and forwards it to the receiver. We assume only the second-order statistics of  $h_i$  is known at the  $i$ -th relay node, then  $b$  can be chosen as  $b = \sqrt{LP_r/\text{Tr}(\mathbb{E}[\mathbf{y}_i\mathbf{y}_i^H])}$ . At the second  $L$ -time slots, the source node sends the signal  $\sqrt{P_s}\mathbf{s}_b$ . Therefore, the destination node receives a superposition of the relay transmission and the source transmission according to

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

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

For convenience in the presentation, we normalize  $\mathbf{y}_b$  by  $w = 1/\sqrt{N_d}$  and denote the received signal vector as  $\mathbf{y} = [\mathbf{y}_a^T \ w\mathbf{y}_b^T]^T$ . Thus, the effective input-output relation for the two-hop transmission with precoding is summarized as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} = \mathbf{H}\mathbf{P}\mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{x}$  is the original transmitted signal vector;  $\mathbf{n} = [\mathbf{n}_a^T \ w\mathbf{n}_e^T]^T$  is i.i.d. complex Gaussian channel noise vector with zero mean and unit variance, *i.e.*,  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ ;  $\mathbf{H}$  is the effective channel matrix of the two-hop relay channel

$$\mathbf{H} = \sqrt{P_s} \begin{bmatrix} h_0\mathbf{I} & \mathbf{0} \\ wbh_i g_i\mathbf{I} & wh_0\mathbf{I} \end{bmatrix}. \quad (2)$$

Our precoding scheme is thus the design of matrix  $\mathbf{P}$  to maximize the mutual information with finite-alphabet constraints. Note that for the proposed algorithm to be effectively

implemented in practice, the destination estimates effective channel matrices of relay networks through pilot assisted channel estimation. Then the destination node selects one relay for cooperation and provides the corresponding effective channel to the source node via a feedback channel. Considering the special structure of the channel matrix (2), the amount of the feedback can be very small. After signal feedback, the source node utilizes the proposed precoding algorithm to optimize the network performance.

### III. MUTUAL INFORMATION FOR RELAY NETWORKS

We consider the conventional equiprobable discrete constellations such as  $M$ -ary PSK, PAM, or QAM, where  $M$  is the number of points in the signal constellation. The mutual information between  $\mathbf{x}$  and  $\mathbf{y}$ , 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 (3), where  $\|\cdot\|$  denotes Euclidean norm of a vector; both  $\mathbf{x}_m$  and  $\mathbf{x}_k$  contain  $2L$  symbols, taken independently from the  $M$ -ary signal constellation [7], [8].

*Proposition 1:* Let  $\mathbf{U}$  be a unitary matrix, and the following relationships hold:

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

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

*Proof:* See proof in [9].  $\square$

Proposition 1 implies 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 is unchanged when either transmitted signal  $\mathbf{x}$  or received signal  $\mathbf{y}$  is rotated by a unitary matrix. The case of finite inputs does not follow the same rule. Therefore, it provides us a new opportunity to improve the system performance.

The optimization of the linear precoding matrix  $\mathbf{P}$  is carried out over all  $2L \times 2L$  complex precoding matrices under transmit power constraint, which can be cast as a constrained nonlinear optimization problem

$$\begin{aligned} &\text{maximize} \quad \mathcal{I}(\mathbf{x}; \mathbf{y}) \\ &\text{subject to} \quad \text{Tr}\{\mathbb{E}[\mathbf{s}\mathbf{s}^H]\} = \text{Tr}(\mathbf{P}\mathbf{P}^H) \leq 2L. \end{aligned} \quad (6)$$

*Proposition 2 (Necessary Condition):* The optimal precoding matrix  $\mathbf{P}^*$  for the optimization problem (6) satisfies the following condition [7], [8]

$$\mu\mathbf{P}^* = \mathbf{H}^H\mathbf{H}\mathbf{P}^*\mathbf{E}, \quad (7)$$

where  $\mu$  is chosen to satisfy the power constraint, and  $\mathbf{E}$  is the minimum mean square error (MMSE) matrix given by (9).

*Proof:* See proof in [9].  $\square$

It is important to note that Proposition 2 gives a necessary condition satisfied by any critical points, since the optimization problem (6) is nonconvex. It is possible to develop an algorithm based on the gradient of the Lagrangian as proposed in [8], [10]. However, this kind of algorithms can be stuck at a local maximum, which is influenced by the starting point and may be far away from the optimal solution. This fact will be shown via an example in the sequel.

#### IV. PRECODER DESIGN TO MAXIMIZE THE MUTUAL INFORMATION

##### A. Optimal Left Singular Vector

We start by characterizing the dependence of mutual information  $\mathcal{I}(\mathbf{x}; \mathbf{y})$  on precoder  $\mathbf{P}$ . Consider the singular value decomposition (SVD) of the  $2L \times 2L$  channel matrix  $\mathbf{H} = \mathbf{U}_H \text{Diag}(\boldsymbol{\sigma}) \mathbf{V}_H^H$ , where  $\mathbf{U}_H$  and  $\mathbf{V}_H$  are unitary matrices, and the vector  $\boldsymbol{\sigma}$  contains nonnegative entries in decreasing order. Note that the equivalent channel matrix  $\mathbf{H}$  defined in (2) is full rank for any nonzero channel gain  $h_0$ . We also consider the SVD of the precoding matrix  $\mathbf{P} = \mathbf{U}_P \text{Diag}(\sqrt{\boldsymbol{\lambda}}) \mathbf{V}_P^H$  and define  $\mathbf{U} = \mathbf{U}_P$  and  $\mathbf{V} = \mathbf{V}_P^H$ , where  $\mathbf{U}$  and  $\mathbf{V}$  are named as the left and right singular vectors, respectively; the vector  $\boldsymbol{\lambda}$  is nonnegative constrained by transmit power.

*Proposition 3:* The mutual information depends on the precoding matrix  $\mathbf{P}$  only through  $\mathbf{P}^H \mathbf{H}^H \mathbf{H} \mathbf{P}$ . For a given  $\mathbf{P}^H \mathbf{H}^H \mathbf{H} \mathbf{P}$ , we can always choose the precoding matrix of the form  $\mathbf{P} = \mathbf{V}_H \text{Diag}(\sqrt{\boldsymbol{\lambda}}) \mathbf{V}$  in order to minimize the transmit power  $\text{Tr}(\mathbf{P} \mathbf{P}^H)$ , i.e., the left singular vector of  $\mathbf{P}$  coincides with the right singular vector of  $\mathbf{H}$ .

*Proof:* See proof in [9].  $\square$

From the results in Proposition 1 and 3, it is possible to simplify the channel matrix (1) to

$$\mathbf{y} = \text{Diag}(\boldsymbol{\sigma}) \text{Diag}(\sqrt{\boldsymbol{\lambda}}) \mathbf{V} \mathbf{x} + \mathbf{n}. \quad (10)$$

Now our discussion will be based on this simplified channel model (10). The optimization variables are power allocation vector  $\boldsymbol{\lambda}$  and right singular vector  $\mathbf{V}$ , which are the focuses of the next two subsections. In the sequel, we will use  $\mathcal{I}(\boldsymbol{\lambda})$  and  $\mathcal{I}(\mathbf{V})$  to emphasize the dependence of mutual information on variables  $\boldsymbol{\lambda}$  and  $\mathbf{V}$ , respectively.

##### B. Optimal Power Allocation

Given a right singular vector of the precoder, we consider the following optimization problem over the power allocation vector  $\boldsymbol{\lambda}$

$$\begin{aligned} & \text{maximize} && \mathcal{I}(\boldsymbol{\lambda}) \\ & \text{subject to} && \mathbf{1}^T \boldsymbol{\lambda} \leq 2L \\ & && \boldsymbol{\lambda} \succeq \mathbf{0}, \end{aligned} \quad (11)$$

where  $\mathbf{1}$  denotes a column vector with all entries one.

*Proposition 4:* The mutual information is a concave function of the squared singular values of the precoder,  $\boldsymbol{\lambda}$ , i.e., the mutual information Hessian with respect to the power allocation vector satisfies  $\mathcal{H}_{\boldsymbol{\lambda}} \mathcal{I}(\boldsymbol{\lambda}) \preceq \mathbf{0}$ . Moreover, the Jacobian of the mutual information with respect to the power allocation vector  $\boldsymbol{\lambda}$  is given by

$$\nabla_{\boldsymbol{\lambda}} \mathcal{I}(\boldsymbol{\lambda}) = \mathbf{R} \cdot \text{vec} \left( \text{Diag}^2(\boldsymbol{\sigma}) \mathbf{V} \mathbf{E} \mathbf{V}^H \right), \quad (12)$$

where  $\mathbf{R} \in \mathbb{R}^{2L \times 4L^2}$  is a reduction matrix given by

$$[\mathbf{R}]_{i, 2L(j-1)+k} = \delta_{ijk}, \quad i, j, k \in [1, 2L]. \quad (13)$$

*Proof:* See proof in [9].  $\square$

The concavity result in Proposition 4 extends the Hessian and concavity results in [11, Theorem 5] from real-valued signal model to a generalized complex-valued case. It ensures to find the global optimal power allocation vector given a right singular vector  $\mathbf{V}$ , and the gradient result in (12) provides the possibility to develop a steepest descent type algorithm to achieve the global optimum [12].

We first rewrite the problem (11) using the *barrier* method:

$$\text{minimize} \quad f(\boldsymbol{\lambda}) = -\mathcal{I}(\boldsymbol{\lambda}) + \sum_{i=1}^{2L} \phi(-\lambda_i) + \phi(\mathbf{1}^T \boldsymbol{\lambda} - 2L), \quad (14)$$

where  $\phi(u)$  is the *logarithmic barrier* function, which approximates an indicator illustrating whether constraints are violated

$$\phi(u) = \begin{cases} -(1/t) \log(-u), & u < 0 \\ +\infty, & u \geq 0 \end{cases} \quad (15)$$

with the parameter  $t > 0$  setting the accuracy of the approximation. The gradient of objective function (14) is

$$\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda}) = -\mathbf{R} \cdot \text{vec} \left( \text{Diag}^2(\boldsymbol{\sigma}) \mathbf{V} \mathbf{E} \mathbf{V}^H \right) - \frac{1}{t} \left( \mathbf{q} - \frac{\mathbf{1}}{2L - \mathbf{1}^T \boldsymbol{\lambda}} \right), \quad (16)$$

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

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

Then it is necessary to decide a positive step size  $\gamma \in \mathbb{R}$  so that  $f(\boldsymbol{\lambda} + \gamma \Delta \boldsymbol{\lambda}) < f(\boldsymbol{\lambda})$ . The backtracking line search conditions [12] states that  $\gamma$  should be chosen to satisfy the inequalities

$$f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda} + \gamma \Delta \boldsymbol{\lambda}) \geq \frac{1}{2} \gamma \|\Delta \boldsymbol{\lambda}\|^2, \quad (17)$$

$$f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda} + 2\gamma \Delta \boldsymbol{\lambda}) < \gamma \|\Delta \boldsymbol{\lambda}\|^2. \quad (18)$$

The above ideas can be summarized as the following algorithm, which ensures to converge to the optimal power allocation vector because of the concavity.

*Algorithm 1:* Steepest Descent to Maximize the Mutual Information Over Power Allocation Vector

- 1) Given a feasible  $\boldsymbol{\lambda}$ ,  $t := t^{(0)} > 0$ ,  $\alpha > 1$ , tolerance  $\epsilon > 0$ .
- 2) Compute the gradient of  $f$  at  $\boldsymbol{\lambda}$ ,  $\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$ , as (16) and the descent direction  $\Delta \boldsymbol{\lambda} = -\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$ . Set the step size  $\gamma := 1$ .
- 3) Evaluate  $\|\Delta \boldsymbol{\lambda}\|^2$ . If it is sufficiently small, then go to Step 7.
- 4) If  $f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda} + 2\gamma \Delta \boldsymbol{\lambda}) \geq \gamma \|\Delta \boldsymbol{\lambda}\|^2$ , then set  $\gamma := 2\gamma$ , and repeat Step 4.
- 5) If  $f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda} + \gamma \Delta \boldsymbol{\lambda}) < \frac{1}{2} \gamma \|\Delta \boldsymbol{\lambda}\|^2$ , then set  $\gamma := \frac{1}{2} \gamma$ , and repeat Step 5.
- 6) Set  $\boldsymbol{\lambda} := \boldsymbol{\lambda} + \gamma \Delta \boldsymbol{\lambda}$ . Go to Step 2.
- 7) Stop if  $1/t < \epsilon$ , else  $t := \alpha t$ , and go to step 2.

### C. Optimization Over Right Singular Vector

This section considers an alternative optimization problem for maximizing the mutual information over the right singular vector  $\mathbf{V}$  for a given power allocation vector,

$$\begin{aligned} & \text{maximize} \quad \mathcal{I}(\mathbf{V}) \\ & \text{subject to} \quad \mathbf{V}^H \mathbf{V} = \mathbf{V} \mathbf{V}^H = \mathbf{I}. \end{aligned} \quad (19)$$

This unitary matrix constrained problem can be formulated as an unconstrained one in a constrained search space

$$\text{minimize} \quad g(\mathbf{V}) \quad (20)$$

where we define the function  $g(\mathbf{V})$  as  $-\mathcal{I}(\mathbf{V})$ , and with domain restricted to the feasible set:

$$\text{dom } g = \{\mathbf{V} \in \text{St}(n)\}, \quad (21)$$

in which the set  $\text{St}(n)$  is *complex Stiefel manifold* [13]

$$\text{St}(n) = \{\mathbf{V} \in \mathbb{C}^{n \times n} | \mathbf{V}^H \mathbf{V} = \mathbf{I}\}. \quad (22)$$

Associated with each point  $\mathbf{V} \in \text{St}(n)$  is a vector space called *tangent space*, which is formed by all the tangent vectors at the point  $\mathbf{V}$ .

*Proposition 5:* The gradient of the mutual information on the tangent space is

$$\begin{aligned} \nabla_{\mathbf{V}} g(\mathbf{V}) = & -\text{Diag}^2(\boldsymbol{\sigma}) \text{Diag}(\boldsymbol{\lambda}) \mathbf{V} \mathbf{E} \\ & + \mathbf{V} \mathbf{E} \mathbf{V}^H \text{Diag}^2(\boldsymbol{\sigma}) \text{Diag}(\boldsymbol{\lambda}) \mathbf{V}. \end{aligned} \quad (23)$$

*Proof:* See proof in [9].  $\square$

Utilizing the gradient on the tangent space as the descent direction has been suggested in [13], *i.e.*,  $\Delta \mathbf{V} = -\nabla_{\mathbf{V}} g(\mathbf{V})$ ; however, moving towards the direction on the tangent space may lost the unitary property. Therefore, it needs to be restored in each step via projection.

The projection of an arbitrary matrix  $\mathbf{W} \in \mathbb{C}^{n \times n}$  onto the Stiefel manifold  $\pi(\mathbf{W})$  is defined to be the point on the Stiefel manifold closest to  $\mathbf{W}$  in the Euclidean norm

$$\pi(\mathbf{W}) = \arg \min_{\mathbf{Q} \in \text{St}(n)} \|\mathbf{W} - \mathbf{Q}\|^2. \quad (24)$$

*Proposition 6 (Projection):* Let  $\mathbf{W} \in \mathbb{C}^{n \times n}$  be a full rank matrix. If the SVD of  $\mathbf{W}$  is  $\mathbf{W} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H$ , the projection is unique, which is given by  $\pi(\mathbf{W}) = \mathbf{U} \mathbf{W} \mathbf{V}^H$ .

*Proof:* See proof in [9].  $\square$

Combining the search direction and projection method with the line search conditions in (17) and (18), we are able to develop the optimization algorithm to maximize the mutual information over the right singular vector  $\mathbf{V}$ .

*Algorithm 2:* Steepest Descent to 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 gradient of  $g$  at  $\mathbf{V}$ ,  $\nabla_{\mathbf{V}} g(\mathbf{V})$ , as (23) and the descent direction  $\Delta \mathbf{V} = -\nabla_{\mathbf{V}} g(\mathbf{V})$ . Set the step size  $\gamma := 1$ .
- 3) Evaluate  $\|\Delta \mathbf{V}\|^2 = \text{Tr}\{(\Delta \mathbf{V})^H \Delta \mathbf{V}\}$ . If it is sufficiently small, then stop.
- 4) If  $g(\mathbf{V}) - g(\pi(\mathbf{V} + 2\gamma \Delta \mathbf{V})) \geq \gamma \|\Delta \mathbf{V}\|^2$ , then set  $\gamma := 2\gamma$ , and repeat Step 4.
- 5) If  $g(\mathbf{V}) - g(\pi(\mathbf{V} + \gamma \Delta \mathbf{V})) < \frac{1}{2} \gamma \|\Delta \mathbf{V}\|^2$ , then set  $\gamma := \frac{1}{2} \gamma$ , and repeat Step 5.
- 6) Set  $\mathbf{V} := \pi(\mathbf{V} + \gamma \Delta \mathbf{V})$ . Go to Step 2.

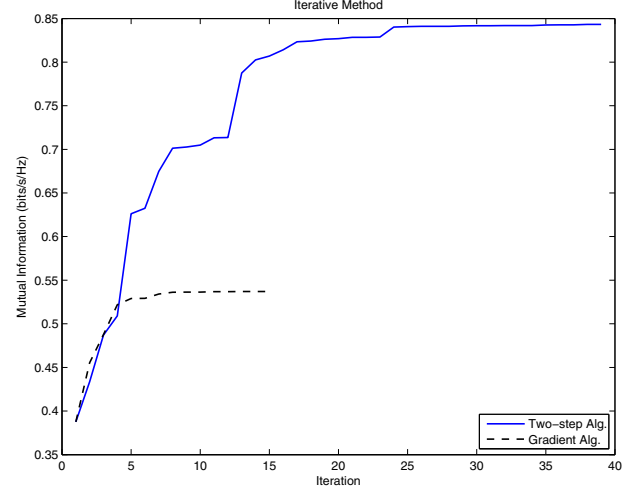


Fig. 1. Evolution of the mutual information as the linear precoder is iteratively optimized with the two-step algorithm and gradient algorithm.

### D. Two-Step Approach to Optimize Precoder

Now we are ready to develop a complete two-step approach to maximize the mutual information over a generalized precoding matrix  $\mathbf{P}$  via combining Proposition 3 and Algorithm 1 and 2.

*Algorithm 3:* Two-Step Algorithm to Maximize the Mutual Information Over a Generalized Precoding Matrix

- 1) Set the left singular vector of the precoder  $\mathbf{U} := \mathbf{V}_H$ , and give a feasible  $\boldsymbol{\lambda}$  and  $\mathbf{V}$ .
- 2) *Update power allocation vector:* Run Algorithm 1 given  $\mathbf{V}$ .
- 3) *Update right singular vector:* Run Algorithm 2 given  $\boldsymbol{\lambda}$ .
- 4) Go to Step 2 until convergence.

## V. APPLICATIONS

We consider a single-relay network with the block length  $L = 1$  and the channel coefficient  $h_0 = 0.4$ ,  $h_1 = 1.2$  and  $g_1 = -0.9j$ . We assume the same transmit power at the source and relay node (*i.e.*,  $P_s = P_r = P$ ), and the SNR is 3 dB. when the elements of the transmitted signal  $\mathbf{x}$  is drawn independently from BPSK constellations, the mutual information is bounded by 1 bit/s/Hz as shown in (3).

The convergence of the proposed approach is illustrated in Fig. 1. We also show the convergence of algorithms proposed in [8], [10]. From Fig. 1, it is shown that the direct gradient method is stuck at a local maximum (0.53 bit/s/Hz). The reason for such performance is that the optimization problem is not convex in general. In contrast, the proposed two-step algorithm exploits the characterization of the optimal solution, which leads to a solution with the optimal left singular vector, the optimal power allocation vector (for a given right singular vector), and the local optimal right singular vector from an arbitrary start point. Hence, the algorithm is able to converge to a much higher value (0.85 bit/s/Hz) with about 60 percent improvement. Note that the progress of the proposed method has a staircase shape, with each stair associated with either the iteration for  $t$ , named as outer iteration in [12], or the change

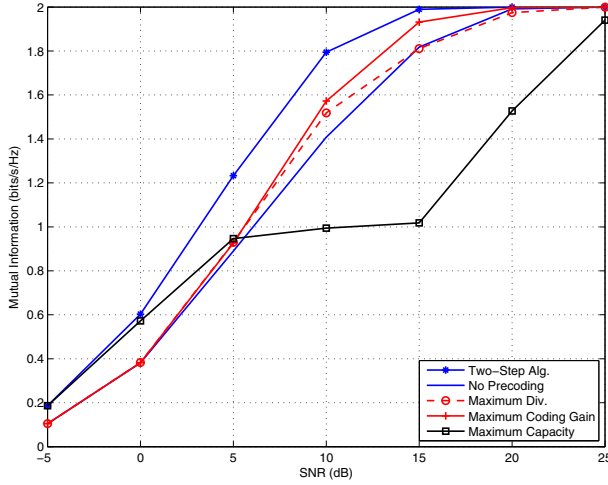


Fig. 2. Mutual Information of relay networks with QPSK inputs.

between the optimizations of the power allocation vector and the right singular vector.

The performance of the proposed algorithm is shown in Fig. 2, in which the information symbol  $x$  is modulated as QPSK, and the channel is the same as the above case. For the sake of completeness, we also show the performance corresponding to the case of no precoding, maximum diversity design in [14], maximum coding gain design in [14], [15], and maximum capacity design with Gaussian inputs assumption in [4]. From Fig. 2, we have following several observations.

The method based on maximizing capacity with Gaussian-input assumption may result in a significant *loss* for discrete inputs, especially when the SNR is in medium-to-high regions. The reason comes from the difference in design power allocation vector and right singular vector. For Gaussian inputs, it is always helpful to allocate more power to the stronger subchannels and less power to the weaker subchannels to maximize the capacity. However, this does not work for the case of finite inputs. Since the mutual information of the relay network is upper bounded by  $\log M$  from (3), there is little incentive to allocate more power to subchannels when they are already close to saturation. Moreover, the right singular vector for Gaussian inputs is an arbitrary unitary matrix to maximize the capacity, because the mutual information is unchanged when the input signal is rotated by a unitary matrix.

The maximum coding gain design has better performance than the method of maximum diversity order and no precoding. We should note that the maximum coding design in [14] is only valid for the case of block length  $L = 1$  and QPSK inputs; it is extended to the case of  $L = 1$  and 16-QAM inputs in [15].

The proposed two-step precoder optimization results in significant gain on mutual information in a wide range of SNR region. For example, it is about 2 dB, 4 dB and 10 dB better than the method of maximum coding gain, no precoding and maximum capacity, respectively, when the channel coding rate is  $2/3$ . Moreover, this algorithm is able to be utilized for an arbitrary block length  $L$  and input type.

## VI. SUMMARY AND CONCLUSION

In this paper, we have studied the precoding design for dual-hop AF relay networks. In contrast with the previous work utilizing various design criteria with the idealistic Gaussian-input assumptions, we have formulated the linear precoding design from the standpoint of discrete-constellation inputs. To develop an efficient precoding design algorithm, we have chosen the mutual information as the utility function. Unfortunately, the maximization of this utility function over all possible complex precoding matrix is nonconvex, *i.e.*, the direct optimization on the precoder can be stuck at a local maxima, which is influenced by the starting point and may be far away from the optimal solution. We have exploited the structure of the precoding matrix under finite-alphabet constraint and developed a unified framework to solve this nonconvex optimization problem. We have proposed a two-step iterative algorithm to maximize the mutual information. Numerical examples have shown substantial gains of our proposed approach on mutual information compared to its conventional counterparts.

## REFERENCES

- [1] 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, 2008.
- [2] 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, 2010.
- [3] N. Khajehnouri and A. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 7, p. 3336, 2007.
- [4] 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, 2009.
- [5] 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, 2010.
- [6] A. Lozano, A. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.
- [7] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Globecom*, New Orleans, LA, Nov. 2008, pp. 1–5.
- [8] W. Zeng, M. Wang, C. Xiao, and J. Lu, "On the power allocation for relay networks with finite-alphabet constraints," in *Proc. IEEE Globecom*, Miami, FL, 2010.
- [9] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for relay networks with finite-alphabet inputs: Theory and practice," *submitted to IEEE Trans. Wireless Commun.*, Dec. 2010.
- [10] D. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 141–154, 2006.
- [11] M. Payaro and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 8, pp. 3613–3628, 2009.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [13] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, 2002.
- [14] 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, 2007.
- [15] —, "Optimal precoder for amplify-and-forward half-duplex relay system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, p. 2891, 2008.