Manifold Optimization Algorithms for SWIPT over MIMO Broadcast Channels with Discrete Input Signals

An-An Lu*, Xiqi Gao*, Yahong Rosa Zheng† and Chengshan Xiao†
*National Mobile Communications Research Laboratory, Southeast University
Nanjing 210096, China, Email: {aalu, xqgao}@seu.edu.cn
†Department of Electrical and Computer Engineering, Missouri University of Science and Technology
Rolla, MO 65409, Email: {zhengyr, xiaoc}@mst.edu

***Abstract*—In this paper, the design of linear precoders for simultaneously wireless information and power transfer (SWIPT) over multi-input multi-output (MIMO) broadcast channels with discrete input signals is investigated. The considered system model consists of one base station (BS), one information receiver (IR) and one energy receiver (ER). The design objective is to maximize the input-output mutual information of the IR subject to the harvested energy requirement for the ER. The structure of the optimal precoder is derived by using the methods of manifold optimization, and an algorithm is proposed to find the optimal precoder. Simulation results show that the proposed algorithm can achieve better performance than the time sharing scheme and the optimal precoder designed for Gaussian inputs.**

Keywords*—*SWIPT, MIMO, precoder, manifold optimization.

I. Introduction

Simultaneous wireless information and power transfer (SWIPT) over multi-input multi-output (MIMO) channels has been widely investigated recently [1], [2]. SWIPT provides a promising solution to prolong the lifetime of energy constrained wireless networks. In SWIPT MIMO systems, the information carrying signals are not only used for communication but also for energy extraction. To achieve a good rate-energy tradeoff, the linear precoders at the transmitter should be properly designed.

In [1], the linear precoders for SWIPT over MIMO broadcast channels with Gaussian inputs are investigated. However, discrete constellations, such as quadrature phase-shift keying (QPSK) and 16-quadrature amplitude modulation (16-QAM), have to be used in practical systems. Thus, we investigate in this paper the design of linear precoders for SWIPT over MIMO broadcast channels with discrete input signals. The considered system model consists of one base station (BS), one information receiver (IR) and one energy receiver (ER). More specifically, we consider maximizing the information rate of the IR subject to the constraint that the energy harvested by the ER is larger than a fixed value.

When the ER is absent, the problem becomes the same as the design of optimal linear precoders for point-to-point MIMO transmissions with discrete input signals, which has been well addressed in the literature. In [3], the necessary conditions for the optimal precoder were derived by applying the Karush–Kuhn–Tucker (KKT) conditions. In [4], it was shown that the left singular vectors of the optimal precoder maximizing the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ under the power constraint are the same as the right singular vectors of the channel matrix. In [5], the structure of the optimum precoder was derived based on the results from [3]. In [6], a globally optimal linear precoder was proposed for discrete input signals over complex vector Gaussian channels.

Optimization algorithms on manifold [7] have been widely used in MIMO systems. The mostly addressed manifolds are Stiefel manifolds and Grassmann manifolds. The manifold of unitary matrices is a special kind of complex Stiefel manifold. In this paper, we view the considered problem as a constrained optimization problem over the product manifold of two unitary manifolds and a vector space. The KKT conditions for the constrained optimization problems on Riemannian manifolds have been investigated in [8]. Using the results from [8], we obtain the structure of the optimal linear precoder for the considered optimization problem. Further, we propose an algorithm to find the optimal precoder.

The rest of this article is organized as follows. The problem formulation is presented in Section II. The structure of the linear optimal precoder is provided in Section III. The algorithm for the precoder design is proposed in Section IV. Simulations are contained in Section V. The conclusion is drawn in Section VI.

II. Problem Formulation

A. System Model

We consider a MIMO broadcasting system over frequency-flat fading channels, which consists of one BS, one IR and one ER. The transmitter at the BS is equipped with M antennas. The IR is equipped with N antennas. The ER is equipped with L antennas. The received signal of the IR is given by

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

where $\mathbf{H}$ is the $N \times M$ channel matrix between the transmitter and the IR, $\mathbf{P}$ is the $M \times M$ precoding matrix, $\mathbf{x}$ is the $M \times 1$ transmitted vector for the IR, and $\mathbf{z}$ is a complex Gaussian noise vector distributed as $\mathcal{CN}(0, \sigma_z^2\mathbf{I}_N)$. The energy harvested by the ER is given by $E = \mathrm{tr}\left(\mathbf{P}\mathbf{P}^H\mathbf{G}^H\mathbf{G}\right)$, where $\mathbf{G}$ is the $L \times M$ channel matrix between the transmitter and the ER. We assume that the transmitter knows $\mathbf{H}$ and $\mathbf{G}$ at each fading state, whereas the receivers only know their corresponding channels.

B. Problem Formulation

We assume that $\mathbf{x}$ is equiprobably drawn from discrete constellations, such as QPSK and 16-QAM. Let $\mathcal{I}(\mathbf{x};\mathbf{y})$ denote the input-output mutual information of the channel $\mathbf{H}$. When $\mathbf{HP}$ is perfectly known at the receiver, $\mathcal{I}(\mathbf{x},\mathbf{y})$ is given by

$$\mathcal{I}(\mathbf{x};\mathbf{y}) = M\log D - \frac{1}{D^M}\sum_{j=1}^{D^M}\mathbb{E}_{\mathbf{z}}\{\log_2\sum_{i=1}^{D^M}\exp(-d_{ij})\} \quad (2)$$

where $d_{ij} = \sigma_z^{-2}(\|\mathbf{HP}(\mathbf{x}_i - \mathbf{x}_j) + \mathbf{z}\|^2 - \|\mathbf{z}\|^2)$ and D is the cardinality of the discrete constellations.

Let $\mathbf{P}^\star$ denote the optimal linear precoder that maximize the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ under the power constraint $\mathrm{tr}(\mathbf{PP}^H) \leq P$. We define $E_{\min} = \mathrm{tr}(\mathbf{P}^\star\mathbf{P}^{\star H}\mathbf{G}^H\mathbf{G})$. Let $E_{\max}$ denote the maximal energy that can be harvested by the ER. It is defined as

$$\begin{aligned} E_{\max} = \max_{\mathbf{P}} \quad & \mathrm{tr}(\mathbf{PP}^H\mathbf{G}^H\mathbf{G}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{PP}^H) \leq P. \end{aligned} \quad (3)$$

Let $\mathbf{P} = \mathbf{U}_P\mathbf{\Sigma}_P\mathbf{V}_P^H$ denote the singular value decomposition (SVD) of $\mathbf{P}$. We consider the problem of maximizing the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ subject to meeting the harvested energy requirement of the ER. Moreover, we view the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ as a function over the product manifold $\mathcal{U}_M \times \mathbb{R}^M \times \mathcal{U}_M$, where $\mathcal{U}_M$ denotes the group of unitary matrices of size $M \times M$. Let $\mathcal{M}$ denote the product manifold $\mathcal{U}_M \times \mathbb{R}^M \times \mathcal{U}_M$. The considered optimization problem can be formulated as

$$\begin{aligned} \max_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)\in\mathcal{M}} \quad & \mathcal{I}(\mathbf{x};\mathbf{y}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{\Sigma}_P^2) \leq P \\ & \mathrm{tr}(\mathbf{U}_P\mathbf{\Sigma}_P^2\mathbf{U}_P^H\mathbf{G}^H\mathbf{G}) \geq E \end{aligned} \quad (4)$$

where $E_{\min} \leq E \leq E_{\max}$. The considered problem (4) is a constrained optimization problem over the manifold $\mathcal{M}$.

III. Optimal Linear Precoder Structure

In this section, we provide the structure of the optimal precoder for the problem (4).

A. Gradients on the Manifold $\mathcal{M}$

The gradient of the input-output mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ at $(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)$ on the Euclidean space $\mathbb{C}^{M\times M} \times \mathbb{R}^M \times \mathbb{C}^{M\times M}$ is constituted as

$$\nabla_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{I} = \left\langle \frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*}, \frac{\partial\mathcal{I}}{\partial\mathbf{\Sigma}_P}, \frac{\partial\mathcal{I}}{\partial\mathbf{V}_P^*} \right\rangle \quad (5)$$

where $\frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*}$, $\frac{\partial\mathcal{I}}{\partial\mathbf{\Sigma}_P}$ and $\frac{\partial\mathcal{I}}{\partial\mathbf{V}_P^*}$ denote the Euclidean gradients of the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{U}_P$, $\mathbf{\Sigma}_P$ and $\mathbf{V}_P$, respectively. Let $\mathbf{W}$ denote $\mathbf{P}^H\mathbf{H}^H\mathbf{HP}$. From Theorem 1 of [6], we obtain that

$$\frac{\partial\mathcal{I}}{\partial\mathbf{W}^*} = \mathbf{\Phi} \quad (6)$$

where $\mathbf{\Phi} = \mathbb{E}\{(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^H\}$ is the minimum mean-square error (MMSE) matrix and $\hat{\mathbf{x}}$ is the conditional mean $\mathbb{E}\{\mathbf{x}|\mathbf{y}\}$. Since the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ depends on $\mathbf{U}_P$, $\mathbf{\Sigma}_P$ and $\mathbf{V}_P$ through $\mathbf{W} = \mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P\mathbf{V}_P^H$, we can obtain the Euclidean gradients of $\mathbf{U}_P$, $\mathbf{\Sigma}_P$ and $\mathbf{V}_P$ from $\frac{\partial\mathcal{I}}{\partial\mathbf{W}^*}$. Using a method similar to that in Lemma 4 of [6], we obtain from (6) the gradients

$$\frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*} = \mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P \quad (7)$$

$$\begin{aligned} \frac{\partial\mathcal{I}}{\partial\mathbf{\Sigma}_P} = {} & \mathrm{diag}(\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P \\ & + \mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P) \end{aligned} \quad (8)$$

$$\frac{\partial\mathcal{I}}{\partial\mathbf{V}_P^*} = \mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P. \quad (9)$$

The gradient of the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ on the manifold $\mathcal{M}$ at $(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)$ is the projection of $\nabla_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{I}$ onto the tangent space $T_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{M}$ of the product manifold $\mathcal{M}$ at $(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)$. The tangent space $T_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{M}$ can be decomposed as the product of tangent spaces, *i.e.*, [9]

$$T_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{M} = T_{\mathbf{U}_P}\mathcal{U}_M \times \mathbb{R}^M \times T_{\mathbf{V}_P}\mathcal{U}_M \quad (10)$$

where

$$T_{\mathbf{U}_P}\mathcal{U}_M = \{\mathbf{X} \in \mathbb{C}^{M\times M} | \mathbf{X}^H\mathbf{U}_P + \mathbf{U}_P^H\mathbf{X} = \mathbf{0}\} \quad (11)$$

and $T_{\mathbf{V}_P}\mathcal{U}_M$ is similarly defined. Thus, the projection of $\nabla_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{I}$ onto $T_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)}\mathcal{M}$ is the product of the projections of $\frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*}$, $\frac{\partial\mathcal{I}}{\partial\mathbf{\Sigma}_P}$ and $\frac{\partial\mathcal{I}}{\partial\mathbf{V}_P^*}$ onto $\mathcal{U}_M$, $\mathbb{R}^M$ and $\mathcal{U}_M$. Let $\mathrm{grad}\mathcal{I}_{\mathbf{V}_P} \in T_{\mathbf{V}_P}\mathcal{U}_M$ and $\mathrm{grad}\mathcal{I}_{\mathbf{U}_P} \in T_{\mathbf{U}_P}\mathcal{U}_M$ denote the gradients of the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{V}_P$ and $\mathbf{U}_P$ on $\mathcal{U}_M$, respectively. The gradient of the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ at $(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)$ on the product manifold $\mathcal{M}$ consists of the gradients of the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ on each submanifold, *i.e.*,

$$\mathrm{grad}\mathcal{I}_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)} = \left\langle \mathrm{grad}\mathcal{I}_{\mathbf{U}_P}, \frac{\partial\mathcal{I}}{\partial\mathbf{\Sigma}_P}, \mathrm{grad}\mathcal{I}_{\mathbf{V}_P} \right\rangle. \quad (12)$$

The gradient $\mathrm{grad}\mathcal{I}_{\mathbf{U}_P}$ is the projection of $\frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*}$ onto the tangent space $T_{\mathbf{U}_P}\mathcal{U}_M$ at $\mathbf{U}_P$. It is computed as [7]

$$\mathrm{grad}\mathcal{I}_{\mathbf{U}_P} = \frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*} - \mathbf{U}_P(\frac{\partial\mathcal{I}}{\partial\mathbf{U}_P^*})^H\mathbf{U}_P. \quad (13)$$

From (7) and (13), we obtain

$$\begin{aligned} \mathrm{grad}\mathcal{I}_{\mathbf{U}_P} = {} & \mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P \\ & -\mathbf{U}_P\mathbf{\Sigma}_P\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P. \end{aligned} \quad (14)$$

Similarly, the gradient $\mathrm{grad}\mathcal{I}_{\mathbf{V}_P}$ is obtained as

$$\begin{aligned} \mathrm{grad}\mathcal{I}_{\mathbf{V}_P} = {} & \mathbf{\Phi}\mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P \\ & -\mathbf{V}_P\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P. \end{aligned} \quad (15)$$

B. Optimal Linear Precoder Structure

In this subsection, we derive the structure of the optimal precoder for the problem (4). First, we investigate the conditions for $\mathbf{V}_P$ at the critical points. Since the constraints are not imposed on $\mathbf{V}_P$, we have $\mathrm{grad}\mathcal{I}_{\mathbf{V}_P} = \mathbf{0}$ at the critical points. Using this condition, we obtain the following theorem.

Theorem 1. *The matrices* $\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P$ *and*

$$\mathbf{\Sigma}_P\mathbf{U}_P^H\mathbf{H}^H\mathbf{HU}_P\mathbf{\Sigma}_P$$

have the same eigenvectors at the critical points of the problem (4).

Proof: The proof is provided in Appendix A. ■

When the second constraint of the problem (4) is removed, the problem becomes to maximize the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ under the power constraint $\mathrm{tr}(\mathbf{\Sigma}_P^2) \le P$. In such case, we also have $\mathrm{grad}\mathcal{I}_{\mathbf{U}_P} = \mathbf{0}$ since there is also no constraint related to $\mathbf{U}_P$. Let $\mathbf{H}^H\mathbf{H} = \mathbf{V}_H\mathbf{\Sigma}_H^2\mathbf{V}_H^H$ denote the eigen value decomposition (EVD) of the Gram matrix $\mathbf{H}^H\mathbf{H}$. Then, we obtain the following theorem.

Theorem 2. *The matrices $\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P$ and $\mathbf{U}_P^H\mathbf{V}_H\mathbf{\Sigma}_H^2\mathbf{V}_H^H\mathbf{U}_P$ are both diagonal matrices at the critical points of the problem*

$$\begin{aligned} \max_{(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)\in\mathcal{M}} \quad & \mathcal{I}(\mathbf{x};\mathbf{y}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{\Sigma}_P^2) \le P. \end{aligned} \tag{16}$$

Proof: The proof is provided in Appendix B. ■

Theorem 2 provides the structure of the optimal linear precoder maximizing the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ under the power constraint $\mathrm{tr}(\mathbf{\Sigma}_P^2) \le P$. Since $\mathbf{V}_P^H\mathbf{\Phi}\mathbf{V}_P$ and $\mathbf{U}_P^H\mathbf{V}_H\mathbf{\Sigma}_H^2\mathbf{V}_H^H\mathbf{U}_P$ are both diagonal matrices, we obtain that the right singular vectors of the optimal precoder $\mathbf{P}$ are the same as the eigenvectors of the MMSE matrix $\mathbf{\Phi}$, and that the left singular vectors of the optimal precoder $\mathbf{P}$ are the same as the right singular vectors of the channel $\mathbf{H}$. The results obtained from Theorem 2 is the same as that of Theorem 1 in [5]. However, Theorem 2 is different from Theorem 1 of [5] in that we view $(\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P)$ as an element from the product manifold $\mathcal{M}$. Furthermore, we show that the optimal structure comes from $\mathrm{grad}\mathcal{I}_{\mathbf{V}_P} = \mathbf{0}$ and $\mathrm{grad}\mathcal{I}_{\mathbf{U}_P} = \mathbf{0}$. Thus, we can obtain the same result when the form of the constraint is changed as long as the constraint is only imposed on $\mathbf{\Sigma}_P$.

The problem (4) is a constrained optimization problem on the product manifold $\mathcal{M}$. The KKT conditions for the constrained optimization problems on Riemannian manifolds have been investigated in [8]. We define the Lagrangian as

$$\begin{aligned} \mathcal{L}(\mu,\lambda,\mathbf{U}_P,\mathbf{\Sigma}_P,\mathbf{V}_P) = & -\mathcal{I}(\mathbf{x};\mathbf{y}) + \mu(\mathrm{tr}(\mathbf{\Sigma}_P^2) - P) \\ & - \lambda(\mathrm{tr}(\mathbf{U}_P\mathbf{\Sigma}_P^2\mathbf{U}_P^H\mathbf{G}^H\mathbf{G}) - E) \end{aligned} \tag{17}$$

where $\mu \ge 0$ and $\lambda \ge 0$ are the Lagrange multipliers associated with the problem constraints. From Theorem 4.1 of [8], we obtain the KKT conditions as

$$\mathrm{grad}\mathcal{L}_{\mathbf{U}_P^\star} = \mathbf{0} \tag{18}$$

$$\mathrm{grad}\mathcal{L}_{\mathbf{V}_P^\star} = \mathbf{0} \tag{19}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{\Sigma}_P^\star} = \mathbf{0} \tag{20}$$

$$\mu^\star(\mathrm{tr}((\mathbf{\Sigma}_P^\star)^2) - P) = 0 \tag{21}$$

$$\mathrm{tr}((\mathbf{\Sigma}_P^\star)^2) - P \le 0 \tag{22}$$

$$\lambda^\star(\mathrm{tr}(\mathbf{U}_P^\star\mathbf{\Sigma}_P^2(\mathbf{U}_P^\star)^H\mathbf{G}^H\mathbf{G}) - E) = 0 \tag{23}$$

$$\mathrm{tr}(\mathbf{U}_P^\star(\mathbf{\Sigma}_P^\star)^2(\mathbf{U}_P^\star)^H\mathbf{G}^H\mathbf{G}) - E \ge 0 \tag{24}$$

where $\mathrm{grad}\mathcal{L}_{\mathbf{U}_P^\star}$ and $\mathrm{grad}\mathcal{L}_{\mathbf{V}_P^\star}$ are the gradients of the Lagrangian with respect to $\mathbf{U}_P^\star$ and $\mathbf{V}_P^\star$ on $\mathcal{U}_M$, respectively. As we can see from the equations (18) to (24), the KKT conditions for the constrained optimization problems over Riemannian manifolds are the extensions of the KKT conditions for traditional constrained optimization problems over Euclidean spaces. Specifically, the Euclidean gradients in the KKT conditions for the traditional constrained optimization problems are replaced by the gradients on the Riemannian manifold. Using (18) to (24), we obtain the following theorem.

Theorem 3. *The optimal precoding matrix $\mathbf{P}^\star$ for the problem* (4) *has the following structure*

$$\mathbf{P}^\star = \mathbf{T}_G^{-1/2}\mathbf{U}_F\mathbf{\Sigma}_F\mathbf{V}_F^H \tag{25}$$

where $\mathbf{T}_G = \mu^\star\mathbf{I} - \lambda^\star\mathbf{G}^H\mathbf{G}$, $\mathrm{tr}(\mathbf{\Sigma}_F^2) \le \mu^\star P - \lambda^\star E$, the columns of $\mathbf{U}_F$ are the eigenvectors of $\mathbf{T}_G^{-1/2}\mathbf{H}^H\mathbf{H}\mathbf{T}_G^{-1/2}$, and $\mathbf{V}_F^H\mathbf{\Phi}\mathbf{V}_F$ is a diagonal matrix.

Proof: The proof is provided in Appendix C. ■

Let g_1 be the maximal eigenvalue of $\mathbf{G}^H\mathbf{G}$. Using a method similar to that in Lemma A.1 of [1], we can obtain $\mu^\star > \lambda^\star g_1$. Furthermore, it is easy to verify that Theorem 3 also holds when $\mathbf{T}_G = \mathbf{I} - (\lambda^\star/\mu^\star)\mathbf{G}^H\mathbf{G}$ and $\mathrm{tr}(\mathbf{\Sigma}_F^2) \le P - (\lambda^\star/\mu^\star)E$.

IV. Proposed Algorithm for Precoder Design

In this section, we proposed an algorithm for the precoder design based on the above obtained optimal structure.

A. Equivalent Equality Constrained Problem

Let β be a real number with $0 < \beta < g_1^{-1}$, $\mathbf{T}_G = \mathbf{I} - \beta\mathbf{G}^H\mathbf{G}$ and $\mathbf{U}_F$ be the eigenvectors of $\mathbf{T}_G^{-1/2}\mathbf{H}^H\mathbf{H}\mathbf{T}_G^{-1/2}$. We introduce the parametrization that $\mathbf{P} = \mathbf{T}_G^{-1/2}\mathbf{U}_F\mathbf{\Sigma}_F\mathbf{V}_F^H$, where $\mathbf{\Sigma}_F \in \mathbb{R}^M$ and $\mathbf{V}_F \in \mathcal{U}_M$. The optimization problem (4) is equivalent to

$$\begin{aligned} \max_{\beta,\mathbf{\Sigma}_F,\mathbf{V}_F} \quad & \mathcal{I}(\mathbf{x};\mathbf{y}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}_F^2) \le P \\ & \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}_F^2) \ge E \end{aligned} \tag{26}$$

where $\mathbf{A} = \mathbf{U}_F^H(\mathbf{I} - \beta\mathbf{G}^H\mathbf{G})^{-1}\mathbf{U}_F$ and $\mathbf{B} = \mathbf{U}_F^H\mathbf{G}^H\mathbf{G}(\mathbf{I} - \beta\mathbf{G}^H\mathbf{G})^{-1}\mathbf{U}_F$. Both the constraints of the problem (26) are inequality constraints. In the following theorem, we show that they can be replaced with equality constraints.

Theorem 4. *The optimization problem* (4) *is equivalent to the equality constrained optimization problem*

$$\begin{aligned} \max_{\beta,\mathbf{\Sigma}_F,\mathbf{V}_F} \quad & \mathcal{I}(\mathbf{x};\mathbf{y}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}_F^2) = P \\ & \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}_F^2) = E \end{aligned} \tag{27}$$

where $E_{\min} \le E \le E_{\max}$.

Proof: The proof is provided in Appendix D. ■

B. Two-Stage Algorithm for Precoder Design

The optimization problem (27) is very difficult to solve directly. Instead, we propose a two-stage algorithm. First, we maximize the mutual information $\mathcal{I}(\mathbf{x};\mathbf{y})$ under the constraint with fixed β. Then, we choose a good β.

When β is fixed, $\mathcal{I}(\mathbf{x};\mathbf{y})$ can be seen as the input-output mutual information of the receive model

$$\mathbf{y} = \mathbf{\Sigma}_{H_T}\mathbf{\Sigma}_F\mathbf{V}_F^H\mathbf{x} + \mathbf{z} \tag{28}$$

where $\mathbf{\Sigma}_{H_T}$ is a diagonal matrix including the singular values of $\mathbf{H}\mathbf{T}_G^{-1/2}$. From Theorems 1 and 2 of [6], we obtain $\mathcal{I}(\mathbf{x};\mathbf{y})$ is a concave function of $\mathbf{V}_F\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2\mathbf{V}_F^H$ and also a concave function of $\mathbf{\Sigma}_F^2$. Furthermore, we develop an algorithm similar to that in [6] to find the optimum of $\mathcal{I}(\mathbf{x};\mathbf{y})$.

Let $\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$ denote the manifold of positive semidefinite matrices with fixed eigenvalues $\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2$. Let $\mathbf{N}_F \in \mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$ and $\mathbf{N}_F = \mathbf{V}_F\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2\mathbf{V}_F^H$. We use a block coordinate decent method to optimize $\mathbf{N}_F$ and $\mathbf{\Sigma}_F^2$.

The tangent space of $\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$ at $\mathbf{N}_F$ is given by [10]

$$T_{\mathbf{N}_F}\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2) = \{\mathbf{N}_F\mathbf{\Omega} - \mathbf{\Omega}\mathbf{N}_F | \mathbf{\Omega}^H = -\mathbf{\Omega}\}. \tag{29}$$

The gradient of $\mathcal{I}(\mathbf{x};\mathbf{y})$ on $\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$ at $\mathbf{N}_F$ is the projection of the Euclidean gradient $\mathbf{\Phi}$ onto $T_{\mathbf{N}_F}\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$. It is given by [10]

$$\mathrm{grad}\mathcal{I}_{\mathbf{N}_F} = \mathbf{N}_F(\mathbf{N}_F\mathbf{\Phi} - \mathbf{\Phi}\mathbf{N}_F) - (\mathbf{N}_F\mathbf{\Phi} - \mathbf{\Phi}\mathbf{N}_F)\mathbf{N}_F. \tag{30}$$

The projected gradient of $\mathcal{I}(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{\Sigma}_F^2$ under the linear constraints $\mathrm{tr}(\mathbf{A}\mathbf{\Sigma}_F^2) = P$ and $\mathrm{tr}(\mathbf{B}\mathbf{\Sigma}_F^2) = E$ is given by

$$\mathbf{D}_F = \frac{\partial \mathcal{I}}{\partial \mathbf{\Sigma}_F^2} - \alpha_1 \mathrm{diag}(\mathbf{A}) - \alpha_2 \mathrm{diag}(\mathbf{B}) \tag{31}$$

where $\frac{\partial \mathcal{I}}{\partial \mathbf{\Sigma}_F^2} = \mathrm{diag}(\mathbf{V}_F^H\mathbf{\Phi}\mathbf{V}_F\mathbf{\Sigma}_{H_T}^2)$, and α_1 and α_2 are obtained through the Gram-Schmidt process.

We now present an algorithm for the precoder design when β is fixed. The details of the projections used are provided in the following subsection.

Algorithm 1: Finding the optimal precoder when β is fixed:

Step 1: Compute $\mathbf{T}_G = \mathbf{I} - \beta\mathbf{G}^H\mathbf{G}$. Apply EVD to decompose $\mathbf{T}_G^{-1/2}\mathbf{H}^H\mathbf{H}\mathbf{T}_G^{-1/2} = \mathbf{U}_{H_T}\mathbf{\Sigma}_{H_T}^2\mathbf{U}_{H_T}^H$.

Step 2: Randomly generate an initial value of $\mathbf{V}_F$. Calculate an initial feasible $\mathbf{\Sigma}_F^2$ by the projection $\pi_{\mathbf{\Sigma}_F^2}(\cdot)$.

Step 3: Calculate the gradient $\mathrm{grad}\mathcal{I}_{\mathbf{N}_F}$ on $\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$ at $\mathbf{N}_F$ according to (30).

Step 4: Update $\mathbf{N}_F$ by $\mathbf{N}_F = \pi_{\mathbf{N}_F}(\mathbf{N}_F + \mu_{\mathbf{N}_F}\mathrm{grad}\mathcal{I}_{\mathbf{N}_F})$ where $\mu_{\mathbf{N}_F}$ is the step size determined by the backtracking line search, and $\pi_{\mathbf{N}_F}(\cdot)$ is the projection onto $\mathcal{M}(\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2)$.

Step 5: Calculate the projected gradient $\mathbf{D}_F$ for $\mathbf{\Sigma}_F^2$ according to (31).

Step 6: Update $\mathbf{\Sigma}_F^2$ by $\pi_{\mathbf{\Sigma}_F^2}(\mathbf{\Sigma}_F^2 + \mu_{\mathbf{\Sigma}_F^2}\mathbf{D}_F)$, where $\mu_{\mathbf{\Sigma}_F^2}$ is the step size determined by the backtracking line search, and $\pi_{\mathbf{\Sigma}_F^2}(\cdot)$ is the projection onto the constraint set.

Repeat Step 3 through Step 6 until convergence or until a preset target is reached. Then an optimal precoder is obtained as $\mathbf{T}_G^{-1/2}\mathbf{U}_{H_T}\mathbf{\Sigma}_F\mathbf{V}_F^H$.

Let $g(\beta)$ denote the maximal mutual information obtained by Algorithm 1 when β is fixed. The problem now is equivalent to find the optimal β that maximize $g(\beta)$. In the simulations, we find that $g(\beta)$ are always quasiconcave functions. Thus, we conjecture that $g(\beta)$ is a quasiconcave function and propose to use the golden section method [11] to find the optimal $\beta^\star$. The details of the golden section method are omitted here due to space limit. Furthermore, the optimal precoder of our algorithm is that provided by Algorithm 1 when $\beta = \beta^\star$.

C. The Projections $\pi_{\mathbf{N}_F}$ and $\pi_{\mathbf{\Sigma}_F^2}$

In this subsection, we introduce the two projections used in the previous subsection.

Let $\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^H$ denote the EVD of $\mathbf{N} + \mu_{\mathbf{N}}\mathrm{grad}\mathcal{I}_{\mathbf{N}}$, where $\mathbf{\Lambda}^2$ and $\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2$ are similarly ordered. Using Theorem 3.9 of [12], we obtain the projection $\pi_{\mathbf{N}_F}(\cdot)$ as

$$\pi_{\mathbf{N}_F}(\mathbf{N}_F + \mu_{\mathbf{N}_F}\mathrm{grad}\mathcal{I}_{\mathbf{N}_F}) = \mathbf{U}\mathbf{\Sigma}_{H_T}^2\mathbf{\Sigma}_F^2\mathbf{U}^H. \tag{32}$$

The projection $\pi_{\mathbf{\Sigma}_F^2}(\mathbf{\Sigma}_X^2)(\cdot)$ is defined as

$$\begin{aligned} \pi_{\mathbf{\Sigma}_F^2}(\mathbf{\Sigma}_X^2) = \arg\min_{\mathbf{\Sigma}_F^2} & \|\mathbf{\Sigma}_F^2 - \mathbf{\Sigma}_X^2\|^2 \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}_F^2) = P \\ & \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}_F^2) = E \\ & \mathbf{\Sigma}_F^2 \succeq 0. \end{aligned} \tag{33}$$

The alternating projection [13], which is an algorithm projecting a point onto the intersection of two convex sets, can be used to realize $\pi_{\mathbf{\Sigma}_F^2}(\cdot)$. Let D denote the set $\{\mathbf{\Sigma}_F^2 : \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}_F^2) = P, \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}_F^2) = E\}$ and C denote the set of positive real diagonal matrices. The projection onto the set D is obtained by [13]

$$P_D(\mathbf{\Sigma}_X^2) = \mathbf{\Sigma}_X^2 - \nu_1\mathrm{diag}(\mathbf{A}) - \nu_2\mathrm{diag}(\mathbf{B}) \tag{34}$$

where ν_1 and ν_2 are the solutions of the equation

$$\mathbf{G}[\nu_1 \ \nu_2] = [\mathrm{tr}(\mathbf{\Sigma}_X^2\mathbf{A}) - P \quad \mathrm{tr}(\mathbf{\Sigma}_X^2\mathbf{B}) - E] \tag{35}$$

and

$$\mathbf{G} = \begin{pmatrix} \mathrm{tr}(\mathrm{diag}(\mathbf{A})\mathrm{diag}(\mathbf{A})) & \mathrm{tr}(\mathrm{diag}(\mathbf{A})\mathrm{diag}(\mathbf{B})) \\ \mathrm{tr}(\mathrm{diag}(\mathbf{B})\mathrm{diag}(\mathbf{A})) & \mathrm{tr}(\mathrm{diag}(\mathbf{B})\mathrm{diag}(\mathbf{B})) \end{pmatrix}. \tag{36}$$

The projection onto the set C is given by

$$P_C(\mathbf{\Sigma}_X^2) = \sum_{i=1}^{M} \max([\mathbf{\Sigma}_X^2]_{ii}, 0)\mathbf{e}_i\mathbf{e}_i^H \tag{37}$$

where $[\mathbf{\Sigma}_X^2]_{ii}$ is the i-th diagonal element $[\mathbf{\Sigma}_X^2]$, and $\mathbf{e}_i$ is the column vector with 1 in the i-th row and 0 in all other rows. Finally, the projection $\pi_{\mathbf{\Sigma}_F^2}(\cdot)$ is given by

$$\pi_{\mathbf{\Sigma}_F^2}(\cdot) = P_C(P_D(\cdots P_C(P_D(\cdot))\cdots)). \tag{38}$$

V. Simulation Results

In this section, we provide simulation results to show the performance of the proposed algorithm for the precoder design. The entries of the matrices $\mathbf{H}$ and $\mathbf{G}$ are randomly generated as identically distributed (i.i.d.) complex Gaussian elements with zero mean and variance 2. The average SNR is given by $\mathrm{SNR} = \frac{1}{M\sigma_z^2}\mathrm{tr}(\mathbf{H}\mathbf{H}^H)$.

We first compare the performance of the proposed algorithm with that of the time sharing scheme under three

$$\mathbf{H} = \begin{pmatrix} -0.5901-0.5385i & 1.5378+0.6440i & -0.4849+0.1216i & 0.0784+0.4684i \\ -2.5604-1.0424i & 2.1721+0.4292i & 0.5338+1.1214i & -0.9200+0.7599i \\ 0.4636-1.3011i & -1.0104-1.3755i & 1.0347-0.6243i & -0.7008-0.8872i \\ 0.5450+0.2953i & -0.3199+1.0883i & 0.4973+0.7091i & -0.6183+0.2224i \end{pmatrix} \tag{39}$$

$$\mathbf{G} = \begin{pmatrix} 0.7582+0.4127i & 1.3232+1.0504i & 0.0790+1.1727i & -1.8051-2.5281i \\ 1.8164+0.3343i & -0.0998+1.0792i & 0.6373-0.4052i & 0.3432-0.3784i \\ -0.8237+0.6564i & 0.8165+1.5711i & -0.0449+0.4414i & -0.6656-0.0992i \\ 1.6773-0.5160i & -0.4451-0.5630i & -0.1227+0.8114i & 0.1749-1.7912i \end{pmatrix} \tag{40}$$

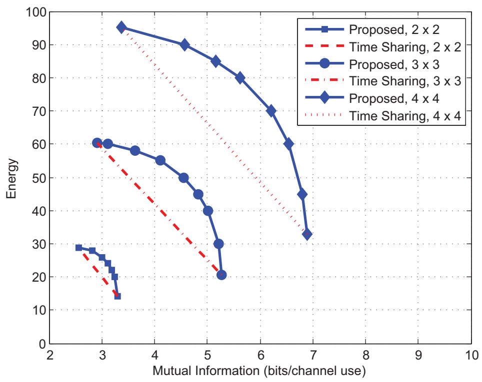

Fig. 1. Mutual information-energy tradeoff for three MIMO Broadcast channels at SNR= 5dB with QPSK inputs.

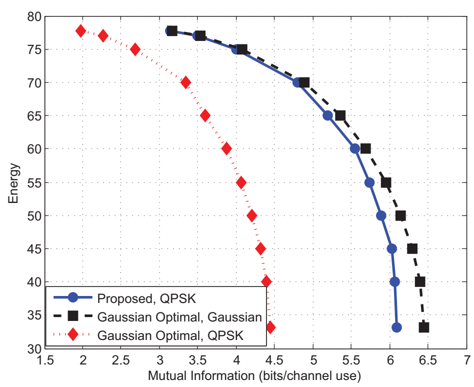

Fig. 2. Mutual information-energy tradeoff for a MIMO Broadcast channel at SNR= 3.75dB with QPSK and Gaussian inputs.

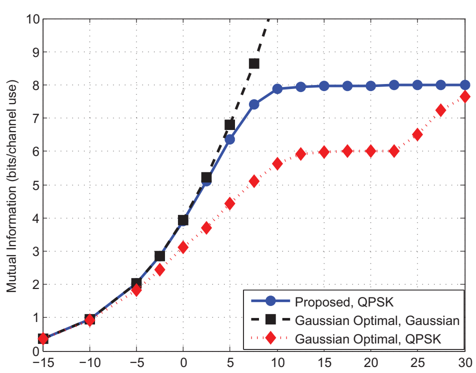

Fig. 3. Mutual information versus SNR for a MIMO Broadcast channel at $E = 55$ with QPSK and Gaussian inputs.

different scenarios. In the time sharing scheme, the precoder is obtained by time sharing between the precoder maximizing the energy of the ER and the precoder maximizing the mutual information of the IR. Fig. 1 presents the mutual information-energy tradeoff of the two algorithms when $N = M = L = 2$, $N = M = L = 3$ and $N = M = L = 4$. The details of $\mathbf{H}$ and $\mathbf{G}$ are omitted here due to space limit. The modulation used is QPSK. As shown in Fig. 1, the proposed algorithm can achieve better mutual information-energy tradeoff than the time sharing scheme except for the two boundary points.

We then compare the performance of the proposed precoding algorithm with that of the optimal precoder designed for Gaussian inputs from [1]. For brevity, we call the later precoder the Gaussian optimal precoder. We consider an example where $\mathbf{H}$ and $\mathbf{G}$ are given in (37) and (38) at the top of this page. Figs. 2 and 3 plot the mutual information-energy tradeoff at SNR= 3.75dB and the mutual information versus SNR at $E = 55$ for the two precoders with QPSK inputs, respectively. As benchmarks, we also plot the results for the Gaussian optimal precoder with Gaussian inputs. From Figs.2 and 3, we observe that the proposed algorithm significantly outperforms the Gaussian optimal precoder when QPSK inputs are used. Furthermore, the performance of the proposed algorithm is very close to that of the Gaussian optimal precoder with Gaussian inputs.

VI. Conclusion

In this paper, we investigated the design of linear precoders for SWIPT over MIMO broadcast channels with discrete input signals. The considered problem is viewed as a constrained optimization problem over a product manifold. Using the KKT conditions for the manifold optimization problems, we obtained the structure of the linear optimal precoder. Furthermore, we proposed an algorithm to find the optimal precoder. Simulation results showed that the proposed algorithm can achieve better performance than the time sharing scheme and the Gaussian optimal precoder when the Gaussian inputs are replaced by discrete input signals.

Appendix A
Proof of Theorem 1

We have $\operatorname{grad} f_{\mathbf{V}_P} = \mathbf{0}$ if $(\mathbf{U}_P, \mathbf{\Sigma}_P, \mathbf{V}_P)$ is a critical point of the problem (4). Thus, from (15) we obtain

$$\begin{aligned}&(\mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P)(\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P) \\ &\quad = (\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P)(\mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P).\end{aligned} \tag{41}$$

The above equation indicates that $\mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P$ commutes with $\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P$. Furthermore, from Theorem 9-33 of

[14] we obtain $\mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P$ and $\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P$ have the same eigenvectors.

Appendix B
Proof of Theorem 2

For the problem (16), we have both $\mathrm{grad} f_{\mathbf{V}_P} = \mathbf{0}$ and $\mathrm{grad} f_{\mathbf{U}_P} = \mathbf{0}$ if $(\mathbf{U}_P, \mathbf{\Sigma}_P, \mathbf{V}_P)$ is a critical point. If $\mathrm{grad} f_{\mathbf{U}_P} = \mathbf{0}$, we can obtain

$$(\mathbf{U}_P^H \mathbf{V}_H \mathbf{\Sigma}_H^2 \mathbf{V}_H^H \mathbf{U}_P)(\mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P \mathbf{\Sigma}_P) = (\mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P \mathbf{\Sigma}_P)(\mathbf{U}_P^H \mathbf{V}_H \mathbf{\Sigma}_H^2 \mathbf{V}_H^H \mathbf{U}_P). \quad (42)$$

From (41) and (42), we have that the matrix $\mathbf{\Sigma}_P^2$ commutes with the matrix $\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{V}_H \mathbf{\Sigma}_H^2 \mathbf{V}_H^H \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P$ and thus they have the same eigenvectors. It follows that $\mathbf{U}_P^H \mathbf{V}_H \mathbf{\Sigma}_H^2 \mathbf{V}_H^H \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P$ is a diagonal matrix. Then, we have both $\mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P$ and $\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{V}_H \mathbf{\Sigma}_H^2 \mathbf{V}_H^H \mathbf{U}_P \mathbf{\Sigma}_P$ are diagonal matrices.

Appendix C
Proof of Theorem 3

Using steps similar to that deriving $\mathrm{grad}\mathcal{I}_{\mathbf{U}_P}$, we can obtain $\mathrm{grad}\mathcal{L}_{\mathbf{U}_P}$ and $\mathrm{grad}\mathcal{L}_{\mathbf{V}_P}$. According to $\mathrm{grad}\mathcal{L}_{\mathbf{U}_P} = \mathbf{0}$ and $\mathrm{grad}\mathcal{L}_{\mathbf{V}_P} = \mathbf{0}$, we obtain that

$$(\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P + \lambda \mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{G}^H \mathbf{G} \mathbf{U}_P \mathbf{\Sigma}_P) \mathbf{\Sigma}_P^2 = \mathbf{\Sigma}_P^2 (\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P + \lambda \mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{G}^H \mathbf{G} \mathbf{U}_P \mathbf{\Sigma}_P). \quad (43)$$

Thus, we obtain

$$\mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} \mathbf{V}_P + \lambda \mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{G}^H \mathbf{G} \mathbf{U}_P \mathbf{\Sigma}_P$$

is a diagonal matrix since it commutes with $\mathbf{\Sigma}_P^2$. Furthermore, from $\frac{\partial \mathcal{L}}{\partial \mathbf{\Sigma}_P} = \mathbf{0}$, we obtain

$$\mathbf{V}_P \mathbf{\Sigma}_P \mathbf{U}_P^H \mathbf{H}^H \mathbf{H} \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H \mathbf{\Phi} = \mathbf{V}_P \mathbf{\Sigma}_P \mathbf{U}_P^H (\mu \mathbf{I} - \lambda \mathbf{G}^H \mathbf{G}) \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H. \quad (44)$$

Let $\mathbf{T}_G = \mu \mathbf{I} - \lambda \mathbf{G}^H \mathbf{G}$ and $\mathbf{U}_F \mathbf{\Sigma}_F \mathbf{V}_F^H = \mathbf{T}_G^{1/2} \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^H$ we obtain

$$\mathbf{\Sigma}_F \mathbf{U}_F^H \mathbf{T}_G^{-1/2} \mathbf{H}^H \mathbf{H} \mathbf{T}_G^{-1/2} \mathbf{U}_F \mathbf{\Sigma}_F \mathbf{V}_F^H \mathbf{\Phi} \mathbf{V}_F = \mathbf{\Sigma}_F^2. \quad (45)$$

Furthermore, it can be verified that $\mathbf{V}_F^H \mathbf{\Phi} \mathbf{V}_F$ commutes with $\mathbf{\Sigma}_F \mathbf{U}_F^H \mathbf{T}_G^{-1/2} \mathbf{H}^H \mathbf{H} \mathbf{T}_G^{-1/2} \mathbf{U}_F \mathbf{\Sigma}_F$. Thus, they are both diagonal matrices, and we obtain the columns of $\mathbf{U}_F$ are the eigenvectors of $\mathbf{T}_G^{-1/2} \mathbf{H}^H \mathbf{H} \mathbf{T}_G^{-1/2}$.

Appendix D
Proof of Theorem 4

We prove this theorem by proving that both the constraints of the optimization problem (4) are active. For any matrices $(\mathbf{U}_P, \mathbf{\Sigma}_P, \mathbf{V}_P)$ satisfying $\mathrm{tr}(\mathbf{\Sigma}_P^2) < P$ and $\mathrm{tr}(\mathbf{U}_P \mathbf{\Sigma}_P^2 \mathbf{U}_P \mathbf{G}^H \mathbf{G}) \geq E$, we can always choose $c_0 > 1$ that $\mathrm{tr}(c_0 \mathbf{\Sigma}_P^2) = P$ such that $\mathcal{I}(\mathbf{x}; \mathbf{y})$ becomes larger and $\mathrm{tr}(c_0 \mathbf{U}_P \mathbf{\Sigma}_P^2 \mathbf{U}_P \mathbf{G}^H \mathbf{G}) > E$. Thus, the solution of the optimization problem (4) remains the same when the first constraint becomes an equality constraint. We then prove that the second constraint of (4) is active for the optimal precoder by using the method of proof by contradiction. We assume that the optimal solution $(\mathbf{U}_P^\star, \mathbf{\Sigma}_P^\star, \mathbf{V}_P^\star)$ is achieved when the constraint $\mathrm{tr}(\mathbf{U}_P \mathbf{\Sigma}_P^2 \mathbf{U}_P^H \mathbf{G}^H \mathbf{G}) \geq E$ is inactive. Thus, $(\mathbf{U}_P^\star, \mathbf{\Sigma}_P^\star, \mathbf{V}_P^\star)$ should be a local maximum of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ under the constraint $\mathrm{tr}(\mathbf{\Sigma}_P^2) = P$. From Theorem 2, we know that the left singular vectors of $\mathbf{P}^* = \mathbf{U}_P^\star \mathbf{\Sigma}_P^\star (\mathbf{V}_P^\star)^H$ are the same as the right singular vectors of $\mathbf{H}$. However, for such case, we can always find a $(\mathbf{U}_P^\star, \mathbf{\Sigma}_P, \mathbf{V}_P)$ in the neighborhood of $(\mathbf{U}_P^\star, \mathbf{\Sigma}_P^\star, \mathbf{V}_P^\star)$ to make the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ larger using the method from [6], since $\mathbf{P}^\star$ is not the global maximum of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ under the constraint $\mathrm{tr}(\mathbf{\Sigma}_P^2) = P$. Thus, $\mathbf{P}^\star$ cannot be the optimal solution and we obtain the problem (4) is equivalent to (27).

Acknowledgment

The work of A.-A. Lu and X. Q. Gao was supported in part by National Natural Science Foundation of China under Grants 61320106003, 61471113, 61521061 and 61631018, the China High-Tech 863 Plan under Grants 2015AA01A701 and 2014AA01A704, National Science and Technology Major Project of China under Grant 2014ZX03003006-003, and the Huawei Cooperation Project. The work of Y. R. Zheng and C. Xiao was supported in part by USA National Science Foundation under Grants ECCS-1231848, ECCS-1408316 and ECCS-1539316. Part of this work was carried out while A.-A. Lu was visiting Missouri University of Science and Technology, Rolla, MO.

References

[1] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, 2013.

[2] J. Zhang, C. Yuen, C.-K. Wen, S. Jin, K.-K. Wong, and H. Zhu, "Large system secrecy rate analysis for SWIPT MIMO wiretap channels," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 74–85, 2016.

[3] F. Pérez-Cruz, M. R. 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, 2010.

[4] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, 2003.

[5] M. M. Lamarca Orozco, "Linear precoding for mutual information maximization in MIMO systems," in *6th IEEE International Symposium on Wireless Communication Systems*, 2010, pp. 26–30.

[6] 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, 2011.

[7] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[8] W. H. Yang, L.-H. Zhang, and R. Song, "Optimality conditions for the nonlinear programming problems on Riemannian manifolds," *Optimization Online*, 2012.

[9] M. Chu, N. Del Buono, L. Lopez, and T. Politi, "On the low-rank approximation of data on the unit sphere," *SIAM Journal on Matrix Analysis and Applications*, vol. 27, no. 1, pp. 46–60, 2005.

[10] U. Helmke and J. B. Moore, *Optimization and dynamical systems*. Springer Science & Business Media, 2012.

[11] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[12] P.-A. Absil and J. Malick, "Projection-like retractions on matrix manifolds," *SIAM Journal on Optimization*, vol. 22, no. 1, pp. 135–158, 2012.

[13] S. Boyd and J. Dattorro, "Alternating projections," 2003.

[14] S. Perlis, *Theory of matrices*. Courier Corporation, 1991.