Low-Complexity Precoding for Spatial Modulation

Peng Cheng†, Zhuo Chen*, J. Andrew Zhang‡, Yonghui Li†, and Branka Vucetic†
†School of Electrical and Information Engineering, The University of Sydney, Australia
*CSIRO DATA61, Australia
‡School of Computing and Communications, University of Technology Sydney, Australia
Email: peng.cheng@sydney.edu.au

***Abstract*—In this paper, we investigate linear precoding for spatial modulation (SM) over multiple-input-multiple-output (MIMO) fading channels. With channel state information available at the transmitter, our focus is to maximize the minimum Euclidean distance among all candidates of SM symbols. We prove that the precoder design is a large-scale non-convex quadratically constrained quadratic program (QCQP) problem. However, the conventional methods, such as semi-definite relaxation and iterative concave-convex process, cannot tackle this challenging problem effectively or efficiently. To address this issue, we leverage augmented Lagrangian and dual ascent techniques, and transform the original large-scale non-convex QCQP problem into a sequence of subproblems. These subproblems can be solved in an iterative manner efficiently. Numerical results show that the proposed method can significantly improve the system error performance relative to the SM without precoding, and features extremely fast convergence rate with very low computational complexity.**

I. Introduction

Spatial modulation (SM) [1] has been recently proposed as a novel energy-efficient and low-complexity scheme for multiple-input-multiple-output (MIMO) transmission [2]. The philosophy behind SM is to activate only a single antenna at each time slot, but simultaneously use the original signal constellation and the additional space of the antenna index to convey bits of information. The main advantages of SM over the conventional layered space-time architecture are the introduction of energy-efficient single-radio frequency chain at the transmitter and the avoidance of inter-channel interference (ICI) at the receiver. To harvest this benefit, a large number of papers strived to design low-complexity SM detectors, such as sphere decoding [3], the ordered-blocked minimum-mean-squared-error [4], and the enhanced Bayesian compressive sensing EBCS [5].

To further improve SM system performance, it is not surprising to see some efforts on preprocessing at the transmitter. For example, an adaptive SM method [6] was proposed to select the optimal transmission mode to improve the error performance while maintaining the target average transmission rate. The use of transmit antenna subset selection in SM systems was also reported in [7]. In general, these methods are relatively complicated to be implemented in practice.

With channel state information (CSI) available at the transmitter, another leap forward of SM improvement comes with the introduction of linear transmit precoding methods based on the principle of the maximum minimum Euclidean distance (MMD) [1], [8]–[10] , due to its simplicity. In [8], the phase alignment technique was applied to SM systems to improve the transmit diversity, but this technique is limited to multiple-input single-output (MISO) channels. The pre-scaling optimization based on semidefinite relaxation (SDR) was proposed in [9]; however, it is only applicable to the special case of space shift keying of SM. By quantizing the amplitude and phase of precoding weights restricted to two transmit antennas, the precoder designs in [1] can achieve sub-optimal error performance. More recently, the MMD based linear precoder [10] for SM was designed by using an iterative concave-convex process. Though the method can achieve much better error performance, it needs to solve multiple convex optimization subproblems before reaching convergence, and the convergence rate is found to be very slow, resulting in high computational complexity. This motivates us to design a low-complexity and high-performance precoding method for SM.

In this paper, we propose a novel linear precoding method for SM systems based on the MMD criterion. We first prove that the precoder design to be a large-scale non-convex quadratically constrained quadratic program (QCQP) problem, which is highly challenging. We propose to address this QCQP problem by leveraging augmented Lagrangian and dual ascent techniques. The proposed algorithm transforms the original large-scale non-convex QCQP problem into a sequence of unconstraint subproblems in an iterative manner, while each subproblem can be efficiently solved by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [11]. Numerical results show that our proposed algorithm can significantly improve the error performance of SM, and features extremely fast convergence rate with very low computational complexity.

The rest of this paper is organized as follows. Section II presents the SM system model. Section III describes the precoding design criterion and conventional precoding methods. Section IV describes the proposed precoding method. In Section V, simulation results are provided to validate the benefits of our proposed method. Finally, conclusions are drawn in Section VI [1].

[1] *Notation*: $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, and $\mathbf{0}_N$ denotes an $N \times N$ all-zero matrix. $\|\mathbf{a}\|$ denotes the 2-norm of a vector $\mathbf{a}$. $\mathrm{Tr}\,(\mathbf{A})$, $\mathrm{rank}\,(\mathbf{A})$, and $\|\mathbf{A}\|_F$ denotes the trace, the rank, and the Frobenius norm of a matrix $\mathbf{A}$, respectively. $[x]^+$ represents $\max(x, 0)$. $\mathrm{Re}\{\mathbf{x}\}$ and $\mathrm{Im}\{\mathbf{x}\}$ represent the real and imaginary parts of $\mathbf{x}$, respectively. $\mathbf{\Lambda} = \mathrm{diag}\,(\mathbf{a})$ changes a vector $\mathbf{a}$ into a diagonal matrix $\mathbf{\Lambda}$. $\mathcal{CN}\,(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the complex vector normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{\Sigma}$. In addition, all the subscripts and indices in this paper begin with zero.

II. Preliminary

A. System Model

Consider a MIMO system with N_t transmit and N_r receive antennas. For ease of exposition, we assume $N_t = 2^n$ with n a positive integer. In contrast to the conventional spatial multiplexing, SM uses both the spatial and signal constellations to convey information bits. At each time slot, the first $n = \log_2 N_t$ bits of information are mapped to a spatial constellation point drawn from the set with the cardinality N_t

$$\mathcal{S}_{\text{spatial}} = \{\mathbf{e}_0, \mathbf{e}_1, \cdots, \mathbf{e}_{N_t-1}\}, \tag{1}$$

where $\mathbf{e}_n \in \mathbb{R}^{N_t},\ n = 0, \cdots, N_t-1$ is the n-th column of $\mathbf{I}_{N_t}$. In other words, only a single transmit antenna is activated at each time slot. Then, the last $m = \log_2 M$ bits of information are mapped to a signal constellation point drawn from the set with the cardinality M

$$\mathcal{S}_{\text{signal}} = \{s_0, s_1, \cdots, s_{M-1}\}, \tag{2}$$

where $s_m \in \mathbb{C},\ m = 0, \cdots, M-1$ is a power normalized M-PSK or M-QAM symbol.

The resulting transmit codebook $\mathcal{S}$ with the cardinality $N_t M$ is the Cartesian product of $\mathcal{S}_{\text{spatial}}$ and $\mathcal{S}_{\text{signal}}$. Specifically,

$$\mathcal{S} = \{s_0\mathbf{e}_0, s_1\mathbf{e}_0, \cdots, s_{M-1}\mathbf{e}_0, \cdots, s_{M-1}\mathbf{e}_{N_t-1}\}. \tag{3}$$

A transmitted SM symbol $\mathbf{x} \in \mathbb{C}^{N_t}$ is chosen from $\mathcal{S}$, i.e., $\mathbf{x} = s_m\mathbf{e}_n$. With channel state information (CSI) available at the transmitter, $\mathbf{x}$ is precoded by **a diagonal matrix** $\mathbf{Q} = \text{diag}(\mathbf{q})$ with precoding weights $\mathbf{q} = [q_0, \cdots, q_{N_t-1}]^T \in \mathbb{C}^{N_t}$. Different from the conventional full precoding matrix in spatial multiplexing, SM precoding needs a diagonal matrix as a precoder to preserve the property of single antenna activation. Clearly, there are N_t unknown precoding weights to be optimized to enhance the error performance.

Assuming a quasi-static frequency-flat fading channel, the received signal vector $\mathbf{y} \in \mathbb{C}^{N_r}$ is given by

$$\mathbf{y} = \mathbf{HQx} + \mathbf{w}, \tag{4}$$

where $\mathbf{H} = [\mathbf{h}_0, \cdots, \mathbf{h}_{N_t}] \in \mathbb{C}^{N_r \times N_t}$ denotes the flat-fading MIMO channel matrix, whose entries follow an i.i.d circularly symmetric complex Gaussian distribution $\mathcal{CN}(0,1)$, and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_r})$ is the additive noise vector.

B. Maximum Likelihood Receiver

For SM systems, the ML detector provides the optimal performance by exhaustively searching through all candidates of SM symbols $\mathbf{x}$ in the codebook $\mathcal{S}$ as

$$\hat{\mathbf{x}}_{\text{ML}} = \arg\min_{\mathbf{x}\in\mathcal{S}} \|\mathbf{y} - \mathbf{HQx}\|_2^2. \tag{5}$$

Based on the ML detector, the error performance for a given channel $\mathbf{H}$ can be approximated by the sum of the pariwise error probability given by [1]

$$P_e \le P_e^o = \sum_{i=1}^{N_t M} \sum_{j=1, j\neq i}^{N_t M} Q\left(\sqrt{\frac{1}{2\sigma^2} d_{i,j}(\mathbf{q})}\right), \tag{6}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}dy$, and $d_{i,j}(\mathbf{p}) = \|\mathbf{HQ}(\mathbf{x}_i - \mathbf{x}_j)\|^2$ is the squared Euclidian distance between two SM symbols $\mathbf{x}_i$ and $\mathbf{x}_j$.

At high signal-to-noise ratio (SNR), P_e^o in (6) can be simplified as

$$P_e^o = \lambda \cdot Q\left(\sqrt{\frac{1}{2\sigma^2} d_{\min}(\mathbf{q})}\right), \tag{7}$$

where λ is the number of neighbor points [12], and $d_{\min} = \min_{\forall i,j,i\neq j} d_{i,j}(\mathbf{q})$ is the minimum squared Euclidian distance among the codebook $\mathcal{S}$.

III. Precoding Design Criterion

The objective of the precoding design is to minimize the bit error rate (BER) of SM systems over MIMO fading channels. Since P_e^o in (7) is a monotonically decreasing function of $d_{\min}$, the precoding design can be formulated as the following MMD problem

$$\begin{aligned} (\mathbf{P0}) \qquad & \max_{\mathbf{q}\in\mathbb{C}^{N_t}} \ d_{\min}(\mathbf{q}) \\ & s.t. \ \ \|\mathbf{q}\|^2 \le P_t, \end{aligned}$$

where P_t is the total power constraint at the transmitter.

We now derive a more detailed form of the objective function in $(\mathbf{P0})$. The squared Euclidian distance $d_{i,j}(\mathbf{q})$ in (6) can be calculated as

$$\begin{aligned} d_{i,j}(\mathbf{q}) &= \|\mathbf{HQ}(\mathbf{x}_i - \mathbf{x}_j)\|^2 \\ &= (\mathbf{x}_i - \mathbf{x}_j)^H \mathbf{Q}^H \mathbf{H}^H \mathbf{HQ}(\mathbf{x}_i - \mathbf{x}_j) \\ &\overset{(a)}{=} \text{Tr}(\mathbf{Q}^H\mathbf{H}^H\mathbf{HQ}\Delta\mathbf{x}_{i,j}) \\ &\overset{(b)}{=} \mathbf{q}^H(\mathbf{H}^H\mathbf{H} \odot \Delta\mathbf{x}_{i,j}^T)\mathbf{q} \\ &= \mathbf{q}^H\mathbf{R}_{i,j}\mathbf{q}, \end{aligned} \tag{8}$$

where $\Delta\mathbf{x}_{i,j} \triangleq (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^H$, $\mathbf{R}_{i,j} = \mathbf{q}^H(\mathbf{H}^H\mathbf{H} \odot \Delta\mathbf{x}_{i,j}^T)\mathbf{q}$, $\overset{(a)}{=}$ uses the trace property $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$ for matrices $\mathbf{A}$ and $\mathbf{B}$, and $\overset{(b)}{=}$ uses the matrix rule in [13] where $\odot$ denotes the Hadamard product. At this stage, $(\mathbf{P0})$ can be rewritten as

$$\begin{aligned} (\mathbf{P0}) \qquad & \max_{\mathbf{q}\in\mathbb{C}^{N_t}} \min_{\forall i,j,i\neq j} \ \mathbf{q}^H\mathbf{R}_{i,j}\mathbf{q} \\ & s.t. \ \|\mathbf{q}\|^2 \le P_t. \end{aligned}$$

By introducing an auxiliary variable t, the equivalent epigraph form [14] of $(\mathbf{P0})$ can be shown as

$$\begin{aligned} (\mathbf{P1}) \qquad & \max \ \ t \\ & s.t. \ \ \mathbf{q}^H\mathbf{R}_{i,j}\mathbf{q} \ge t, \ \forall i,j,i\neq j \\ & \qquad \|\mathbf{q}\|^2 \le P_t. \end{aligned}$$

We can further show that $(\mathbf{P1})$ is equivalent to the following problem

$$\begin{aligned} (\mathbf{P2}) \qquad & \min_{\mathbf{q}\in\mathbb{C}^{N_t}} \ \ \|\mathbf{q}\|^2 \\ & s.t. \quad \mathbf{q}^H\mathbf{R}_{i,j}\mathbf{q} \ge d_{\min}, \ \forall i,j,i\neq j, \end{aligned}$$

where $d_{\min}$ is the desired minimum squared distance. The rationale behind $(\mathbf{P2})$ is to guarantee the minimum squared distance among the codebook $\mathcal{S}$, while pursuing the minimum power usage as the objective.

Taking a close look at $(\mathbf{P2})$, we can treat it as a large-scale non-convex QCQP problem. It can be calculated that the number of the quadratic constraints in $(\mathbf{P2})$ is $N_tM\times(N_tM-1)/2$, which is very large for a large number of antennas and a high-order modulation. For example, in a moderate SM system with $N_t=8$ with $M=16$, there are 8128 quadratic constraints. Naturally, it is extremely difficult to obtain the global optimal solution to such problem. A prevailing method to tackle a small-scale non-convex QCQP problem is to approximate it as a convex semi-definite programming (SDP) problem via semidefinite relaxation. By setting $\mathbf{C}=\mathbf{q}\mathbf{q}^H$ and relaxing the problem through dropping the rank 1 constraint of the matrix $\mathbf{C}$, a globally optimal $\hat{\mathbf{C}}$ can be obtained by many convex optimization methods such as the interior point method [14]. When $\mathrm{rank}(\hat{\mathbf{C}})=1$, optimal $\mathbf{q}$ is derived through the decomposition of $\hat{\mathbf{C}}$. However, with the increased number of quadratic constraints, the probability of achieving rank-one solutions by SDR is extremely low, making the method ineffective.

Recently, the authors in [10] proposed a convex relaxation method to approximate $(\mathbf{P2})$. Following [10], the approximated MMD problem (AMMD) is given by

$$\begin{aligned}&\min_{\mathbf{q}\in\mathbb{C}^{N_t}} \quad \|\mathbf{q}\|^2\\ &\quad s.t.\ \mathrm{Re}\big\{2(\mathbf{q}^k)^H\mathbf{R}_{i,j}\mathbf{q}-(\mathbf{q}^k)^H\mathbf{R}_{i,j}\mathbf{q}^k\big\}\geq d_{\min}.\end{aligned} \tag{9}$$

Given $\mathbf{q}^k$, the solution to (9) is $\mathbf{q}^{k+1}$. As $2(\mathbf{q}^k)^H\mathbf{R}_{i,j}\mathbf{q}-(\mathbf{q}^k)^H\mathbf{R}_{i,j}\mathbf{q}^k$ is an affine function of $\mathbf{q}$, (9) is a convex optimization problem. Initializing an point $\mathbf{q}_0$ and iteratively solving (9) for K times until convergence, we can obtain a sequence of solutions $\{\mathbf{q}^k\}_{k=1}^K$, and the last one $\mathbf{q}^K$ serves as the approximated solution to $(\mathbf{P2})$. The AMMD precoding method is able to achieve a favorable error performance in the simulations; however, our further investigations found that the weakness of this method lies in its high computational complexity. On the one hand, the complexity of the primal-dual interior point method to solve (9) each time is $\mathcal{O}\left(N_t^2N_r\right)+\mathcal{O}\left(N_t^4\right)$, which increases dramatically with increasing N_t. On the other hand, we need to solve (9) for K times but K is usually very large. Therefore, the aforementioned two issues impose a great burden in the real-time precoding design at the transmitter. This motivates us to design a novel low-complexity precoding method for SM.

IV. The Proposed Precoder Design

In this section, we develop a novel algorithm to address $(\mathbf{P2})$ by leveraging augmented Lagrangian together with dual ascent techniques. We first review the basic principle of the dual ascent technique [14].

Algorithm 1 Proposed Precoding Algorithm

Initialization: Initialize $\mathbf{I}_{2N_t}\leftarrow\mathbf{m}^0,\forall i,j,i\neq j$, $0.5\leftarrow\lambda_{i,j}$, $10\leftarrow\mu^0$. **Repeat**

1) Update the primal vector $\mathbf{m}^{k+1}$
$$\mathbf{m}^{k+1}=\arg\min_{\mathbf{m}} L(\mathbf{m},\boldsymbol{\lambda}^k,\mu^k). \tag{16}$$
2) Update the dual vector $\boldsymbol{\lambda}^{k+1}$
$$\lambda_{i,j}^{k+1}=\left[\lambda_{i,j}^k-\mu^k((\mathbf{m}^k)^T\mathbf{G}_{i,j}\mathbf{m}^k-d_{\min})\right]^+. \tag{17}$$
3) Update the penalty parameter μ^{k+1}
$$\mu^{k+1}=2\mu^k. \tag{18}$$
4) Set $k\leftarrow k+1$.

Until convergence criterion is met.

A. Dual Ascent

Consider an equality-constrained convex optimization problem

$$\begin{aligned}&\min_{\mathbf{x}\in\mathbb{R}} f(\mathbf{x})\\ &s.t.\quad \mathbf{A}\mathbf{x}=\mathbf{b}.\end{aligned} \tag{10}$$

The Lagrangian corresponding to (10) can be written as

$$L(\mathbf{x},\boldsymbol{\lambda})=f(\mathbf{x})+\boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x}-\mathbf{b}), \tag{11}$$

where $\boldsymbol{\lambda}$ is the dual variable or Lagrange multiplier. The dual function of (11) can be shown as

$$g(\boldsymbol{\lambda})=\min_{\mathbf{x}\in\mathbb{R}} L(\mathbf{x},\boldsymbol{\lambda})=-f^*(-\mathbf{A}^T\boldsymbol{\lambda})-\mathbf{b}^T\boldsymbol{\lambda}, \tag{12}$$

where $f^*(\cdot)$ is the convex conjugate of $f(\cdot)$. The dual problem with respect to the original problem in (10) is

$$\max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}). \tag{13}$$

Assuming that strong duality holds, the optimal values of the primal and dual problems are the same. In this case, we can recover a primal optimal point $\mathbf{x}^\star$ from a dual optimal point $\boldsymbol{\lambda}^\star$ as

$$\mathbf{x}^\star=\arg\min_{\mathbf{x}} L(\mathbf{x},\boldsymbol{\lambda}^\star), \tag{14}$$

provided there is only one minimizer of $L\left(\mathbf{x},\boldsymbol{\lambda}^\star\right)$. At this stage, the dual ascent method consists of primal and dual vectors iterating updates

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} L(\mathbf{x},\boldsymbol{\lambda}^k) \tag{15a}$$
$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k+\alpha^k(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}) \tag{15b}$$

where $\alpha^k>0$ is a step size , and the superscript is the iteration counter. The first step (15a) is an $\mathbf{x}$-minimization step, and the second step (15b) is a dual variable update. With appropriate choice of k, the dual function increases in each step, i.e., $g\left(\boldsymbol{\lambda}^{k+1}\right)>g\left(\boldsymbol{\lambda}^k\right)$.

B. Augmented Lagrangian

We now develop the corresponding dual ascent method in terms of augmented Lagrangian capable of bringing robustness to the dual ascent method. Instead of having the equality constraint in (10), here we need to deal with a massive number of inequality constraints in $(\mathbf{P2})$. For convenience, we first convert $(\mathbf{P2})$ to a real-valued form; this yields a real vector $\mathbf{m} = \left[\mathrm{Re}\{\mathbf{q}\}^T \ \mathrm{Im}\{\mathbf{q}\}^T\right]^T \in \mathbb{R}^{2N_t \times 2N_t}$, and a real matrix $\mathbf{G}_{i,j}$

$$\mathbf{G}_{i,j} = \begin{bmatrix} \mathrm{Re}\{\mathbf{R}_{i,j}\} & -\mathrm{Im}\{\mathbf{R}_{i,j}\} \\ \mathrm{Im}\{\mathbf{R}_{i,j}\} & \mathrm{Re}\{\mathbf{R}_{i,j}\} \end{bmatrix}. \tag{19}$$

In this case, $(\mathbf{P2})$ can be equivalently illustrated as

$$(\mathbf{P3}) \qquad \min_{\mathbf{m}\in\mathbb{R}^{2N_t}} \quad \mathbf{m}^T\mathbf{m}$$
$$s.t. \quad \mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} \geq d_{\min}, \ \forall i, j, i \neq j.$$

The augmented Lagrangian for $(\mathbf{P3})$ can be shown as

$$L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}, \mu) = \mathbf{m}^T\mathbf{m} - \sum_{i=1}^{N_tM} \sum_{j=1, j\neq i}^{N_tM} \lambda_{i,j} \left(\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} - s_{i,j} - d_{\min}\right) + \frac{\mu}{2} \sum_{i=1}^{N_tM} \sum_{j=1, j\neq i}^{N_tM} \left(\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} - s_{i,j} - d_{\min}\right)^2, \tag{20}$$

where $\mu > 0$ is the penalty parameter, and $\boldsymbol{\lambda} \triangleq \{\lambda_{i,j}\}$ is the dual vector. Note that we transform the inequality constraints $\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} \geq d_{\min}$ in $(\mathbf{P3})$ to the equality constraints by introducing a slack vector $\mathbf{s} \triangleq \{s_{i,j}\}$ and showing $\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} = d_{\min} + s_{i,j}$ with $s_{i,j} \geq 0$.

In the following, we elaborate on the details of updating the primal vector $\mathbf{m}$. Following the same spirit in (15a), we first minimize the augmented Lagrangian $L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}^k, \mu^k)$ with respect to $\mathbf{m}$ and $\mathbf{s}$ at the iteration k, given by

$$\min_{\mathbf{m},\mathbf{s}} L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}^k, \mu^k) \tag{21a}$$
$$s.t. \quad \mathbf{s} \geq \mathbf{0}. \tag{21b}$$

Since each $s_{i,j}$ occurs in just two terms of (21a) , which is in fact a convex quadratic function with respect to each of these slack variables. We can therefore perform an explicit minimization in (21a) with respect to each of the $s_{i,j}$ separately. With $\nabla_{\mathbf{s}} L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}^k, \mu^k) = \mathbf{0}$, we have

$$s_{i,j} = \mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} - d_{\min} - \frac{\lambda_{i,j}^k}{\mu^k}. \tag{22}$$

If this unconstrained minimizer is smaller than the lower bound of 0, then since (21a) is convex in $s_{i,j}$, the optimal value of $s_{i,j}$ in (21a) is 0. Therefore, the optimal value of $s_{i,j}$ is given by

$$s_{i,j} = \left[\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} - d_{\min} - \frac{\lambda_{i,j}^k}{\mu^k}, 0\right]^+. \tag{23}$$

In this case, it is readily to substitute (23) into $L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}^k, \mu^k)$, obtaining an equivalent form $L(\mathbf{m}, \boldsymbol{\lambda}^k, \mu^k)$ without $\mathbf{s}$. To minimize the augmented Lagrangian $L(\mathbf{m}, \boldsymbol{\lambda}^k, \mu^k)$, we employ a limited-memory BFGS algorithm [11] which only requires computation of the gradient of $L(\mathbf{m}, \boldsymbol{\lambda}^k, \mu^k)$ with respect to $\mathbf{m}$, $\nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}^k, \mu^k)$, given by

$$\nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k) = 2\mathbf{m} - \sum_{i=0}^{N_tM-1} \sum_{j=0, j\neq i}^{N_tM-1} \psi\left(\mathbf{m}^T\mathbf{G}_{i,j}\mathbf{m} - d_{\min}, \boldsymbol{\lambda}_k, \mu_k\right), \tag{24}$$

where the function $\psi(z, a, b)$ is defined as

$$\psi(z, a, b) \triangleq \begin{cases} (a - bz)\nabla z, & z - \frac{a}{b} \leq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Once the approximate solution $\mathbf{m}^{k+1}$ is obtained, we use the following formula to update the Lagrange multipliers and the penalty parameter

$$\lambda_{i,j}^{k+1} = \left[\lambda_{i,j}^k - \mu^k((\mathbf{m}^k)^T\mathbf{G}_{i,j}\mathbf{m}^k - d_{\min})\right]^+ \tag{26}$$
$$\mu^{k+1} = 2\mu^k. \tag{27}$$

The algorithm procedure is summarized in Algorithm 1.

V. Simulation Results and Discussion

In this section, we present numerical comparisons between our proposed augmented Lagrangian (AL) based method with the AMMD precoding method [10] and phase rotation aided precoding method (PRP) [1]. The ideal CSI is assumed to be available at the transmitter. According to [10], the AMMD precoding can achieve the optimal BER performance among all the linear precoding methods in the literature.

In Fig. 1, we compare the convergence rates of the proposed method and the AMMD method by showing the probability mass function (PMF) of the number of iterations. We consider the SM-MIMO system with $N_t = 8$, $N_r = 4$, and QPSK. In

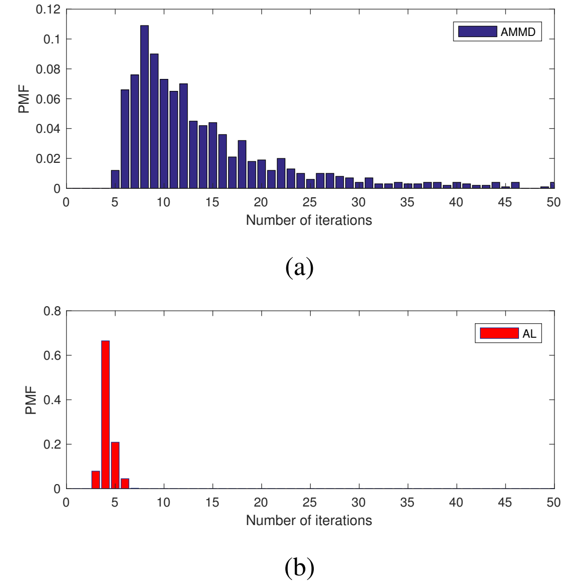

Fig. 1. Convergence performance for SM systems with $N_t = 8$, $N_r = 4$, and QPSK. (a) AMMD; (b) AL.

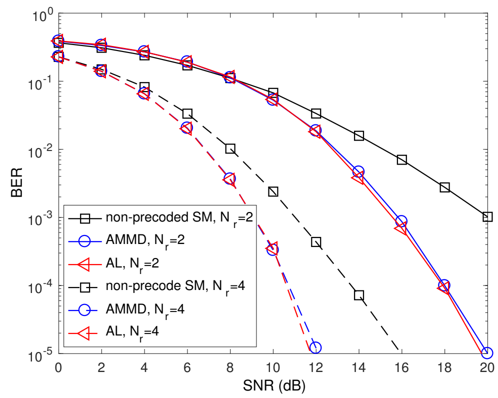

Fig. 2. Comparison of BER performance for different SM systems with QPSK and $N_t = 8$.

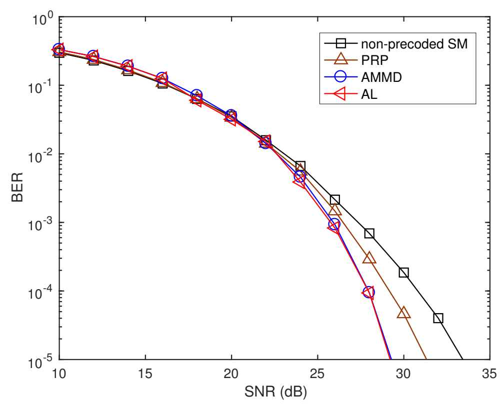

Fig. 3. Comparison of BER performance for different SM systems with 16-PSK and $N_t = 8$.

the simulation, the convergence threshold is $\epsilon = 10^{-5}$ for both methods, and 10^5 trials were conducted. By comparing Fig. 1(a) with Fig. 1(b), it is clear that AL converges much faster than AMMD. This is consistent with our previous discussion on the convergence rate of AMMD, based on solving (9) for many times. In contrast, the convergence rate of the proposed method, based on solving (16). On the other hand, the complexity of the primal-dual interior point algorithm to solve (9) each time is $\mathcal{O}\left(N_t^2 N_r\right) + \mathcal{O}\left(N_t^4\right)$. By contrast, the complexity of the BFGS algorithm [11] to solve (16) each time is only $\mathcal{O}\left(4N_t^2\right)$. Therefore, the proposed method features extremely fast convergence rate with very low computational complexity compared to AMMD.

The BER performance comparison between the AMMD method and the proposed method with QPSK is presented in Fig. 2, where $N_t = 8$, $N_r = 2$ (solid lines), and $N_r = 4$ (dashed lines). It is clearly shown that the proposed method achieves much better error performance than the non-precoded SM. For example, AL is 4 dB superior to the non-precoded SM with $N_r = 2$ at BER=10^{-3} and $N_r = 4$ at BER=10^{-5}, respectively. Note that the proposed method achieves almost the same error performance with AMMD for both $N_r = 2$ and $N_r = 4$.

Simulation was also run for increased modulation level (16-PSK), and the BER performance is presented for AMMD, PRP, and the proposed method in Fig. 3, where $N_t = 8$, $N_r = 2$. Note that 16-PSK is used here because it was found in [12] that SM associated with constant-envelope modulation provides better performance than amplitude modulation schemes. It can be observed that, similar to Fig. 2, the relative superiority of AMMD and AL also holds.

From the simulation results in Figs. 1-3, we can conclude that the proposed method maintains the same favorable error performance as the AMMD method, but reduces the complexity by orders of magnitude. Therefore, the proposed method is highly promising for practical implementations.

VI. Conclusion

In this paper, we proposed a novel precoding method for SM systems based on the MMD criterion. We first prove that the precoder design is a large-scale non-convex QCQP problem. We then propose to address this QCQP problem by leveraging augmented Lagrangian and dual ascent techniques. Numerical results show that our proposed method can significantly improve the error performance of SM, and features extremely fast convergence rate with very low computational complexity.

References

[1] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 6–26, May 2014.

[2] M. Di Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple-antenna wireless systems: a survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, Dec. 2011.

[3] A. Younis, S. Sinanovic, M. Di Renzo, R. Mesleh, and H. Haas, "Generalised sphere decoding for spatial modulation," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2805–2815, Jul. 2013.

[4] Y. Xiao, Z. Yang, L. Dan, P. Yang, L. Yin, and W. Xiang, "Low-complexity signal detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 403–406, Mar. 2014.

[5] C. Wang, P. Cheng, Z. Chen, J. A. Zhang, Y. Xiao, and L. Gui, "Near-ML low-complexity detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 618–621, Mar. 2016.

[6] P. Yang, Y. Xiao, Y. Yu, and S. Li, "Adaptive spatial modulation for wireless mimo transmission systems," *IEEE Commun. Lett.*, vol. 15, no. 6, pp. 602–604, Jun. 2011.

[7] K. Ntontin, M. D. Renzo, A. I. Perez-Neira, and C. Verikoukis, "A low-complexity method for antenna selection in spatial modulation systems," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2312–2315, Dec. 2013.

[8] C. Masouros, "Improving the diversity of spatial modulation in MISO channels by phase alignment," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 729–732, May 2014.

[9] A. Garcia-Rodriguez, C. Masouros, and L. Hanzo, "Pre-scaling optimization for space shift keying based on semidefinite relaxation," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4231–4243, Nov. 2015.

[10] M. C. Lee, W. H. Chung, and T. S. Lee, "Generalized precoder design formulation and iterative algorithm for spatial modulation in MIMO systems with CSIT," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1230–1244, Apr. 2015.

[11] A. Antoniou and W.-S. Lu, *Practical optimization: algorithms and engineering applications*. Springer Science & Business Media, 2007.

[12] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.

[13] X. Zhang, *Matrix analysis and applications*. Tsinghua and Springer Publishing house, Beijing, 2004.

[14] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.