

“© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# A Unified Precoding Scheme for Generalized Spatial Modulation

Peng Cheng, Zhuo Chen, J. Andrew Zhang, Yonghui Li, and Branka Vucetic

**Abstract**—Generalized spatial modulation (GSM) activates  $n_t$  out of  $N_t$  ( $1 \leq n_t < N_t$ ) available transmit antennas, and information is conveyed through  $n_t$  modulated symbols as well as the index of the  $n_t$  activated antennas. GSM strikes an attractive tradeoff between spectrum efficiency and energy efficiency. Linear precoding that exploits channel state information at the transmitter enhances the system error performance. For GSM with  $n_t = 1$  (the traditional SM), the existing precoding methods suffer from high computational complexity. On the other hand, GSM precoding for  $n_t \geq 2$  is not thoroughly investigated in the open literature. In this paper, we develop a unified precoding design for GSM systems, which universally works for all  $n_t$  values. Based on the maximum minimum Euclidean distance criterion, we find that the precoding design can be formulated as a large-scale non-convex quadratically constrained quadratic program (QCQP) problem. Then we transform this challenging problem into a sequence of unconstrained subproblems by leveraging augmented Lagrangian and dual ascent techniques. These subproblems can be solved in an iterative manner efficiently. Numerical results show that the proposed method can substantially improve the system error performance relative to the GSM without precoding, and features extremely fast convergence rate with a very low computational complexity.

**Index Terms**—Spatial modulation, generalized spatial modulation, linear precoding, maximum minimum Euclidean distance, augmented Lagrangian.

## I. INTRODUCTION

Generalized spatial modulation (GSM) [1]–[6] has recently emerged as a novel low-complexity multiple-input-multiple-output (MIMO) scheme to achieve an attractive tradeoff between spectrum efficiency and energy efficiency. The philosophy behind a GSM system with  $N_t$  transmit antennas is that, at each time, only  $n_t$  ( $1 \leq n_t < N_t$ ) antennas are activated for transmission. As such, a GSM system conveys information through both the original signal constellation and the spatial constellation [5] (antenna index space). The GSM with  $n_t = 1$  [1]–[3] is usually referred to as spatial modulation (SM) in the literature. SM eliminates inter-channel interference (ICI), rules out the need for inter-antenna synchronization, and only requires a single radio frequency (RF) chain [7]. Note that space shift keying (SSK) [7] can be considered as a special case of SM without signal constellation. In contrast, GSM with

$n_t \geq 2$  [4]–[6] introduces spatial multiplexing by activating more than one antenna at each time slot, and the spectrum efficiency is further enhanced.

To exploit the benefits of GSM, the maximum likelihood (ML) detection algorithm, which jointly searches all the possible transmit antenna combinations (TACs) and the modulated symbols, achieves the optimal error performance. This is usually at the cost of prohibitively high computational complexity. For practical implementation, many low-complexity sub-optimal detectors such as sphere decoding [8], the ordered-blocked minimum-mean-squared-error [9], Gaussian approximation [10], and the enhanced Bayesian compressive sensing [11] have been designed. Generally, these algorithms represent different levels of tradeoff between error performance and computational complexity. In addition to the processing at the receiver side, preprocessing at the transmitter to improve the performance has been considered in many papers, mainly for SM systems ( $n_t = 1$ ). For example, link adaptations schemes [12]–[14] were proposed, where transmit parameters are dynamically adapted to the changing channel conditions. Another approach is to develop antenna selection methods [15]–[17] to introduce transmit diversity in SM. Besides, the power allocation between the pilot and data was optimized in [18] for maximizing the capacity of SM systems, and the error performance of SM was further optimized in [19] by finding the optimal combination of the number of available transmit antennas and the signal constellation size.

With channel state information (CSI) available at the transmitter, linear precoding potentially provides another systematic avenue to improving GSM error performance. In the literature, precoding efforts are mainly focused on SM ( $n_t = 1$ ). Conceptually speaking, the precoding design for SM is fundamentally different from conventional MIMO precoding [20], [21]. SM precoding aims to maximize the separation of the effective channel vectors for different active antennas, while conventional MIMO precoding targets the alignment of the transmitted signal with the channel matrix. Under the umbrella of the maximum minimum Euclidean distance (MMD) criterion [5], various linear diagonal transmit precoding methods have been reported. In [22], the phase alignment technique was applied to SM systems to improve the transmit diversity, but it is limited to multiple-input single-output (MISO) channels. The pre-scaling optimization based on semidefinite relaxation (SDR) was proposed in [23]; however, it is only applicable to SSK systems. A closed-form solution was derived for SM systems only in the case of two transmit antennas [24]. By quantizing the amplitude and phase of the precoding weights, the transmit precoding (TPC) design in [2] enables a brute force search

P. Cheng, Y. Li, and B. Vucetic are with School of Electrical and Information Engineering, The University of Sydney, Australia (e-mail: peng.cheng@sydney.edu.au; yonghui.li@sydney.edu.au; branka.vucetic@sydney.edu.au). Z. Chen is with CSIRO DATA61, Australia (e-mail: zhuo.chen@csiro.au). J. A. Zhang is with School of Computing and Communications, University of Technology Sydney, Australia (email: andrew.zhang@uts.edu.au).

The work of Y. Li was supported by ARC under Grant DP150104019 and NSFC under Grant 61531006 and Grant 61772233.

to optimize the complex precoding weights. To improve the computational efficiency, the real and imaginary parts of the precoding weights are separately optimized in [25] and [26], respectively. An obvious disadvantage associated with the MMD criterion lies in its non-convex nature. A recent paper [27] relaxes such a non-convex MMD problem into a sequence of convex subproblems, which can be solved by various convex optimization methods [28]. The precoding design in [27] can achieve a significantly improved error performance; however, it needs to solve multiple convex optimization subproblems before reaching convergence, and the convergence rate is found by simulations to be very slow, resulting in a very high computational complexity and processing delay.

In contrast to the abundant efforts in SM precoding, until now little work is known on linear precoding for GSM with  $n_t \geq 2$ , other than [27] which extends the linear diagonal precoding method for SM to a simplified version of GSM, where multiple active antennas transmit the same symbol without spatial multiplexing to avoid ICI. To the best of our knowledge, the precoding design for GSM with  $n_t \geq 2$  remains an open challenge, including design criterion and precoding process.

In summary, the existing GSM precoding design suffers from high computational complexity for  $n_t = 1$  and no known solutions exists for  $n_t \geq 2$ . This serves as the original motivation for our work in this paper, where we take a holistic view on linear GSM precoding, and develop a unified precoding scheme over flat Rayleigh fading channels. Against this background, our novel contributions can be summarized as follows.

- 1) To establish the mathematical tools for the unified precoding design, we first consider the case of SM ( $n_t = 1$ ). A diagonal precoding structure is used to maintain the single antenna activation. By transforming the original precoding design criterion based on MMD, we formulate the precoder design as a large-scale non-convex quadratically constrained quadratic program (QCQP) problem, which is highly challenging to solve. By leveraging augmented Lagrangian and dual ascent techniques [29], we transform this non-convex QCQP problem into a sequence of unconstrained subproblems in an iterative manner, while each subproblem can be efficiently solved by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [30]. This developed method is capable of handling a large number of inequality constraints in the non-convex QCQP problem.
- 2) As a generalization of the design philosophy for  $n_t = 1$ , we consider GSM with  $n_t \geq 2$ , and design a full square precoding matrix set for all the TACs. Spatial multiplexing is assumed across  $n_t$  antennas. We find that the precoding design for  $n_t \geq 2$  could also be formulated as a large-scale QCQP problem, which can be efficiently addressed by the proposed method for SM. Consequently, we jointly optimize all the precoding weights in a very efficient way.
- 3) We further prove that our proposed precoding designs for both cases aforementioned follow the same design approach and share the same optimization solution to large-

scale QCQP problems. Therefore, the proposed precoding design provides a unified solution to GSM with  $1 \leq n_t < N_t$ . This precoding scheme applies to any number of activated antennas  $n_t$ , and features low computational complexity. Numerical results show that our methods can substantially improve the error performance of GSM with  $1 \leq n_t < N_t$ , and have extremely fast convergence rate with a very low computational complexity.

In this paper, we limit our discussions to flat fading channels. However, with the well-known techniques such as orthogonal frequency-division multiplexing (OFDM) or single-carrier transmissions [31], the proposed precoding matrix design method could be easily extended to frequency-selective fading channels. Moreover, our proposed scheme could also find other potential applications such as massive MIMO and millimeter wave transmission [32].

The rest of this paper is organized as follows. Section II presents the system model of a SM system. Section III describes the precoding design formulation for SM. Section IV describes the proposed precoding method for SM systems. Section V describes the proposed precoding methods for GSM systems. In Section VI, extensive simulation results are provided to validate the benefits of our methods. Finally, conclusions are drawn in Section VII.

*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  $\mathbf{a}$ .  $\text{Tr}(\mathbf{A})$  denotes the trace of a matrix  $\mathbf{A}$ .  $\otimes$  denotes the Kronecker product.  $[\mathbf{A}]_{k,n}$  represents the element in the  $k$ -th row and  $n$ -th column of a matrix  $\mathbf{A}$ .  $[x]^+$  represents  $\max(x, 0)$ .  $\text{Re}\{\mathbf{x}\}$  and  $\text{Im}\{\mathbf{x}\}$  represent the real and imaginary parts of  $\mathbf{x}$ , respectively.  $\nabla$  denotes the gradient of a function.  $\mathbf{\Lambda} = \text{diag}(\mathbf{a})$  changes a vector  $\mathbf{a}$  into a diagonal matrix  $\mathbf{\Lambda}$ .  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the complex vector normal distribution with a mean vector  $\boldsymbol{\mu}$  and a covariance matrix  $\boldsymbol{\Sigma}$ .

## II. PRELIMINARY

### A. System Model

We commence by introducing the SM concept, and then generalize it to the GSM scheme in Section V. Consider a MIMO system with  $N_t$  transmit and  $N_r$  receive antennas shown in Fig. 1. For ease of exposition,  $N_t = 2^n$  is assumed with  $n$  a positive integer. In contrast to the conventional spatial multiplexing, the fundamental operational principle of SM is to use both the spatial and signal constellations to convey information. At each time slot, the first  $b_1 = \log_2 N_t$  bits are mapped to a spatial constellation point drawn from the set with the cardinality  $N_t$

$$\mathcal{S}_{\text{spatial}} = \{\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_i, \dots, \mathbf{e}_{N_t-1}\}, \quad (1)$$

where  $\mathbf{e}_i \in \mathbb{R}^{N_t}$ ,  $i = 0, \dots, N_t - 1$ , is the  $i$ -th column of  $\mathbf{I}_{N_t}$ . In other words, only a single transmit antenna with index  $i$  is activated at each time slot. Then, the last  $b_2 = \log_2 M$  bits are mapped to a signal constellation point drawn from the set with the cardinality  $M$

$$\mathcal{S}_{\text{signal}} = \{s_0, s_1, \dots, s_m, \dots, s_{M-1}\}, \quad (2)$$

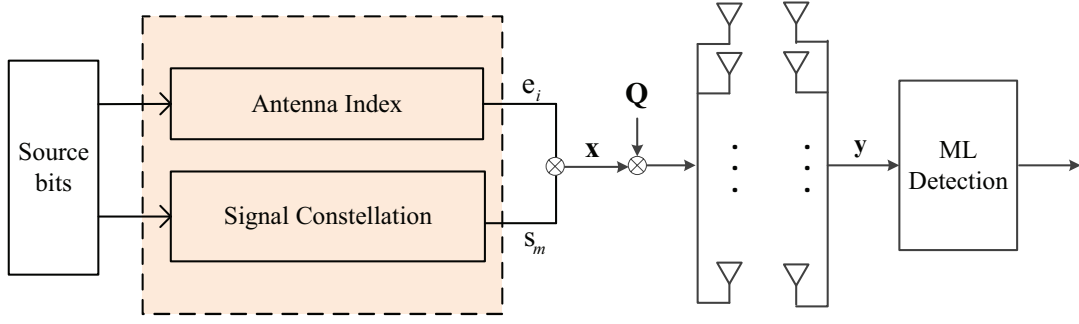


Fig. 1. The linear precoded SM transmission system based on ML detection.

where  $s_m \in \mathbb{C}$ ,  $m = 0, \dots, M-1$ , is a power normalized  $M$ -level modulated 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, we have

$$\mathcal{S} = \{s_0 \mathbf{e}_0, s_1 \mathbf{e}_0, \dots, s_{M-1} \mathbf{e}_0, \dots, s_{M-1} \mathbf{e}_{N_t-1}\}. \quad (3)$$

Then the SM symbol  $\mathbf{x} = s_m \mathbf{e}_i \in \mathbb{C}^{N_t}$  is precoded by a diagonal matrix  $\mathbf{Q} = \text{diag}(\mathbf{q})$  with precoding weights  $\mathbf{q} = [q_0, \dots, q_{N_t-1}]^T \in \mathbb{C}^{N_t}$ . Note that the precoding weights are independent of signal constellation. Different from the conventional MIMO precoding, SM precoding needs a diagonal precoding matrix due to the single antenna activation. Clearly, there are  $N_t$  unknown precoding weights to be optimized to enhance the error performance of SM.

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

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\mathbf{Q}\mathbf{x} + \mathbf{w} \\ &= q_i \mathbf{h}_i s_m + \mathbf{w}, \end{aligned} \quad (4)$$

where  $\mathbf{H} = [\mathbf{h}_0, \dots, \mathbf{h}_i, \dots, \mathbf{h}_{N_t-1}] \in \mathbb{C}^{N_r \times N_t}$  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

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{y} - \mathbf{H}\mathbf{Q}\mathbf{x}\|^2. \quad (5)$$

In other words, the ML detector jointly decodes an SM symbol, including both the transmit antenna index and the transmitted symbol. Based on the ML detector, the error performance for a given channel  $\mathbf{H}$  can be approximated by the sum of the pairwise error probability given by [2]

$$P_e \leq P_e^o = \sum_{i=0}^{N_t M-1} \sum_{j=0, j \neq i}^{N_t M-1} Q \left( \sqrt{\frac{1}{2\sigma^2} d_{i,j}(\mathbf{q})} \right), \quad (6)$$

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

symbols  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . At high signal-to-noise ratios (SNRs),  $P_e^o$  in (6) can be simplified as

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

where  $\lambda$  is the number of neighbor points [5], and  $d_{\min} = \min_{i,j, i \neq j} d_{i,j}(\mathbf{q})$  is the minimum squared Euclidian distance among the codebook  $\mathcal{S}$ .

### III. PRECODING DESIGN FORMULATION FOR SM

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

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

where  $P_t$  is the total power constraint at the transmitter.

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

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

where  $\Delta \mathbf{X}_{i,j} \triangleq (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^H$ ,  $\mathbf{R}_{i,j} = \mathbf{H}^H \mathbf{H} \odot \Delta \mathbf{X}_{i,j}^T$ , step (a) uses the trace property  $\text{Tr}(\mathbf{A}\mathbf{B}) = \text{Tr}(\mathbf{B}\mathbf{A})$  for matrices  $\mathbf{A}$  and  $\mathbf{B}$ , and step (b) uses the matrix rule in [33, Eq. (1.10.6)] where  $\odot$  denotes the Hadamard product. As both  $\mathbf{H}^H \mathbf{H}$  and  $\Delta \mathbf{X}_{i,j}$  are positive semidefinite, based on [33, Page 104, Property 9], we find that  $\mathbf{R}_{i,j} = \mathbf{H}^H \mathbf{H} \odot \Delta \mathbf{X}_{i,j}^T$  is positive semidefinite. At this stage, (P0) can be rewritten as

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

*Remark 1:* Note that the precoding optimization problem in [27] is formulated to consider both signal and spatial

constellations. In contrast, our formulation, like most of the papers on SM precoding such as [2], [25], [26], only considers the spatial constellation. This means that the problem in [27] has the same optimization structure as (P0) (both take the form of quadratic functions), and the only difference lies in the dimension of the precoding weights to be optimized ( $MN_t$  in [27] and  $N_t$  in our case). For brevity, in the following we use (P0) to represent the original MMD problem.

By introducing an auxiliary variable  $t$ , we have the equivalent epigraph form [28] of (P0) as

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

However, (P1) is a non-convex problem<sup>1</sup>, and therefore it is extremely difficult to find the optimal solution to (P1). A convex relaxation method to approximate (P1) was proposed in [27]. The core idea of the convex relaxation is to linearize  $\mathbf{q}^H \mathbf{R}_{i,j} \mathbf{q}$  by writing  $\mathbf{q}^H \mathbf{R}_{i,j} \mathbf{q}$  as  $\text{Re}\{2\mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q} - \mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q}_k\}$  where  $\mathbf{q}_k$  represents the solution to the last iteration which will be explained later. The approximated MMD (AMMD) problem in [27] is given by

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & \text{Re}\{2\mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q} - \mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q}_k\} \geq t, \quad \forall i, j, i \neq j, \\ & \|\mathbf{q}\|^2 \leq P_t. \end{aligned} \quad (9)$$

As  $\text{Re}\{2\mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q} - \mathbf{q}_k^H \mathbf{R}_{i,j} \mathbf{q}_k\}$  is an affine function of  $\mathbf{q}$ , clearly (9) is a convex optimization problem that can be solved by the primal-dual interior point method [28]. Initializing a point  $\mathbf{q}_0$  and iteratively solving (9)  $K_{\text{AMMD}}$  iterations (times) until convergence, we can obtain a sequence of solutions  $\{\mathbf{q}_k\}_{k=1}^{K_{\text{AMMD}}}$ , and the last vector  $\mathbf{q}_{K_{\text{AMMD}}}$  serves as the approximated solution to (P1).

The AMMD precoding method is able to achieve a favorable error performance in the simulations; however, our investigations found that the main limitation of this method lies in its high computational complexity. The complexity of the primal-dual interior point method to solve (9) is  $\mathcal{O}(N_t^2 N_r) + \mathcal{O}(M^2 N_t^4)$ . On the other hand, we need to solve (9)  $K_{\text{AMMD}}$  iterations, and  $K_{\text{AMMD}}$  is found in simulations to be very large. Therefore, the combined complexity imposes a great burden in the real-time precoding design at the transmitter.

*Remark 2:* As discussed in *Remark 1*, the original AMMD precoding method [27] considers both signal and spatial constellations. In [27], there are  $N_t M$  unknown precoding weights to be optimized. Therefore, the complexity<sup>2</sup> to solve a similar form as in (9) at each iteration is  $\mathcal{O}(M^2 N_t^2 N_r) + \mathcal{O}(M^4 N_t^4)$ . Our numerical results (not shown here due to space limitation) show that the original AMMD has the similar error performance as that presented in (9). Therefore, the precoding problem considered in this paper does not need to consider the signal constellation. This also reduces the complexity involved in the design.

<sup>1</sup>If the constraints are  $\mathbf{q}^H \mathbf{R}_{i,j} \mathbf{q} \leq t$ , then the problem becomes convex.

<sup>2</sup>Note that the three-order items were omitted in [27].

Let us review (P1) in an equivalent way. By the transformation in Appendix A, we can show that (P1) is equivalent to the following problem

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

where  $d$  is the target distance. The rationale behind (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 (P2), we identify (P2) as a large-scale non-convex QCQP problem; it can be calculated that the number of the quadratic constraints in (P2) is  $N_t M \times (N_t M - 1)/2$ , which is very large for large  $N_t$  and  $M$ . For example, for a moderate SM system with  $N_t = 8$  with  $M = 16$ , there are 8128 quadratic constraints. Naturally, it is extremely difficult to obtain the globally optimal solution to such a problem. Note that for a small-scale non-convex QCQP problem, a popular method known as semidefinite relaxation (SDR) [34] has been well established. SDR approximates this QCQP problem as a convex semi-definite programming problem via 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 [28]. When  $\text{rank}(\hat{\mathbf{C}}) = 1$ , 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 or low rank solutions by SDR is extremely low, making the method not effective.

Consequently, to materialize the optimization problem in (P2), we will develop an effective and efficient solution to the large-scale QCQP problem (P2).

#### IV. THE PROPOSED PRECODING DESIGN FOR SM SYSTEMS

In this section, we develop a novel algorithm to address (P2) by leveraging augmented Lagrangian together with the dual ascent technique [29]. The quasi-Newton method BFGS [35] is incorporated in the proposed algorithm. We first review the basic principle of the dual ascent technique.

##### A. Dual Ascent

Consider an equality-constrained convex optimization problem

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

The Lagrangian corresponding to (10) can be written as

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

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

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

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

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

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  $\mathbf{x}^*$  from a dual optimal point  $\boldsymbol{\lambda}^*$  as

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

provided there is only one minimizer of  $L(\mathbf{x}, \boldsymbol{\lambda}^*)$ . At this stage, the dual ascent method consists of primal and dual vectors, which are iteratively updated as below

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}_k) \quad (15a)$$

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \alpha_k (\mathbf{A}\mathbf{x}_{k+1} - \mathbf{b}), \quad (15b)$$

where  $\alpha_k > 0$  is a step size, and the subscript 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.,  $g(\boldsymbol{\lambda}_{k+1}) > g(\boldsymbol{\lambda}_k)$ .

### B. Augmented Lagrangian Precoding Method

We now develop a new dual ascent method in terms of augmented Lagrangian (AL) 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 (P2). For convenience, we first convert (P2) to a real-valued form; this yields a real vector  $\mathbf{m} = [\text{Re}\{\mathbf{q}\}^T \text{Im}\{\mathbf{q}\}^T]^T \in \mathbb{R}^{2N_t \times 2N_t}$ , and a real matrix  $\mathbf{G}_{i,j}$

$$\mathbf{G}_{i,j} = \begin{bmatrix} \text{Re}\{\mathbf{R}_{i,j}\} & -\text{Im}\{\mathbf{R}_{i,j}\} \\ \text{Im}\{\mathbf{R}_{i,j}\} & \text{Re}\{\mathbf{R}_{i,j}\} \end{bmatrix}. \quad (16)$$

In this case, (P2) can be equivalently represented as

$$\begin{aligned} (\text{P3}) \quad & \min_{\mathbf{m} \in \mathbb{R}^{2N_t}} \mathbf{m}^T \mathbf{m} \\ \text{s.t.} \quad & \mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} \geq d, \forall i, j, i \neq j. \end{aligned}$$

The augmented Lagrangian for (P3) can be written as

$$\begin{aligned} L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}, \mu) &= \mathbf{m}^T \mathbf{m} - \sum_{i=0}^{N_t M-1} \sum_{j=0, j \neq i}^{N_t M-1} \lambda_{i,j} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - s_{i,j} - d) \\ &+ \frac{\mu}{2} \sum_{i=0}^{N_t M-1} \sum_{j=0, j \neq i}^{N_t M-1} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - s_{i,j} - d)^2, \end{aligned} \quad (17)$$

where  $\boldsymbol{\lambda} \triangleq \{\lambda_{i,j}\}$  is the dual vector,  $\sum_{i=1}^{N_t M} \sum_{j=1, j \neq i}^{N_t M} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - s_{i,j} - d)^2$  is the penalty item, and  $\mu > 0$  is the penalty parameter. Note that we transform the inequality constraints  $\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} \geq d$  in (P3) to the equality constraints by introducing a slack vector  $\mathbf{s} \triangleq \{s_{i,j}\}$  and having  $\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} = d + s_{i,j}$  with  $s_{i,j} \geq 0$ .

Next, we elaborate on the details of updating the primal vector  $\mathbf{m}$ . Following a similar approach as 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  $k$ -th iteration, given by

$$\min_{\mathbf{m}, \mathbf{s}} L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}_k, \mu_k) \quad (18a)$$

$$\text{s.t. } \mathbf{s} \geq \mathbf{0}. \quad (18b)$$

With reference to (17), it is clear that for each iteration,  $s_{i,j}$  occurs in just two terms of (18a), which is in fact a convex quadratic function with respect to each of these slack variables. Therefore, the minimization process in (18a) can be carried out 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

$$s_{i,j} = \mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d - \frac{\lambda_{i,j,k}}{\mu_k}. \quad (19)$$

If this unconstrained minimizer is smaller than the lower bound of 0, then since (18a) is convex in  $s_{i,j}$ , the optimal value of  $s_{i,j}$  in (18a) is 0. The optimal value of  $s_{i,j}$  is therefore given by

$$s_{i,j} = \left[ \mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d - \frac{\lambda_{i,j,k}}{\mu_k}, 0 \right]^+. \quad (20)$$

Substituting (20) into  $L(\mathbf{m}, \mathbf{s}, \boldsymbol{\lambda}_k, \mu_k)$ , we can obtain an equivalent form  $L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  without  $\mathbf{s}$ . The items in (17) can be written by (21) at the bottom of the page. Consequently,  $L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  can be written in a compact form as

$$\begin{aligned} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k) &= \mathbf{m}^T \mathbf{m} + \sum_{i=0}^{N_t M-1} \sum_{j=0, j \neq i}^{N_t M-1} \phi(\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d, \lambda_{i,j,k}, \mu_k), \end{aligned} \quad (22)$$

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

$$\phi(z, a, b) \triangleq \begin{cases} -az + \frac{b}{2} z^2, & z - \frac{a}{b} \leq 0 \\ -\frac{1}{2b} a^2, & \text{otherwise.} \end{cases} \quad (23)$$

With the constraints removed, to minimize the augmented Lagrangian  $L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  in (22), we take advantage of the BFGS algorithm [35], one type of quasi-Newton method. The main advantage of BFGS is that it only requires the computation of the gradient of  $L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  with respect to  $\mathbf{m}$ ,  $\nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$ . In contrast, the Newton method requires the calculation of the inverse of Hessian matrix of second derivatives, which usually incurs prohibitively high computational complexity. In the context of BFGS,  $\nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  can be calculated as

$$\begin{aligned} \nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k) &= 2\mathbf{m} - \sum_{i=0}^{N_t M-1} \sum_{j=0, j \neq i}^{N_t M-1} \psi(\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d, \lambda_{i,j,k}, \mu_k), \end{aligned} \quad (24)$$

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

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

and  $\nabla_{\mathbf{m}} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d) = 2\mathbf{G}_{i,j} \mathbf{m}$ . Now with  $\nabla_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  in (24), the BFGS algorithm to minimize

$L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  works as follows. The iteration process can be shown as

$$\mathbf{m}_{l+1} = \mathbf{m}_l - \alpha_l \mathbf{S}_l \mathbf{g}_l, \quad (26)$$

where  $\mathbf{S}_l$  is an approximation to the inverse Hessian matrix  $[\nabla^2 L(\mathbf{m}_l, \boldsymbol{\lambda}_k, \mu_k)]^{-1}$ ,  $\mathbf{g}_l = \nabla_{\mathbf{m}} L(\mathbf{m}_l, \boldsymbol{\lambda}_k, \mu_k)$ , and  $\alpha_k$  is a step size. Defining the vectors

$$\boldsymbol{\delta}_l = \mathbf{m}_{l+1} - \mathbf{m}_l, \quad \boldsymbol{\gamma}_l = \mathbf{g}_{l+1} - \mathbf{g}_l, \quad (27)$$

the BFGS algorithm requires the approximation matrix  $\mathbf{S}_{l+1}$  to satisfy

$$\mathbf{S}_{l+1} \boldsymbol{\gamma}_l = \boldsymbol{\delta}_l, \quad (28)$$

and the matrix  $\mathbf{S}_{l+1}$  is updated by

$$\mathbf{S}_{l+1} = \mathbf{S}_l + \left(1 + \frac{\boldsymbol{\gamma}_l^T \mathbf{S}_l \boldsymbol{\gamma}_l}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l}\right) \frac{\boldsymbol{\delta}_l \boldsymbol{\delta}_l^T}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l} - \frac{\boldsymbol{\delta}_l \boldsymbol{\gamma}_l^T \mathbf{S}_l + \mathbf{S}_l \boldsymbol{\gamma}_l \boldsymbol{\delta}_l^T}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l}. \quad (29)$$

The detailed procedure for BFGS is summarized in **Algorithm 2**.

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

$$\begin{aligned} \lambda_{i,j,k+1} &= [\lambda_{i,j,k} - \mu_k (\mathbf{m}_{k+1}^T \mathbf{G}_{i,j} \mathbf{m}_{k+1} - d)]^+ \quad (30) \\ \mu_{k+1} &= \rho \mu_k. \quad (31) \end{aligned}$$

where the parameter  $\rho > 0$ . The procedure for the proposed AL precoding method<sup>3</sup> is summarized in **Algorithm 1**.

The property of the dual ascent technique [29] guarantees that  $L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k)$  increases at each iteration until convergence. Due to the non-convex nature of the MMD criterion and the scale of inequity constraints, by theory only local optimum can be guaranteed, and the gap between local and global optimums is unknown. However, in practice, we tested a large number of different initial points in simulations, and we found that the solutions achieve similar values of  $d_{\min}$  and BER performance, which indicates that the proposed method is not sensitive to the initial points. The performance and convergence of the proposed method will be elaborated in Section VI.

## V. PRECODING DESIGN FOR GSM SYSTEMS WITH $n_t \geq 2$

In this section, we investigate the precoding design problem for GSM systems with  $n_t \geq 2$ .  $N = 2^{\lfloor \log_2 \binom{N_t}{n_t} \rfloor}$  TACs can implicitly convey  $b_1 = \log_2 N$  bits. Let  $L_n = (l_n^{(0)}, l_n^{(1)}, \dots, l_n^{(n_t-1)})$  be the indicator for the  $n$ -th TAC ( $n = 0, \dots, N-1$ ), where  $l_n^{(a)}$  is the index of the  $a$ -th

<sup>3</sup>Note that the result  $\mathbf{q}$  from **Algorithm 1** will be scaled to meet the total power constraint at the transmitter.

---

### Algorithm 1 Proposed Precoding Algorithm Starting from (P3)

---

**Initialization:** Initialize  $\mathbf{1}_{2N_t} \leftarrow \mathbf{m}_0, \forall i, j, i \neq j, 0.5 \leftarrow \lambda_{i,j}, 10 \leftarrow \mu_0, 2 \leftarrow \rho$ , and  $0 \leftarrow k$ . **Repeat**

1) Update the primal vector  $\mathbf{m}_{k+1}$

$$\mathbf{m}_{k+1} = \arg \min_{\mathbf{m}} L(\mathbf{m}, \boldsymbol{\lambda}_k, \mu_k). \quad (32)$$

2) Update the dual vector  $\boldsymbol{\lambda}_{k+1}$

$$\lambda_{i,j,k+1} = [\lambda_{i,j,k} - \mu_k (\mathbf{m}_{k+1}^T \mathbf{G}_{i,j} \mathbf{m}_{k+1} - d)]^+. \quad (33)$$

3) Update the penalty parameter  $\mu_{k+1}$

$$\mu_{k+1} = \rho \mu_k. \quad (34)$$

4) Set  $k \leftarrow k + 1$ .

**Until** convergence criterion is met.

---



---

### Algorithm 2 The BFGS Algorithm for Solving (32)

---

**Initialization:** Initialize  $\mathbf{I}_{2N_t} \leftarrow \mathbf{S}_0$  and  $0 \leftarrow l$ . Define  $\mathbf{g}_l = \nabla_{\mathbf{m}} L(\mathbf{m}_l, \boldsymbol{\lambda}_k, \mu_k)$ . **Repeat**

1) Compute search direction  $\mathbf{d}_l$

$$\mathbf{d}_l = -\mathbf{S}_l \mathbf{g}_l \quad (35)$$

2) Update  $\mathbf{m}_{l+1}$

$$\mathbf{m}_{l+1} = \mathbf{m}_l + \alpha_k \mathbf{d}_k, \quad (36)$$

where  $\alpha_k$  is computed by a line search procedure to satisfy the Wolfe conditions [30].

3) Define  $\boldsymbol{\delta}_l = \mathbf{m}_{l+1} - \mathbf{m}_l$  and  $\boldsymbol{\gamma}_l = \mathbf{g}_{l+1} - \mathbf{g}_l$ .

4) Update  $\mathbf{S}_{l+1}$  by

$$\mathbf{S}_{l+1} = \mathbf{S}_l + \left(1 + \frac{\boldsymbol{\gamma}_l^T \mathbf{S}_l \boldsymbol{\gamma}_l}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l}\right) \frac{\boldsymbol{\delta}_l \boldsymbol{\delta}_l^T}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l} - \frac{\boldsymbol{\delta}_l \boldsymbol{\gamma}_l^T \mathbf{S}_l + \mathbf{S}_l \boldsymbol{\gamma}_l \boldsymbol{\delta}_l^T}{\boldsymbol{\gamma}_l^T \boldsymbol{\delta}_l}. \quad (37)$$

5) Set  $l \leftarrow l + 1$ .

**Until** convergence criterion is met.

---

active antenna ( $a = 0, \dots, n_t - 1$ ). Explicitly,  $n_t$  active antennas convey  $n_t$   $M$ -ary modulated symbols, or equivalently,  $b_2 = n_t \log_2 M$  bits. The transmit vector  $\mathbf{x}$  is expressed as  $\mathbf{x} = [\dots, 0, s_1, 0, \dots, 0, s_2, 0, \dots, 0, s_{n_t}, 0, \dots]^T \in \mathbb{C}^{N_t \times 1}$ , where  $\mathbf{x}$  has exactly  $n_t$  nonzero elements and each  $s_i$  ( $i = 1, \dots, n_t$ ) is drawn from the signal constellation set  $\mathcal{S}_{\text{signal}}$ . Therefore, an overall  $b_1 + b_2$  bits are transmitted per symbol period. Next, we propose two different precoding designs based on the MMD criterion.

---


$$\begin{aligned} & -\lambda_{i,j,k} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - s_{i,j} - d) + \frac{\mu_k}{2} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - s_{i,j} - d)^2 \\ & = \begin{cases} -\lambda_{i,j,k} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d) + \frac{\mu_k}{2} (\mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d)^2, & \mathbf{m}^T \mathbf{G}_{i,j} \mathbf{m} - d - \frac{\lambda_{i,j,k}}{\mu_k} \leq 0 \\ -\frac{1}{2\mu_k} \lambda_{i,j,k}^2, & \text{otherwise.} \end{cases} \end{aligned} \quad (21)$$

### A. Diagonal Precoding Design

The first design is to, as what we have done for SM, use a diagonal matrix  $\mathbf{Q}_n = \text{diag}(0, q_n^{(0)}, \dots, 0, q_n^{(a-1)}, \dots, 0)$  as a precoding matrix. Collecting all the unknown precoding weights as a vector  $\mathbf{q} \in \mathbb{C}^{N n_t}$  given by

$$\mathbf{q} = [q_0^{(0)}, \dots, q_0^{(n_t-1)}, \dots, q_{N-1}^{(n_t-1)}]^T, \quad (38)$$

and following the precoding process in SM, we can derive a similar form as (P2), which can be addressed by our AL method in Section IV in a very efficient way.

### B. Full Precoding Design

The proposed second design has a different precoder structure from the first design. The multiplexing of signal constellation symbols on active antennas results in ICI among active antennas. Therefore, the precoding design does not need to preserve the ICI-free property like SM, and the diagonal structure might not provide the best solution. Instead, we consider a general case with a full square precoding matrix. Specifically, we have a precoding matrix set  $\mathcal{Q} = \{\mathbf{Q}_0, \dots, \mathbf{Q}_{N-1}\}$  with  $\mathbf{Q}_n \in \mathbb{C}^{n_t \times n_t}$  designed for the  $n$ -th TAC. In this case, the received signal can be shown by (39) at the bottom of the page, where  $\mathbf{H}_{L_n}$  represents the columns in  $\mathbf{H}$  corresponding to  $L_n$ , and  $\mathbf{x}_{L_n}$  represents the transmit symbols in the active antennas corresponding to  $L_n$ .

Considering that (39) has two summations, which are not easy to handle, we rebuild the channel matrix and the transmitted signal in a new form. We first construct a new channel matrix as

$$\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_{L_0}, \dots, \tilde{\mathbf{H}}_{L_n}, \dots, \tilde{\mathbf{H}}_{L_{N-1}}] \in \mathbb{C}^{N_r \times n_t^2 N} \quad (40)$$

in which

$$\begin{aligned} \tilde{\mathbf{H}}_{L_n} &= [\mathbf{H}_{(l_n^{(0)})}, \dots, \mathbf{H}_{(l_n^{(0)})}, \dots, \mathbf{H}_{(l_n^{(n_t-1)})}, \dots, \mathbf{H}_{(l_n^{(n_t-1)})}] \\ &= \mathbf{H}_{L_n} \otimes \mathbf{1}^{1 \times n_t}, \end{aligned} \quad (41)$$

where each  $\mathbf{H}_{(l_n^{(a)})}$  repeats  $n_t$  times. Meanwhile, the transmitted signal is rearranged as

$$\tilde{\mathbf{x}}_i = [\mathbf{0}, \dots, \tilde{\mathbf{x}}_{L_n}^T, \dots, \mathbf{0}]^T \in \mathbb{C}^{n_t^2 N}, \quad i = 0, \dots, M^{n_t} N, \quad (42)$$

where

$$\tilde{\mathbf{x}}_{L_n} = \underbrace{[\mathbf{x}_{L_n}, \dots, \mathbf{x}_{L_n}, \dots, \mathbf{x}_{L_n}]_{n_t}}^T. \quad (43)$$

In (43),  $\mathbf{x}_{L_n}$  also repeats  $n_t$  times. Besides, the precoding matrices in  $\mathcal{Q}$  are collected together as

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}'_0 & & & \\ & \ddots & & \\ & & \mathbf{Q}'_n & \\ & & & \ddots \\ & & & & \mathbf{Q}'_{N-1} \end{bmatrix}, \quad (44)$$

where  $\mathbf{Q}'_n = \text{diag}(\text{vec}(\mathbf{Q}_n^T)) = \text{diag}([p_n^{(0,0)}, p_n^{(0,1)}, \dots, p_n^{(n_t-1, n_t-1)}]^T)$ .

Following the procedure in (40)-(44), we can rewrite (39) by (45) at the bottom of the page. At this point, it is interesting to find that (45) has the same structure as (4). More specifically,  $\tilde{\mathbf{Q}}$  is a diagonal matrix with  $n_t^2 N$  unknown precoding weights in its diagonal, while  $\tilde{\mathbf{x}}_i$  only has  $n_t^2$  non-zero elements, all incorporated in  $\tilde{\mathbf{x}}_{L_n}$  corresponding to the  $n$ -th TAC. Therefore, the corresponding squared Euclidian distance  $d_{i,j}(\mathcal{Q})$  can be given by (c.f. (8))

$$\begin{aligned} d_{i,j}(\mathcal{Q}) &= \|\tilde{\mathbf{H}}\tilde{\mathbf{Q}}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)\|^2 \\ &= (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^H \tilde{\mathbf{Q}}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \tilde{\mathbf{Q}} (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j) \\ &= (\mathbf{q}')^H \mathbf{R}'_{i,j} \mathbf{q}', \end{aligned} \quad (46)$$

where  $\mathbf{R}'_{i,j} = \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \odot ((\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^H)^T$ , and  $\mathbf{q}' = \text{Diag}(\tilde{\mathbf{Q}}) \in \mathbb{C}^{n_t^2 N}$  includes all the precoding weights to be jointly optimized. Similarly, the precoding design problem can be formulated as

$$\begin{aligned} (\mathbf{P4}) \quad & \min_{\mathbf{q}' \in \mathbb{C}^{n_t^2 N}} \|\mathbf{q}'\|^2 \\ & \text{s.t.} \quad (\mathbf{q}')^H \mathbf{R}'_{i,j} \mathbf{q}' \geq d, \quad \forall i, j, i \neq j. \end{aligned}$$

Here, we can shed light on the fundamental properties of (P4).

- (P4) for  $n_t \geq 2$  and (P2) for  $n_t = 1$  share the similar structure. The only difference lies in the dimensions of the solution, with  $n_t^2 N$  precoding weights for  $n_t \geq 2$  systems, and  $N_t$  for  $n_t = 1$ . Both (P4) and (P2) can be classified as a large-scale non-convex QCQP problem.
- As (P4) is a large-scale non-convex QCQP problem, it can also be addressed by our AL method in Section IV.

---


$$\begin{aligned} \mathbf{y} &= \mathbf{H}_{L_n} \mathbf{Q}_n \mathbf{x}_{L_n} + \mathbf{w} \\ &= [\mathbf{H}_{(l_n^{(0)})}, \dots, \mathbf{H}_{(l_n^{(n_t-1)})}] \begin{bmatrix} q_n^{(0,0)} & \dots & q_n^{(0, n_t-1)} \\ \vdots & \ddots & \vdots \\ q_n^{(n_t-1, 0)} & \dots & q_n^{(n_t-1, n_t-1)} \end{bmatrix} \begin{bmatrix} x_{(l_n^{(0)})} \\ \vdots \\ x_{(l_n^{(n_t-1)})} \end{bmatrix} + \mathbf{w} \\ &= \sum_{a=0}^{n_t-1} \sum_{b=0}^{n_t-1} q_n^{(a,b)} \mathbf{H}_{(l_n^{(a)})} x_{(l_n^{(b)})} + \mathbf{w}, \end{aligned} \quad (39)$$



TABLE I  
A COMPARISON OF COMPUTATIONAL COMPLEXITY FOR  $N_t = 8$ ,  $N_r = 4$ , AND QPSK.

Precoding Schemes	Complexity Order	Operation Numbers
Diagonal TPC [2]	$\mathcal{O}(L_1 L_2 N_t^2 M^2 N_r)$	147456
AMMD [27]	$\mathcal{O}(K_{\text{AMMD}} M^2 N_t^2 N_r) + \mathcal{O}(K_{\text{AMMD}} M^4 N_t^4)$	15790080
AL	$\mathcal{O}(N_t^2 M^2) + \mathcal{O}(N_t^2 N_r) + \mathcal{O}(K_{\text{AL}} N_t^4 M^2) + \mathcal{O}(K_{\text{AL}} N_t^2 n_{\text{BFGS}})$	276224

- It is clear that the full square matrix  $\mathbf{Q}_n$  is developed for  $n_t \geq 2$ . However, if we let  $n_t = 1$ ,  $\mathbf{Q}_n$  reduces to a single scalar  $q_n$  while  $\tilde{\mathbf{H}}$  and  $\tilde{\mathbf{x}}_i$  in (45) become  $\mathbf{h}_n$  and  $s_m$  in (4), respectively. This clearly demonstrates that the approach we developed is a unified one, equally applicable to all  $1 \leq n_t < N_t$ , although the cases of  $n_t = 1$  and  $n_t \geq 2$  are treated separately in Sections IV and V, respectively.

In a nutshell, the proposed second precoding design provides a unified solution to all GSM systems.

## VI. NUMERICAL RESULTS AND DISCUSSION

### A. Computational Complexity

In this subsection, we first evaluate and compare the computational complexity of the proposed AL precoding method and the AMMD precoding method [27] for SM systems. As stated in *Remark 1*, for AMMD, the main complexity consist of two aspects: 1) the computational complexity of the convex solving process at each iteration (solving a similar form like (9)), and 2) the number of iterations required for the convergence. By contrast, the main complexity for AL lies in: 1) the complexity in Algorithm 1 at each iteration, and 2) the number of iterations required for the convergence.

According to [27], the complexity of the primal-dual interior point algorithm to solve a similar form like (9) at each iteration is  $\mathcal{O}(M^2 N_t^2 N_r) + \mathcal{O}(M^4 N_t^4)$ . By contrast, for Algorithm 1, each iteration incurs the complexity  $\mathcal{O}(N_t^4 M^2) + \mathcal{O}(N_t^2 n_{\text{BFGS}})$ , which counts for the complexity of the BFGS algorithm with  $n_{\text{BFGS}}$  being the number of iterations [35]. In addition, we should consider the problem formulation complexity for Algorithm 1, which comes from the calculation of the gradient in (24) and updating the dual vector. This problem formulation complexity is  $\mathcal{O}(N_t^2 M^2) + \mathcal{O}(N_t^2 N_r)$ , which is small relative to the complexity at each iteration.

It is not difficult to find that AL has lower computational complexity than AMMD at each iteration.

We then compare the convergence rates of the AL method and the AMMD method by showing the probability mass function (PMF) of the number of iterations. In Fig. 2, we consider the SM system with  $N_t = 8$ ,  $N_r = 4$ , and quadrature phase-shift keying (QPSK). In the simulation, the convergence threshold is  $\epsilon = 10^{-5}$  for both methods, and  $10^5$  simulation trials were conducted to calculate the PMFs. By comparing Fig. 2(a) with Fig. 2(b), it is clear that the AL method converges much faster than AMMD. More specifically, for AMMD more than 90% simulation trials require at least 30 iterations to converge. This is consistent with our previous discussion on the slow convergence rate of AMMD. The average number of iterations for AMMD in the simulation is  $K_{\text{AMMD}} = 15$ . In sharp contrast, for AL all the simulation trials converge within 7 iterations, and the average number of iterations is only  $K_{\text{AL}} = 4$ .

Following the discussion in the previous two paragraphs, we can conclude that the AL precoding method achieves much lower computational complexity than the AMMD method. In Table I, we summarize the complexity orders of three different linear precoding methods. Specifically, we consider the SM system with  $N_t = 8$ ,  $N_r = 4$ , and QPSK over uncorrelated Rayleigh fading channels. In the calculation of operation numbers, the quantization levels of amplitude and phase for the diagonal TPC [2] are  $L_1 = 6$  and  $L_2 = 6$ , respectively. In addition, consistent with the numerical results in Fig. 2, we set  $K_{\text{AMMD}} = 15$  for the AMMD method and  $K_{\text{AL}} = 4$  and  $n_{\text{BFGS}} = 50$  for the AL precoding method. It is clearly shown that compared to AMMD, the proposed method can reduce the complexity by orders of magnitude.

### B. Performance for GSM with $n_t = 1$ (SM)

In this subsection, we present numerical comparisons between our AL precoding method, the AMMD precoding

$$\begin{aligned}
 \mathbf{y} &= \mathbf{H}_{L_n} \mathbf{Q}_n \mathbf{x}_{L_n} + \mathbf{w} \\
 &= \begin{bmatrix} \tilde{\mathbf{H}}_{L_0}, \dots, \tilde{\mathbf{H}}_{L_n}, \dots, \tilde{\mathbf{H}}_{L_{N-1}} \end{bmatrix} \begin{bmatrix} \mathbf{Q}'_0 & & & \\ & \ddots & & \\ & & \mathbf{Q}'_n & \\ & & & \ddots \\ & & & & \mathbf{Q}'_{N-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \vdots \\ \tilde{\mathbf{x}}_{L_n} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \mathbf{w} \\
 &= \tilde{\mathbf{H}} \tilde{\mathbf{Q}} \tilde{\mathbf{x}}_i + \mathbf{w}.
 \end{aligned} \tag{45}$$

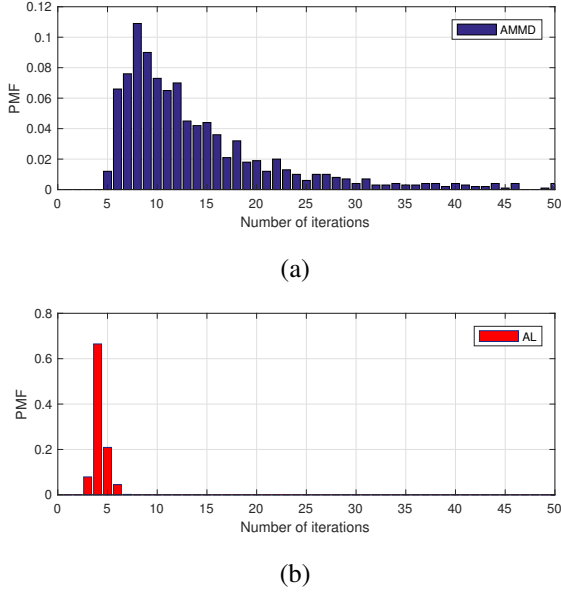


Fig. 2. Convergence performance for precoded SM systems with  $N_t = 8$ ,  $N_r = 4$ , and QPSK. (a) AMMD; (b) AL.

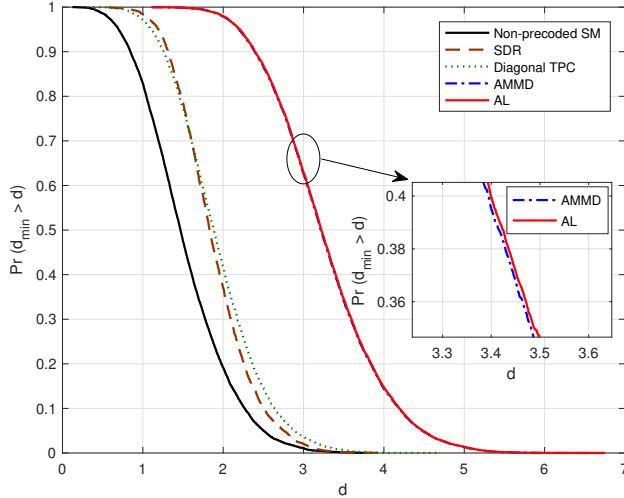


Fig. 3. The CCDF of the minimum distance  $d_{\min}$  for different precoding methods with QPSK and  $N_t = 8$ ,  $N_r = 4$ .

method [27], and the diagonal TPC method [2] over uncorrelated Rayleigh fading channels. The non-precoded SM is included for reference purpose. We also consider the SDR with randomization algorithm [34] to solve the large-scale non-convex QCQP problem in (P2). According to [27], the AMMD precoding can achieve the optimal BER performance among all the known linear precoding methods. The ideal CSI is assumed to be available at the transmitter. In the following figures, SNR is defined as the average received signal power over noise power at each receive antenna.

In Fig. 3, we first compare the complementary cumulative distribution function (CCDF) of the distance  $d_{\min}$  (c.f. (7)) for different linear precoding methods based on the MMD criterion. We consider the SM system with  $N_t = 8$ ,  $N_r = 4$ , and

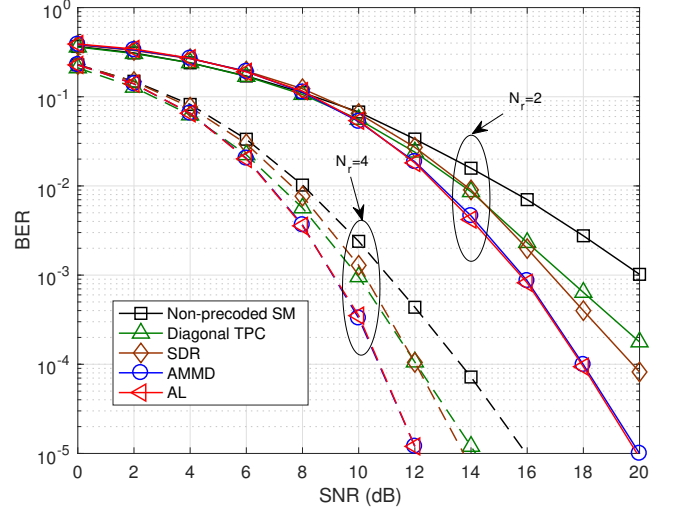


Fig. 4. BER comparison between different precoded SM systems with QPSK and  $N_t = 8$ .

QPSK. First, we can find that either the diagonal TPC method or the SDR with randomization method cannot dramatically improve  $d_{\min}$  relative to the original SM without precoding. The reasons behind this can be explained as follows. Diagonal TPC quantizes the amplitude and phase of precoding weights, which is a suboptimal solution. On the other hand, although SDR with randomization is a computationally efficient approximation approach to many non-convex QCQP problem with a small number of constraints [34], its performance will usually degrade with increasing number of constraints. By contrast, it is clear that both AMMD and our AL precoding method can significantly improve  $d_{\min}$  in comparison to non-precoded SM, and they achieve the same CCDF performance. The probability  $\Pr(d_{\min} > 2)$  for non-precoded SM, SDR, diagonal TPC, AMMD, and the AL precoding method is 0.17, 0.36, 0.4, 0.98, and 0.98, respectively. As  $d_{\min}$  dominates the error performance, it is expected that AMMD and AL can substantially improve the system error performance relative to the SM without precoding.

The BER performance comparison with QPSK is presented in Fig. 4, where  $N_t = 8$ ,  $N_r = 2$  (solid lines), and  $N_r = 4$  (dashed lines). Clearly, a better error performance can be obtained by increasing  $N_r$  from 2 to 4 for all the methods. Consistent with the CCDF of  $d_{\min}$  in Fig. 3, the AL method achieves a similar BER performance to AMMD, and significantly outperforms the other three methods. For example, at  $\text{BER}=10^{-5}$ , AL and AMMD achieve a 4 dB gain compared to the non-precoded SM, and a 2 dB gain compared to SDR and diagonal TPC.

Simulation was also performed for increased modulation level (16-PSK) in Fig. 5, where  $N_t = 8$ ,  $N_r = 4$ . Note that 16-PSK rather than 16-quadrature amplitude modulation (16-QAM) is used here because it was found in [5] that SM associated with constant-envelope modulation usually provides better performance than amplitude modulation schemes. It can be observed that, similar to Fig. 4, the relative superiority of

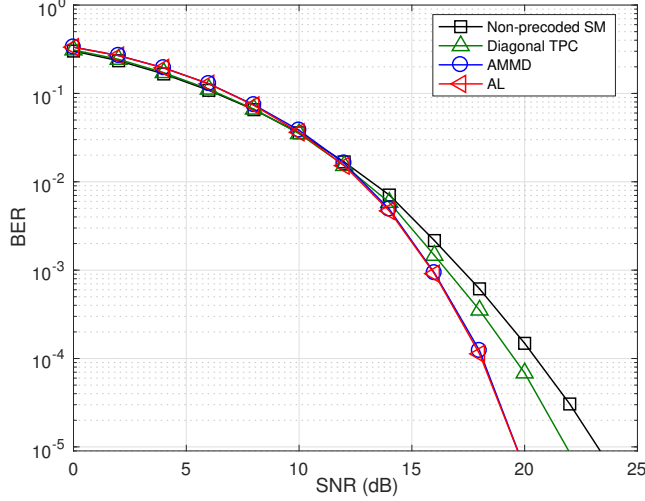


Fig. 5. BER comparison between different precoded SM systems with 16-PSK and  $N_t = 8, N_r = 4$ .

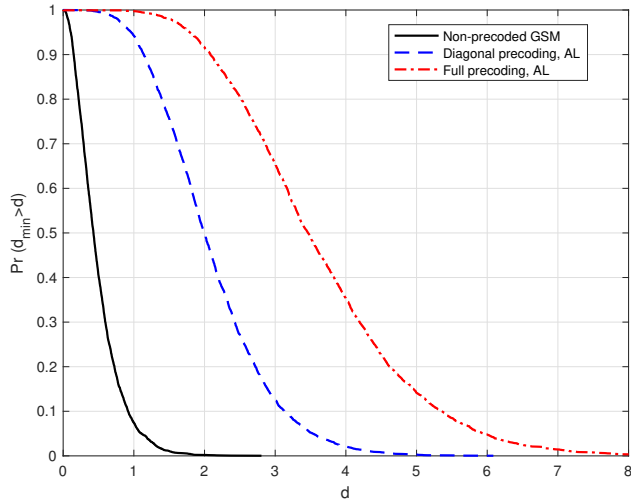


Fig. 6. The CCDF of the minimum distance  $d_{\min}$  for different precoding methods with BPSK and  $N_t = 5, N_r = 2, n_t = 2$ .

AL and AMMD also holds.

From the simulation results in Figs. 2-5 and Table I, we can conclude that both the AL and AMMD methods achieve a similar error performance, which is notably improved compared to that of the unprecoded SM. However, the AL method has a reduced complexity by orders of magnitude relative to AMMD.

### C. Performance for GSM with $n_t \geq 2$

In this subsection, the numerical results are presented for our proposed precoding design for GSM with  $n_t \geq 2$ . Numerical comparisons are carried out among three schemes, including non-precoded GSM, our diagonal precoding design with AL, and our full precoding design with AL.

In Fig. 6, we compare the CCDF of the distance  $d_{\min}$  for our precoding designs over uncorrelated Rayleigh fading channels.

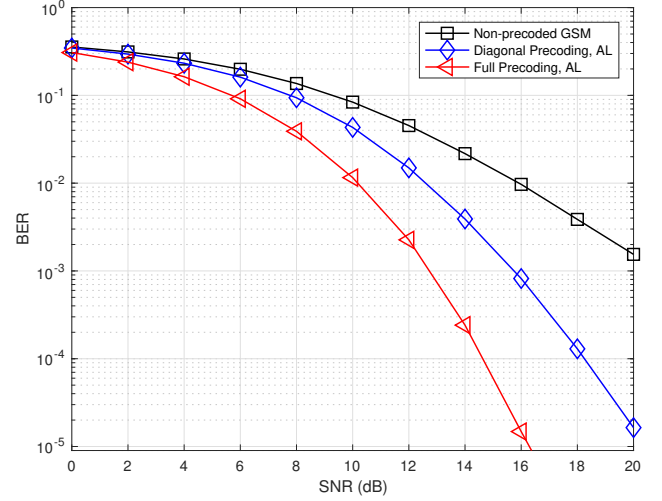


Fig. 7. Comparison of BER performance for different GSM systems with BPSK and  $N_t = 5, N_r = 2, n_t = 2$ , uncorrelated Rayleigh channels.

In this specific example, we have  $N_t = 5, N_r = 2$ , and  $n_t = 2$ , which leads to a total of  $N = 8$  TACs. Since GSM incurs a relatively high computational complexity at the receiver [11], we employ BPSK for a low-complexity implementation. It is shown that our new diagonal and full precoding designs can substantially improve  $d_{\min}$  in comparison to non-precoded GSM. The probability  $\Pr(d_{\min} > 2)$  for non-precoded GSM, the diagonal precoding design, and the full precoding design is 0.01, 0.13, and 0.92, respectively, which clearly indicates the full square precoding offers the best solution to GSM based on the MMD criterion. Furthermore, the improvement of  $d_{\min}$  is much more significant than that in Fig. 3 for SM. The reason is that there are only  $N_t$  precoding weights to be optimized for SM while  $n_t^2 N$  precoding weights for GSM with the full precoding design. Compared to SM, the computational complexity for GSM with the full precoding design increases to  $\mathcal{O}(n_t^4 N^2 M^2) + \mathcal{O}(n_t^4 N^2 N_r) + \mathcal{O}(K_{\text{AL}} n_t^8 N^4 M^2) + \mathcal{O}(K_{\text{AL}} n_t^4 N^2 n_{\text{BFGS}})$ , which is still affordable for small  $n_t$ .

In Fig. 7, we compare the unprecoded GSM error performance with the new precoding designs in uncorrelated Rayleigh fading channels. The simulation parameters are  $N_t = 5, N_r = 2$ , and  $n_t = 2$ , which are the same with those in Fig. 6. It is shown that the new full precoding design notably outperforms the diagonal precoding design and the conventional unprecoded GSM. For example, at  $\text{BER} = 10^{-3}$ , the full precoding design is more than 7 dB superior to the non-precoded one and 3 dB better than the diagonal precoding design. The huge performance gains come from the fact that the new full precoding design has the capability of jointly optimizing the total of  $n_t^2 N = 32$  precoding weights. It is also worth noting that the precoding gain can enlarge with increasing  $n_t$ ; however, a large  $n_t$  will degrade the energy efficiency of GSM systems at the transmitter while increasing the computational complexity at the receiver. Note that Fig. 4 and Fig. 7 have the same spectrum efficiency (5 bits/s/Hz), the BER performance for non-precoded SM with

$n_t = 1$  is slightly better than non-coded GSM; however, with precoding, we can find that precoded GSM significantly outperforms precoded SM. This demonstrates that precoding is vital in GSM implementations.

To demonstrate the robustness of the new precoding design in different channel conditions, in Fig. 8 we consider the error performance of GSM over highly correlated Rayleigh fading channels where  $N_t = 5$ ,  $N_r = 2$ , and  $n_t = 2$ . The channel matrix can be represented by  $\mathbf{H} = \mathbf{\Theta}_R^{1/2} \mathbf{H}_{i.i.d} \mathbf{\Theta}_T^{1/2}$  with  $[\mathbf{\Theta}_R]_{i,j} = \rho^{|i-j|}$  and  $[\mathbf{\Theta}_T]_{i,j} = \rho^{|i-j|}$  being the correlation matrices for the receive and transmit antennas, respectively, while  $\mathbf{H}_{i.i.d}$  represents an i.i.d. Rayleigh fading channel. Here, the uncorrelated ( $\rho = 0$ ), correlated ( $\rho = 0.5$ ) and highly correlated ( $\rho = 0.8$ ) MIMO channels are considered. It is found that the full precoding design always has an effective precoding gain for both scenarios. For example, at  $\text{BER}=10^{-2}$ , the new precoding design achieves an SNR gain of 6.5 dB for  $\rho = 0.5$ , and around 7 dB for  $\rho = 0.8$ . The results validate the robustness of the new precoding design against channel correlation.

In Fig. 9, we further evaluate the error performance of our design in the presence of imperfect CSI. Such imperfection originates from channel estimation and/or feedback errors. The model of imperfect CSI is given by

$$\mathbf{H}' = \mathbf{H} + \mathbf{H}_r$$

where  $\mathbf{H}_r$  is the error matrix whose entries follow an i.i.d. circularly symmetric complex Gaussian distribution  $\mathcal{CN}(0, \sigma_r^2)$ . The error performance of the new design with different error levels is compared in the figure, where  $N_t = 5$ ,  $N_r = 2$ , and  $n_t = 2$ . As expected, the error performance degrades with imperfect CSI; however, the full precoding design can still achieve an effective precoding gain compared to non-coded GSM. For example, at  $\text{BER}=10^{-4}$ , the new precoding method achieves an SNR gain of about 5.0 dB and 6.3 dB for  $\sigma_r^2 = 0.3$  and  $\sigma_r^2 = 0.1$ , respectively, which demonstrates the superiority of our new design.

From the simulation results in Figs. 6-9, we can conclude that the new full precoding design can notably improve the error performance of GSM without precoding, while offering the advantage of fast convergence rate with a low computational complexity provided by the AL method in Section IV. Therefore, the new design is highly beneficial to practical implementations.

## VII. CONCLUSION

In this paper, we proposed a unified precoding scheme for GSM systems based on the MMD criterion, which universally works for any number of activated antennas. We found that the precoding design can be formulated as a large-scale non-convex quadratically constrained quadratic program (QCQP) problem. To address this challenging problem, we transformed it into a sequence of unconstrained subproblems by leveraging augmented Lagrangian and dual ascent techniques. These subproblems can be solved in an iterative manner efficiently. Numerical results show that the proposed method can substantially improve the system error performance relative to

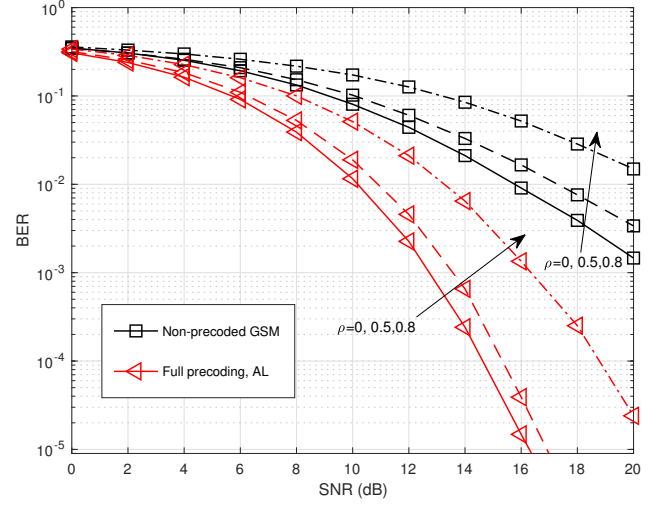


Fig. 8. Comparison of BER performance for different GSM systems with BPSK and  $N_t = 5$ ,  $N_r = 2$ ,  $n_t = 2$ , correlated Rayleigh channels.

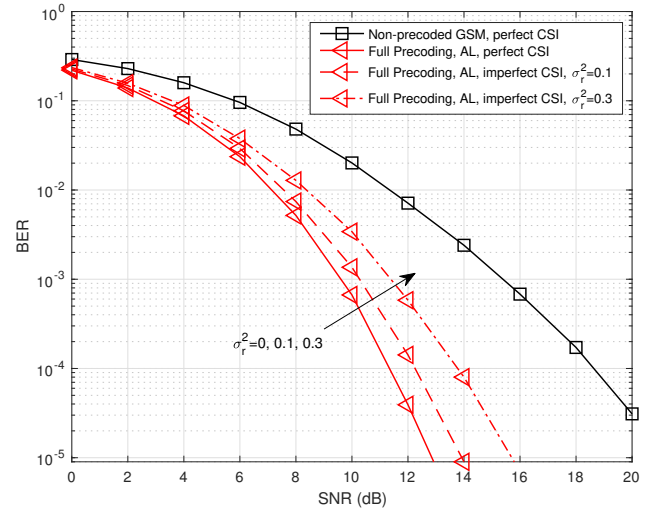


Fig. 9. Comparison of BER performance for different GSM systems with BPSK and  $N_t = 5$ ,  $N_r = 3$ ,  $n_t = 2$ , imperfect CSI

the GSM without precoding, and has an extremely fast convergence rate with a very low computational complexity.

## APPENDIX A

### PROOF OF THE EQUIVALENCE BETWEEN (P1) AND (P2)

Basically, we use a similar technique to that in [36] to prove this equivalence. We denote (P1) with the power constraint  $P_t$  as  $\mathcal{P}_1(P_t)$  and (P2) with the constraint  $d$  as  $\mathcal{P}_2(d)$ . Let the optimal solution of  $\mathcal{P}_2(d)$  be  $\mathbf{q}_2$  with the power  $P_s$ . In this case, we can easily show that the problem  $\mathcal{P}_2(P_t d / P_s)$  has the optimal solution  $\mathbf{q}'_2 = \sqrt{P_t / P_s} \mathbf{q}_2$  with the power  $P_t$ . Furthermore, we have

$$(\mathbf{q}'_2)^H \mathbf{R}_{i,j} \mathbf{q}'_2 \geq P_t d / P_s. \quad (47)$$

Obviously,  $\mathbf{q}_2'$  is feasible in  $\mathcal{P}_1(P_t)$ , which means the optimal solution  $\mathbf{q}_1$  of  $\mathcal{P}_1(P_t)$  must satisfy

$$\mathbf{q}_1^H \mathbf{R}_{i,j} \mathbf{q}_1 \geq P_t d/P_s. \quad (48)$$

This implies that  $\mathbf{q}_1$  is feasible and optimal in  $\mathcal{P}_2(P_t d/P_s)$ . As such,  $\mathbf{q}_1$  and  $\mathbf{q}_2$  is equivalent in  $\mathcal{P}_2(P_t d/P_s)$ , and thus equivalent in  $\mathcal{P}_1(P_t)$ . Then, we can reach the conclusion that (P1) is equivalent to (P2).

## REFERENCES

- [1] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.
- [2] 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.
- [3] 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.
- [4] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605–1615, Apr. 2012.
- [5] 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.
- [6] T. L. Narasimhan, P. Raviteja, and A. Chockalingam, "Generalized spatial modulation in large-scale multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3764–3779, Jul. 2015.
- [7] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3692–3703, Jul. 2009.
- [8] 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.
- [9] 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.
- [10] C. T. Lin, W. R. Wu, and C. Y. Liu, "Low-complexity ML detectors for generalized spatial modulation systems," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4214–4230, Nov. 2015.
- [11] 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.
- [12] 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.
- [13] P. Yang, Y. Xiao, L. Li, Q. Tang, Y. Yu, and S. Li, "Link adaptation for spatial modulation with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3808–3813, Oct. 2012.
- [14] P. Yang, Y. Xiao, Y. Yu, L. Li, Q. Tang, and S. Li, "Simplified adaptive spatial modulation for limited-feedback MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2656–2666, Jul. 2013.
- [15] 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.
- [16] N. Pillay and H. Xu, "Comments on "Antenna selection in spatial modulation systems,"" *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1681–1683, Sep. 2013.
- [17] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Antenna selection in spatial modulation systems," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 521–524, Mar. 2013.
- [18] —, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 112–125, Jan. 2014.
- [19] M. Maleki, H. R. Bahrami, A. Alizadeh, and N. H. Tran, "On the performance of spatial modulation: Optimal constellation breakdown," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 144–157, Jan. 2014.
- [20] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge university press, 2003.
- [21] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [22] 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.
- [23] 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.
- [24] P. Yang, Y. L. Guan, Y. Xiao, M. D. Renzo, S. Li, and L. Hanzo, "Transmit precoded spatial modulation: Maximizing the minimum Euclidean distance versus minimizing the bit error ratio," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2054–2068, Mar. 2016.
- [25] P. Yang, Y. Xiao, S. Li, and L. Hanzo, "A low-complexity power allocation algorithm for multiple-input-multiple-output spatial modulation systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1819–1825, Mar. 2016.
- [26] C. Masouros and L. Hanzo, "Constellation randomization achieves transmit diversity for single-RF spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8101–8111, Oct. 2016.
- [27] 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.
- [28] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [30] A. Antoniou and W.-S. Lu, *Practical optimization: algorithms and engineering applications*. Springer Science & Business Media, 2007.
- [31] P. Yang, Y. Xiao, Y. L. Guan, K. V. S. Hari, A. Chockalingam, S. Sugiura, H. Haas, M. D. Renzo, C. Masouros, Z. Liu, L. Xiao, S. Li, and L. Hanzo, "Single-carrier SM-MIMO: A promising design for broadband large-scale antenna systems," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1687–1716, 2016.
- [32] P. Yang, Y. Xiao, Y. L. Guan, Z. Liu, S. Li, and W. Xiang, "Adaptive SM-MIMO for mmWave communications with reduced RF chains," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1472–1485, Jul. 2017.
- [33] X. Zhang, *Matrix analysis and applications*. Tsinghua and Springer Publishing house, Beijing, 2004.
- [34] Z. Q. Luo, W. K. Ma, A. M. C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal. Proc. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [35] A. Skajaa, "Limited memory BFGS for nonsmooth optimization," *Courant Institute of Mathematical Science, New York, Master's thesis*, 2010.
- [36] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.



**Peng Cheng** (M'12) received the B.S. and M.S. degrees with great honors in communication and information systems from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2006 and 2009 and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2013. He is currently a Research Fellow at the University of Sydney, Australia. His current research interests include machine learning, network MIMO, OFDM and compressive sensing theory. He is a recipient of the Scholarship Award for Excellent

Doctoral Student granted by Ministry of Education and the Excellent Doctoral Dissertation of Shanghai Municipality.





**Zhuo Chen** (S01M05) received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1997 and the M.S. and Ph.D. degrees from the School of Electrical and Information Engineering, University of Sydney, Sydney, Australia, in 2001 and 2004, respectively. He is currently a Senior Research Scientist with Commonwealth Scientific and Industrial Research Organization (CSIRO), Sydney, Australia. His research interests include wireless communications, multiple-input-multiple-output systems, error control

coding techniques, and wireless sensor networks.



**Branka Vucetic** is an ARC Laureate Fellow and Director of the Centre of Excellence for IoT and Telecommunications at the University of Sydney. Her current research work is in wireless networks and the Internet of Things. In the area of wireless networks, she works on communication system design for millimeter wave frequency bands. In the area of the Internet of Things, Vucetic works on providing wireless connectivity for mission critical applications. Branka Vucetic is a Fellow of IEEE, the Australian Academy of Technological Sciences

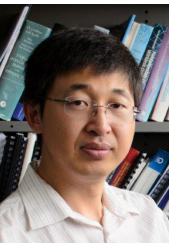
and Engineering and the Australian Academy of Science.



**J. Andrew Zhang** (M'04-SM'11) received the B.Sc. degree from Xi'an JiaoTong University, China, in 1996, the M.Sc. degree from Nanjing University of Posts and Telecommunications, China, in 1999, and the Ph.D. degree from the Australian National University, in 2004.

Currently, Dr. Zhang is an associate Professor in School of Computing and Communications, University of Technology Sydney, Australia. He was a researcher with Data61, CSIRO, Australia from 2010 to 2016, the Networked Systems, NICTA, Australia

from 2004 to 2010, and ZTE Corp, Nanjing, China from 1999 to 2001. Dr. Zhang's research interests are in the area of signal processing for wireless communications and sensing, and autonomous vehicular networks. He has published more than 100 papers in leading international Journals and conference proceedings, and has won 4 best paper awards for his work. He is a recipient of CSIRO Chairman's Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multi-gigabit wireless communications.



**Yonghui Li** (M04-SM09) received his PhD degree in November 2002 from Beijing University of Aeronautics and Astronautics. From 1999- 2003, he was affiliated with Linkair Communication Inc, where he held a position of project manager with responsibility for the design of physical layer solutions for the LAS-CDMA system. Since 2003, he has been with the Centre of Excellence in Telecommunications, the University of Sydney, Australia. He is now a Professor in School of Electrical and Information Engineering, University of Sydney. He is the recipient of the Australian Queen Elizabeth II Fellowship in 2008 and the Australian Future Fellowship in 2012.

His current research interests are in the area of wireless communications, with a particular focus on MIMO, millimeter wave communications, machine to machine communications, coding techniques and cooperative communications. He holds a number of patents granted and pending in these fields. He is now an editor for IEEE transactions on communications and IEEE transactions on vehicular technology. He was also the guest editor for IEEE JSAC Special issue on Millimeter Wave Communications for Future Mobile Networks. He received the best paper awards from IEEE International Conference on Communications (ICC) 2014, IEEE PIMRC 2017, and IEEE Wireless Days Conferences (WD) 2014.