MU-MIMO Symbol-Level Precoding for QAM Constellations with Maximum Likelihood Receivers

Xiao Tong, *Graduate Student Member, IEEE*, Ang Li, *Senior Member, IEEE*, Lei Lei, *Senior Member, IEEE*, Xiaoyan Hu, *Member, IEEE*, Fuwang Dong, *Member, IEEE*, Symeon Chatzinotas, *Fellow, IEEE* and Christos Masouros, *Fellow, IEEE*

***Abstract*—In this paper, we investigate symbol-level precoding (SLP) and efficient decoding techniques for downlink transmission, where we focus on scenarios where the base station (BS) transmits multiple quadrature amplitude modulation (QAM) constellation streams to users equipped with multiple receive antennas. We begin by formulating a symbol-level joint design scheme aimed at collaboratively optimizing the transmit precoding and receive combining matrices. This coupled problem is addressed by employing the alternating optimization (AO) method, and closed-form solutions are derived by analyzing the obtained two subproblems. Furthermore, to address the dependence of the receive combining matrix on the transmit signals, we switch to maximum likelihood detection (MLD) method for decoding. Notably, we have demonstrated that the smallest singular value of the precoding matrix significantly impacts the performance of MLD method. Specifically, a lower value of the smallest singular value results in degraded detection performance. Additionally, we show that the traditional SLP matrix is rank-one, making it infeasible to directly apply MLD at the receiver end. To circumvent this limitation, we propose a novel symbol-level smallest singular value maximization problem, termed SSVMP, to enable SLP in systems where users employ the MLD decoding approach. Moreover, to reduce the number of variables to be optimized, we further derive a more generic semidefinite programming (SDP)-based optimization problem. Numerical results validate the effectiveness of our proposed schemes and demonstrate that they significantly outperform the traditional block diagonalization (BD)-based method.**

***Index Terms*—symbol-level precoding (SLP), MU-MIMO, maximum likelihood detection (MLD), quadrature amplitude modulation (QAM).**

I. Introduction

With the worldwide deployment of fifth-generation (5G) mobile communication networks, both academia and industry have envisioned the roadmap to forthcoming generation of wireless communication systems, specifically the sixth generation (6G), which aims to achieve massive, hyper-reliable, and low-latency communications [1], [2]. Multiple-input multiple-output (MIMO) communication represents one of the most promising technologies for addressing these evolving requirements of modern wireless networks. It leverages the capabilities of multiple antennas deployed on both the transmitter and receiver sides to simultaneously transmit multiple data streams to multiple users (MU) [3]. This spatial multiplexing approach increases communication capacity and enhances spectral efficiency. However, a primary limitation of MIMO systems is the interference arising from the simultaneous transmission of multiple signals using the same time and frequency resources. Therefore, it is crucial for both the transmitter and receiver to exploit the additional degrees of freedom offered by multiple antennas to mitigate the interference [4]–[8] and improve decoding efficiency [27]–[34].

A. Previous Work on Transmit Precoding

At the transmitter side, precoding has been extensively studied as an effective strategy to mitigate interference and enhance communication performance. Traditional precoding schemes utilize channel state information (CSI) to design the precoding matrix that maximizes some system performance while adhering to the power budget. Representative examples include zero-forcing (ZF)-based approaches [4], [5], as well as block diagonalization (BD)-based methods [6], [7]. However, in these schemes, the multi-user interference (MUI) is treated as detrimental to communication systems and needs to be suppressed, overlooking the potential benefits of instantaneous interference. This interference can be further exploited through appropriate precoding design, known as symbol-level precoding (SLP) [9], [10].

Unlike traditional block-level precoding (BLP), which treats interference as detrimental to system performance, the superiority of the SLP method lies in its ability to identify instantaneous interference into either constructive interference (CI) [11] or destructive interference (DI) [12], and DI can be further converted into CI by the precoding design. As a further development, optimization-based SLP has been implemented to achieve enhanced communication performance. In [13] and [14], the authors proposed closed-form SLP solutions for MU multiple-input single-output (MISO) downlink communication

X. Tong and X. Hu are with the School of Information and Communications Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: xiao.tong.2023@stu.xjtu.edu.cn; xiaoyanhu@xjtu.edu.cn).

A. Li and L. Lei are with the School of Information and Communications Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: ang.li.2020@xjtu.edu.cn; lei.lei@xjtu.edu.cn).

F. Dong is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: dongfuwang@hrbeu.edu.cn).

S. Chatzinotas is with Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg (e-mail: symeon.chatzinotas@uni.lu).

C. Masouros is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: c.masouros@ucl.ac.uk).

systems using phase shift keying (PSK) and quadrature amplitude modulation (QAM) constellation symbols, respectively, where the bit error rate (BER) performance was improved. However, the optimization of the precoding matrix on the symbol level introduces a considerable computational burden for MU-MISO communication systems. To alleviate the computational costs, a low-complexity approach was proposed in [15], where the users were grouped and a performance-complexity trade-off was achieved by transforming the intra-group interference into CI while suppressing the inter-group interference.

In addition, the SLP method exploits interference to achieve communication performance comparable to that of conventional BLP while requiring less transmit power, thereby significantly improving overall system energy efficiency [16], [17]. The enhanced energy efficiency of the SLP technique makes it well-suited for integrated sensing and communications (ISAC) systems, as it ensures reliable communication while enabling more transmit power to be allocated for sensing compared to the BLP method, thereby improving overall ISAC performance [18]–[20]. In contrast to the aforementioned SLP approaches that aim to maximize the CI effect, a more intuitive problem formulation is to minimize the symbol error rate (SER) [21], [22], wherein the SER expression has been derived and efficient algorithms have been proposed to deal with the non-convex problems.

Notably, most current research on SLP predominantly concentrates on MU-MISO downlink systems, i.e., with single antenna receivers, while only limited works addressing the difficulties associated with MU-MIMO SLP issues, where users are equipped multiple antennas [23]–[26]. The key distinction in MIMO systems lies in the need for additional receive decoding techniques at the user side, particularly with the SLP method, where interference plays a different role compared to traditional approaches as it has been transformed into desired signals. The joint design of CI combiner and precoder was first considered in [23], but only one data stream was transmitted for each user. In [24], the closed-form transmit SLP matrix and receive combining matrix were derived for MU-MIMO systems employing PSK modulated symbols, where a novel regularized interference rejection combiner (RIRC) receiver was proposed for signal decoding. In [25], [26], the authors considered a joint SLP and linear receive combining design problem to minimize the SER for MU-MIMO systems using QAM constellation symbols. However, the dependency of the receive combining on the transmit signals was not fully addressed. Given the above analysis, SLP for MU-MIMO systems remains insufficiently explored, and therefore there is a need for more practical symbol-level schemes to be studied for MU-MIMO systems when using multi-level modulations such as QAM.

B. Previous Work on Receive Detection

At the receiver side, MIMO detection techniques are employed to detect the received signals [27], [28]. These methods can be generally categorized into two main types: linear and non-linear. Typical linear schemes offer lower computational complexity, such as ZF [29], minimum mean-squared error (MMSE) [30] and maximum asymptotic-multiuser-efficiency (MAME) [31]. However, the performance of these linear detection schemes deteriorates significantly compared to the optimal maximum likelihood detection (MLD) method [32].

MLD is a non-linear detection method that returns the optimal detection performance. However, the disadvantage of this method is that its computational complexity increases exponentially as the modulation order or the number of data streams increases. Given the increasing demand for achieving excellent transmission performance with low computational complexity, schemes that offer better performance-complexity tradeoff have been investigated [33], [34]. A novel QR decomposition based MLD (QR-MLD) method was proposed in [33], where the channel matrix was transformed into an upper triangular matrix via QR decomposition, leading to a reduction in the number of signal candidates that need to be searched. Furthermore, the authors proposed the QR decomposition with *M-algorithm* (QRM-MLD) method in [34], which allowed for a further complexity reduction. Although the QRM-MLD method significantly decreases the decoding computational complexity compared to the MLD and QR-MLD methods, this improvement comes at the expense of decoding accuracy. In summary, full MLD is suitable for low-order modulation and a small number of streams, QR-MLD is effective for moderate system configurations, and QRM-MLD is preferred for large-scale scenarios requiring a balance between performance and complexity.

Based on the aforementioned analysis, the MLD method is applied at the user side to decode the received signals in this paper for the following reasons: 1) MLD method provides promising performance, and a favorable trade-off between performance and complexity can be attained with the alternative QRM-MLD method; 2) This method enhances the generality of our proposed SLP schemes, as it requires no modification to receiver architectures; 3) Both MLD and QRM-MLD methods have been widely adopted in commercial devices, thereby increasing the practicality of the SLP schemes proposed in this paper [35].

C. Contributions

In this study, we propose practical SLP schemes and efficient decoding strategies specifically designed for MU-MIMO communication systems. Unlike [14] which addresses MU-MISO scenarios with single-antenna users, our work considers a more realistic MU-MIMO setting involving multi-antenna users, where receiver processing becomes essential for fully harnessing the benefits of SLP. In contrast to [24] which focuses on interference suppression through RIRC at the receiver side, our approach exploits interference constructively by aligning the SLP design with the MLD decoding criterion. The main contributions of this paper are summarized as follows:

1) We first formulate a symbol-level joint design framework that aims to collaboratively optimize the transmit precoding and receive combining matrices for MU-MIMO systems with QAM modulated data symbols,

where we maximize the CI effect of the outer constellation symbols while maintaining the performance of the inner constellation symbols. To solve the coupled optimization problem, we utilize the alternating optimization (AO) method to decouple the variables, transforming the original problem into two convex subproblems.

2) By analyzing the Lagrangian and Karush-Kuhn-Tucker (KKT) conditions, and applying mathematical analysis to the subproblems, we derive the optimal structures for the transmit precoding and receive combining matrices in closed form, expressed as functions of the dual variables. These subproblmes are further simplified into Quadratic Programming (QP) problems that can be solved in low complexity. Additionally, we conduct an analysis of the convergence and computational complexity associated with the proposed AO algorithm.
3) To address the reliance of the receive combining matrix on the transmit symbols, we further utilize the MLD method at the user side to decode the received signals. Importantly, we demonstrate that the performance of MLD method deteriorates as the smallest singular value of the transmit precoding matrix decreases. Furthermore, we have shown that the traditional SLP matrix is rank-one, making it infeasible to directly apply the MLD method at the users. Additionally, we highlight that the SLP method presents challenges for independent MLD decoding for each user. Thus, it is essential to incorporate independent decoding constraint into the optimization problem.
4) We propose a novel optimization problem to maximize the lower bound of the smallest singular value of the transmit SLP matrix, subject to the CI constraints, independent decoding requirement and the transmit power budget. Moreover, we derive a generic semi-definite programming (SDP)-based problem to reduce the number of variables to be optimized by omitting the CI constraints and focusing solely on optimizing the smallest singular value of the transmit precoding matrix.

Simulation results demonstrate that the proposed schemes achieve significant performance improvements over conventional BD-based approach. Our proposed singular value optimization problems outperform the joint design scheme, which benefits from the superiority of MLD over linear detection algorithms. The SDP-based problem experiences only a minor performance loss, but benefits from fewer variables need to be optimized.

D. Organization and Notations

The structure of this paper is as follows. We first introduce the system model and CI metric for QAM constellation, and present the formulated optimization problem in Section II. Then we derive the closed-form solutions for the joint optimization problem in Section III. Section IV analyzes the performance of the MLD method and the properties of the SLP matrix, proposing two symbol-level singular value optimization problems. In Section V, we analyze the convergence and computational complexity of the proposed algorithms. Our simulation findings are given in Section VI. Finally, we conclude the paper in Section VII.

Notations: $\mathbb{C}^{M\times N}$ or $\mathbb{R}^{M\times N}$ ($\mathbb{C}^{M\times 1}$ or $\mathbb{R}^{M\times 1}$) represents a complex-valued or real-valued $M \times N$ matrix ($M \times 1$ vector). For any given matrix $\mathbf{A}$, $(\mathbf{A})^T$, $(\mathbf{A})^*$, $(\mathbf{A})^H$, $(\mathbf{A})^{-1}$ and $\mathrm{Tr}(\mathbf{A})$ denote the transpose, complex conjugate, complex conjugate transpose, inverse and trace of $\mathbf{A}$, respectively. $\Re(\cdot)$ and $\Im(\cdot)$ signify the real and imaginary parts of a complex number, respectively, and $\|\cdot\|_2$ represents the l_2-norm. $card\{\cdot\}$ denotes the cardinality of a set and $\otimes$ is the Kronecker product. $\mathbf{e}_i$ denotes the i-th column of the identity matrix. $\mathrm{diag}(\mathbf{a})$ denotes a square diagonal matrix with the elements of vector $\mathbf{a}$ on the main diagonal.

II. System Model and Problem Formulation

A. System Model

We study the transceiver design problem for a downlink MU-MIMO communication system, where the BS and the downlink users are equipped with multiple antennas. Specifically, the BS has N_T transmit antennas, and each user is equipped with N_R receive antennas. The BS simultaneously serves K downlink communication users, and transmits L data streams to each user. The number of transmit and receive antennas satisfies that $N_T \geq KL$ and $L \leq N_R$. In this paper, we primarily focus on the transceiver design, assuming that perfect CSI is known [24]–[26]. Let $\mathbf{P}_k \in \mathbb{C}^{N_T\times L}$ and $\mathbf{s}_k \in \mathbb{C}^{L\times 1}$ denote the transmit precoding matrix and signal vector for the k-th user, respectively, where the entries of symbol vector are selected from a normalized QAM constellation [14]. Based on these assumptions, the transmit signal can be expressed as

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \sum_{k=1}^{K} \mathbf{P}_k \mathbf{s}_k, \tag{1}$$

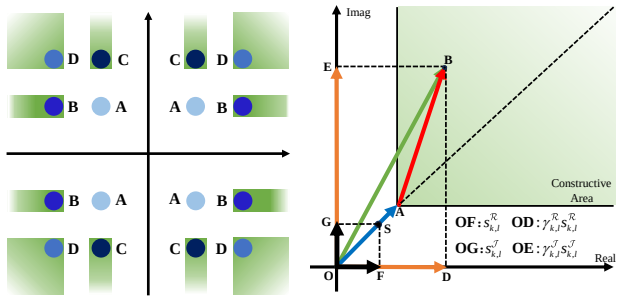
where $\mathbf{x}$ is the overall transmit signal vector for the users, $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_K] \in \mathbb{C}^{N_T\times KL}$ and $\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_K^T\right]^T \in \mathbb{C}^{KL\times 1}$. Furthermore, let $\mathbf{H}_k \in \mathbb{C}^{N_R\times N_T}$ and $\mathbf{n}_k \in \mathbb{C}^{N_R\times 1}$ denote the channel matrix between the k-th user and the BS and received complex Gaussian noise vector, respectively, the received signal at the k-th user can be expressed as

$$\begin{aligned}\mathbf{y}_k &= \mathbf{H}_k\mathbf{x} + \mathbf{n}_k \\ &= \mathbf{H}_k\mathbf{P}_k\mathbf{s}_k + \mathbf{H}_k \sum_{i=1, i\neq k}^{K} \mathbf{P}_i\mathbf{s}_i + \mathbf{n}_k,\end{aligned} \tag{2}$$

where $\mathbf{n}_k$ is a complex Gaussian vector and follows the distribution $\mathbf{n}_k \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$.

When the number of data streams is less than the number of receive antennas, the k-th user employs a receive combining matrix $\mathbf{W}_k \in \mathbb{C}^{L\times N_R}$ to decode the received signal. Then, the decoded signal for the k-th user can be expressed as

$$\hat{\mathbf{s}}_k = \mathbf{W}_k\mathbf{H}_k\mathbf{P}_k\mathbf{s}_k + \mathbf{W}_k\mathbf{H}_k \sum_{i=1, i\neq k}^{K} \mathbf{P}_i\mathbf{s}_i + \mathbf{W}_k\mathbf{n}_k, \tag{3}$$



(a) (b)

Fig. 1. Constellation point categorization and CI metric for 16QAM.

B. Constructive Interference

To improve the completeness of this paper and facilitate understanding, a concise review of CI metric for QAM modulations is provided. For illustration, Fig. 1(a) shows the entire constellation map of the 16QAM constellation, and Fig. 1(b) depicts one quadrant of a nominal 16QAM constellation in detail. Based on whether the CI can be utilized, the 16QAM constellation can be classified into four types of constellation points [38]:

1) 'A': inner points that cannot use CI;
2) 'B': real outer points that can use CI;
3) 'C': imaginary outer points that can use CI;
4) 'D': corner outer points that can use CI.

Accordingly, the 'symbol-scaling' metric discussed in [14] is presented here. Specifically, the constellation points and noiseless received signals can be mathematically decomposed into

$$\begin{aligned} s_{k,l} &= s^{\mathcal{R}}_{k,l} + s^{\mathcal{I}}_{k,l} \\ \mathbf{w}_{k,l}\mathbf{H}_k\mathbf{P}\mathbf{s} &= \gamma^{\mathcal{R}}_{k,l}s^{\mathcal{R}}_{k,l} + \gamma^{\mathcal{I}}_{k,l}s^{\mathcal{I}}_{k,l}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \end{aligned}$$

where $\mathbf{w}_{k,l} \in \mathbb{C}^{1\times N_R}$ denotes the l-th row of the receive combining matrix $\mathbf{W}_k$ and $s_{k,l}$ represents the l-th data symbol for the k-th user. $s^{\mathcal{R}}_{k,l} = \Re\{s_{k,l}\}$ and $s^{\mathcal{I}}_{k,l} = j\cdot\Im\{s_{k,l}\}$ are the real and imaginary detection boundary bases of $s_{k,l}$, respectively. The real-valued scaling parameters $\gamma^{\mathcal{R}}_{k,l} \geq 0$ and $\gamma^{\mathcal{I}}_{k,l} \geq 0$ are further introduced to maximize the CI effect, and improved communication performance can be attained with an increased value of $\min\left\{\gamma^{\mathcal{R}}_{k,l}, \gamma^{\mathcal{I}}_{k,l}\right\}$. To concisely express the CI constraints, we define

$$\boldsymbol{\gamma}_{k,l} = \left[\gamma^{\mathcal{R}}_{k,l}, \gamma^{\mathcal{I}}_{k,l}\right]^T, \bar{\mathbf{s}}_{k,l} = \left[s^{\mathcal{R}}_{k,l}, s^{\mathcal{I}}_{k,l}\right]^T, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \tag{4}$$

then $\mathbf{w}_{k,l}\mathbf{H}_k\mathbf{P}\mathbf{s}$ can be further simplified as

$$\mathbf{w}_{k,l}\mathbf{H}_k\mathbf{P}\mathbf{s} = \boldsymbol{\gamma}^T_{k,l}\bar{\mathbf{s}}_{k,l}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \tag{5}$$

For notational clarity, we define two sets: $\mathcal{O}$ and $\mathcal{I}$. Set $\mathcal{O}$ comprises the outer points that can utilize CI, such as the real parts of Type 'B' and 'D' points as well as the imaginary parts of Type 'C' and 'D' points. Conversely, set $\mathcal{I}$ comprises the inner points that cannot utilize CI, such as the real parts of Type 'A' and 'C' points and imaginary parts of Type 'A' and 'B' points. Accordingly, the CI constraints can be expressed as

$$\gamma^{\mathcal{O}}_{k,l} \geq t, \gamma^{\mathcal{I}}_{k,l} = t, \forall \gamma^{\mathcal{O}}_{k,l} \in \mathcal{O}, \forall \gamma^{\mathcal{I}}_{k,l} \in \mathcal{I}, \tag{6}$$

where t is rescaling factor needs to be optimized. A larger value of t indicates that the transmit signal is pushed away from the decision boundary, which can lead to improved communication performance [10], [14].

C. Problem Formulation

Consistent with [14], the interference on the inner points that cannot be utilized is destructive, whereas the interference on the outer points is considered to be constructive. Following [24], we propose a joint optimization approach for both the transmit precoding and receive combining matrices to enhance the CI effect for all users, where the CI effect of the outer points is maximized while the performance of the inner points is preserved. For our considered MU-MIMO system, the joint design problem can be formulated as

$$\begin{aligned} \text{P1}: \max_{\mathbf{W}_k,\mathbf{P},t,\boldsymbol{\Gamma}} \quad & t \\ s.t. \quad & \mathbf{C1}: \mathbf{W}_k\mathbf{H}_k\mathbf{P}\mathbf{s} = \mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k, \forall k \in \mathcal{K} \\ & \mathbf{C2}: t - \gamma^{\mathcal{O}}_m \leq 0, \forall \gamma^{\mathcal{O}}_m \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma^{\mathcal{I}}_n = 0, \forall \gamma^{\mathcal{I}}_n \in \mathcal{I} \\ & \mathbf{C4}: \|\mathbf{P}\mathbf{s}\|^2_2 \leq p \\ & \mathbf{C5}: \|\mathbf{W}_k\|^2_2 \leq 1, \forall k \in \mathcal{K}, \end{aligned} \tag{7}$$

where $\mathbf{U}_1 = \mathbf{I}_L \otimes [1,1] \in \mathbb{R}^{L\times 2L}$ is introduced as an intermediate matrix to align the dimensionality of variables, $\boldsymbol{\Gamma}_k = \left[\boldsymbol{\gamma}^T_{k,1}, \boldsymbol{\gamma}^T_{k,2}, \cdots, \boldsymbol{\gamma}^T_{k,L}\right]^T \in \mathbb{R}^{2L\times 1}$ is the real-valued scaling vector and $\bar{\mathbf{s}}_k = \left[\bar{\mathbf{s}}^T_{k,1}, \bar{\mathbf{s}}^T_{k,2}, \cdots, \bar{\mathbf{s}}^T_{k,L}\right]^T \in \mathbb{C}^{2L\times 1}$ is the signal basis vector. **C1**-**C3** represent the CI constraints for QAM symbols, and we can obtain

$$\begin{aligned} &\mathcal{O} \cup \mathcal{I} = \left\{\gamma^{\mathcal{R}}_{1,1}, \gamma^{\mathcal{I}}_{1,1}, \cdots, \gamma^{\mathcal{R}}_{k,l}, \gamma^{\mathcal{I}}_{k,l}, \cdots, \gamma^{\mathcal{R}}_{K,L}, \gamma^{\mathcal{I}}_{K,L}\right\}, \\ &card\left\{\mathcal{O}\right\} + card\left\{\mathcal{I}\right\} = 2KL. \end{aligned} \tag{8}$$

C4 represents the transmit power budget at the BS and **C5** denotes the receive combining power budget. It can be observed from (3) that $\mathbf{W}_k$ is multiplied by the noise, this operation ensures that **C5** will not affect the communication performance while also making the formulated problem bounded and feasible.

III. Proposed Joint Design Scheme

In this section, we aim to solve the formulated joint design problem P1. It can be observed that the matrix variables are coupled in **C1**, which makes the problem intractable. Therefore, the AO algorithm is employed to decouple the variables and iteratively optimize the transmit precoding matrix and the receive combining matrix. Specifically, P1 can be transformed into two subproblems by fixing $\mathbf{P}$ or $\mathbf{W}_k$, allowing for more efficient algorithms to separately optimize the subproblems with simpler structures.

A. Optimization on Transmit Precoding Matrix

First, for a given $\mathbf{W}_k$, we optimize the transmit precoding matrix $\mathbf{P}$. Fortunately, the formulated subproblem resembles the optimization problem $\mathcal{P}_3$ presented in [14], which indicates the closed-form solution of the optimal SLP matrix in MISO systems can be extended to MIMO systems. Furthermore, the modified iterative algorithm proposed in [14] can be applied to enhance the proposed AO algorithm's convergence speed.

To be more specific, when $\mathbf{W}_k$ is held constant, P1 can be restructured to

$$\begin{aligned} \text{P2}: \max_{\mathbf{P},t,\mathbf{\Gamma}} \ & t \\ s.t. \ & \mathbf{C1}: \mathbf{GPs} = \mathbf{U}\text{diag}\left(\mathbf{\Gamma}\right)\bar{\mathbf{s}} \\ & \mathbf{C2}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \\ & \mathbf{C4}: \|\mathbf{Ps}\|_2^2 \le p, \end{aligned} \tag{9}$$

where $\mathbf{G} = \mathbf{WH} \in \mathbb{C}^{KL\times N_T}$, $\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_K^T\right]^T \in \mathbb{C}^{KN_R\times N_T}$ is the total channel matrix and $\mathbf{W} \in \mathbb{C}^{KL\times KN_R}$ is a diagonal matrix composed of $\mathbf{W}_k$ and expressed as

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{W}_K \end{bmatrix}.$$

$\bar{\mathbf{s}} = \left[\bar{\mathbf{s}}_1^T, \bar{\mathbf{s}}_2^T, \cdots, \bar{\mathbf{s}}_K^T\right]^T \in \mathbb{C}^{2KL\times 1}$ represents the total signal basis vector, $\mathbf{\Gamma} = \left[\mathbf{\Gamma}_1^T, \mathbf{\Gamma}_2^T, \cdots, \mathbf{\Gamma}_k^T\right]^T \in \mathbb{R}^{2KL\times 1}$ is the total real-valued scaling vector and $\mathbf{U} = \mathbf{I}_{KL} \otimes [1, 1] \in \mathbb{R}^{KL\times 2KL}$ is an intermediate matrix. The closed-form SLP matrix $\mathbf{P}$ can be expressed as [14]

$$\begin{aligned} &\mathbf{P} \\ &= \frac{1}{KL}\mathbf{G}^H\left(\mathbf{GG}^H\right)^{-1}\mathbf{U}\text{diag}\left(\sqrt{\frac{p}{\mathbf{u}^T\tilde{\mathbf{V}}^{-1}\mathbf{u}}}\mathbf{E}^{-1}\tilde{\mathbf{V}}^{-1}\mathbf{u}\right)\bar{\mathbf{s}}\hat{\mathbf{s}}^T, \end{aligned} \tag{10}$$

where

$$\hat{\mathbf{s}} = \left[\frac{1}{s_1}, \frac{1}{s_2}, \cdots, \frac{1}{s_{KL}}\right] \in \mathbb{C}^{KL\times 1}, \tag{11a}$$

$$\mathbf{u} = \left[\mu_1, \mu_2, \cdots, \mu_{card\{O\}}, \nu_1, \nu_2, \cdots, \nu_{card\{I\}}\right]^T, \tag{11b}$$

$\mathbf{u} \in \mathbb{R}^{2KL\times 1}$ is a dual variable vector associated with $2KL$ constraints and can be calculated by solving the following QP optimization problem:

$$\begin{aligned} \text{P3}: \min_{\mathbf{u}} \ & \mathbf{u}^T\tilde{\mathbf{V}}^{-1}\mathbf{u} \\ s.t. \ & \mathbf{C1}: \mathbf{1}^T\mathbf{u} - 1 = 0 \\ & \mathbf{C2}: \mu_m \ge 0, \forall m \in \{1, 2, \cdots, card\{\mathcal{O}\}\}, \end{aligned} \tag{12}$$

where $\tilde{\mathbf{V}} = \mathbf{EVE}^T$, $\mathbf{V} = \Re\{\mathbf{T}\}$. $\mathbf{T} \in \mathbb{C}^{2KL\times 2KL}$ and $\mathbf{E} \in \mathbb{R}^{2KL\times 2KL}$ are expressed as

$$\mathbf{T} = \text{diag}\left(\bar{\mathbf{s}}^H\right)\mathbf{U}^H\left(\mathbf{GG}^H\right)^{-1}\mathbf{U}\text{diag}\left(\bar{\mathbf{s}}\right), \tag{13a}$$

$$\mathbf{E} = \left[\mathbf{e}_{L(\tilde{s}_1)}, \mathbf{e}_{L(\tilde{s}_2)}, \cdots, \mathbf{e}_{L(\tilde{s}_{2KL})}\right]^T. \tag{13b}$$

Here, $L(\cdot)$ is a '**Locater**' function that returns the index of $\tilde{s}_m$ in $\bar{\mathbf{s}}$, defined as

$$L(\tilde{s}_m) = k, \text{ if } \tilde{s}_m = \bar{s}_k. \tag{14}$$

$\mathbf{E}$ is an invertible matrix introduced to rearrange the columns and rows of the matrix $\mathbf{V}$ for notational and mathematical convenience [14].

B. Optimization on Receive Combining Matrix

After optimizing the SLP matrix $\mathbf{P}$, the combining matrix $\mathbf{W}_k$ can be updated with the obtained $\mathbf{P}$, and a total number of K combining matrices can be optimized in parallel. Specifically, given a known $\mathbf{P}$, the optimization on a specific $\mathbf{W}_k$ can be reformulated as

$$\begin{aligned} \text{P4}: \max_{\mathbf{W}_k,t,\mathbf{\Gamma}} \ & t \\ s.t. \ & \mathbf{C1}: \mathbf{W}_k\mathbf{r}_k = \mathbf{U}_1\text{diag}\left(\mathbf{\Gamma}_k\right)\bar{\mathbf{s}}_k, \forall k \in \mathcal{K} \\ & \mathbf{C2}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \\ & \mathbf{C4}: \|\mathbf{W}_k\|_2^2 \le 1, \forall k \in \mathcal{K}, \end{aligned} \tag{15}$$

where $\mathbf{r}_k = \mathbf{H}_k\mathbf{Ps} \in \mathbb{C}^{N_R\times 1}$. P4 is a second-order-cone programming (SOCP) problem and solved readily. Furthermore, the optimal $\mathbf{W}_k$ is shown to exhibit a closed-form expression based on the following proposition.

Proposition 1: The optimal receive combining matrix $\mathbf{W}_k$ for P4 can be expressed as

$$\mathbf{W}_k = \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\mathbf{U}_1\text{diag}\left(\sqrt{\frac{\mathbf{r}_k^H\mathbf{r}_k}{\mathbf{u}_1^T\tilde{\mathbf{V}}_1^{-1}\mathbf{u}_1}}\mathbf{E}^{-1}\tilde{\mathbf{V}}_1^{-1}\mathbf{u}_1\right)\bar{\mathbf{s}}_k\mathbf{r}_k^H, \tag{16}$$

where $\tilde{\mathbf{V}}_1 = \mathbf{E}_1\mathbf{V}_1\mathbf{E}_1^T$, $\mathbf{V}_1 = \Re\left(\mathbf{T}_1\right)$. $\mathbf{T}_1 \in \mathbb{C}^{2L\times 2L}$ and $\mathbf{E}_1 \in \mathbb{R}^{2L\times 2L}$ are expressed as.

$$\mathbf{T}_1 = \text{diag}\left(\bar{\mathbf{s}}_k^H\right)\mathbf{U}_1^H\mathbf{U}_1\text{diag}\left(\bar{\mathbf{s}}_k\right), \tag{17a}$$

$$\mathbf{E}_1 = \left[\mathbf{e}_{L(\tilde{s}_1)}, \mathbf{e}_{L(\tilde{s}_2)}, \cdots, \mathbf{e}_{L(\tilde{s}_{2L})}\right]^T. \tag{17b}$$

Proof: See Appendix A.

Furthermore, $\mathbf{u}_1 \in \mathbb{R}^{2L\times 1}$ is obtained by solving the following QP optimization problem:

$$\begin{aligned} \text{P5}: \min_{\mathbf{u}_1} \ & \mathbf{u}_1^T\tilde{\mathbf{V}}_1^{-1}\mathbf{u}_1 \\ s.t. \ & \mathbf{C1}: \mathbf{1}^T\mathbf{u}_1 - 1 = 0 \\ & \mathbf{C2}: \mu_m \ge 0, \forall m \in \{1, 2, \cdots, card\{\mathcal{O}\}\}. \end{aligned} \tag{18}$$

The detailed derivation process of P5 is similar to that of $\mathcal{P}_5$ in [14], which can be found in *Proposition* 2 presented there. Furthermore, P5 is a convex QP problem in the same form as P3 and can be effectively solved.

C. Alternating Optimization Algorithm

The optimal solutions to P1 can be obtained by iteratively solving the QP problems P3 and P5, until convergence. It can be observed that the number of variables to be optimized in P2, P3, P4 and P5 are N_TKL, $2KL$, N_RKL and $2KL$,

Algorithm 1 AO algorithm used to update $\mathbf{P}$ and $\mathbf{W}_k$

Input : CSI matrix $\mathbf{H}$, transmit signal vector $\mathbf{s}$, and convergence threshold κ.

Output : precoding matrix $\mathbf{P}$, combining matrix $\mathbf{W}_k$.

Initialize $\mathbf{P}^{(1)}$, $n=1$, $\delta=1$ and $\kappa=10^{-5}$.

while $\delta > \kappa$ **do**

Calculate $\mathbf{Q}_1 = \tilde{\mathbf{V}}_1^{-1}$

Obtain $\mathbf{u}_1$ via Algorithm 1 proposed in [14]

Obtain $\mathbf{W}_k^{(n+1)}$ via (50)

Calculate $\mathbf{Q} = \tilde{\mathbf{V}}^{-1}$

Obtain $\mathbf{u}$ via Algorithm 1 proposed in [14]

Calculate $\mathbf{P}^{(n+1)}$ and $t^{(n+1)}$ via (10) and (9)

Obtain $\delta = \left| t^{(n+1)} - t^{(n)} \right|$

Update $n = n+1$

Until convergence

$\mathbf{P} = \mathbf{P}^{(n)}$, $\mathbf{W}_k = \mathbf{W}_k^{(n)}$.

respectively, which demonstrates that the computational complexity of the two subproblems is reduced. Furthermore, P3 and P5 exhibit the same form as $\mathcal{P}_5$ in [14], which indicates the generic iterative algorithm proposed therein also can be used to update our variables, further accelerating the solving process. Based on the preceding discussion, the proposed AO algorithm is summarized in **Algorithm 1**.

In our proposed AO algorithm, we iteratively update the transmit precoding matrix and receive combining matrix, the algorithm converges with the increase of the number of iterations. Furthermore, the iterative algorithm introduced in [14] is used to effectively solve the two QP subproblems to obtain the dual variables. Additional convergence analysis of the proposed AO algorithm, along with the computational complexity of the calculation of the variables are presented in Section V.

IV. Symbol-Level Precoding for MLD Decoding

Although the aforementioned joint optimization scheme can achieve excellent communication performance, the reliance of the receive combining matrix on the transmit data symbols makes it inapplicable in practice. To address this challenge and meanwhile ensure promising decoding performance, in this section we consider the SLP design at the BS when users employ the MLD decoding method. We first show that the traditional SLP design is not applicable in such case because the rank-one property of the SLP matrix, followed by the proposition of a novel SLP solution tailored for MLD that offers a promising performance.

A. Performance Analysis of MLD Method

The MLD method is a non-linear detection scheme and is regarded as the optimal signal detection method for MIMO systems [28]. However, if the transmit precoding matrix is rank-deficient, the performance of MLD will deteriorate severely. In this section, we demonstrate that the smallest singular value of the precoding matrix is a crucial factor that affects the system performance when users employ MLD as the decoding method, which motivates our subsequent design.

In a generic MU-MIMO system where users employ MLD method to decode the received signals, the estimation process can be expressed as

$$\begin{aligned}\mathbf{s}^{\star} &= \underset{\tilde{\mathbf{s}}}{\arg\ \min} \left\|\mathbf{y} - \mathbf{M}\tilde{\mathbf{s}}\right\|_2^2 \\ &= \underset{\tilde{\mathbf{s}}}{\arg\ \min} \left\|\mathbf{M}\left(\mathbf{s}-\tilde{\mathbf{s}}\right) + \mathbf{n}\right\|_2^2 \\ &\overset{(a)}{=} \underset{\tilde{\mathbf{s}}}{\arg\ \min} \left\{ \mathrm{Tr}\left[\left(\mathbf{s}-\tilde{\mathbf{s}}\right)\left(\mathbf{s}-\tilde{\mathbf{s}}\right)^H \mathbf{M}^H\mathbf{M} \right] + \sigma^2 \right\} \\ &= \underset{\tilde{\mathbf{s}}}{\arg\ \min} \left\{ \mathrm{Tr}\left[\tilde{\mathbf{S}}\tilde{\mathbf{M}} \right] + \sigma^2 \right\},\end{aligned} \tag{19}$$

where the subscript k for the k-th user is omitted for clarity. Here, $\mathbf{M} = \mathbf{HP}$ denotes the equivalent transmit-receive channel matrix, $\tilde{\mathbf{s}}$ represents the candidate symbol vector selected from the constellation book. Additionally, we define $\tilde{\mathbf{S}} = \left(\mathbf{s}-\tilde{\mathbf{s}}\right)\left(\mathbf{s}-\tilde{\mathbf{s}}\right)^H$ and $\tilde{\mathbf{M}} = \mathbf{M}^H\mathbf{M}$. The step (a) can be achieved because the transmit symbols and received noise are independent. Furthermore, we introduce *Von Neumann's trace inequality* here for subsequent derivation.

Lemma 1 (*Von Neumann's trace inequality*): Let $\mathbf{A}$ and $\mathbf{B}$ be the N-dimensional Hermitian positive semi-definite matrices. Denote the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ as $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_N(\mathbf{A})$ and $\lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{B}) \geq \cdots \geq \lambda_N(\mathbf{B})$, respectively, which are arranged in a non-decreasing order. Then we have

$$\mathrm{Tr}\left[\mathbf{AB}\right] \geq \sum_{n=1}^{N} \lambda_n(\mathbf{A})\,\lambda_{N-n+1}(\mathbf{B}). \tag{20}$$

Based on *Lemma 1*, we obtain:

$$\mathrm{Tr}\left[\tilde{\mathbf{S}}\tilde{\mathbf{M}}\right] \geq \sum_{l=1}^{L} \lambda_l\left(\tilde{\mathbf{S}}\right)\lambda_{L-l+1}\left(\tilde{\mathbf{M}}\right) \overset{(b)}{=} \lambda_1\left(\tilde{\mathbf{S}}\right)\lambda_L\left(\tilde{\mathbf{M}}\right), \tag{21}$$

where the step (b) follows from the fact that $rank(\tilde{\mathbf{S}}) = 1$. According to (19), under the same SNR conditions, a larger value of $\mathrm{Tr}(\tilde{\mathbf{S}}\tilde{\mathbf{M}})$ is more beneficial for correctly decoding the transmit symbols from the received signals that contain noise, which can be achieved by maximizing $\lambda_1(\tilde{\mathbf{S}})\lambda_L(\tilde{\mathbf{M}})$, the lower bound of $\mathrm{Tr}(\tilde{\mathbf{S}}\tilde{\mathbf{M}})$. Moreover, $\lambda_1(\tilde{\mathbf{S}})$ remains constant once the candidate symbol vector is selected, and since $\tilde{\mathbf{M}}$ is comprised of the optimized transmit precoding matrix $\mathbf{P}$, the singular values of $\mathbf{P}$ will affect the MLD performance. $\lambda_L(\tilde{\mathbf{M}})$ can be expressed as

$$\lambda_L\left(\tilde{\mathbf{M}}\right) = \lambda_L\left(\mathbf{P}^H\mathbf{H}^H\mathbf{HP}\right). \tag{22}$$

To guarantee a promising MLD performance, $\mathbf{P}$ must be a full-rank matrix, with the smallest singular value being maximized. This is due to the fact that a larger smallest singular value of $\mathbf{P}$ will yield a correspondingly larger smallest singular value of $\tilde{\mathbf{M}}$, thereby enhancing MLD performance. Based on the above analysis, our primary optimization objective for the subsequent problem is to maximize the smallest singular value of $\mathbf{P}$.

B. SLP Matrix Analysis

In this section, we show analytically that the traditional SLP approach returns a rank-one precoding matrix, as stated in *Corollary 1*.

Corollary 1: In P2, the optimized transmit SLP matrix $\mathbf{P}$ is rank-one.

Proof: We begin by transforming the power constraint in P2, where $\mathbf{Ps}$ can be decomposed as follows

$$\mathbf{Ps} = \sum_{i=1}^{KL} \mathbf{p}_i s_i, \forall i \in \mathcal{KL}. \tag{23}$$

In P2, $\mathbf{Ps}$ can be regarded as an overall vector variable, therefore, the allocation of power across each individual component $\mathbf{p}_i$ associated with the symbol s_i does not influence the attainment of the optimal solution. Thus, each $\mathbf{p}_i s_i$ can be treated as identical, which results in

$$\|\mathbf{Ps}\|_2^2 = \|KL\mathbf{p}_i s_i\|_2^2 = K^2L^2 s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i = KL \sum_{i=1}^{KL} s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i, \tag{24}$$

then the power constraint is equivalent to

$$\sum_{i=1}^{KL} s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i \le \frac{p}{KL}. \tag{25}$$

By replacing **C4** in P2 with (25), P2 can be further transformed into

$$\begin{aligned} \text{P6}: \min_{\mathbf{P},t,\mathbf{\Gamma}} \quad & -t \\ s.t. \quad & \mathbf{C1}: \mathbf{g}_{k,l} \sum_{i=1}^{KL} \mathbf{p}_i s_i = \boldsymbol{\gamma}_{k,l}^T \bar{\mathbf{s}}_{k,l}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L} \\ & \mathbf{C2}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \\ & \mathbf{C4}: \sum_{i=1}^{KL} s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i \le \frac{p}{KL}, \end{aligned} \tag{26}$$

where $\mathbf{g}_{k,l} = \mathbf{w}_{k,l}\mathbf{H}_k$. The Lagrangian function of P6 can be expressed as

$$\begin{aligned} &\mathcal{L}\left(\mathbf{p}_i, t, \alpha_{k,l}, \theta_m, \vartheta_n, \alpha_0\right) \\ &= -t + \sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{k,l} \left(\mathbf{g}_{k,l} \sum_{i=1}^{KL} \mathbf{p}_i s_i - \boldsymbol{\gamma}_{k,l}^T \bar{\mathbf{s}}_{k,l} \right) \\ &\quad + \sum_{m=1}^{card\{O\}} \theta_m \left(t - \gamma_m^O\right) + \sum_{n=1}^{card\{I\}} \vartheta_n \left(t - \gamma_n^I\right) \\ &\quad + \alpha_0 \left(\sum_{i=1}^{KL} s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i \le \frac{p}{KL} \right), \end{aligned} \tag{27}$$

where $\alpha_{k,l}$, $\theta_m \ge 0$, ϑ_n and $\alpha_0 \ge 0$ represent the introduced dual variables, and each $\alpha_{k,l}$ and ϑ_n can be complex. Furthermore, the KKT conditions of (27) can be derived as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}_i} = \left(\sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{k,l}\mathbf{g}_{k,l} \right) s_i + \alpha_0 s_i s_i^* \mathbf{p}_i^H = \mathbf{0}, \forall i \in \mathcal{KL} \tag{28a}$$

$$\frac{\partial \mathcal{L}}{\partial t} = -1 + \sum_{m=1}^{card\{O\}} \theta_m + \sum_{n=1}^{card\{I\}} \vartheta_n = 0 \tag{28b}$$

$$\mathbf{g}_{k,l} \sum_{i=1}^{KL} \mathbf{p}_i s_i - \boldsymbol{\gamma}_{k,l}^T \bar{\mathbf{s}}_{k,l} = 0, \forall k \in \mathcal{K}, \forall l \in \mathcal{L} \tag{28c}$$

$$\theta_m \left(t - \gamma_m^O\right) = 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \tag{28d}$$

$$t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \tag{28e}$$

$$\alpha_0 \left(\sum_{i=1}^{KL} s_i^* \mathbf{p}_i^H \mathbf{p}_i s_i \le \frac{p}{KL} \right) = 0. \tag{28f}$$

It can be observed that $\alpha_0 \neq 0$ based on (28a), and given $\alpha_0 \ge 0$, it follows that $\alpha_0 > 0$. This indicates the power constraint must be satisfied with equality when optimality is achieved. Therefore, $\mathbf{p}_i^H$ can be expressed as

$$\mathbf{p}_i^H = -\frac{1}{\alpha_0 s_i^*} \left(\sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{k,l}\mathbf{g}_{k,l} \right) = \frac{1}{s_i^*} \left(\sum_{k=1}^{K}\sum_{l=1}^{L} \chi_{k,l}\mathbf{g}_{k,l} \right), \tag{29}$$

where we define

$$\chi_{k,l} = -\frac{\alpha_{k,l}}{\alpha_0}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. \tag{30}$$

Based on (29) and by taking the conjugate transpose, $\mathbf{p}_i$ can be derived as follows:

$$\mathbf{p}_i = \left(\sum_{k=1}^{K}\sum_{l=1}^{L} \chi_{k,l}^* \mathbf{g}_{k,l}^H \right) \frac{1}{s_i}, \forall i \in \mathcal{KL}, \tag{31}$$

which further leads to

$$\mathbf{p}_i s_i = \sum_{k=1}^{K}\sum_{l=1}^{L} \chi_{k,l}^* \mathbf{g}_{k,l}^H, \forall i \in \mathcal{KL}. \tag{32}$$

It can be observed from (32) that $\mathbf{p}_i s_i$ is a constant vector. This indicates that each column of the precoding matrix satisfies: $\mathbf{p}_i = \mathbf{p}_k \cdot \frac{s_k}{s_i}, \forall i, k$, which subsequently result in $rank(\mathbf{P}) = 1$.

Based on the preceding analysis, it is clear that the MLD method necessitates that the transmit precoding matrix possesses a larger smallest singular value to enhance the decoding efficiency. However, we demonstrate that the SLP matrix is rank-one, which indicates that the MLD method cannot be directly applied to decode the signals transmitted by the SLP scheme. This limitation motivates us to propose more practical SLP-based schemes, which are tailored for MU-MIMO systems where users employ the MLD decoding approach.

C. Independent MLD Decoding

In a typical MU-MIMO communication system, users independently decode their received signals using the MLD method. This is feasible because traditional BLP schemes are designed to eliminate or minimize inter-user interference, and the received signals contains only the desired signal and noise,

where the inter-stream interference within the user itself is handled by the MLD procedure. More precisely, the received signal for the k-th user with the traditional BD precoding can be expressed as

$$\begin{aligned}\mathbf{y}_k^{\mathrm{BD}} &= \mathbf{H}_k \sum_{k=1}^{K} \mathbf{P}_k^{\mathrm{BD}} \mathbf{s}_k + \mathbf{n}_k \\ &\overset{(c)}{=} \mathbf{H}_k \mathbf{P}_k^{\mathrm{BD}} \mathbf{s}_k + \mathbf{n}_k,\end{aligned} \tag{33}$$

where step (c) is achieved since the precoding matrix $\mathbf{P}_k^{\mathrm{BD}}$ for the k-th user lies in the null space of the other users' channel matrices. Subsequently, the MLD principle shown in (19) can be used to decode the received signal as

$$\begin{aligned}\mathbf{s}_k^{\star} &= \arg \min_{\tilde{\mathbf{s}}_k} \left\| \mathbf{H}_k \mathbf{P}_k^{\mathrm{BD}} \mathbf{s}_k - \mathbf{H}_k \mathbf{P}_k^{\mathrm{BD}} \tilde{\mathbf{s}}_k + \mathbf{n}_k \right\| \\ &= \arg \min_{\tilde{\mathbf{s}}_k} \left\| \mathbf{H}_k \mathbf{P}_k^{\mathrm{BD}} \left(\mathbf{s}_k - \tilde{\mathbf{s}}_k \right) + \mathbf{n}_k \right\|.\end{aligned} \tag{34}$$

It is evident that $\mathbf{s}_k^{\star} = \mathbf{s}_k$. There are two main reasons why the MLD method can achieve correct symbol recovery when the BS uses BD precoding: 1) BD approach completely eliminates inter-user interference; 2) both the channel matrix and the BD precoding matrix are full-rank, thereby ensuring that the symbol space remains fully searchable under the MLD criterion. It is worth noting that when either the channel matrix or the precoding matrix is rank-deficient, this indicates that some rows of the matrix are correlated. Such correlation leads to non-uniqueness in the variables minimizing the MLD criterion, thereby degrading the performance of the MLD method.

Next, we specifically analyze the challenges encountered by the receiver when employing the MLD method under the scenario where SLP is used at the transmitter. To be specific, the received signal for the k-th user can be expressed as

$$\mathbf{y}_k^{\mathrm{SLP}} = \mathbf{H}_k \sum_{i=1}^{K} \mathbf{P}_i^{\mathrm{SLP}} \mathbf{s}_i + \mathbf{n}_k. \tag{35}$$

Unlike BD precoding, the SLP approach transforms interference into CI. As a result, the presence of these interference in the received signal invalidates the MLD method. To be more specific, the MLD criterion in this case can be expressed as

$$\mathbf{s}_k^{\star} = \arg \min_{\tilde{\mathbf{s}}_k} \left\| \underbrace{\mathbf{H}_k \mathbf{P}_k^{\mathrm{SLP}} \left(\mathbf{s}_k - \tilde{\mathbf{s}}_k \right)}_{\text{Decoded Signal}} + \underbrace{\sum_{i=1, i \neq k}^{K} \mathbf{P}_i^{\mathrm{SLP}} \mathbf{s}_i}_{\text{Constructive Interference}} + \mathbf{n}_k \right\|, \tag{36}$$

where the first term corresponds to the 'Decoded Signal' component, which is intended to be used for symbol detection. The second term represents the 'Constructive Interference', which has been transformed into a constructive component to improve the received signal quality. However, this component cannot be directly utilized for symbol detection, as each user has access only to its own precoding matrix. In contrast to the BD method, this 'Constructive Interference' may become destructive when using the MLD detection technique. As a result, the optimal solution does not necessarily coincide with $\mathbf{s}_k$.

Based on the foregoing analysis, the use of a general SLP matrix leads to the failure of the MLD method. Nevertheless, it remains an open question whether employing the closed-form precoding matrix derived in (10) alters this outcome. Next, we proceed to address this question.

When the closed-form precoding matrix derived in (10) is used to transmit the symbols, the received signal for the k-th user can be further simplified into

$$\begin{aligned}\mathbf{y}_k^{\mathrm{SLP}} &= \mathbf{H}_k \sum_{k=1}^{K} \mathbf{P}_k^{\mathrm{SLP}} \mathbf{s}_k + \mathbf{n}_k \\ &\overset{(d)}{=} K \mathbf{H}_k \mathbf{P}_k^{\mathrm{SLP}} \mathbf{s}_k + \mathbf{n}_k,\end{aligned} \tag{37}$$

where step (d) is implemented because $\mathbf{p}_i^{\mathrm{SLP}} s_i$ is a constant vector for QAM modulated symbols as proved above in (32), which leads to $\mathbf{P}_i^{\mathrm{SLP}} \mathbf{s}_i = \mathbf{P}_j^{\mathrm{SLP}} \mathbf{s}_j, i \neq j$. While it may seem that the MLD method can be applied by simply scaling the received signal by a factor K, this assumption does not hold under closer examination. As shown in Section IV-B, the precoding matrix is rank-one, implying that its row vectors are related with each other. This dependency leads to the existence of multiple minimizers in the MLD criterion, or worse, causes the minimizer to deviate from the actual transmitted symbol vector, thereby rendering the MLD method ineffective for accurate signal detection. Based on this analysis, it becomes evident that the conventional SLP criterion is fundamentally incompatible with the MLD framework. To overcome this challenge, we next propose an SLP scheme that is explicitly designed to be compatible with the MLD criterion in MU-MIMO systems.

D. Proposed CSI-Free SLP-MLD Design

Based on the aforementioned analysis, we know that the smallest singular value of the SLP matrix should be maximized to enable MLD decoding in the users. However, in a typical MU-MIMO communication system, $\mathbf{P}$ is not a square matrix, which complicates the optimization of its smallest singular value. Therefore, we propose to decompose $\mathbf{P}$ into a Hermitian matrix and a normal matrix. The lower bound of the smallest singular value of $\mathbf{P}$ can be obtained by maximizing the smallest singular value of the corresponding Hermitian matrix. Consequently, a full-rank SLP matrix can be derived and the MLD method can be applied to decode the received signals.

Specifically, given that $\mathbf{P}$ is a $N_T \times KL$ dimensional matrix and $N_t \geq KL$ for MU-MIMO systems, $\mathbf{P}$ can be decomposed into two parts as

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix}, \tag{38}$$

where $\mathbf{P}_1 \in \mathbb{C}^{KL \times KL}$ is assumed to be a Hermitian matrix and $\mathbf{P}_2$ is a normal matrix of dimensions $(N_T - KL) \times KL$. Based on the aforementioned analysis, $\mathbf{P}$ will be full-rank if $\mathbf{P}_1$ is optimized to be a full-rank matrix. Furthermore, according to the fundamental rank properties [39], it can be established that

$$\sigma_{\min}(\mathbf{P}) \geq \sigma_{\min}(\mathbf{P}_1) = \lambda_{\min}(\mathbf{P}_1), \tag{39}$$

where $\sigma_{\min}(\mathbf{P})$ represents the smallest singular value of $\mathbf{P}$. It can be observed that the smallest eigenvalue of $\mathbf{P}_1$ is the lower bound of the smallest singular value of $\mathbf{P}$. Therefore, the task of maximizing the smallest singular value of $\mathbf{P}$ can be transformed into maximizing the smallest eigenvalue of $\mathbf{P}_1$. Thus, the smallest singular value maximization problem can be expressed as

$$\begin{aligned}\text{P7}: \max_{\mathbf{P}_1,\mathbf{P}_2,t,\mathbf{\Gamma}} \ & \lambda_{\min}(\mathbf{P}_1) \\ s.t. \quad & \mathbf{C1}: \mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix}, \mathbf{P}_1 \text{ is Hermitian matrix} \\ & \mathbf{C2}: \mathbf{GPs} = \mathbf{U}\text{diag}(\mathbf{\Gamma})\,\bar{\mathbf{s}} \\ & \mathbf{C3}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C4}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \\ & \mathbf{C5}: \|\mathbf{Ps}\|_2^2 \le p \\ & \mathbf{C6}: \ \mathbf{Ps} = K\mathbf{P}_k\mathbf{s}_k, \forall k \in \mathcal{K}, \end{aligned} \tag{40}$$

where $\mathbf{C1}$ ensures that $\mathbf{P}_1$ is a Hermitian matrix, $\mathbf{C2}-\mathbf{C4}$ transform the interference into the useful signal and $\mathbf{C6}$ leverages the benefits of the SLP method while effectively aligning them with the MLD criterion to realize the user independent MLD decoding. P7 is a convex optimization problem and can be directly solved by using the CVX tool.

Furthermore, the analysis of the MLD method in Section IV-A reveals that a larger smallest singular value of the transmit precoding matrix is more beneficial to the decoding efficiency of the users. Therefore, the CI constraints $\mathbf{C2}$-$\mathbf{C4}$, which aim to transform multi-user interference into CI are no longer necessary for optimizing the singular values, as users utilize the MLD method to decode received signals. Thus, we can reformulate P7 into

$$\begin{aligned}\text{P8}: \max_{\mathbf{P}_1,\mathbf{P}_2,z} \ & z \\ s.t. \quad & \mathbf{C1}: \|\mathbf{P}_1\mathbf{s}\|_2^2 + \|\mathbf{P}_2\mathbf{s}\|_2^2 \le p \\ & \mathbf{C2}: \mathbf{P}_1\mathbf{s} = K\mathbf{P}_{k,1}\mathbf{s}_k, \\ & \qquad \mathbf{P}_2\mathbf{s} = K\mathbf{P}_{k,2}\mathbf{s}_k, \forall k \in \mathcal{K} \\ & \mathbf{C3}: \begin{bmatrix} \mathcal{R}(\mathbf{P}_1) & -\mathcal{I}(\mathbf{P}_1) \\ \mathcal{I}(\mathbf{P}_1) & \mathcal{R}(\mathbf{P}_1) \end{bmatrix} - z\mathbf{I} \succeq \mathbf{0}, \end{aligned} \tag{41}$$

where $\mathbf{P}_{k,1}$ is a square matrix composed of the first L rows of $\mathbf{P}_k$, and $\mathbf{P}_{k,2}$ includes the last $N_T - L$ rows of $\mathbf{P}_k$, and the objective function z represents the optimized smallest singular value of $\mathbf{P}_1$. It can be observed that only the transmit symbols are utilized to design the SLP matrix, while the CSI matrices are omitted. This approach not only simplifies the optimization problem but also mitigates the performance losses associated with inaccurate channel estimation.

Notably, our goal is to maximize the smallest singular value of $\mathbf{P}_1$, which is uncorrelated with $\mathbf{P}_2$. Therefore, more power should be allocated to $\mathbf{P}_1$ and the variables can be decoupled by decomposing P8 into two subproblems as

$$\begin{aligned}\text{P9.1}: \min_{\mathbf{P}_2} \ & \|\mathbf{P}_2\mathbf{s}\|_2^2 \\ s.t. \quad & \mathbf{C1}: \mathbf{P}_2\mathbf{s} = K\mathbf{P}_{2,k}\mathbf{s}_k, \forall k \in \mathcal{K}, \end{aligned} \tag{42}$$

and

$$\begin{aligned}\text{P9.2}: \max_{\mathbf{P}_1,z} \ & z \\ s.t. \quad & \mathbf{C1}: \|\mathbf{P}_1\mathbf{s}\|_2^2 \le p \\ & \mathbf{C2}: \mathbf{P}_1\mathbf{s} = K\mathbf{P}_{1,k}\mathbf{s}_k, \forall k \in \mathcal{K} \\ & \mathbf{C3}: \begin{bmatrix} \mathcal{R}(\mathbf{P}_1) & -\mathcal{I}(\mathbf{P}_1) \\ \mathcal{I}(\mathbf{P}_1) & \mathcal{R}(\mathbf{P}_1) \end{bmatrix} - z\mathbf{I} \succeq \mathbf{0}. \end{aligned} \tag{43}$$

It is evident that the optimal solution to P9.1 is $\mathbf{P}_2^* = \mathbf{0}$, which further indicates that only $\mathbf{P}_1$ needs to be optimized. Moreover, P9.2 is a simpler convex problem with fewer variables, and can be solved by using the CVX tool.

Although P9.2 includes fewer variables that need to be optimized compared to P7, which simplifies the problem, this advantage may result in a degradation of communication performance. This degradation primarily arises from the omission of the CI constraints. First, these constraints convert interference into useful signals, thereby enhancing the receiver's decoding capability. Second, they ensure that $\mathbf{P}_2$ is a non-zero matrix, which can potentially increase the singular values of the precoding matrix $\mathbf{P}$.

V. Convergence and Computational Complexity Analysis

In this section, we analyze the convergence of the proposed AO algorithm for the joint design scheme. Subsequently, we derive the computational complexity of the proposed joint design scheme, the SSVMP scheme, the SDP-based scheme and the MLD method.

A. Convergence Analysis

The solution to our proposed joint design scheme is attained through the AO method, and the convergence of this algorithm is ensured for two key reasons: 1) After each iteration, the objective function value exhibits a monotonically increasing behavior. This characteristic guarantees that the algorithm consistently progresses toward an improved solution with each iteration [24]. 2) The proposed problem has an upper bound on the objective function value, determined by the CI constraints. These two observations ensure convergence, thereby maintaining the stability of the optimization process and preventing it from exceeding permissible boundaries [40]. The convergence is also validated by our numerical results to be shown in Section VI.

B. Computational Complexity Analysis

We begin by analyzing the computational complexity associated with the proposed **Algorithm 1**, where the transmit precoding matrix and the receive combining matrix are iteratively updated. First, the iterative algorithm introduced in [14] effectively solves the two QP subproblems to obtain the dual variables $\mathbf{u}_1$ and $\mathbf{u}$. Both QP problems have the same number of variables, and the primary computational complexity arises from matrix inversion, leading to a computational complexity

of $\mathcal{O}((2KL)^3)$ for each subproblem. Second, the most computationally expensive operation within the calculation of the two closed-form solutions is also the matrix inversion, which has a complexity of approximately $\mathcal{O}(8K^3L^3)$. Consequently, if we denote N_{iter} as the total number of iterations, the overall computational complexity of **Algorithm 1** can be expressed as $N_{iter}(2\mathcal{O}((2KL)^3) + \mathcal{O}(8K^3L^3))$ [24].

Besides using the receive combining matrix, we propose to utilize the MLD method to decode the received signals at the receivers. The computational complexity of the MLD method is KL^{M_c}, where M_c represents the modulation order of QAM constellation , and each user independently decodes the data symbols. It is evident that the computational complexity increases exponentially, leading to a significant computational burden in MU-MIMO systems. To address this issue, the QRM-MLD method is considered as an alternative, with a computational complexity of $K\left[M_c + M\left(L-1\right)\right]$, where M denotes the number of surviving symbol candidates that must be calculated starting from the second stage. Clearly, QRM-MLD method significantly reduces the computational complexity.

Additionally, two novel symbol-level singular value optimization problems are introduced to enhance the decoding efficiency. SSVMP in (40) is a convex matrix block optimization problem that can be directly solved by using the CVX tool, resulting in a computational complexity of $\mathcal{O}((N_TKL)^{3.5})$, where the number of variables is N_TKL. To further reduce computational complexity, an SDP-based problem is introduced, which requires only the calculation of $\mathbf{P}_1$. The computational complexity is $\mathcal{O}((KL)^7)$, with the number of variables reduced to K^2L^2.

VI. Simulation Results

In this section, we provide a comparative analysis of the numerical results for the proposed joint design scheme alongside two singular value optimization problems, in contrast to the traditional SLP and BD precoding schemes paired with MLD estimation at the receivers. MonteCarlo simulations are utilized as the evaluation methodology. For each time slot, we set $P_T = 1\,\mathrm{W}$, and the transmit SNR is defined as $\rho = 1/\sigma^2$. The elements of the channel matrix $\mathbf{H}$ are assumed to follow a standard complex Gaussian distribution, specifically $\mathbf{H}_{m,n} \sim \mathcal{CN}(0,1)$. To ensure clarity, the following abbreviations will be consistently used throughout this section:

1) 'Traditional SLP': traditional optimization-based SLP scheme based on P2.
2) 'BD': traditional BD scheme as proposed in [6];
3) 'Joint Design': proposed joint symbol-level transmit precoding and receive combining scheme based on **Algorithm 1**;
4) 'SSVMP': proposed smallest singular value maximization problem (SSVMP) based on P7;
5) 'SDP': proposed SDP-based singular value optimization problem based on P9.2;
6) 'MLD', 'QR-MLD' or 'QRM-MLD': the MLD, QR-MLD or QRM-MLD methods used at the user side to decode the receive signals.

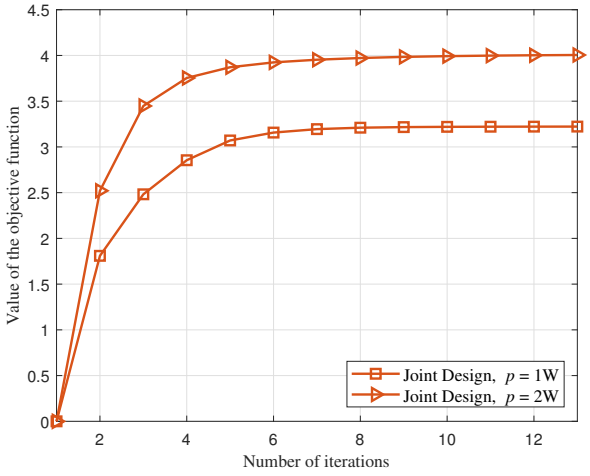


Fig. 2. Convergence for the proposed Joint Design scheme, 16QAM, $N_T = 16, N_R = 8, L = 4$ and $K = 2$.

The proposed AO algorithm is used to solve the 'Joint Design' problem and Fig. 2 illustrates the convergence of this approach using a 16QAM constellation with $N_T = 16, N_R = 8, L = 4$ and $K = 2$. In each iteration, the optimized objective t between the current and the previous iteration for different transmit power settings is distinguished, and it increases with the transmit power. This is because more power can be used to increase the distance between the detection threshold and the CI region. We also observe that the proposed algorithm converges only within several iterations. This finding aligns well with the convergence analysis detailed in Section V-A, reinforcing the validity of the proposed approach.

In Fig. 3 and Fig. 4, we compare the BER performance of the proposed against the 'Traditional SLP' scheme when using various decoding techniques for 16QAM and 64QAM constellations, with parameters set to $N_T = 16, N_R = 8, L = 4$ and $K = 2$. It can be observed that the BER performance of the 'Traditional SLP + MLD' scheme experiences significant degradation, which validates the analysis proposed in Section IV-A and Section IV-B and highlights the necessity of our work. Furthermore, our proposed schemes achieves promising BER performance, this is attributed to the utilization of symbol-by-symbol optimization and the MLD method. In both figures, the 'SSVMP' scheme achieves the best performance, and the 'SDP' scheme incurs only an acceptable performance loss compared to the 'SSVMP' scheme. In addition, the 'QR-MLD' method can achieve the same performance with 'MLD' method with lower computational complexity. To further reduce the computational complexity, the QRM-MLD method is a more practical and efficient alternative. Therefore, we evaluate the BER performance of the 'QRM-MLD' method for clarity, setting $M = 8$ and $M = 16$. Notably, when $M = 16$, the 'MLD' method is equivalent to the 'QRM-MLD' method for 16QAM constellation. It can be observed that the 'QRM-MLD' method exhibits a slight performance loss compared to the 'MLD' method, and a larger value of M leads to better performance, this indicates there is a trade-off

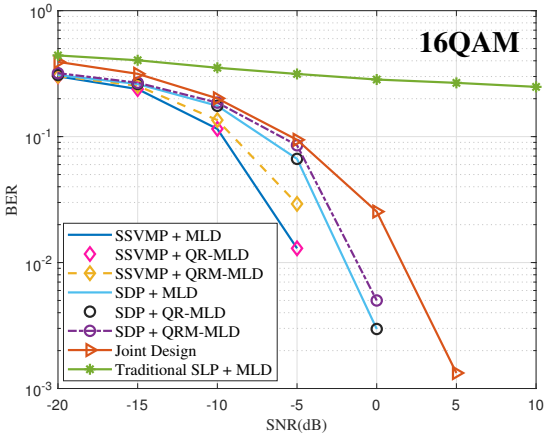


Fig. 3. BER v.s. SNR, 16QAM, $N_T = 16, N_R = 8, L = 4, K = 2$ and $M = 8$.

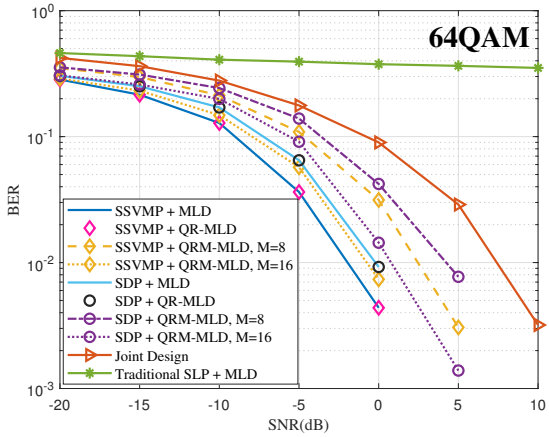


Fig. 4. BER v.s. SNR, 64QAM, $N_T = 16, N_R = 8, L = 4, K = 2$ and $M = 8$ *or* 16.

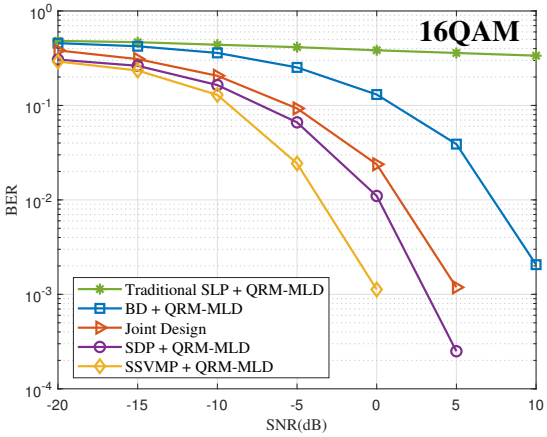


Fig. 5. BER v.s. SNR, 16QAM, $N_T = 32, N_R = 8, L = 4, K = 2$ and $M = 8$.

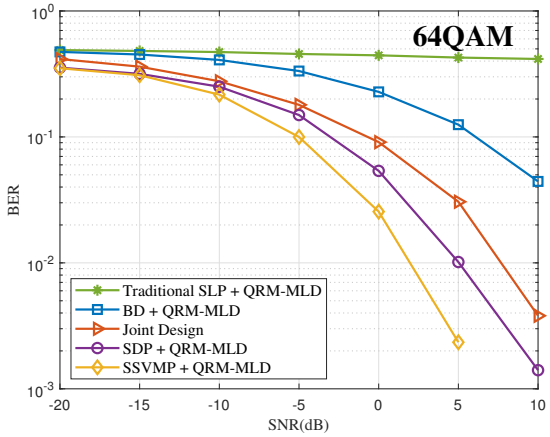


Fig. 6. BER v.s. SNR, 64QAM, $N_T = 32, N_R = 8, L = 4, K = 2$ and $M = 8$.

between the performance and complexity. To further verify the effectiveness of the proposed schemes, subsequent simulations will directly employ the 'QRM-MLD' method. Additionally, Fig. 3 achieves a lower BER compared to Fig. 4, this is due to the smaller distance between symbols and there are more inner points at higher modulation orders, which increases the likelihood of decoding errors.

In Fig. 5 and Fig. 6, we compare the BER performance of our proposed schemes to 'Traditional SLP' and 'BD' schemes employing 16QAM and 64QAM constellations. In both figures, $N_T = 32, N_R = 8, L = 4, K = 2$ and $M = 8$. It can be observed that the BER decreases with the increase of the SNR, and higher BER gains can be achieved in the high SNR regime. It's evident that the proposed SLP schemes significantly outperform the 'Traditional SLP' and 'BD' schemes, and the performance of the 'Traditional SLP + QRM-MLD' completely fails, which indicates that traditional SLP approach is not compatible with MLD method. Furthermore, the 'SSVMP' and 'SDP' schemes demonstrate a more substantial performance improvement over the 'Joint Design' scheme, benefiting from the SLP structure optimized specifically for the MLD method. In addition, Fig. 5 exhibits a lower BER than Fig. 6 at the same SNR levels, which can be attributed to the transmission of symbols with a reduced modulation order.

Fig. 7 shows the BER performance with respect to the number of data streams per user employing 16QAM constellation. In the figure, the solid and the dashed lines represent $L = 4$ and $L = 8$, respectively, with $N_T = 32, N_R = 16, K = 2$. Similar BER performance is observed across the different schemes. However, the BER increases as the number of data streams grows under the same schemes, this can be attributed to the reduction in design flexibility resulting from the increased number of data streams. Furthermore, it can be observed that when the number of data streams is relatively large, the performance difference between the 'SSVMP' scheme and the 'SDP' scheme decreases, indicating that the advantages offered by the 'SSVMP' approach become less pronounced under conditions of high data stream density. This observation renders the low-complexity 'SDP' scheme more practical.

Fig. 8 depicts the BER performance of various precoding

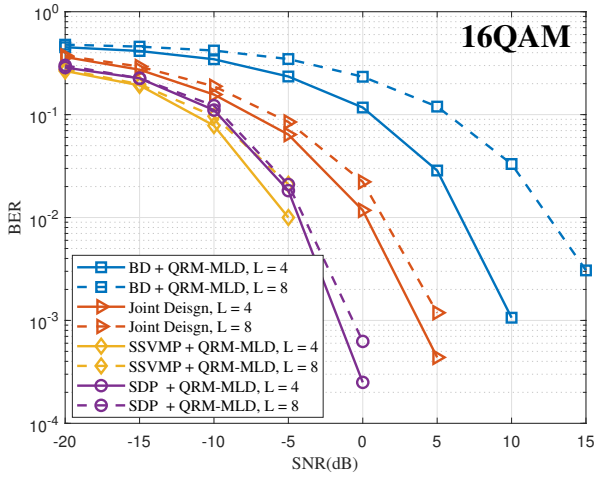


Fig. 7. BER v.s. SNR for different number of data streams per user, 16QAM, $N_T = 32, N_R = 16$ and $K = 2$.

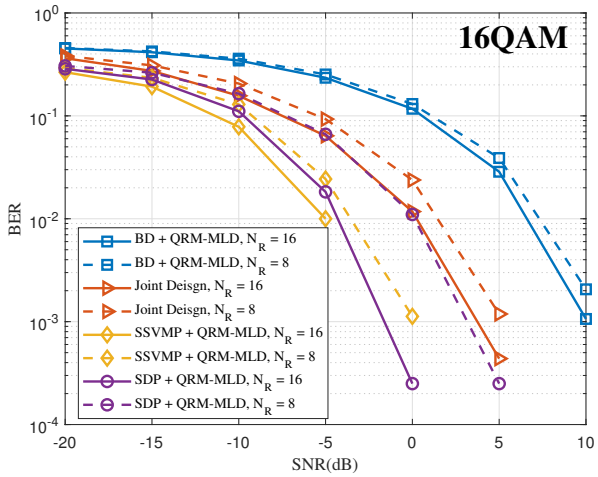


Fig. 8. BER v.s. SNR for different number of receive antennas, 16QAM, $N_T = 32, L = 4$ and $K = 2$.

approaches with respect to the number of the receive antennas while using 16QAM constellation, with $N_T = 32, L = 4, K = 2$. In the figure, the solid and the dashed lines represent $N_R = 16$ and $N_R = 8$, respectively. It can be observed that BER decreases with the increase of the number of the receive antennas. This is because when users are equipped with more receive antennas, the rank of the channel matrix $\mathbf{H}_k$ increases, which improves the estimation performance of the MLD method. Furthermore, an increased number of receive antennas provides enhanced design flexibility, facilitating a more effective implementation of the 'Joint Design' scheme.

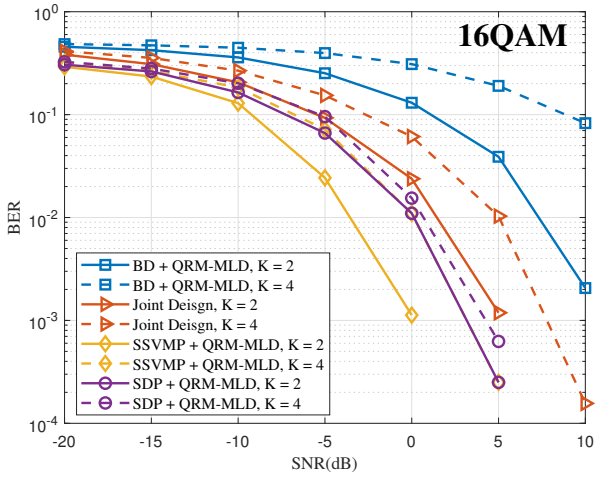


Fig. 9. BER v.s. SNR for different number of users, 16QAM, $N_T = 32, N_R = 8$ and $L = 4$.

Fig. 9 illustrates the relationship between BER and SNR as the number of users varies, where 16QAM constellation is used and $N_T = 32, N_R = 8, L = 4$. In the figure, the solid and the dashed lines represent $K = 2$ and $K = 4$, respectively. It's evident that the increase in the number of users clearly leads to a rise in the BER. This is due to the fact that as more users are added, there are fewer design freedoms available, reaching a critical point when $N_T = KL$. At this stage, the one-to-one relationship between transmit antennas and data streams imposes limitations on performance.

In addition, to provide a comprehensive comparison and facilitate a clear understanding of the practical implications of our work, Table I presents a summary of all the schemes discussed in this paper, including their advantages, disadvantages, and suitable application scenarios.

VII. Conclusion

This paper studies symbol-level transmit precoding and receive decoding for MU-MIMO systems with QAM constellations. A joint design to optimize transmit precoding and receive combining matrices is proposed, and the optimal variable structures are derived. MLD is employed at the receiver to

TABLE I
COMPARISON OF DIFFERENT PRECODING SCHEMES IN MU-MIMO SYSTEMS

Schemes	Advantages	Disadvantages	Applicable Scenarios
Joint Design	Closed-form expression available	Require iterative procedures	MU-MIMO
SSVMP+MLD	Near-optimal performance under MLD	Increased computational complexity	MU-MIMO
SDP+MLD	Reduced computational complexity	Moderate performance degradation	MU-MIMO
Traditional SLP	Exploits interference constructively	Incompatible with MLD principle	MU-MISO
BD	Simple and easy to implement	Performance bottleneck exists	MU-MIMO

address the dependence of the combining matrix on transmit symbols. To overcome the rank-deficiency of the SLP matrix, two symbol-level singular value optimization problems are introduced to enhance decoding efficiency. Numerical results show that the proposed schemes are MLD-compatible and outperform the conventional BD-based precoding methods. Future work includes robust design under channel uncertainty and low-complexity, interference-aware detection.

Appendix A
Proof for Proposition 1

We first use the row vectors $\mathbf{w}_{k,l}$ to replace the original matrix variables $\mathbf{W}_k$, then we can express P4 as

$$\begin{aligned}\text{P10}: \min_{\mathbf{w}_{k,l},t,\boldsymbol{\Gamma}} \quad & -t \\ s.t. \quad & \mathbf{C1}: \mathbf{w}_{k,l}\mathbf{r}_k = \boldsymbol{\gamma}_{k,l}^T\bar{\mathbf{s}}_{k,l}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L} \\ & \mathbf{C2}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \\ & \mathbf{C4}: \sum_{l=1}^{L} \mathbf{w}_{k,l}\mathbf{w}_{k,l}^H \le 1, \forall k \in \mathcal{K},\end{aligned} \tag{44}$$

where $\mathbf{w}_{k,l}$ is the l-th row of the receive combining matrix $\mathbf{W}_k$. The Lagrangian function of P10 can be expressed as

$$\begin{aligned}&\mathcal{L}\left(\mathbf{w}_{k,l}, t, \delta_{k,l}, \mu_m, \nu_n, \varphi_k\right) \\ &= -t + \sum_{k=1}^{K}\sum_{l=1}^{L}\delta_{k,l}\left(\mathbf{w}_{k,l}\mathbf{r}_k - \boldsymbol{\gamma}_{k,l}^T\bar{\mathbf{s}}_{k,l}\right) \\ &\quad + \sum_{m=1}^{card\{\mathcal{O}\}}\mu_m\left(t - \gamma_m^{\mathcal{O}}\right) + \sum_{n=1}^{card\{\mathcal{I}\}}\nu_n\left(t - \gamma_n^{\mathcal{I}}\right) \\ &\quad + \sum_{k=1}^{K}\varphi_k\left(\sum_{l=1}^{L}\mathbf{w}_{k,l}\mathbf{w}_{k,l}^H - 1\right),\end{aligned} \tag{45}$$

where $\delta_{k,l}$, $\mu_m \ge 0$, ν_n and $\varphi_k \ge 0$. Each $\delta_{k,l}$ and ν_n may be complex since they are dual variables corresponding to the equality constraints. Based on (45), the KKT conditions can be derived as

$$\frac{\partial\mathcal{L}}{\partial\mathbf{w}_{k,l}} = \delta_{k,l}\mathbf{r}_k + \varphi_k\mathbf{w}_{k,l}^H = \mathbf{0}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L} \tag{46a}$$

$$\frac{\partial\mathcal{L}}{\partial t} = -1 + \sum_{m=1}^{card\{\mathcal{O}\}}\mu_m + \sum_{n=1}^{card\{\mathcal{I}\}}\nu_n = 0 \tag{46b}$$

$$\mathbf{w}_{k,l}\mathbf{r}_k - \boldsymbol{\gamma}_{k,l}^T\bar{\mathbf{s}}_{k,l} = 0, \forall k \in \mathcal{K}, \forall l \in \mathcal{L} \tag{46c}$$

$$\mu_m\left(t - \gamma_m^{\mathcal{O}}\right) = 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \tag{46d}$$

$$t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I} \tag{46e}$$

$$\varphi_k\left(\sum_{l=1}^{L}\mathbf{w}_{k,l}\mathbf{w}_{k,l}^H - 1\right) = 0. \tag{46f}$$

In order to satisfy conditions (46a) and $\varphi_k \ge 0$, it can be concluded that $\varphi_k > 0$. Therefore, $\mathbf{w}_{k,l}$ can be derived as

$$\begin{aligned}&\mathbf{w}_{k,l}^H = -\frac{\delta_{k,l}}{\varphi_k}\mathbf{r}_k = \zeta_{k,l}\mathbf{r}_k \\ \Rightarrow\ &\mathbf{w}_{k,l} = \zeta_{k,l}^*\mathbf{r}_k^H, \forall k \in \mathcal{K}, \forall l \in \mathcal{L},\end{aligned} \tag{47}$$

where we define $\zeta_{k,l} \triangleq -\dfrac{\delta_{k,l}}{\varphi_k}$. The receive combining matrix can be further expressed as

$$\mathbf{W}_k = \begin{bmatrix}\mathbf{w}_{k,1}\\ \mathbf{w}_{k,2}\\ \vdots \\ \mathbf{w}_{k,L}\end{bmatrix} = \begin{bmatrix}\zeta_{k,1}^*\mathbf{r}_k^H\\ \zeta_{k,2}^*\mathbf{r}_k^H\\ \vdots \\ \zeta_{k,L}^*\mathbf{r}_k^H\end{bmatrix} = \begin{bmatrix}\zeta_{k,1}^*\\ \zeta_{k,2}^*\\ \vdots \\ \zeta_{k,L}^*\end{bmatrix}\mathbf{r}_k^H = \boldsymbol{\zeta}_k^*\mathbf{r}_k^H, \tag{48}$$

where $\boldsymbol{\zeta}_k = [\zeta_{k,1}, \zeta_{k,2}, \cdots, \zeta_{k,L}]^T$. According to constraint **C1** in P4, we have

$$\begin{aligned}&\mathbf{W}_k\mathbf{r}_k = \mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k \\ \Rightarrow\ &\boldsymbol{\zeta}_k^*\mathbf{r}_k^H\mathbf{r}_k = \mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k \\ \Rightarrow\ &\boldsymbol{\zeta}_k^* = \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k.\end{aligned} \tag{49}$$

Using (49), we can obtain the optimal receive combining matrix $\mathbf{W}_k$ is a function of $\boldsymbol{\Gamma}_k$ can be expressed as

$$\mathbf{W}_k = \boldsymbol{\zeta}_k^*\mathbf{r}_k^H = \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k\mathbf{r}_k^H. \tag{50}$$

Based on above observation, we conclude that $\varphi_k > 0$, therefore, the normalized requirement should be strictly active when optimality is achieved. Then, we could have

$$\begin{aligned}&\|\mathbf{W}_k\|_2^2 = 1 \\ &\Rightarrow \text{Tr}\left[\frac{1}{\left(\mathbf{r}_k^H\mathbf{r}_k\right)^2}\mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k\mathbf{r}_k^H\mathbf{r}_k\bar{\mathbf{s}}_k^H\text{diag}\left(\boldsymbol{\Gamma}_k\right)\mathbf{U}_1^H\right] = 1 \\ &\Rightarrow \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\text{Tr}\left[\mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k\bar{\mathbf{s}}_k^H\text{diag}\left(\boldsymbol{\Gamma}_k\right)\mathbf{U}_1^H\right] = 1 \\ &\Rightarrow \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\bar{\mathbf{s}}_k^H\text{diag}\left(\boldsymbol{\Gamma}_k\right)\mathbf{U}_1^H\mathbf{U}_1\text{diag}\left(\boldsymbol{\Gamma}_k\right)\bar{\mathbf{s}}_k = 1 \\ &\Rightarrow \frac{1}{\mathbf{r}_k^H\mathbf{r}_k}\boldsymbol{\Gamma}_k^T\underbrace{\text{diag}\left(\bar{\mathbf{s}}_k^H\right)\mathbf{U}_1^H\mathbf{U}_1\text{diag}\left(\bar{\mathbf{s}}_k\right)}_{\mathbf{T}_1}\boldsymbol{\Gamma}_k = 1 \\ &\Rightarrow \boldsymbol{\Gamma}_k^T\mathbf{T}_1\boldsymbol{\Gamma}_k = \mathbf{r}_k^H\mathbf{r}_k,\end{aligned} \tag{51}$$

where $\mathbf{T}_1 = \text{diag}\left(\bar{\mathbf{s}}_k^H\right)\mathbf{U}_1^H\mathbf{U}_1\text{diag}\left(\bar{\mathbf{s}}_k\right)$. Since $\mathbf{T}_1$ is a positive semi-definite Hermitian matrix and each entry of $\boldsymbol{\Gamma}_k$ is real, (51) can be further transformed into

$$\boldsymbol{\Gamma}_k^T\mathbf{T}_1\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^T\Re\left(\mathbf{T}_1\right)\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^T\mathbf{V}_1\boldsymbol{\Gamma}_k = \mathbf{r}_k^H\mathbf{r}_k, \tag{52}$$

where $\mathbf{V}_1 = \Re\left(\mathbf{T}_1\right)$ is a symmetric matrix. Therefore, P4 can be further transformed into an equivalent optimization on $\boldsymbol{\Gamma}$, given by

$$\begin{aligned}\text{P11}: \min_{t,\boldsymbol{\Gamma}} \quad & -t \\ s.t. \quad & \mathbf{C1}: \boldsymbol{\Gamma}_k^T\mathbf{V}_1\boldsymbol{\Gamma}_k = \mathbf{r}_k^H\mathbf{r}_k, \forall k \in \mathcal{K} \\ & \mathbf{C2}: t - \gamma_m^{\mathcal{O}} \le 0, \forall \gamma_m^{\mathcal{O}} \in \mathcal{O} \\ & \mathbf{C3}: t - \gamma_n^{\mathcal{I}} = 0, \forall \gamma_n^{\mathcal{I}} \in \mathcal{I}.\end{aligned} \tag{53}$$

The optimal receive combining matrix for P4 can be obtained by substituting the solution of P11 into (50). Additionally, P11 has the same form as $\mathcal{P}_4$ in [14], then $\boldsymbol{\Gamma}_k$ can be further derived based on *Proposition* 2 presented in [14]. By substituting the expression for $\boldsymbol{\Gamma}_k$ into $\mathbf{W}_k$ in (50), the closed-form structure (16) is obtained, which completes the proof.

REFERENCES

[1] H. Lu et al., "A Tutorial On Near-Field XL-MIMO Communications Towards 6G," *IEEE Commun. Surveys Tuts.*, Early Access, 2024.

[2] F. Dong, F. Liu, Y. Cui, W. Wang, K. Han and Z. Wang, "Sensing as a Service in 6G Perceptive Networks: A Unified Framework for ISAC Resource Allocation," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3522-3536, May 2023.

[3] Z. Wang et al., "A Tutorial on Extremely Large-Scale MIMO for 6G: Fundamentals, Signal Processing, and Applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560-1605, 2024.

[4] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9,pp. 4409–4418, Sep. 2008.

[5] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[6] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[7] K. Zu, R. C. de Lamare, and M. Haardt, "Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 10,pp. 4232–4242, Oct. 2013.

[8] S. Lu et al., "Random ISAC Signals Deserve Dedicated Precoding," *IEEE Trans. Signal Process.*, vol. 72, pp. 3453-3469, 2024.

[9] M. Alodeh et al., "Symbol-Level and Multicast Precoding for Multiuser Multiantenna Downlink: A State-of-the-Art, Classification, and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1733-1757, 2018.

[10] A. Li et al., "A tutorial on interference exploitation via symbol-level precoding: Overview, state-of-the-art and future directions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 796–839, 3rd Quart., 2020.

[11] C. Masouros and E. Alsusa, "A Novel Transmitter-Based Selective-Precoding Technique for DS/CDMA Systems," in *2007 IEEE International Conference on Communications*, Glasgow, UK, pp. 2829-2834, 2007.

[12] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1396–1404, Mar. 2009.

[13] A. Li and C. Masouros, "Interference exploitation precoding made practical: Optimal closed-form solutions for PSK modulations," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7661–7676, 2018.

[14] A. Li, C. Masouros, B. Vucetic, Y. Li, and A. L. Swindlehurst, "Interference exploitation precoding for multi-level modulations: Closed-form solutions," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 291–308, 2021.

[15] Z. Xiao, R. Liu, M. Li, Y. Liu, and Q. Liu, "Low-complexity designs of symbol-level precoding for MU-MISO systems," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4624–4639, Jul. 2022.

[16] Y. Xu, F. Fang, S. Wang and D. Cai, "Energy Efficiency Maximization for Downlink NOMA Designs via Symbol-Level Precoding," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 14974-14987, Oct. 2024.

[17] S. Zhang, B. Di and H. Zhang, "Energy Efficient Symbol Level Precoding Design in a RIS Assisted Communication System," *IEEE Commun. Lett.*, vol. 28, no. 5, pp. 1141-1145, May 2024.

[18] Z. Liao and F. Liu, "Symbol-Level Precoding for Integrated Sensing and Communications: A Faster-Than-Nyquist Approach," *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3210-3214, Dec. 2023.

[19] N. Babu, A. Kosasih, C. Masouros and E. Björnson, "Symbol-Level Precoding for Near-Field ISAC," *IEEE Commun. Lett.*, vol. 28, no. 9, pp. 2041-2045, Sept. 2024.

[20] Y. Chen, F. Liu, Z. Liao and F. Dong, "Symbol-Level Precoding for MIMO ISAC Transmission Based on Interference Exploitation," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 283-287, Feb. 2024.

[21] E. S. P. Lopes, L. T. N. Landau and A. Mezghani, "Minimum Symbol Error Probability Discrete Symbol Level Precoding for MU-MIMO Systems With PSK Modulation," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5935-5949, Oct. 2023.

[22] Y. Wang, H. Hou, W. Wang and X. Yi, "Symbol-Level Precoding for Average SER Minimization in Multiuser MISO Systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1103-1107, April 2024.

[23] Z. Wei, C. Masouros and T. Xu, "Constructive Interference based Joint Combiner and Precoder Design in Multiuser MIMO Systems," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, Montreal, QC, Canada, 2021.

[24] X. Tong, A. Li, L. Lei, F. Liu and F. Dong, "Symbol-Level Precoding for MU-MIMO System With RIRC Receiver," *IEEE Trans. Commun.*, vol. 72, no. 5, pp. 2820-2834, May 2024.

[25] S. Cai, T.-H. Chang, and H. Zhu, "Joint symbol level precoding and receive beamforming for multiuser MIMO downlink," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.

[26] S. Cai, H. Zhu, C. Shen, and T.-H. Chang, "Joint symbol level precoding and receive beamforming optimization for multiuser MIMO downlink," *IEEE Trans. Signal Process.*, vol. 70, pp. 6185–6199, 2022.

[27] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941-1988, Fourthquarter 2015.

[28] K. Kato, K. Fukawa, R. Yamada, H. Suzuki and S. Suyama, "Low-Complexity MIMO Signal Detection Employing Multistream Constrained Search," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1217-1230, Feb 2018.

[29] M. Matthaiou, N. D. Chatzidiamantis, G. K. Karagiannidis and J. A. Nossek, "ZF Detectors over Correlated K Fading MIMO Channels," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1591-1603, June 2011.

[30] Y. Jiang, M. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. Inf. Theory.*, vol. 57, no. 4, pp. 2008–2026, Apr 2011.

[31] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory.*, vol. 35, no. 1, pp. 123–136, Jan 1989.

[32] X. Zhu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 187–191, Feb 2002.

[33] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory.*, vol. 49, no. 10, pp. 2389–2402, Oct 2003.

[34] K. J. Kim and J. Yue, "Joint channel estimation and data detection algorithms for MIMO-OFDM systems," in *Proc. 36th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, USA, pp. 1857–1861, Nov 2002.

[35] K. Higuchi, H. Kawai, N. Maeda, H. Taoka and M. Sawahashi, "Experiments on real-time 1-Gb/s packet transmission using MLD-based signal detection in MIMO-OFDM broadband radio access," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 6, pp. 1141-1153, June 2006.

[36] Z. Xiao, R. Liu, M. Li, Y. Liu and Q. Liu, "Low-Complexity Designs of Symbol-Level Precoding for MU-MISO Systems," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4624-4639, July 2022.

[37] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, Jul. 2015.

[38] M. Alodeh, S. Chatzinotas and B. Ottersten, "Symbol-Level Multiuser MISO Precoding for Multi-Level Adaptive Modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5511-5524, Aug. 2017.

[39] X. Zhang, Matrix analysis and application. Tsinghua, C.H.N.: Tsinghua Univ. Press, 2013.

[40] L. Vandenberghe and S. Boyd, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.