Exploiting Power Amplifier Nonlinearities through Symbol-Level Interference Exploitation Precoding in the MU-MIMO Downlink

Guorui Wei, *Student Member, IEEE*, Ang Li, *Senior Member, IEEE*, and Christos Masouros, *Senior Member, IEEE*

***Abstract*—In this paper, we study the interference exploitation precoding in the presence of distortion from nonlinear power amplifiers (PAs) in multi-user multiple-input single-output (MU-MISO) downlink communication systems. We consider the memoryless polynomial model of nonlinear PAs, which is incorporated into the symbol-level precoding (SLP) design to allow the PA nonlinearities in the constructive interference (CI) exploitation. The optimization problem that aims to enhance the signal-to-interference-plus-noise ratio (SINR) without investing additional transmit signal power is formulated for both PSK and QAM signaling. Since the original optimization problem is nonconvex, we first introduce auxiliary variables to transform the optimization problem and adopt the alternating optimization framework for the new optimization problem. For non-convex subproblems, additional auxiliary variables are introduced and several approximations are employed to transform the problem into a semidefinite programming (SDP) form, where the semidefinite relaxation (SDR) method is adopted to obtain feasible solutions. In order to reduce the computational cost of the iterative algorithm, we further propose a low-complexity algorithm for the original PA-aware SLP optimization problem. Numerical results verify the superiority of our proposed PA-aware SLP approach in the presence of nonlinear PAs in the MU-MISO downlink in terms of the error-rate performance over the state-of-the-art.**

***Index Terms*—MIMO, symbol-level precoding, constructive interference, nonlinear power amplifier, optimization.**

I. Introduction

With the rapid expansion of the application field of wireless communication technology, the wireless communication devices and wireless data services have grown dramatically during the past decades. According to the global mobile data traffic forecast by Cisco [1], the global mobile data traffic will continue to grow at an annual rate of more than 60% in the coming years. Therefore, the future mobile communication technology needs to meet this rapid growth of mobile data traffic. Massive multiple-input multiple-output (MIMO) technology can significantly improve the spectral efficiency and the capacity of wireless communication systems [2], which has attracted considerable attention in both the academic research and the wireless industry in recent years.

For the ideal fully-digital massive MIMO architecture, without considering the limited resolution of the digital-to-analog converters (DACs) and the nonlinearity of the power amplifiers (PAs), research has shown that a simple frame-level linear precoding scheme can achieve transmission performance close to Shannon's theoretical limit [3]. However, downlink transmission of MIMO systems requires the base station (BS) to equip each transmit antenna element with a separate radio frequency (RF) chain (including a pair of high-resolution DACs, a linear PA, mixers, and so on). Since massive MIMO system contains hundreds or thousands of antenna elements, the fully-digital massive MIMO BS architecture presents high hardware complexity and power consumption, which have a significant impact on the performance and power efficiency of the communication system [4]. Hence, fully-digital massive MIMO may not be preferable in practical implementation, and the study for reducing the power consumption of BS in massive MIMO plays an important role for their widespread deployment in cellular mobile communication systems.

To provide the trade-off between achievable spectral efficieny and power consumption in massive MIMO architecture, the hybrid analog/digital architecture, where the MIMO signal processing is divided into analog and digital domains, can reduce the number of required RF chains and attracted extensive attention from both academia and industry [5], [6]. In addition to the hybrid architecture, employing low-resolution DACs instead of high-resolution DACs in massive MIMO architecture can effectively reduce the power consumption of BS by reducing the power consumption per RF chain instead of reducing the number of RF chains [7], [8]. However, low-resolution DAC introduces severe signal distortion to massive MIMO systems, which is difficult to compensate by traditional precoding schemes and usually requires symbol-level processing to achieve satisfactory performance [9]–[11].

Similar to the low-bit DAC architecture, employing power-efficient nonlinear PAs in massive MIMO system can also reduce the power consumption of each RF chain, thereby improving the energy efficiency of massive MIMO communication systems. However, in the traditional multi-antenna system, due to the limited linear region of nonlinear PAs, the transmit signals with a high peak-to-average power ratio (PAPR) will introduce non-negligible signal distortions, which has negative effects on the performance of the communication system. A number of studies have been carried out on PAPR problem, which can be divided into two directions: a) to keep the signal power constant by designing the constant envelope precoding (CEP) scheme [12]–[18], which can be classified as SLP schemes; b) to reduce the PAPR of the transmit signal by optimizing the frame-level precoding matrix [19]–[24]. By limiting the amplitude of the transmit signal to a constant value, the CEP can completely eliminate the

performance loss introduced by nonlinear PAs. To be more specific, [12] first studied the CEP in single-user massive communication system, while the optimal CEP scheme in the single-user case is further given in [13]. Based on [12], the joint design scheme of CEP and antenna selection in single-user scene was considered in [14], which presented a new idea to design the precoding scheme from a geometric perspective. [15] extended the CEP to multiuser communication system for the first time, where a nonlinear least squares precoding optimization problem was constructed to minimize the multi-user interference (MUI), followed by an efficient iterative solution algorithm. [16] presented the optimization of CEP based on the cross-entropy algorithm, which further improves the performance of the scheme in [15].

In addition to the CEP, there exist other low-PAPR transmission schemes which relax the strict CE constraint by allowing the maximum PAPR to a certain value. For example, a low-PAPR transmission scheme based on minimizing the dynamic range of the transmit signals was presented in [19]. In [20], a zero forcing (ZF) scheme was proposed, which reduced the PAPR by limiting the input power of each antenna. Based on the nonlinear vector perturbation (VP) precoding scheme, [21] proposed a method to optimize the transmit signal norm, which achieved a compromise among transmission performance, PAPR and computational complexity. [22] introduced the internal back-off scheme to limit the maximum input power of the PAs such that they work in the linear region. [23] proposed symbol-level precoding (SLP) for MU-MISO systems with nonlinear PAs, which can minimize the transmit power of each antenna or minimize the spatial PAPR (SPAPR). By studying the relationship between the PAPR and the bit-error-rate (BER) at users, [24] introduced PAPR as an additional constraint into the optimization problem that aimed for error-rate minimization, where a distributed solution was proposed. Despite the above precoding schemes for nonlinear PAs, few of them have taken the response characteristics of the nonlinear PAs and their specific impact on the wireless transmission signal into consideration.

Very recently, studies that utilize PA's response characteristics for the precoding design in the presence of nonlinear PAs have appeared in the literature, rather than only reducing the PAPR of the transmit signal. [25] used a clipping function to model the response characteristics of nonlinear PAs, where the precoder was designed to resist the MUI and the nonlinearities of PA. [26] suggested a distortion-aware beamforming (DAB) algorithm for MU-MISO communication system, where an iterative algorithm is proposed for rate maximization. [27] investigated a PA-aware precoding scheme in massive MU-MIMO downlink system, and developed an efficient algorithm to reduce the MUI and PA nonlinearity. In a single-user MISO communication system with nonlinear PAs at BS, [28] developed a power control method and a precoding scheme that maximized the received SINR, where an iterative precoding algorithm was presented. [29] jointly optimized the precoding and power allocation strategy to maximize the achievable sum rate of MU-MIMO systems. [30] investigated the effect of PA nonlinearity for the downlink MU-MIMO orthogonal frequency division multiplexing (OFDM) system in a correlated channel, and derived the analytical signal to distortion, interference and noise ratio (SINDR).

Although some performance benefits can be observed in [25]–[30] by considering the nonlinear PA characteristics, these schemes view the MUI in the communication system and the signal distortion introduced nonlinear PA as harmful interference that need to be eliminated. Nevertheless, such design criterion may be sub-optimal since it has been shown that known interference can further benefit the system performance, achieved by constructive interference (CI) and SLP [31]. In [32], the interference in communication system was categorized into constructive interference (CI) and destructive interference (DI) from symbol level for the first time. Based on this concept, [33] proposed a modified zero-forcing (ZF) precoding, which only forces DI to be zero and exploits CI. In [34], a correlation rotation scheme was further proposed, in which it is shown that all interference can become CI through manipulating and rotating DI. In order to relax the strict phase rotation constraints in [34] and improve performance, an SLP scheme based on CI convex optimization was presented in [35] and [36], which also introduced the concept of constructive region. However, optimization-based CI precoding methods may be computationally inefficient since the convex optimization problem needs to be solved symbol by symbol. In [37], optimal and low-complexity solution for CI precoding based on SINR balancing optimization was designed, where an iterative precoder was formed to achieve a compromise between performance and complexity. In addition, [38] further presented a general form for multi-level modulation. Because of the performance benefits that CI can offer, CI-based precoding design has been applied to massive MIMO systems with limited hardware. For example, [39] proposed several transmit beamforming schemes for the massive MIMO downlink with 1-bit DACs based on CI, which greatly improves the performance of low-resolution massive MIMO system. [17] introduced a CI-based CEP method for generic PSK modulations. Furthermore, a low-complexity manifold algorithm for the CI-based CEP was presented in [18].

In this paper, we investigate the potential of CI for the precoder design in the presence of nonlinear PAs. By incorporating the response characteristics of the nonlinear PAs, we study how interference exploitation SLP can help alleviate the distortion brought from nonlinear PAs, and propose an iterative algorithm based on alternating optimization and SDR to obtain a near-optimal solution. A low-complexity algorithm is also designed to address the performance-complexity tradeoffs. Moreover, since power-efficient PAs are employed and their nonlinearity is already taken into account in the precoder design, there is no need to consider the specific impact of PAPR in our work. For the sake of clarity, we list this paper's contributions as follows:

1) We focus on the interference exploitation SLP design for the downlink transmission of MU-MISO communication systems. Specifically, we construct an SINR balancing SLP optimization problem and introduces the response characteristics of nonlinear PAs into the precoder design as elements that can be exploited constructively.

2) For PSK modulation, the non-convex optimization problem is firstly transformed by introducing auxiliary variables, and subsequently the iterative algorithm based on alternating optimization framework is presented for the new optimization problem. For subproblems that are still non-convex, additional auxiliary variables are introduced and several approximations are employed to transform the problem into the SDP form, where SDR is adopted to obtain feasible solutions.
3) For QAM modulation, we show that the non-convex problem has a similar solution framework as PSK modulation. The major difference between QAM modulation and PSK modulation is the mathematical CI condition, where for QAM modulation only the outer constellation points can exploit CI while all the interference for the inner constellation points are seen as destructive. Thanks to the similar problem structure, the proposed iterative algorithm can be extended to the PA-aware SLP design with QAM modulation.
4) We further propose a low-complexity algorithm to reduce the computational complexity for both PSK-modulated and QAM-modulated PA-aware SLP. By substituting the higher-order terms in the PA polynomial model with the initial CI precoded vector, a relaxed version of the original non-convex optimization problem is obtained, thus allowing a flexible iterative procedure.

Simulation results show that the proposed iterative algorithm based on the alternating optimization framework converges within a few iterations, which indicates that the iterative algorithm can approximate the optimal solution to the original problem. It is also observed that the proposed low-complexity algorithm achieves promising performance-complexity trade-offs in small-scale MU-MISO systems, and offers a near-optimal performance in large-scale MU-MISO systems in lower complexity. Moreover, it is shown that both of our proposed interference exploitation precoding considering the response characteristics of nonlinear PA are superior to traditional precoding schemes in terms of error-rate performance.

The remainder of this paper is organized as follows. In Section II, the system model and nonlinear PA model are introduced, and the conception of CI is reviewed. Section III introduces the proposed PA-aware SLP formulation for PSK and QAM modulation respectively, including mathematical CI conditions and the original optimization problem formulation. The iterative algorithm based on alternating optimization, the problem-dependent Gaussian randomization and the low-complexity algorithm are presented in Section IV. The complexity analysis of the proposed iterative algorithm and the low-complexity algorithm are both discussed in Section V. Numerical results of the proposed algorithms are shown in Section VI, and Section VII concludes the paper.

Notations: Lowercase, lowercase boldface and uppercase boldface letters denote scalar, vectors and matrices, respectively. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $tr\{\cdot\}$ denote conjugate, transposition, conjugate transposition and trace of a matrix, respectively. $diag(\cdot)$ is the transformation of a column vector into a diagonal matrix. $\lceil\cdot\rceil$ represents the ceiling function. $\mathbf{a}_i$ denotes the i-th term of vector $\mathbf{a}$. $|\cdot|$ denotes the absolute value of a real number or the modulus of a complex number, and $\|\cdot\|_2$ denotes the ℓ_2-norm. $\mathbb{C}^{n\times n}$ and $\mathbb{R}^{n\times n}$ represent the sets of $n\times n$ complex- and real-valued matrices, respectively. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ respectively denote the real and imaginary part of a complex scalar, vector or matrix. $\jmath$ denotes the imaginary unit, $\mathbf{I}_K$ denotes the $K\times K$ identity matrix, and $\mathbf{e}_i$ represents the i-th column of the identity matrix. $\mathbf{0}_K$ represents the K-dimensional zero vector.

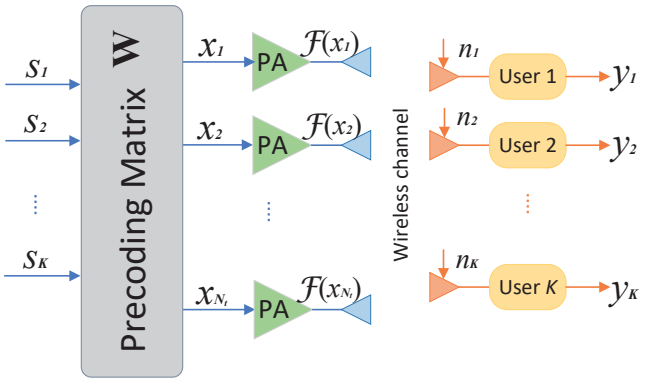

Fig. 1. MU-MISO downlink with nonlinear PA communication system model

II. System Model and Constructive Interference

A. System Model

We consider a downlink MU-MISO communication system as shown in Fig. 1. The BS with N_t transmit antennas communicates with K single-antenna users on the same time-frequency resources simultaneously, where $K \leq N_t$. We assume that all the nonlinear PAs at the BS have the same response characteristics, which is known to the BS.

We express the k-th user's received signal as

$$y_k = \mathbf{h}_k^T \mathcal{F}(\mathbf{x}) + n_k, \tag{1}$$

where, $\mathbf{h}_k \in \mathbb{C}^{N_t\times 1}$ represents the flat-fading channel vector between the BS and user k. Throughout the paper, perfect CSI is assumed. $\mathcal{F}(\cdot) : \mathbb{C} \to \mathbb{C}$ represents the transfer function of nonlinear PAs. $\mathbf{x} = \mathcal{W}(\mathbf{s}) \in \mathbb{C}^{N_t\times 1}$ is the precoded signal, $\mathcal{W}(\cdot) : \mathbb{C} \to \mathbb{C}$ represents the precoder and $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T \in \mathbb{C}^{K\times 1}$ is the data symbol vector. n_k is the standard complex additive Gaussian noise at the k-th user with zero mean and variance σ^2.

B. Nonlinear PA Model

In this paper, we adopt the memoryless polynomial model to modeling the behavior of nonlinear PAs, which has long been commonly used to approximate nonlinear PAs [26]–[30]. Specifically, the output signal of the nonlinear PAs can be expressed as

$$\mathcal{F}(\mathbf{x}) = \sum_{p=1}^{P} \beta_p diag(|\mathbf{x}|^{p-1})\mathbf{x}, \forall P \in \mathbb{Z}^+, \tag{2}$$

where β_p is the coefficient corresponding to the p-th order component, $|\mathbf{x}|$ represents the modulus of $\mathbf{x}$ operated on the element-wise. The first-order term ($p = 1$) describes the linear relationship between the input and output signals, and the other higher-order terms ($p > 1$) describe the nonlinear relationship between input and output signals [40], [41]. The advantages of the polynomial PA model are twofolds: a) the response characteristics of arbitrarily memoryless PAs can be approximated accurately by adjusting the modle parameters β_p [42]; b) the model is simple and mathematically tractable. In addition, since it has been shown in [41] that the even-order terms only contributes to the out-of-band distortion and lead to spectrum regrowth, we omit the even-order terms in the subsequent derivations. The transfer function of the memoryless polynomial model can be simplified as

$$\hat{\mathcal{F}}(\mathbf{x}) = \sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x}, \forall \hat{P} \in \mathbb{Z}^+, \tag{3}$$

where P and $\hat{P}$ satisfies $\hat{P} = \lceil P/2 \rceil$. By substituting (3) into (1), y_k can be further written as

$$y_k = \mathbf{h}_k^T \sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x} + n_k, \tag{4}$$

C. Constructive Interference

Interference exists in the multi-antenna communication system, which makes the received signal deviate from the nominal constellation points in both amplitude and phase. Observing the interference from the instantaneous point of view, [32] shows that interference can be divided into two types: constructive interference (CI) and destructive interference (DI). CI is the interference that push the noiseless received signal away from all of their corresponding decision boundaries of the modulated-symbol constellation, which thus contributes to the useful signal power [31].

The CI condition for the strict phase rotation metric refers to the case where the interference signals are strictly aligned to data symbols of interest by controlling and rotating the phases of the interference [34]. The concept of CI region has been further introduced in [35], and it is the region in the complex plane where the received signal falls when CI is achieved. Based on this, [35] proposed the CI condition for the non-strict phase rotation metric, under which the phase of the interference signals may not be necessarily strictly aligned to that of the data symbols of interest. It should be noted that the non-strict phase rotation is a relaxed CI metric, which gains further performance improvements, and the strict phase rotation is suboptimal. In addition, both the strict phase rotation and the non-strict phase rotation can only be applied to PSK modulation, but not to QAM modulation. The reason is that PSK modulation only modulates phase, whereas QAM modulation modulates both amplitude and phase, so that only the real or imaginary part of the outer constellation points can exploit CI. Consequently, the CI metric of QAM modulation has been discussed in [36], [43], where the symbol-scaling metric is introduced.

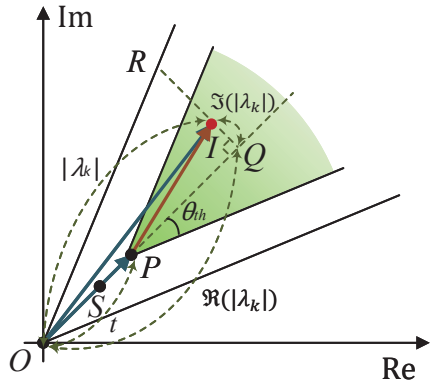

Fig. 2. 8PSK, 'non-strict phase rotation', [28]

III. PA-aware SLP Formulation

A. PSK Modulation

1) Non-Strict Phase-Rotation CI Metric: In Fig.2, the CI condition for non-strict phase rotation metric is shown by depicting the first quadrant of an 8PSK constellation as an example. Without loss of generality, we use point S to represent the nominal constellation point of the desired data symbol for user k, then $\vec{OS} = s_k$, and introduce $\vec{OP} = t \cdot s_k$ to represent the desired data symbol satisfying power constraint. Therefore, $t = \frac{|\vec{OP}|}{|\vec{OS}|}$ is the distance between the detection thresholds and the CI region, and a larger value of t leads to a better error-rate performance. Let $\vec{OI}$ represent the noiseless received signal for user k, i.e., $\vec{OI} = \mathbf{h}_k^T\mathbf{x}$, then $\vec{PI} = \vec{OI} - \vec{OP}$ is the interference signal.

We introduce a complex auxiliary variable λ_k to represent the scaling effect on both the amplitude and phase of the data symbol after experiencing the wireless channel, i.e.,

$$\vec{OI} = \mathbf{h}_k^T\mathbf{x} = \lambda_k s_k. \tag{5}$$

The projection of λ_k in the direction of $\vec{OS}$ is its real part, i.e., $\vec{OQ} = \Re(\lambda_k)$, and the projection perpendicular to $\vec{OS}$ is its imaginary part, i.e., $\vec{IQ} = \Im(\lambda_k)$. In order for $\vec{OI}$ to fall into the 'green' CI region, we need

$$\theta_{\angle POI} \leq \theta_{th}, \tag{6}$$

where $\theta_{th} = \frac{\pi}{M}$ for M-PSK. (6) can be further expressed as a function of λ_k as

$$[\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \tag{7}$$

where $\mathbb{K} = \{1, 2, \ldots, K\}$. (7) is therefore the mathematical CI condition for non-strict phase rotation to be used subsequently for PSK modulation.

2) Problem Formulation: In this paper, we consider SLP design based on CI for PA nonlinearity. Specifically, we focus on the CI-based SINR balancing problem where we aim to maximize the power scaling parameter t. Accordingly, we construct the optimization problem of the SLP design for M-

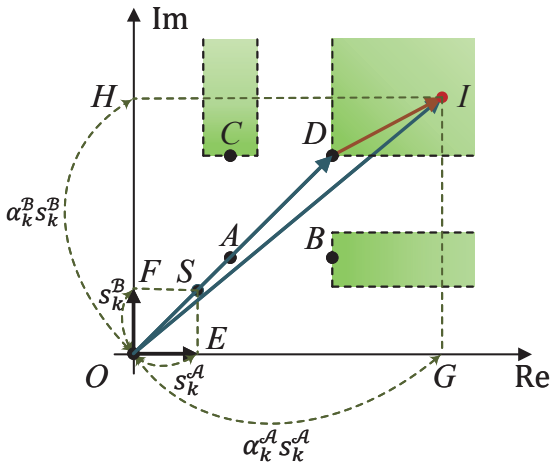

Fig. 3. 16QAM, 'symbol-scaling', [29]

PSK modulation as

$$\begin{aligned}
&\mathcal{P}_1: \max_{\mathbf{x}} \ t \\
&s.t. \\
&\mathcal{C}_1: \mathbf{h}_k^T \left[\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x}\right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\
&\mathcal{C}_2: [\Re(\lambda_k) - t]\tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\
&\mathcal{C}_3: \left\|\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x}\right\|_2^2 \leq p_0.
\end{aligned} \tag{8}$$

In $\mathcal{P}_1$, $\mathcal{C}_1$ and $\mathcal{C}_2$ jointly represent the CI condition for non-strict phase rotation, $\mathcal{C}_3$ indicates that the maximum transmit power does not exceed p_0. The above optimization problem $\mathcal{P}_1$ is non-convex and difficult to directly handle, owing to the existence of high-order terms in the memoryless polynomial PA model.

B. QAM modulation

1) Symbol-scaling CI Metric: Fig.3 depicts a quarter of 16QAM constellation as the example to characterize the CI condition for the symbol-scaling metric. The constellation points A, B, C and D respectively represent the four received signal types in the first quadrant of the 16QAM constellation and the green area indicates the CI region. With loss of generality, denoting $\vec{OS} = s_k$ as a nominal constellation point that is the intended data symbol for user k, we decompose the $\vec{OS}$ along the detection thresholds as

$$\vec{OS} = \vec{OE} + \vec{OF} \Rightarrow \quad s_k = s_k^{\mathcal{A}} + s_k^{\mathcal{B}}, \tag{9}$$

where $s_k^{\mathcal{A}} = \Re(s_k)$ and $s_k^{\mathcal{B}} = \jmath \cdot \Im(s_k)$ represent a set of bases derived from the decomposition of the $\vec{OS}$ along the detection threshold direction. Following the similar approach to (9), we decompose the noiseless received signal $\vec{OI}$ in the same direction to yield $\vec{OG}$ and $\vec{OH}$, given by

$$\vec{OI} = \mathbf{h}_k^T \mathbf{x} = \vec{OG} + \vec{OH}. \tag{10}$$

Based on which, we introduce a set of real auxiliary scalars, $\alpha_k^{\mathcal{A}}$ and $\alpha_k^{\mathcal{B}}$, to represent $\vec{OI}$ with the bases $s_k^{\mathcal{A}}$ and $s_k^{\mathcal{B}}$ obtained above as

$$\vec{OI} = \alpha_k^{\mathcal{A}} s_k^{\mathcal{A}} + \alpha_k^{\mathcal{B}} s_k^{\mathcal{B}} = \mathbf{\Omega}_k^T \mathbf{s}_k, \tag{11}$$

where $\mathbf{\Omega}_k$ and $\mathbf{s}_k$ are given by

$$\mathbf{\Omega}_k^T = \begin{bmatrix} \alpha_k^{\mathcal{A}} & \alpha_k^{\mathcal{B}} \end{bmatrix}^T, \mathbf{s}_k = \begin{bmatrix} s_k^{\mathcal{A}} & s_k^{\mathcal{B}} \end{bmatrix}^T. \tag{12}$$

It is observed that the values of $\mathbf{\Omega}_k$ directly indicate the effect of the CI. Since only the real or imaginary part of the outer constellation points can exploit CI, then in Fig.3, the CI region only includes the real part of constellation point type B, the imaginary part of type C and the real and imaginary part of type D. We refer to $\vec{OD} = t \cdot s_k$ as the user k's desired data symbol satisfying power constraint, where the power scaling parameter t is the distance between the detection thresholds and the CI region, then $\vec{DI} = \vec{OI} - \vec{OD}$ can represent the interference signal of user k. Similarly, $\vec{OD}$ can be written in the form of $\mathbf{s}_k$ as

$$\vec{OD} = t \cdot s_k = t \cdot (s_k^{\mathcal{A}} + s_k^{\mathcal{B}}). \tag{13}$$

Subsequently, for the purpose of forcing the noiseless received signal $\vec{OI}$ to fall into the CI region, the CI condition for symbol-scaling metric to be used for QAM modulation can be expressed as

$$\begin{aligned}
&t \leq \alpha_l^{\mathbb{O}}, \forall \alpha_l^{\mathbb{O}} \in \mathbb{O}, \\
&t = \alpha_m^{\mathbb{I}}, \forall \alpha_m^{\mathbb{I}} \in \mathbb{I},
\end{aligned} \tag{14}$$

where set $\mathbb{O}$ includes the real scalars corresponding to the real or imaginary part of the constellation points that can be scaled, and set $\mathbb{I}$ includes the real scalars corresponding to the real or imaginary part of the constellation points that cannot be scaled. Accordingly, we obtain

$$\begin{aligned}
&card\{\mathbb{O}\} + card\{\mathbb{I}\} = 2K, \\
&\mathbb{O} \cup \mathbb{I} = \left\{\alpha_1^{\mathcal{A}}, \alpha_1^{\mathcal{B}}, \alpha_2^{\mathcal{A}}, \alpha_2^{\mathcal{B}}, \cdots, \alpha_K^{\mathcal{A}}, \alpha_K^{\mathcal{B}}\right\}.
\end{aligned} \tag{15}$$

2) Problem Formulation: The optimization problem of PA-aware SLP for QAM modulation that maximizes the CI effect for the outer constellation points while maintaining the performance for the inner constellation points, can be constructed as

$$\begin{aligned}
&\mathcal{P}_2: \max_{\mathbf{x}} \ t \\
&s.t. \\
&\mathcal{C}_1: \mathbf{h}_k^T \left[\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x}\right] = \mathbf{\Omega}_k^T \mathbf{s}_k, \forall k \in \mathbb{K}, \\
&\mathcal{C}_2: t \leq \alpha_l^{\mathbb{O}}, \forall \alpha_l^{\mathbb{O}} \in \mathbb{O}, \\
&\mathcal{C}_3: t = \alpha_m^{\mathbb{I}}, \forall \alpha_m^{\mathbb{I}} \in \mathbb{I}, \\
&\mathcal{C}_4: \left\|\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}|^{2p-2})\mathbf{x}\right\|_2^2 \leq p_0.
\end{aligned} \tag{16}$$

In $\mathcal{P}_2$, $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ represent the CI condition for QAM modulation, and $\mathcal{C}_4$ indicates the available transmit power of

antennas. Due to the high-order terms in $\mathcal{C}_1$ and $\mathcal{C}_4$, $\mathcal{P}_2$ is non-convex. It is observed that $\mathcal{P}_2$ and $\mathcal{P}_1$ are similar in terms of problem formulation and all the optimization variables in $\mathcal{P}_2$ are consistent with those in $\mathcal{P}_1$, and the only difference between them is the CI condition, which is convex and thus will not affect the subsequent problem transformation and solution.

IV. Proposed Solutions

In order to obtain feasible and near-optimal solutions to $\mathcal{P}_1(\mathcal{P}_2)$, we propose an iterative algorithm which can obtain a near-optimal solution of $\mathcal{P}_1(\mathcal{P}_2)$, as well as a low-complexity algorithm which can offer a performance-complexity tradeoff in this section. We consider $\mathcal{P}_1$ as an example in the following problem derivation.

A. The Iterative Algorithm via Alternating Optimization

We begin by introducing $\mathbf{x}_p$ and performing the following variable substitution for $\mathbf{x}$ in $\mathcal{P}_1$:

$$\begin{cases} \mathbf{x}_1 = \mathbf{x} \in \mathbb{C}^{N_t \times 1}, \\ |\mathbf{x}_p|^{2p-2} = |\mathbf{x}|^{2p-2} \in \mathbb{C}^{N_t \times 1}, \forall p \in \mathbb{P}, p{>}1, \end{cases} \tag{17}$$

where $\mathbb{P} = \{1, 2, \ldots, \hat{P}\}$. Inserting (17) into $\mathcal{P}_1$ yields

$$\begin{aligned} \mathcal{P}_3 : &\max_{\mathbf{x}_p} \ t \\ &s.t. \\ &\mathcal{C}_1 : \mathbf{h}_k^T \left[\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}_p|^{2p-2}) \mathbf{x}_1 \right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\ &\mathcal{C}_2 : [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\ &\mathcal{C}_3 : \left\| \sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}_p|^{2p-2}) \mathbf{x}_1 \right\|_2^2 \leq p_0, \\ &\mathcal{C}_4 : \mathbf{x}_1 = \mathbf{x}_2 = \ldots = \mathbf{x}_{\hat{P}}. \end{aligned} \tag{18}$$

With the addition of $\mathcal{C}_4$, the solutions to $\mathcal{P}_3$ and $\mathcal{P}_1$ are guaranteed to be equivalent, and the variables to be optimized are changed from $\mathbf{x}$ to $\mathbf{x}_p, \forall p \in \mathbb{P}$.

The above optimization problem is still difficult to directly solve. Therefore, we adopt the following iterative procedure which sequentially updates each $\mathbf{x}_p$ that corresponds to $|\mathbf{x}|^{2p-2}$ from lower-order terms to higher-order terms, where in each iteration, we optimize $\mathbf{x}_p$ and maintain other $\mathbf{x}_q, (q \neq p)$ fixed. We repeat this process until $\mathbf{x}_1 \approx \mathbf{x}_2 \approx \ldots \approx \mathbf{x}_{\hat{P}}$.

1) Optimization on $\mathbf{x}_1$: Given $\mathbf{x}_p, p = 2, 3, \ldots, \hat{P}$, the optimization for $\mathbf{x}_1$ can be expressed as

$$\begin{aligned} \mathcal{P}_4 : &\max_{\mathbf{x}_1} \ t \\ &s.t. \\ &\mathcal{C}_1 : \mathbf{h}_k^T \left[\sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}_p|^{2p-2}) \mathbf{x}_1 \right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\ &\mathcal{C}_2 : [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\ &\mathcal{C}_3 : \left\| \sum_{p=1}^{\hat{P}} \beta_{2p-1} diag(|\mathbf{x}_p|^{2p-2}) \mathbf{x}_1 \right\|_2^2 \leq p_0, \\ &\mathcal{C}_4 : \|\mathbf{x}_1 - \mathbf{x}_p\|_2^2 \leq \epsilon_1, \forall p \in \mathbb{P}, p \neq 1, \end{aligned} \tag{19}$$

where ϵ_1 is a small nonnegative parameter which provides a relaxed version of $\mathcal{C}_4$. With the above approximations, $\mathcal{P}_4$ becomes convex and can be solved efficiently with existing convex optimization tools. By solving $\mathcal{P}_4$, we obtain the optimal $\mathbf{x}_1^*$, and update the $\mathbf{x}_1$ to the obtained $\mathbf{x}_1^*$.

To sequentially update each $\mathbf{x}_p, 1{<}p{<}\hat{P}$, we additionally introduce $\mathbf{Q}_{p'} = \sum_{\forall p \in \mathbb{P}, p \neq p'} \beta_{2p-1} diag(|\mathbf{x}_p|^{2p-2}) \mathbf{x}_1$ that is irrelevant to the variable $\mathbf{x}_{p'}$ to be optimized.

2) Optimization on $\mathbf{x}_2$: We substitute $\mathbf{x}_1$ into $\mathcal{P}_2$. Consequently, the optimization for $\mathbf{x}_2$ with fixed $\mathbf{x}_p, p \neq 2$ can be expressed as

$$\begin{aligned} \mathcal{P}_5 : &\max_{\mathbf{x}_2} \ t \\ &s.t. \\ &\mathcal{C}_1 : \mathbf{h}_k^T \left[\beta_3 diag(|\mathbf{x}_2|^2) \mathbf{x}_1 + \mathbf{Q}_2 \right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\ &\mathcal{C}_2 : [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\ &\mathcal{C}_3 : \left\| \beta_3 diag(|\mathbf{x}_2|^2) \mathbf{x}_1 + \mathbf{Q}_2 \right\|_2^2 \leq p_0, \\ &\mathcal{C}_4 : \|\mathbf{x}_2 - \mathbf{x}_p\|_2^2 \leq \epsilon_1, \forall p \in \mathbb{P}, p \neq 2. \end{aligned} \tag{20}$$

Based on the fact that $diag(|\mathbf{x}_2|^2)\mathbf{x}_1 = diag(\mathbf{x}_1)|\mathbf{x}_2|^2$, $\mathcal{P}_5$ can be further expressed as

$$\begin{aligned} \mathcal{P}_6 : &\max_{\mathbf{x}_2} \ t \\ &s.t. \\ &\mathcal{C}_1 : \mathbf{h}_k^T \left[\beta_3 diag(\mathbf{x}_1) |\mathbf{x}_2|^2 + \mathbf{Q}_2 \right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\ &\mathcal{C}_2 : [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\ &\mathcal{C}_3 : \left\| \beta_3 diag(\mathbf{x}_1) |\mathbf{x}_2|^2 + \mathbf{Q}_2 \right\|_2^2 \leq p_0, \\ &\mathcal{C}_4 : \|\mathbf{x}_2 - \mathbf{x}_p\|_2^2 \leq \epsilon_1, \forall p \in \mathbb{P}, p \neq 2. \end{aligned} \tag{21}$$

By introducing $\mathbf{d}_1 = |\mathbf{x}_2|^2 \in \mathbb{R}^{N_t \times 1}$, $\mathcal{P}_6$ can be transformed as

$$\begin{aligned} \mathcal{P}_7 : &\max_{\mathbf{d}_1} \ t \\ &s.t. \\ &\mathcal{C}_1 : \mathbf{h}_k^T \left[\beta_3 diag(\mathbf{x}_1) \mathbf{d}_1 + \mathbf{Q}_2 \right] = \lambda_k s_k, \forall k \in \mathbb{K}, \\ &\mathcal{C}_2 : [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\ &\mathcal{C}_3 : \left\| \beta_3 diag(\mathbf{x}_1) \mathbf{d}_1 + \mathbf{Q}_2 \right\|_2^2 \leq p_0, \\ &\mathcal{C}_4 : \left\| \mathbf{d}_1 - |\mathbf{x}_p|^2 \right\|_2^2 \leq \epsilon_2, \forall p \in \mathbb{P}, p \neq 2, \\ &\mathcal{C}_5 : \mathbf{d}_1(i) \geq 0, i = 1, 2, \cdots, N_t, \end{aligned} \tag{22}$$

where ϵ_2 is another small nonnegative parameter, which leads to the relaxed version of $\mathbf{d}_1 = |\mathbf{x}_2|^2 \approx |\mathbf{x}_p|^2, \forall p \in \mathbb{P}, p \neq 2$. $\mathcal{P}_7$ is convex and belongs to the second-order cone programming (SOCP) problem, which is equivalent to quadratic programming quadratic constraints (QCQP) by squaring each of the constraints [44]. By solving $\mathcal{P}_7$ we can obtain the optimal $\mathbf{d}_1^*$.

With the obtained $\mathbf{d}_1^*$, we need to find a proper $\mathbf{x}_2$ that satisfies $\mathbf{d}_1^* = |\mathbf{x}_2|^2$ and is closest to $\mathbf{x}_p, \forall p \in \mathbb{P}, p \neq 2$ as much as possible. To proceed, we expand $\mathbf{x}_2 \in \mathbb{C}^{N_t \times 1}$ into its real form $\tilde{\mathbf{x}}_2 \in \mathbb{R}^{2N_t \times 1}$, given by

$$\tilde{\mathbf{x}}_2 = \left[\begin{array}{cc} \Re(\mathbf{x}_2) & \Im(\mathbf{x}_2) \end{array} \right]^T, \tag{23}$$

based on the above, we can express the i-th entry in $\mathbf{x}_2$ with $\tilde{\mathbf{x}}_2$ as follows:

$$\mathbf{x}_2(i) = \mathbf{A}_i \tilde{\mathbf{x}}_2, i = 1, 2, \cdots, N_t, \tag{24}$$

where $\mathbf{A}_i$ is a selection matrix, given by

$$\mathbf{A}_i = \left[\begin{array}{cc} \mathbf{e}_i^T & \mathbf{0}_{N_t} \\ \mathbf{0}_{N_t} & \mathbf{e}_i^T \end{array} \right] \in \mathbb{R}^{2 \times 2N_t}, i = 1, 2, \cdots, N_t. \tag{25}$$

Subsequently, we can further express $\mathbf{d}_1$ as

$$\mathbf{d}_1 = |\mathbf{x}_2|^2 = \left[\begin{array}{c} |\mathbf{x}_2(1)|^2 \\ |\mathbf{x}_2(2)|^2 \\ \vdots \\ |\mathbf{x}_2(N_t)|^2 \end{array} \right] = \left[\begin{array}{c} \tilde{\mathbf{x}}_2^T \mathbf{A}_1^T \mathbf{A}_1 \tilde{\mathbf{x}}_2 \\ \tilde{\mathbf{x}}_2^T \mathbf{A}_2^T \mathbf{A}_2 \tilde{\mathbf{x}}_2 \\ \vdots \\ \tilde{\mathbf{x}}_2^T \mathbf{A}_{N_t}^T \mathbf{A}_{N_t} \tilde{\mathbf{x}}_2 \end{array} \right], \tag{26}$$

by expanding each $\mathbf{x}_p, \forall p \in \mathbb{P}, p \neq 2$ into its real form $\tilde{\mathbf{x}}_p \in \mathbb{C}^{2N_t \times 1}, \forall p \in \mathbb{P}, p \neq 2$, and following a similar way, we can construct a real-valued optimization problem $\mathcal{P}_8$ that aims to find a $\tilde{\mathbf{x}}_2$ that is close to other $\tilde{\mathbf{x}}_p$, given by

$$\begin{aligned} \mathcal{P}_8 : & \min_{\tilde{\mathbf{x}}_2} \sum_{\forall p \in \mathbb{P}, p \neq 2} \|\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_2\|_2^2 \\ & s.t.\ \tilde{\mathbf{x}}_2^T \mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{x}}_2 = \mathbf{d}_1^*(i), i = 1, 2, \cdots, N_t, \end{aligned} \tag{27}$$

by observing that the matrices $\mathbf{A}_i^T \mathbf{A}_i, i = 1, 2, \cdots, N_t$ are all real symmetric and positive semi-definite, we propose to reformulate the above problem into its SDP form and adopt SDR to solve the above problem. To be more specific, we introduce $\mathbf{C}_i = \mathbf{A}_i^T \mathbf{A}_i \in \mathbb{S}^{2N_t}$ and transform the constraint of $\mathcal{P}_8$ into

$$\begin{aligned} & \tilde{\mathbf{x}}_2^T \mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{x}}_2 = \mathbf{d}_1^*(i), \\ \Rightarrow & Tr\left\{ \tilde{\mathbf{x}}_2^T \mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{x}}_2 \right\} = \mathbf{d}_1^*(i), \\ \Rightarrow & Tr\left\{ \mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_2 \right\} = \mathbf{d}_1^*(i), \\ \Rightarrow & Tr\left\{ \mathbf{C}_i \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_2 \right\} = \mathbf{d}_1^*(i), i = 1, 2, \cdots, N_t. \end{aligned} \tag{28}$$

Accordingly, we can transform $\mathcal{P}_8$ into

$$\begin{aligned} \mathcal{P}_9 : & \min_{\tilde{\mathbf{x}}_2, r} \sum_{\forall p \in \mathbb{P}, p \neq 2} \|r\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_2\|_2^2 \\ & s.t. \\ & \mathcal{C}_1 : r^2 = 1 \\ & \mathcal{C}_2 : Tr\left\{ \mathbf{C}_i \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_2 \right\} = \mathbf{d}_1^*(i), i = 1, 2, \cdots, N_t, \end{aligned} \tag{29}$$

if $(\tilde{\mathbf{x}}_2^*, r^*)$ is an optimal solution to $\mathcal{P}_9$, then $\tilde{\mathbf{x}}_2^*$ is an optimal solution to $\mathcal{P}_8$ when $r^* = 1$ and $-\tilde{\mathbf{x}}_2^*$ is an optimal solution when $r^* = -1$ [36]. $\mathcal{P}_9$ can then be expressed as a separable QCQP:

$$\begin{aligned} \mathcal{P}_{10} : & \min_{\tilde{\mathbf{x}}_2, r} \sum_{\forall p \in \mathbb{P}, p \neq 2} \left[\begin{array}{c} \tilde{\mathbf{x}}_2 \\ r \end{array} \right]^T \left[\begin{array}{cc} \mathbf{I}_{2N_t} & -\mathbf{x}_p \\ -\mathbf{x}_p^T & \|\tilde{\mathbf{x}}_p\|_2^2 \end{array} \right] \left[\begin{array}{c} \tilde{\mathbf{x}}_2 \\ r \end{array} \right] \\ & s.t. \\ & \mathcal{C}_1 : r^2 = 1, \\ & \mathcal{C}_2 : Tr\left\{ \mathbf{C}_i \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_2 \right\} = \mathbf{d}_1^*(i), i = 1, 2, \cdots, N_t, \end{aligned} \tag{30}$$

by introducing a column vector $\hat{\mathbf{x}}_2 \in \mathbb{R}^{(2N_t+1) \times 1}$, matrices $\hat{\mathbf{X}}_2 \in \mathbb{R}^{(2N_t+1) \times (2N_t+1)}$ and $\mathbf{F}_p \in \mathbb{R}^{2N_t \times 2N_t}, \forall p \in \mathbb{P}$ are given below

$$\begin{aligned} & \hat{\mathbf{x}}_2 = \left[\begin{array}{c} \tilde{\mathbf{x}}_2 \\ r \end{array} \right], \hat{\mathbf{X}}_2 = \hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2^T = \left[\begin{array}{cc} \tilde{\mathbf{x}}_2 \tilde{\mathbf{x}}_2^T & r\tilde{\mathbf{x}}_2 \\ r\tilde{\mathbf{x}}_2^T & r^2 \end{array} \right], \\ & \mathbf{F}_p = \left[\begin{array}{cc} \mathbf{I}_{2N_t} & -\tilde{\mathbf{x}}_p \\ -\tilde{\mathbf{x}}_p^T & \|\tilde{\mathbf{x}}_p\|_2^2 \end{array} \right], \forall p \in \mathbb{P} \end{aligned} \tag{31}$$

therefore, the objective function of $\mathcal{P}_{10}$ can be further expressed as

$$\begin{aligned} & \sum_{\forall p \in \mathbb{P}, p \neq 2} \hat{\mathbf{x}}_2^T \mathbf{F}_p \hat{\mathbf{x}}_2 \\ = & \sum_{\forall p \in \mathbb{P}, p \neq 2} Tr\left\{ \hat{\mathbf{x}}_2^T \mathbf{F}_p \hat{\mathbf{x}}_2 \right\} \\ = & \sum_{\forall p \in \mathbb{P}, p \neq 2} Tr\left\{ \mathbf{F}_p \hat{\mathbf{X}}_2 \right\}, \end{aligned} \tag{32}$$

it should be noted that $\hat{\mathbf{X}}_2 = \hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2^T$ is equivalent to $\hat{\mathbf{x}}_2$ being a rank-one symmetric positive semidefinite (PSD) matrix. Here, we introduce

$$\begin{aligned} & \mathbf{D}_i = \left[\begin{array}{cc} \mathbf{C}_i & 0 \\ 0 & 0 \end{array} \right] \in \mathbb{R}^{(2N_t+1) \times (2N_t+1)}, i = 1, 2, \cdots, N_t, \\ & \mathbf{E} = \left[\begin{array}{cc} \mathbf{0}_{(2N_t+1) \times (2N_t+1)} & 0 \\ 0 & 1 \end{array} \right] \in \mathbb{R}^{(2N_t+1) \times (2N_t+1)}, \end{aligned} \tag{33}$$

based on the above, we can obtain the constraints of $\mathcal{P}_{10}$, given by

$$\begin{aligned} & Tr\left\{ \mathbf{D}_i \hat{\mathbf{X}}_2 \right\} = \mathbf{d}_1^*(i), \\ & Tr\left\{ \mathbf{E} \hat{\mathbf{X}}_2 \right\} = 1,\ i = 1, 2, \cdots, N_t. \end{aligned} \tag{34}$$

Therefore, we obtain the following equivalent formulation of $\mathcal{P}_9$:

$$\begin{aligned} \mathcal{P}_{11} : & \min_{\hat{\mathbf{X}}_2 \in \mathbb{S}^{(2N_t+1)}} \sum_{\forall p \in \mathbb{P}, p \neq 2} Tr\left\{ \mathbf{F}_p \hat{\mathbf{X}}_2 \right\} \\ & s.t. \\ & \mathcal{C}_1 : Tr\left\{ \mathbf{D}_i \hat{\mathbf{X}}_2 \right\} = \mathbf{d}_1^*(i), i = 1, 2, \cdots, N_t, \\ & \mathcal{C}_2 : Tr\left\{ \mathbf{E} \hat{\mathbf{X}}_2 \right\} = 1, \\ & \mathcal{C}_3 : \hat{\mathbf{X}}_2 \succeq 0, \\ & \mathcal{C}_4 : \text{rank}(\hat{\mathbf{X}}_2) = 1. \end{aligned} \tag{35}$$

subsequently, SDR can be used to produce an approximate solution to $\mathcal{P}_{11}$ [45], where we relax the nonconvex rank-one

constraint $\mathcal{C}_4$ in $\mathcal{P}_{11}$, which leads to the SDR form of $\mathcal{P}_{11}$ as

$$\begin{aligned}\mathcal{P}_{12}: &\min_{\hat{\mathbf{X}}_2\in\mathbb{S}^{2N_t+1}} \sum_{\forall p\in\mathbb{P},p\neq 2} Tr\left\{\mathbf{F}_p\hat{\mathbf{X}}_2\right\}\\ &s.t.\\ &\mathcal{C}_1: Tr\left\{\mathbf{D}_i\hat{\mathbf{X}}_2\right\}=\mathbf{d}_1^*(i), i=1,2,\cdots,N_t,\\ &\mathcal{C}_2: Tr\left\{\mathbf{E}\hat{\mathbf{X}}_2\right\}=1,\\ &\mathcal{C}_3: \hat{\mathbf{X}}_2\succeq 0.\end{aligned} \tag{36}$$

Denoting $\hat{\mathbf{X}}_2^*$ as the optimal solution to $\mathcal{P}_{12}$, we can obtain a feasible solution $\tilde{\mathbf{x}}_2$ through the Gaussian randomization method introduced in Section IV-B in the following. With the obtained $\tilde{\mathbf{x}}_2$, $\mathbf{x}_2$ can be obtained as

$$\mathbf{x}_2=\mathbf{U}_1\tilde{\mathbf{x}}_2, \tag{37}$$

where $\mathbf{U}_1=[\mathbf{I}_{N_t};\mathbf{I}_{N_t}\cdot j]\in\mathbb{R}^{2N_t\times N_t}$ is a transformation matrix that recovers its original complex form. For the optimization on the higher-order terms, i.e., $2\hat{P}-1>3$, we adopt a solution similar to that of the above. In the following, we discuss the optimization for $\mathbf{x}_{\tilde{p}}\,(\tilde{p}>2)$.

3) Optimization on $\mathbf{x}_{\tilde{p}}$: By updating each $\mathbf{x}_p$ to the optimal value of the latest iteration optimization, the optimization for $\mathbf{x}_{\tilde{p}}$ can be expressed as

$$\begin{aligned}\mathcal{P}_{13}: &\max_{\mathbf{x}_{\tilde{p}}}\ t\\ &s.t.\\ &\mathcal{C}_1: \mathbf{h}_k^T\left[\beta_{2\tilde{p}-1}diag\left[|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}\right]\mathbf{x}_1+\mathbf{Q}_{\tilde{p}}\right]=\lambda_k s_k,\forall k\in\mathbb{K},\\ &\mathcal{C}_2: [\Re(\lambda_k)-t]\tan\theta_{th}\geq|\Im(\lambda_k)|,\forall k\in\mathbb{K},\\ &\mathcal{C}_3: \left\|\beta_{2\tilde{p}-1}diag\left[|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}\right]\mathbf{x}_1+\mathbf{Q}_{\tilde{p}}\right\|_2^2\leq p_0,\\ &\mathcal{C}_4: \|\mathbf{x}_{\tilde{p}}-\mathbf{x}_p\|_2^2\leq\epsilon_1,\forall p\in\mathbb{P},p\neq\tilde{p}.\end{aligned} \tag{38}$$

By referring to the fact that $diag\left[|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}\right]\mathbf{x}_1 = diag(\mathbf{X}_1)|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}$, and introducing an auxiliary variable $\mathbf{d}_{\tilde{p}-1}=|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}\in\mathbb{C}^{N_t\times 1}$, $\mathcal{P}_{13}$ can be further transformed as

$$\begin{aligned}\mathcal{P}_{14}: &\max_{\mathbf{d}_{\tilde{p}-1}}\ t\\ &s.t.\\ &\mathcal{C}_1: \mathbf{h}_k^T\left[\beta_{2\tilde{p}-1}diag\left[\mathbf{x}_1\right]\mathbf{d}_{\tilde{p}-1}+\mathbf{Q}_{\tilde{p}}\right]=\lambda_k s_k,\forall k\in\mathbb{K},\\ &\mathcal{C}_2: [\Re(\lambda_k)-t]\tan\theta_{th}\geq|\Im(\lambda_k)|,\forall k\in\mathbb{K},\\ &\mathcal{C}_3: \left\|\beta_{2\tilde{p}-1}diag\left[\mathbf{x}_1\right]\mathbf{d}_{\tilde{p}-1}+\mathbf{Q}_{\tilde{p}}\right\|_2^2\leq p_0,\\ &\mathcal{C}_4: \left\|\mathbf{d}_{\tilde{p}-1}-|\mathbf{x}_p|^{2\tilde{p}-2}\right\|_2^2\leq\epsilon_{\tilde{p}},\forall p\in\mathbb{P},p\neq\tilde{p},\\ &\mathcal{C}_5: \mathbf{d}_{\tilde{p}-1}(i)\geq 0, i=1,2,\cdots,N_t,\end{aligned} \tag{39}$$

where $\epsilon_{\tilde{p}}$ is a small nonnegative parameter to relax the constraint $\mathbf{d}_{\tilde{p}-1}=|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}\approx|\mathbf{x}_p|^{2p-2},\forall p\in\mathbb{P},p\neq\tilde{p}$. $\mathcal{P}_{14}$ becomes convex and belongs to the SOCP problem, and by solving it we can obtain the optimal $\mathbf{d}_{\tilde{p}-1}^*$.

With the obtained $\mathbf{d}_{\tilde{p}-1}^*$, we need to find a proper $\mathbf{x}_{\tilde{p}}$ that satisfies $\mathbf{d}_{\tilde{p}-1}=|\mathbf{x}_{\tilde{p}}|^{2\tilde{p}-2}$ and is closest to $\mathbf{x}_p,\forall p\in\mathbb{P},p\neq\tilde{p}$. To proceed, we introduce a variable $\mathbf{g}_{\tilde{p}-2}$, given by

$$\mathbf{g}_{\tilde{p}-2}=\sqrt[\tilde{p}-1]{\mathbf{d}_{\tilde{p}-1}^*}, \tag{40}$$

based on the above, we expand $\mathbf{x}_{\tilde{p}}$ to its real form $\tilde{\mathbf{x}}_{\tilde{p}}$, then $\mathbf{g}_{\tilde{p}-2}$ can be further expressed as

$$\mathbf{g}_{\tilde{p}-2}=|\mathbf{x}_{\tilde{p}}|^2=\begin{bmatrix}\tilde{\mathbf{x}}_{\tilde{p}}^T\mathbf{A}_1^T\mathbf{A}_1\tilde{\mathbf{x}}_{\tilde{p}}\\ \tilde{\mathbf{x}}_{\tilde{p}}^T\mathbf{A}_2^T\mathbf{A}_2\tilde{\mathbf{x}}_{\tilde{p}}\\ \vdots\\ \tilde{\mathbf{x}}_{\tilde{p}}^T\mathbf{A}_{N_t}^T\mathbf{A}_{N_t}\tilde{\mathbf{x}}_{\tilde{p}}\end{bmatrix}\in\mathbb{C}^{N_t\times 1}, \tag{41}$$

subsequently, we can construct a real-value optimization problem $\mathcal{P}_{15}$ that aims to find a $\tilde{\mathbf{x}}_{\tilde{p}}$ that is close to other $\tilde{\mathbf{x}}_p$, given by

$$\begin{aligned}\mathcal{P}_{15}: &\min_{\tilde{\mathbf{x}}_{\tilde{p}}} \sum_{\forall p\in\mathbb{P},p\neq\tilde{p}}\|\tilde{\mathbf{x}}_p-\tilde{\mathbf{x}}_{\tilde{p}}\|_2^2\\ &s.t.\\ &\tilde{\mathbf{x}}_{\tilde{p}}^T\mathbf{A}_i^T\mathbf{A}_i\tilde{\mathbf{x}}_{\tilde{p}}=\mathbf{g}_{\tilde{p}-2}(i), i=1,2,\cdots,N_t,\end{aligned} \tag{42}$$

similarly, we introduce the real equivalent of $\mathbf{x}_{\tilde{p}}$ and a new variable $\hat{\mathbf{X}}_{\tilde{p}}$, given by

$$\hat{\mathbf{x}}_{\tilde{p}}=\begin{bmatrix}\tilde{\mathbf{x}}_{\tilde{p}}\\ r\end{bmatrix}, \hat{\mathbf{X}}_{\tilde{p}}=\hat{\mathbf{x}}_{\tilde{p}}\hat{\mathbf{x}}_{\tilde{p}}^T=\begin{bmatrix}\tilde{\mathbf{x}}_{\tilde{p}}\tilde{\mathbf{x}}_{\tilde{p}}^T & r\tilde{\mathbf{x}}_{\tilde{p}}\\ r\tilde{\mathbf{x}}_{\tilde{p}}^T & r^2\end{bmatrix}, \tag{43}$$

based on the expression for $\mathbf{F}_p$ in (31) and $\mathbf{D}_i$, $\mathbf{E}$ in (33), we transform $\mathcal{P}_{15}$ into its SDP form as

$$\begin{aligned}\mathcal{P}_{16}: &\min_{\hat{\mathbf{x}}_{\tilde{p}}\in\mathbb{S}^{2N_t+1}} \sum_{\forall p\in\mathbb{P},p\neq\tilde{p}} Tr\left\{\mathbf{F}_p\hat{\mathbf{X}}_{\tilde{p}}\right\}\\ &s.t.\\ &\mathcal{C}_1: Tr\left\{\mathbf{D}_i\hat{\mathbf{X}}_{\tilde{p}}\right\}=\mathbf{g}_{\tilde{p}-2}(i), i=1,2,\cdots,N_t,\\ &\mathcal{C}_2: Tr\left\{\mathbf{E}\hat{\mathbf{X}}_{\tilde{p}}\right\}=1,\\ &\mathcal{C}_3: \hat{\mathbf{X}}_{\tilde{p}}\succeq 0,\\ &\mathcal{C}_4: \text{rank}(\hat{\mathbf{X}}_{\tilde{p}})=1,\end{aligned} \tag{44}$$

by relaxing the rank-one constraint $\mathcal{C}_4$, we obtain the SDR form of $\mathcal{P}_{16}$, given by

$$\begin{aligned}\mathcal{P}_{17}: &\min_{\hat{\mathbf{X}}_{\tilde{p}}\in\mathbb{S}^{2N_t+1}} \sum_{\forall p\in\mathbb{P},p\neq\tilde{p}} Tr\left\{\mathbf{F}_p\hat{\mathbf{X}}_{\tilde{p}}\right\}\\ &s.t.\\ &\mathcal{C}_1: Tr\left\{\mathbf{D}_i\hat{\mathbf{X}}_{\tilde{p}}\right\}=\mathbf{g}_{\tilde{p}-2}(i), i=1,2,\cdots,N_t,\\ &\mathcal{C}_2: Tr\left\{\mathbf{E}\hat{\mathbf{X}}_{\tilde{p}}\right\}=1,\\ &\mathcal{C}_3: \hat{\mathbf{X}}_{\tilde{p}}\succeq 0.\end{aligned} \tag{45}$$

Denoting the optimal solution to $\mathcal{P}_{17}$ as $\hat{\mathbf{X}}_{\tilde{p}}^*$, the feasible solution $\tilde{\mathbf{x}}_{\tilde{p}}$ can be obtained from $\hat{\mathbf{X}}_{\tilde{p}}^*$ through the Gaussian randomization method introduced in Section IV-B in the following. $\tilde{\mathbf{x}}_{\tilde{p}}$ can then be converted to $\mathbf{x}_{\tilde{p}}$ by

$$\mathbf{x}_{\tilde{p}}=\mathbf{U}_1\tilde{\mathbf{x}}_{\tilde{p}}. \tag{46}$$

By optimizing each $\mathbf{x}_p$ sequentially until obtaining $\mathbf{x}_{\hat{P}}$, one alternating optimization process is completed. Then, we need to examine whether the results of this iteration satisfy the convergence condition, given by

$$\eta\leq\gamma, \tag{47}$$

where γ is a small nonnegative value and η is the sum of norms among optimization variables, given by

$$\eta = \sum_{p=1}^{P} \sum_{p'=p}^{P} \|\mathbf{x}_p - \mathbf{x}_{p'}\|_2^2, \tag{48}$$

subsequently, (47) provides a relaxed version of $\mathcal{C}_4$ in $\mathcal{P}_3$. If (47) is valid, $\mathbf{x}_1$ is considered as the approximate optimal solution of $\mathbf{x}$, if (47) is invalid, we need to update each $\mathbf{x}_p, \forall p \in \mathbb{P}$ by substituting the results of the previous alternating optimization as the initial values of the new optimization, and repeat the above process (19)-(45) until convergence.

Based on the above problem transformation and the alternating optimization, we present the iterative algorithm for the PSK-modulated PA-aware SLP, given in Algorithm 1, where n_{Iter} represents the number of iterations. Algorithm 1 can be extended to the QAM-modulated PA-aware SLP $\mathcal{P}_2$ by replacing the CI condition for PSK modulation with that for QAM modulation.

Algorithm 1 The proposed iterative PA-aware SLP algorithm based on alternating optimization

Input: $\mathbf{s}, \mathbf{H}, \beta_p, p_0, \epsilon_p, \gamma, \text{Iter}_{max}, \hat{P}$;
Output: $\mathbf{x}$;
Initialize: $\mathbf{x}_p, \eta, n_{\text{Iter}} = 0$;
While $\eta \geq \gamma$ and $n_{\text{Iter}} < \text{Iter}_{max}$ do
  Solve $\mathcal{P}_4$, Obtain $\mathbf{x}_1$;
  Solve $\mathcal{P}_7$ Obtain $\mathbf{d}_1$;
  Solve $\mathcal{P}_{12}$, Obtain $\hat{\mathbf{X}}_2$;
  Calculate $\mathbf{x}_2$;
  While $p' \leq \hat{P}$
    Solve $\mathcal{P}_{14}$, Obtain $\mathbf{d}_{p'-1}$;
    Calculate $\mathbf{g}_{p'-2}$ with (40);
    Solve $\mathcal{P}_{17}$, Obtain $\hat{\mathbf{X}}_{p'}$;
    Calculate $\mathbf{x}_{p'}$;
    Update $p' = p' + 1$;
  Endwhile
  Update $n_{\text{Iter}} = n_{\text{Iter}} + 1$;
  Obtain η with (48);
Endwhile
Obtain $\mathbf{x} = \mathbf{x}_1$.

B. Gaussian Randomization

In this paper, we consider adopting Gaussian randomization to obtain feasible solutions for $\mathcal{P}_{12}$ and $\mathcal{P}_{17}$, which share a same problem form. Gaussian randomization is a effective methods for extracting the solution of the original QCQP problem from solution of the SDR problem [45]. The procedure is based on the typical random procedure provided in [45], and include the problem-dependent design, which is shown below.

Takeing $\mathcal{P}_{17}$ as an example, we start by generating random vector $\boldsymbol{\xi}_l \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{X}}_{\tilde{p}}^*), l = 1, \cdots, L$, where L is the number of randomizations. Since the obtained random vectors do not satisfy the equality constraints of $\mathbf{P}_{17}$, we further construct the $\mathbf{P}_{17}$-feasible vectors as:

$$\begin{aligned}
&\hat{\boldsymbol{\xi}}_l(i) = \frac{\xi(i)}{\sqrt{(\xi^T \mathbf{D}_i \xi)}} \cdot \mathbf{d}_i, i = 1, 2, \cdots, N_t, l = 1, \cdots, L \\
&\hat{\boldsymbol{\xi}}_l(i + N_t) = \frac{\xi(i + N_t)}{\sqrt{(\xi^T \mathbf{D}_i \xi)}} \cdot \mathbf{d}_i, i = 1, 2, \cdots, N_t, l = 1, \cdots, L \\
&\hat{\boldsymbol{\xi}}_l(2N_t + 1) = sign\left\{\xi(2N_t + 1)\right\},
\end{aligned} \tag{49}$$

Through the above modifications, $\hat{\boldsymbol{\xi}}_l^2(i) + \hat{\boldsymbol{\xi}}_l^2(i + N_t) = \mathbf{d}_i$ and $\hat{\boldsymbol{\xi}}_l^2(2N_t + 1) = 1$ can be obtained, and the modified vector used for approximation satisfies the equality constraints. Based on the above, we choose the modified vector that yields the best objective as $\hat{\mathbf{x}}_{\tilde{p}}^*$ through

$$\hat{\mathbf{x}}_{\tilde{p}}^* = arg \min_{l=1,\cdots,L} \sum_{\forall p \in \mathbb{P}, P \neq \tilde{p}} \hat{\boldsymbol{\xi}}^T \mathbf{F}_p \hat{\boldsymbol{\xi}}, \tag{50}$$

$\tilde{\mathbf{x}}_{\tilde{p}}^*$ can then be determined by $\hat{\mathbf{x}}_{\tilde{p}}^*$ can be determine by

$$\tilde{\mathbf{x}}_{\tilde{p}}^* = \hat{\mathbf{x}}_{\tilde{p}}^*(2N_t + 1) \cdot \mathbf{U}_2 \hat{\mathbf{x}}_{\tilde{p}}^* \tag{51}$$

where $\mathbf{U}_2 = [\mathbf{I}_{2N_t}, \mathbf{O}_{2N_t}]$ is a selection matrix.

C. The Low-Complexity Algorithm

While the above iterative algorithm can obtain a near-optimal solution of $\mathcal{P}_1(\mathcal{P}_2)$, the corresponding computational complexity would be high as the alternating optimization framework decomposes $\mathcal{P}_1(\mathcal{P}_2)$ into $(2\hat{P} - 1)n_{\text{Iter}}$ subproblems. Therefore in this subsection, we further propose a low-complexity algorithm, where we consider only optimizing the lower-order terms $\mathbf{x}_l, \forall l \in \mathbb{P}, l \leq N (N < \hat{P})$ by substituting a fixed vector for the higher-order terms $\mathbf{x}_h, \forall h \in \mathbb{P}, h > N$, instead of optimizing $\mathbf{x}_p, p \in \mathbb{P}$ sequentially. The precoded signal of the SLP based on CI [37], [38] is adopted for the fixed vector.

By substitute the higher-order terms $\mathbf{x}_h, \forall h \in \mathbb{P}, h > N$ with the precoded signal vector $\mathbf{x}_{SLP-CI}$ in $\mathcal{P}_3$, we obtain a relaxed version of $\mathcal{P}_3$, which can be written as (52), where N represents the number of $\mathbf{x}_l, \forall l \in \mathbb{P}, l \leq N$ to be optimized.

$$\begin{aligned}
\mathcal{P}_{18}: &\max_{\mathbf{x}_l} \ t \\
&s.t. \\
&\mathcal{C}_1: \mathbf{h}_k^T \left[\sum_{l=1}^{N} \beta_{2l-1} diag(|\mathbf{x}_l|^{2l-2}) + \right. \\
&\left. \sum_{h=N+1}^{\hat{P}} \beta_{2h-1} diag(|\mathbf{x}_{SLP-CI}|^{2h-2}) \right] \mathbf{x}_1 = \lambda_k s_k, \forall k \in \mathbb{K}, \\
&\mathcal{C}_2: [\Re(\lambda_k) - t] \tan\theta_{th} \geq |\Im(\lambda_k)|, \forall k \in \mathbb{K}, \\
&\mathcal{C}_3: \left\| \left[\sum_{l=1}^{N} \beta_{2l-1} diag(|\mathbf{x}_l|^{2l-2}) + \right. \right. \\
&\left. \left. \sum_{h=N+1}^{\hat{P}} \beta_{2h-1} diag(|\mathbf{x}_{SLP-CI}|^{2h-2} \right] \mathbf{x}_1 \right\|_2^2 \leq p_0, \\
&\mathcal{C}_4: \mathbf{x}_1 = \mathbf{x}_2 = \ldots = \mathbf{x}_N.
\end{aligned} \tag{52}$$

It can be observed that compared with $\mathcal{P}_3$, $\mathcal{P}_{18}$ omits some high-order terms that are difficult to handle. When $N = 1$, $\mathcal{P}_{18}$ is simplified to a SOCP problem, which can be directly solved by using standard convex optimization tools or the IPM without iteration, and when $1<N<\hat{P}$, $\mathcal{P}_{18}$ is nonconvex, but it can obtain the near optimal solution using the iterative algorithm provided in the previous subsection by substituting $\hat{P}$ with N. The low-complexity algorithm can also be extended to the QAM-modulated PA-aware SLP by replacing the CI condition for PSK modulation with that for QAM modulation.

V. Computational Complexity Analysis

In this section, we analyze the computational complexity of the proposed iterative algorithm. Since both algorithms are based on optimization, the computational cost is evaluated based on the worst case complexity via the IPM [45], [46].

Without loss of generality, we focus on the iterative algorithm for PSK modulation, while it is very simple to extend to QAM modulation, since they have similar problem formulation.

For the optimization-based iterative algorithm, the complexity is dominated by solving the convex optimization problem $\mathcal{P}_4$, $\mathcal{P}_7$ and $\mathcal{P}_{14}$ via the IPM. For $\mathcal{P}_4$ and $\mathcal{P}_{14}$, which belong to the SOCP problem, they have the same problem formulation. Here, we consider to take $\mathcal{P}_{14}$ as an example and express it in a standard SOCP form as

$$\begin{aligned}
\mathcal{P}_{14}^*: \min_{\mathbf{z}} \; & -\mathbf{c}^T\mathbf{z} \\
s.t. \; & \\
\mathcal{C}_1: \; & \mathbf{c}_{k1}^T\mathbf{z} + \mathbf{l}_{k1} = 0, \forall k \in \mathbb{K}, \\
\mathcal{C}_2: \; & \mathbf{c}_{k2}^T\mathbf{z} \leq 0, \forall k \in \mathbb{K}, \\
\mathcal{C}_3: \; & \mathbf{c}_{k3}^T\mathbf{z} \leq 0, \forall k \in \mathbb{K}, \\
\mathcal{C}_4: \; & \|\mathbf{A}_4\mathbf{z} + \mathbf{Q}_{\tilde{p}}\|_2 \leq \sqrt{p_0} \\
\mathcal{C}_5: \; & \|\mathbf{A}_5\mathbf{z} - \mathbf{b}_{p5}\|_2 \leq \sqrt{\epsilon_{\tilde{p}}}, \forall p \in \mathbb{P}, p \neq \tilde{p}, \\
\mathcal{C}_6: \; & \mathbf{c}_{i6}^T\mathbf{z} \leq 0, i = 1, 2, \cdots, N_t,
\end{aligned} \tag{53}$$

where we introduce

$$\begin{aligned}
&\mathbf{z} = \left[\mathbf{d}_{\tilde{p}-1}^T, t, \Re(\lambda_k), \Im(\lambda_k)\right]^T \in \mathbb{C}^{N_t+1+2K}, \forall k \in \mathbb{K}, \\
&\mathbf{c} = \left[\mathbf{0}_{N_t}^T, 1, \mathbf{0}_{2K}^T\right]^T \in \mathbb{C}^{N_t+1+2K}, \\
&\mathbf{c}_{k1} = \left[\mathbf{h}_k^T \cdot \frac{\beta_{2\tilde{p}-1} diag[\mathbf{x}_1]}{s_k}, 0, -\mathbf{e}_k^T, -\jmath\mathbf{e}_k^T\right] \in \mathbb{C}^{N_t+1+2K}, \\
&\forall k \in \mathbb{K}, \\
&\mathbf{l}_{k1} = \mathbf{h}_k^T \cdot \mathbf{Q}_{\tilde{p}}, \forall k \in \mathbb{K}, \\
&\mathbf{c}_{k2} = \left[\mathbf{0}_{N_t}^T, 1, -\mathbf{e}_k^T, -\frac{1}{tan\theta_{th}} \cdot \mathbf{e}_k^T\right] \in \mathbb{C}^{N_t+1+2K}, \forall k \in \mathbb{K}, \\
&\mathbf{c}_{k3} = \left[\mathbf{0}_{N_t}^T, 1, -\mathbf{e}_k^T, \frac{1}{tan\theta_{th}} \cdot \mathbf{e}_k^T\right] \in \mathbb{C}^{N_t+1+2K}, \forall k \in \mathbb{K}, \\
&\mathbf{A}_4 = \left[\beta_{2\tilde{p}-1} diag[\mathbf{x}_1], \mathbf{0}_{N_t \times (2K+1)}\right] \in \mathbb{C}^{N_t \times (N_t+1+2K)}, \\
&\mathbf{A}_5 = \left[\mathbf{I}_{N_t}, \mathbf{0}_{N_t \times (2K+1)}\right] \in \mathbb{C}^{N_t \times (N_t+1+2K)}, \forall p \in \mathbb{P}, p \neq \tilde{p}, \\
&\mathbf{b}_{p5} = |\mathbf{x}_p|^{2\tilde{p}-2} \in \mathbb{C}^{N_t}, \forall p \in \mathbb{P}, p \neq \tilde{p}, \\
&\mathbf{c}_{i6} = \left[\mathbf{e}_i^T, \mathbf{0}_{2K+1}^T\right] \in \mathbb{C}^{N_t+1+2K}, i = 1, 2, \cdots, N_t.
\end{aligned} \tag{54}$$

Based on [46], the complexity of a generic IPM for solving conic programming consists of two parts: a) the iteration complexity, which is on the order of $\sqrt{k_{IPM}} log(1/\epsilon)$, where k_{IPM} represents the number of cones and ϵ represents the target accuracy of the solutions; b) per-iterative computation cost, which dominated by the formation and the factorization of the coefficient matrix $\mathcal{H}$, which is used to find the searching direction. The $\mathcal{C}_{form}$ and $\mathcal{C}_{fact}$ are the order of the cost of forming $\mathcal{H}$ and dominated by the dimensions of $\mathbf{z}$ and all cones. By combining the two parts, it follows that the complexity of a generic IPM for solving $\mathcal{P}_{14}^*$ is on the order of $\sqrt{k_{IPM}} \cdot (\mathcal{C}_{form} + \mathcal{C}_{fact}) \cdot log(1/\epsilon)$, where $k_{IPM} = N_t + 1 + 2K$. The expression of $\mathcal{C}_{form}$ and $\mathcal{C}_{fact}$ are given by

$$\begin{aligned}
\mathcal{C}_{form} =& (N_t + 2K + 1)(N_t + 2K) + \\
& (N_t + 2K + 1)^2(N_t + 2K) + (N_t + 2K + 1)\hat{P}(N_t^2), \\
\mathcal{C}_{fact} =& (N_t + 1 + 2K)^3.
\end{aligned} \tag{55}$$

The overall complexity of $\mathcal{P}_{14}$ is on the order of (56), which is shown in the bottom of this page.

Following the similar analysis approach, the complexity in solving $\mathcal{P}_4$ can be analyzed. With some inspection, $\mathcal{P}_4$ is less N_t linear cones compared to $\mathcal{P}_{14}$. Therefore, the complexity of $\mathcal{P}_4$ via IPM is on the order of (57), which is shown in the top of this page.

As for $\mathcal{P}_{17}$, we can obtain the worst-case complexity for solving the SDR problem via the IPM by referring to [45],

$$\begin{aligned}
\mathcal{C}_{\mathcal{P}_{14}} =& \mathcal{O}\left\{Iter \cdot \sqrt{N_t + 2K + 1}\left[(N_t + 2K + 1)(N_t + 2K) + (N_t + 2K + 1)^2(N_t + 2K) + \right.\right. \\
& \left.\left. (N_t + 2K + 1)\hat{P}(N_t^2) + (N_t + 1 + 2K)^3\right] \cdot log(1/\epsilon)\right\},
\end{aligned} \tag{56}$$

$$\begin{aligned}
\mathcal{C}_{\mathcal{P}_4} =& \mathcal{O}\left\{Iter \cdot \sqrt{N_t + 2K + 1}\left[(N_t + 2K + 1)(2K) + (N_t + 2K + 1)^2(2K) + (N_t + 2K + 1)\hat{P}(N_t^2) + \right.\right. \\
& \left.\left. (N_t + 1 + 2K)^3\right] \cdot log(1/\epsilon)\right\}.
\end{aligned} \tag{57}$$

given by

$$\mathcal{C}_{\mathcal{P}_{17}} = \mathcal{O}(max\left\{m_{SDR}, n_{SDR}\right\}^4 n_{SDR}^{1/2} log(1/\epsilon)), \quad (58)$$

where m_{SDR} is the number of the quadratic constraints and n_{SDR} represents the dimension of the optimization variable. Based on the construction of $\mathcal{P}_{17}$, we obtain

$$m_{SDR} = N_t + 1, n_{SDR} = 2N_t + 1, \quad (59)$$

which further leads to the expression of $\mathcal{C}_{\mathcal{P}_{12}}$ as

$$\mathcal{C}_{\mathcal{P}_{17}} = \mathcal{O}\left\{(2N_t + 1)^{9/2} log(1/\epsilon)\right\}, \quad (60)$$

With the complexity of each optimization problem, we can obtain the overall complexity of the iterative algorithm based on the structure of the iterative algorithm, given by

$$\mathcal{C}_{Iter} = \left\{\mathcal{C}_{P_4} + (\hat{P} - 1)(\mathcal{C}_{P_{14}} + \mathcal{C}_{P_{17}})\right\} \cdot Iter, \quad (61)$$

where $Iter$ is the average number of iterations. In addition, the complexity of the low-complexity algorithm can be similarly obtained by replacing $'\hat{P}'$ with $'N'$ in (61), and is omitted for brevity.

VI. Numerical Simulation

In this section, the numerical results of the proposed schemes are presented and compared with other conventional precoding schemes in the presence of PA nonlinearities, using Monte Carlo simulations. Both PSK and QAM modulation are considered. In each plot, we assume the total transmit power available as $p_0 = 1$, the transmit SNR per antenna as $\rho = 1/\sigma^2$.We consider to adopt the precoded signal of ZF precoding as the initial point of $\mathbf{x}_p, \forall p \in \mathbb{P}$[1], and set the value of $\epsilon_p, \forall p \in \mathbb{P}$ as follow

$$\begin{cases} \epsilon_1 = \alpha \frac{N_t}{M}, \\ \epsilon_p = \epsilon_{p-1}^2, \forall p \in \mathbb{P}, p \neq 1, \end{cases} \quad (62)$$

where α is the scaling factor. To achieve smooth and efficient convergence, it is always set to 0.8.

The following abbreviations are used throughout this section:

1) 'GR, Iter/LC': the Gaussian randomization method for SDR, the iterative algorithm proposed in Section V-A and the low-complexity algorithm proposed in Section V-B;
2) 'R1A, Iter/LC': the rank-one approximation method for SDR, the iterative algorithm proposed in Section V-A and the low-complexity algorithm proposed in Section V-B;
3) 'ZF': traditional ZF precoding with symbol-level power normalization. The ZF precoding signals are $\mathbf{x}_{ZF} = \frac{\sqrt{p_0}}{f_{ZF}} \cdot \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{s}$, where $f_{ZF} = \left\|\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{s}\right\|_2$ [3];
4) 'RZF': traditional regularization zero-forcing precoding with symbol-level power normalization. The RZF precoded signals are $\mathbf{x}_{RZF} = \frac{\sqrt{p_0}}{f_{RZF}} \cdot \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \frac{K}{\sigma^2} \cdot \mathbf{I})^{-1}\mathbf{s}$, where $f_{RZF} = \left\|\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \frac{K}{\sigma^2} \cdot \mathbf{I})^{-1}\mathbf{s}\right\|_2$ [3];
5) 'MRT': traditional maximum ratio transmission precoding with symbol-level power normalization. The MRT precoded signals are $\mathbf{x}_{MRT} = \frac{\sqrt{p_0}}{f_{MRT}} \cdot \mathbf{H}^H\mathbf{s}$, where $f_{MRT} = \left\|\mathbf{H}^H\mathbf{s}\right\|_2$ [47];
6) 'CEP-CI': constant envelope precoding based on CI using manifold algorithm [18];
7) 'SLP-CI': optimization-based CI SLP for non strict phase rotation based on $\mathcal{P}_5$ in [37] and $\mathcal{P}_3$ in [38];
8) 'PA-aware SLP, Iter/LC': the proposed PA-aware SLP based on CI, the iterative algorithm proposed in Section V-A and the low-complexity algorithm proposed in Section V-B;

[1]The initial point of $\mathbf{x}_p, \forall p \in \mathbb{P}$ can be arbitrarily selected as long as $\mathbf{x}_1 = \mathbf{x}_2 = \cdots = \mathbf{x}_P$ is satisfied, and it is verified that this does not change the final solution of the algorithm when the iteration terminates.

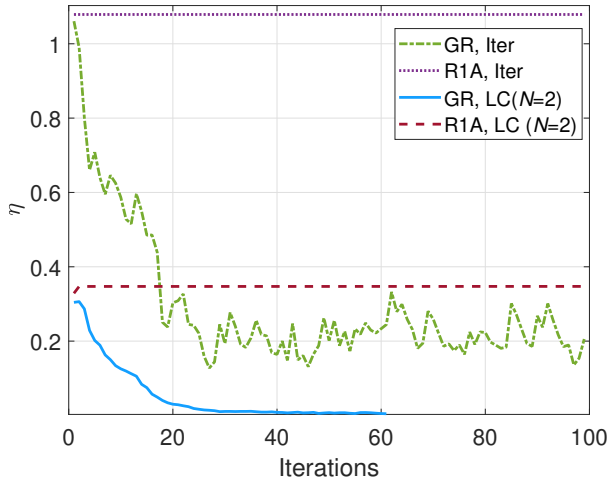

Fig. 4. Convergence behavior of the proposed algorithm, 8PSK, $N_t = K = 8$, SNR= 30dB

Before we present the performance results of the PA-aware SLP, in Fig.4 we first show the convergence behavior of the Algorithm 1 by illustrating the number of iterations to achieve the near-optimality through the GR with that through the R1A method, where we consider 8PSK modulation, $N_t = K = 8$, SNR= 30dB, and set the PAs' coefficients as $\beta_1 = 1.0108 + \jmath 0.0858$, $\beta_3 = 0.0879 - \jmath 0.1583$, $\beta_5 = -1.0992 - \jmath 0.8891$, $\beta_p = 0 (p > 5)$ [48]. As can be observed, the GR based approach is convergent while the R1A based approach is not. This means that, both the proposed iterative algorithm and the low-complexity algorithm can obtain a near-optimal solution by SDR with the GR. In addition, by choosing parameters γ and ϵ_p appropriately, the GR method converges within 60 iterations. In subsequent simulation, we employ the GR method to obtain a feasible solution to SDR problem.

Fig.5 shows the BER performance for the different computational complexity, where we consider the same system setting as Fig.4. Note that 'PA-aware SLP' is solved by the iterative algorithm proposed in Section V-A when $N = 3, (2N - 1 = 5)$ and solved by the low-complexity algorithm proposed in Section V-B when $N < 3$. It is then observed that when $N = 3$, the performance of 'PA-aware SLP' is consistent with that when $N = 2, (2N - 1 = 3)$. This is because the fifth-

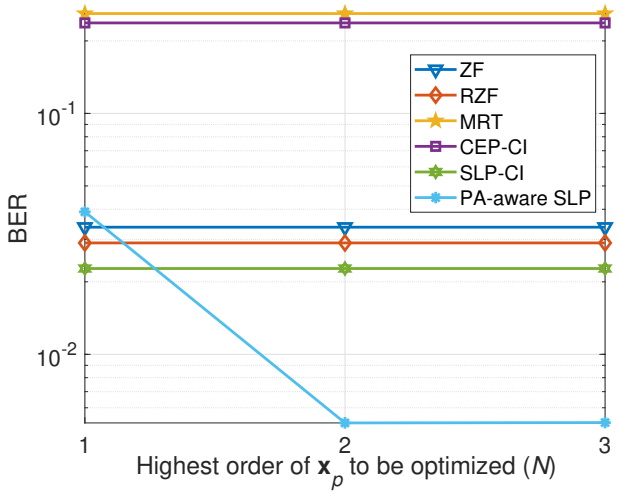

Fig. 5. Uncoded BER v.s. N, 8PSK, $N_t = K = 8$, SNR= 30dB

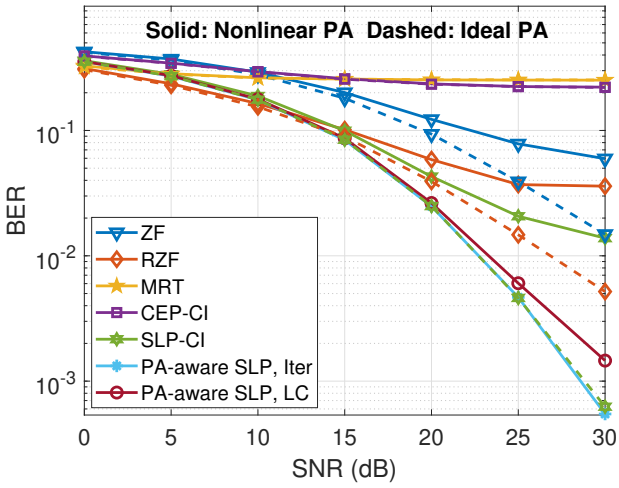

Fig. 6. Uncoded BER v.s. transmit SNR, 8PSK, $N_t = K = 8$

order term of the complex coefficient polynomial PA model used here has negligible impact on the performance of the communication system. Therefore, for simplicity, we consider employing the third-order complex coefficient polynomial PA model at the BS in the subsequent simulation, and assume that $\beta_1 = 1.0108 + \jmath 0.0858$, $\beta_3 = 0.2637 - \jmath 0.4749$, $\beta_p = 0\,(p > 3)$,which come from fitting the fifth order model used above.

In Fig.6, we demonstrate the BER performance for different precoding schemes with 8PSK modulation for the case of $N_t = K = 8$, where the BS with nonlinear PAs and ideal PAs are denoted as solid and dashed lines, respectively. It is observed that 'PA-aware SLP, Iter' with nonlinear PAs in the BS can achieve almost same performance as that of 'SLP-CI' with ideal PAs in the BS, which indicates that 'PA-aware SLP, Iter' can completely eliminate the impact of nonlinear distortion of PAs on communication system. Moreover, 'PA-aware SLP, LC' is inferior to 'PA-aware SLP, Iter' while outperforming 'SLP-CI' with nonlinear PAs in the BS. Both

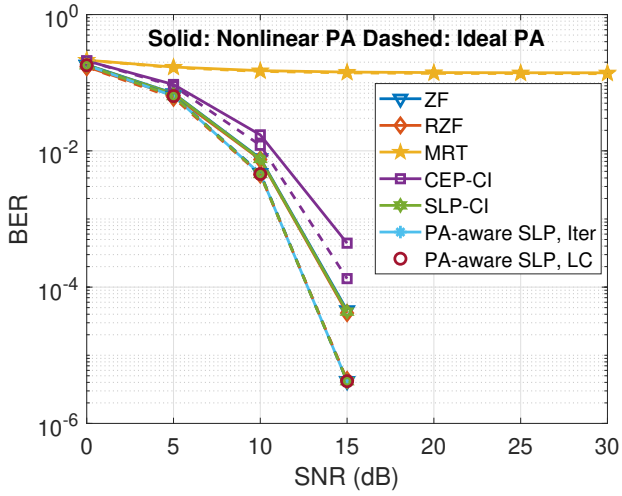

Fig. 7. Uncoded BER v.s. transmit SNR, 8PSK, $N_t = 64$, $K = 20$

of 'PA-aware SLP, Iter' and 'PA-aware SLP, LC' achieve an improved performance over ZF precoding, MRT precoding, 'CEP-CI', and 'SLP-CI' with nonlinear PAs in the BS for all SNRs, while also outperforming RZF precoding scheme at high SNR. CEP performs poorly in small-scale MIMO systems because it cannot handle high residual interference. For 'PA-aware SLP, Iter' at high SNR regime, we observe that a SNR gain of over 14dB compared to ZF precoding, over 10dB compared to RZF precoding, over 8dB compared to 'SLP-CI' and near 2dB compared to 'PA-aware SLP, LC' under nonlinear PAs architecture.

In Fig.7, we demonstrate the BER performance for different precoding schemes with 8PSK modulation in large-scale communication system, where $N_t = 64$, $K = 20$. As can be observed, with the number of the transmit antennas increasing the performance gains of the 'SLP-CI', 'PA-aware SLP, Iter' and 'PA-aware SLP, LC' are no longer significant, however, the performance of 'CEP-CI' is significantly improved. Moreover, 'PA-aware SLP, LC' achieves exactly the same performance as 'PA-aware SLP, Iter', which indicates that the performance loss of 'PA-aware SLP, LC' can be ignored in large-scale systems. When BER reaches 10^{-3}, we observe that 'PA-aware SLP, Iter' and 'PA-aware SLP, LC' can provide a SNR gain of about 2dB compared with ZF precoding, RZF precoding and 'SLP-CI' and near 4dB compared with 'CEP-CI'.

We further simulate the BER performance comparison under 16QAM modulation in Fig.8, where $K = N_t = 8$. It is observed that our proposed PA-aware SLP still has superiority over the traditional precoding schemes and 'SLP-CI'. In addition, 'PA-aware SLP, Iter' achieves a better performance than 'PA-aware SLP, LC'. Comparing Fig.8 with Fig.6, we observe that the performance gain of PA-aware SLP and 'SLP-CI' under 16QAM modulation is slightly less than that under PSK modulation, the loss of the performance gain is due to only the outer constellation points can exploit CI in QAM modulation. In the high SNR regime, 'PA-aware SLP, Iter' can provide a SNR gain of more than 9dB over ZF precoding, 7dB over RZF precoding, 6dB over 'SLP-CI' and 1dB over

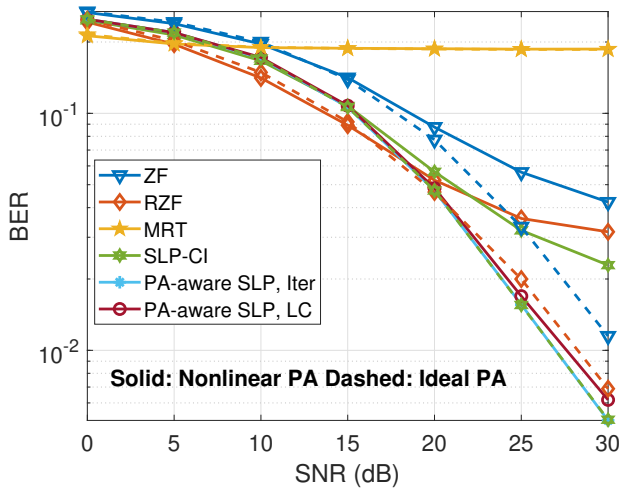

Fig. 8. Uncoded BER v.s. transmit SNR, 16QAM, $N_t = K = 8$

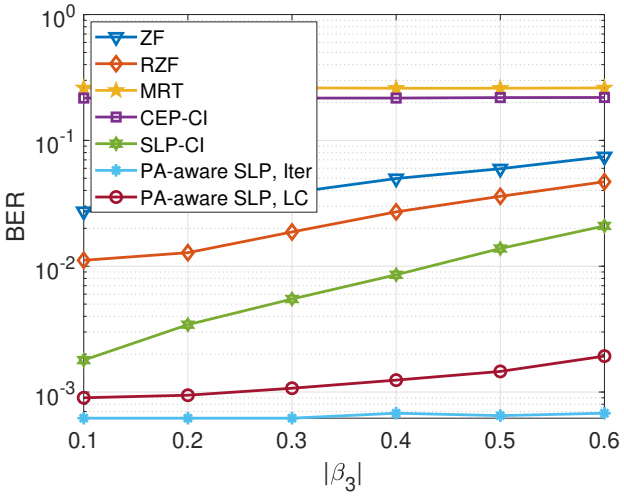

Fig. 10. Uncoded BER v.s. $|\beta_3|$, 8PSK, $N_t = K = 8$, SNR=30dB

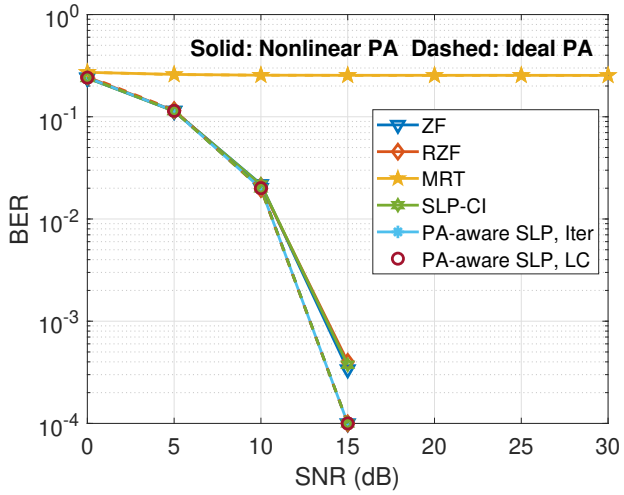

Fig. 9. Uncoded BER v.s. transmit SNR, 16QAM, $N_t = 64$, $K = 20$

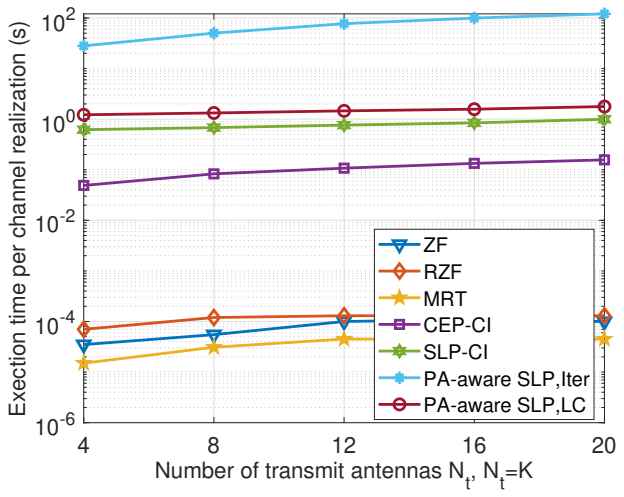

Fig. 11. Execution time v.s. N_t, 8PSK, $K = N_t$

'PA-aware SLP, LC'.

Fig.9 shows the BER performance for different precoding schemes with 16QAM modulation in large-scale communication system, where $N_t = 64$, $K = 20$. Similar to PSK modulation, we observe that the error-rate performance of 'PA-aware SLP, Iter' and 'PA-aware SLP, LC' are exctly the same. In addition, when BER reaches 10^{-3} , 'PA-aware SLP, Iter' and 'PA-aware SLP, LC' can provide a SNR gain of near 2dB compared with ZF precoding, RZF precoding and 'SLP-CI'.

In Fig.10, we present the BER performance of different precoding schemes with respect to the PA nonlinearities by varying the value of β_3. Fig.10 shows that as the absolute value of β_3 becomes larger, the BER performance of the ZF precoding, RZF precoding and 'SLP-CI' increases, which indicates that the enhancement of nonlinear distortion will degrade the performance of the precoding schemes designed under the ideal hardware assumption. In addition, since the low-complexity algorithm omits the optimization of higher-order terms of the PA model in the original optimization problem, we observe that the BER performance of 'PA-aware SLP, LC' also deteriorates with the increase of PA nonlinearity. Due to the error-rate performance of 'CEP-CI' and MRT precoding is heavily influenced by other factors, no significant variation is observed. Meanwhile, the BER of 'PA-aware SLP, Iter' can essentially maintain a constant value when the absolute value of β_3 rises, which shows the superiority of our proposed iterative algorithm to nonlinear distortion.

In order to demonstrate the potential complexity benefits of the proposed iterative algorithm and low-complexity algorithm, we evaluate the computational cost of different precoding schemes with 8PSK modulation in terms of the execution time per channel realization in Fig.11. It is observed that 'PA-aware SLP, Iter' requires more time to obtain the near optimal solution because of the larger required subproblems number than 'PA-aware SLP, LC'. Moreover, 'PA-aware SLP, LC' exhibits additional over 20dB complexity gain than the proposed iterative algorithm, i.e., the proposed low-complexity algorithm is more time-efficient than the proposed iterative

algorithm, which motivates the use of the PA-aware SLP in practice.

VII. Conclusion

In this paper, we study the interference exploitation SLP for a downlink MU-MISO communication system with nonlinear PAs, where both PSK modulation and QAM modulation are considered. By analyzing the non-convex optimization problem of PA-aware SLP problem, we first introduce auxiliary variables to transform the optimization problem into a new form, then propose an iterative algorithm, which can obtain a near-optimal solution to the PA-aware SLP by employing the Gaussian random method to the SDR problem. For the purpose of addressing the performance-complexity tradeoffs, we further present a low-complexity algorithm. Numerical results show that the PA-aware SLP design can well alleviate the performance loss of non-ideal PAs due to nonlinearity, and achieves an improved performance over the benchmark schemes. It is also observed that the performance loss of the low-complexity algorithm becomes marginal in large-scale systems.

References

[1] T. Barnett, S. Jain, U. Andra, and T. Khurana, "Cisco Visual Networking Index (vni) Complete Forecast Update, 2017–2022," *Americas/EMEAR Cisco Knowledge Network (CKN) Presentation*, pp. 1–30, 2018.

[2] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[3] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[4] H. Moazzen, A. Mohammadi, and M. Majidi, "Performance analysis of linear precoded mu-mimo-ofdm systems with nonlinear power amplifiers and correlated channel," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6753–6765, 2019.

[5] S. Han, C.-l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[6] A. Li, C. Masouros, and F. Liu, "Hybrid Analog-Digital Precoding for Interference Exploitation (Invited Paper)," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 812–816.

[7] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel Estimation and Performance Analysis of One-Bit Massive MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4075–4089, 2017.

[8] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of One-Bit Quantized Precoding for the Multiuser Massive MIMO Downlink," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4624–4634, 2017.

[9] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit Processing with Low Resolution D/A-Converters," in *2009 16th IEEE International Conference on Electronics, Circuits and Systems - (ICECS 2009)*, 2009, pp. 683–686.

[10] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE Precoder for Massive MIMO Using 1-bit Quantization," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3381–3385.

[11] H. Jedda, J. A. Nossek, and A. Mezghani, "Minimum BER precoding in 1-Bit massive MIMO systems," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016, pp. 1–5.

[12] S. K. Mohammed and E. G. Larsson, "Single-User Beamforming in Large-Scale MISO Systems with Per-Antenna Constant-Envelope Constraints: The Doughnut Channel," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3992–4005, 2012.

[13] J. Pan and W.-K. Ma, "Constant Envelope Precoding for Single-User Large-Scale MISO Channels: Efficient Precoding and Optimal Designs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 982–995, 2014.

[14] J. Zhang, Y. Huang, J. Wang, B. Ottersten, and L. Yang, "Per-Antenna Constant Envelope Precoding and Antenna Subset Selection: A Geometric Approach," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6089–6104, 2016.

[15] S. K. Mohammed and E. G. Larsson, "Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, 2013.

[16] J.-C. Chen, C.-K. Wen, and K.-K. Wong, "Improved Constant Envelope Multiuser Precoding for Massive MIMO Systems," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1311–1314, 2014.

[17] P. V. Amadori and C. Masouros, "Constant Envelope Precoding by Interference Exploitation in Phase Shift Keying-Modulated Multiuser Transmission," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 538–550, 2017.

[18] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An Efficient Manifold Algorithm for Constructive Interference Based Constant Envelope Precoding," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1542–1546, 2017.

[19] Z. Guo, Y. Yılmaz, and X. Wang, "Transmitter-Centric Channel Estimation and Low-PAPR Precoding for Millimeter-Wave MIMO Systems," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2925–2938, 2016.

[20] F. Boccardi and H. Huang, "Zero-Forcing Precoding for the MIMO Broadcast Channel under Per-Antenna Power Constraints," in *2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, 2006, pp. 1–5.

[21] F. Boccardi and G. Caire, "The p -Sphere Encoder: Peak-Power Reduction by Lattice Precoding for the MIMO Gaussian Broadcast Channel," *IEEE Transactions on Communications*, vol. 54, no. 11, pp. 2085–2091, 2006.

[22] S. V. Zavjalov, D. K. Fadeev, and S. V. Volvenko, "Influence of Input Power Backoff of Nonlinear Power Amplifier on BER Performance of Optimal SEFDM Signals," in *2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2016, pp. 447–450.

[23] D. Spano, M. Alodeh, S. Chatzinotas, and B. Ottersten, "Symbol-Level Precoding for the Nonlinear Multiuser MISO Downlink Channel," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1331–1345, 2018.

[24] W. Peng, L. Zheng, D. Chen, C. Ni, and T. Jiang, "Distributed Precoding for BER Minimization With PAPR Constraint in Uplink Massive MIMO Systems," *IEEE Access*, vol. 6, pp. 6668–6676, 2018.

[25] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Precoding under Instantaneous Per-antenna Peak Power Constraint," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 863–867.

[26] S. R. Aghdam, S. Jacobsson, and T. Eriksson, "Distortion-Aware Linear Precoding for Millimeter-Wave Multiuser MISO Downlink," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.

[27] R. Zayani, H. Shaïek, and D. Roviras, "Efficient Precoding for Massive MIMO Downlink Under PA Nonlinearities," *IEEE Communications Letters*, vol. 23, no. 9, pp. 1611–1615, 2019.

[28] J. Jee, G. Kwon, and H. Park, "Precoding Design and Power Control for SINR Maximization of MISO System With Nonlinear Power Amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14 019–14 024, 2020.

[29] ——, "Joint Precoding and Power Allocation for Multiuser MIMO System With Nonlinear Power Amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8883–8897, 2021.

[30] H. Moazzen, A. Mohammadi, and M. Majidi, "Performance Analysis of Linear Precoded MU-MIMO-OFDM Systems With Nonlinear Power Amplifiers and Correlated Channel," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6753–6765, 2019.

[31] A. Li, D. Spano, J. Krivochiza, S. Domouchtsidis, C. G. Tsinos, C. Masouros, S. Chatzinotas, Y. Li, B. Vucetic, and B. Ottersten, "A Tutorial on Interference Exploitation via Symbol-Level Precoding: Overview, State-of-the-Art and Future Directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 796–839, 2020.

[32] C. Masouros and E. Alsusa, "A Novel Transmitter-Based Selective-Precoding Technique for DS/CDMA Systems," *IEEE Signal Processing Letters*, vol. 14, no. 9, pp. 637–640, 2007.

[33] ——, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1396–1404, 2009.
[34] C. Masouros, "Correlation Rotation Linear Precoding for MIMO Broadcast Communications," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 252–262, 2011.
[35] C. Masouros and G. Zheng, "Exploiting Known Interference as Green Signal Power for Downlink Beamforming Optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3628–3640, 2015.
[36] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Symbol-Level Multiuser MISO Precoding for Multi-Level Adaptive Modulation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5511–5524, 2017.
[37] A. Li and C. Masouros, "Interference Exploitation Precoding Made Practical: Optimal Closed-Form Solutions for PSK Modulations," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7661–7676, 2018.
[38] A. Li, C. Masouros, B. Vucetic, Y. Li, and A. L. Swindlehurst, "Interference Exploitation Precoding for Multi-Level Modulations: Closed-Form Solutions," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 291–308, 2021.
[39] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-Bit DAC Transmission: A Low-Complexity Symbol Scaling Approach," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7559–7575, 2018.
[40] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the Energy Efficiency of MIMO Hybrid Beamforming for Millimeter-Wave Systems With Nonlinear Power Amplifiers," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7208–7221, 2018.
[41] L. Ding and G. Zhou, "Effects of Even-order Nonlinear Terms on Predistortion Linearization," in *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.*, 2002, pp. 1–6.
[42] S. Boyd and L. Chua, "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series," *IEEE Transactions on Circuits and Systems*, vol. 32, no. 11, pp. 1150–1161, 1985.
[43] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive Interference through Symbol Level Precoding for Multi-Level Modulation," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.
[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[45] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.
[46] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks: Tractable Approximations by Conic Optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5690–5705, 2014.
[47] T. Lo, "Maximum Ratio Transmission," *IEEE Transactions on Communications*, vol. 47, no. 10, pp. 1458–1461, 1999.
[48] F. Quadri and A. D. Tete, "FPGA Implementation of Digital Modulation Techniques," in *2013 International Conference on Communication and Signal Processing*, 2013, pp. 913–917.