Parallelizable First-Order Fast Algorithm for Symbol-Level Precoding in Lage-Scale Systems

Junwen Yang*, Ang Li*, Xuewen Liao*‡, and Christos Masouros†
School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China*
National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China‡
Department of Electronic and Electrical Engineering, University College London, London, UK†
Email: jwyang@stu.xjtu.edu.cn*, ang.li.2020@xjtu.edu.cn*, yeplos@mail.xjtu.edu.cn*, c.masouros@ucl.ac.uk†

***Abstract*—We investigate constructive interference (CI)-based symbol-level precoding (SLP) in large-scale systems with massive connectivity of users to minimize the transmit power subject to the instantaneous signal-to-noise-ratio (SINR) and CI constraints. By converting the considered problem into a novel separable formulation, we reveal the existence of separability in SLP, which is therefore well-suited for decomposition. The proximal Jacobian alternating direction method of multipliers (PJ-ADMM) framework is adopted to decompose the reformulated problem into multiple subproblems, which can be solved in parallel with closed-form solutions. We further linearize the second-order terms by approximation, which leads to a parallelizable first-order fast solution to SLP. Our derivations are validated by simulation results, which also show that our algorithm can provide optimal performance with substantially lower computational complexity than state-of-the-art algorithms.**

***Index Terms*—Massive MU-MISO, constructive interference, symbol-level precoding, separability, ADMM, parallel and distributed computing.**

I. Introduction

Multi-antenna technique is an overwhelming paradigm for mainstream wireless communication systems. With the growing requirements on efficiency and reliability, massive multiple-input multiple-output (M-MIMO) that equips large-scale antennas has been recognized as a promising enabler [1], [2]. Precoding is an efficient downlink transmission technique to fully utilize the degree of freedom offered by M-MIMO. Conventional precoding maps the transmit data symbols into transmit signal using channel state information (CSI), therefore classified as linear precoding or block-level precoding (BLP). Such as the zero-forcing (ZF) precoding [3], the regularized ZF (RZF) precoding [4], and the minimum mean-square error (MMSE) precoding [5], which has a closed-form precoding structure. Additionally, the object-oriented linear precoding that optimizes one performance metric subject to some given constraints can achieve a particular systematic objective, which is also widely investigated in the literature. The most representative ones include the signal-to-interference-plus-noise (SINR)-constrained power minimization (PM) precoding [6], the power-constrained max-min SINR balancing (SB) precoding [7]–[9], and the power-constrained weighted sum-rate (WSR) maximization precoding [10], [11].

The above conventional precoding inspects the interference from a statistical view, where interference cannot be controlled accurately, thus acting as an unfavorable factor. On the contrary, from an instantaneous perspective, interference is controllable and can be manipulated to enhance signal detection, which is the so-called constructive interference (CI) [12]–[14]. CI-based precoding converts the known interference into useful signal power with the aid of both CSI and data symbols, thereby also known as symbol-level precoding (SLP). It is also observed that the transmit signal of SLP is a nonlinear transformation of data symbols. The typical objective-oriented precoding schemes in SLP include PM and SB problems, which optimize the corresponding objective in each symbol slot rather than the coherence block. The advantages of CI-based SLP over conventional precoding have been validated in [12]. Despite the performance superiority of SLP, efficient solutions to the underlying optimization problems are demanded by practical communication systems, especially for large-scale settings.

Towards fast and efficient CI precoding solutions, recently several studies have endeavored to push the frontiers, such as the efficient gradient projection algorithm to solve the Lagrangian dual problem of PM-SLP [12], closed-form sub-optimal solutions to PM-SLP [15], [16], derivations of the optimal precoding structure for SB-SLP with iterative algorithms [17], [18], the CI-based block-level precoding (CI-BLP) approach [19], [20], the grouped SLP (G-SLP) approach [21], and the deep learning-based schemes for PM-SLP [22], [23]. Nevertheless, the existing works described above focus on sequential and centralized solutions and ignore the separable nature of PM-SLP, which lead to a conservative complexity reduction to the optimal solutions.

In this paper, we propose a parallelizable first-order fast algorithm for the PM-SLP problem. We first derive the separable structure for the PM-SLP problem by reformulating the canonical problem to convert it into a novel separable equivalent, which is amenable to decomposition methods. To take advantage of the revealed separability, we subsequently decompose the reformulated problem into multiple simpler subproblems leveraging the proximal Jacobian alternating direction method of multipliers (PJ-ADMM) framework. Since the second-order terms in subproblems lead to inefficient solutions with matrix inversion, we linearize them by approximation. Therefore the subproblems can be solved in closed form without matrix inversion. We also call it the parallel inverse-free SLP (PIF-SLP) algorithm. Numerical results demonstrate that the proposed

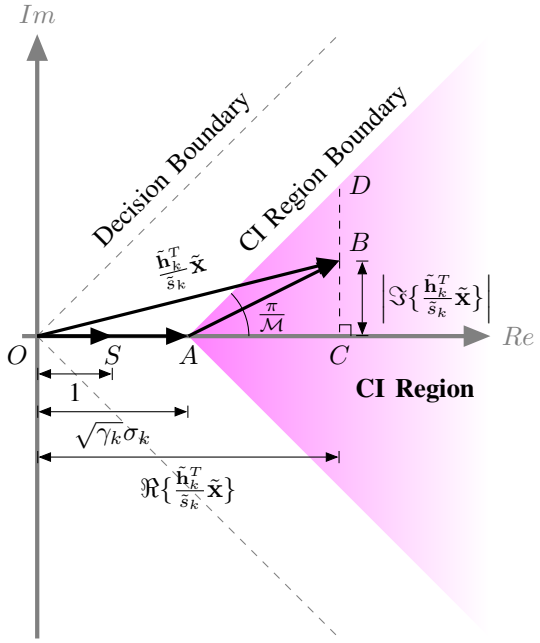


Fig. 1. Illustration of CI regions for a generic $\mathcal{M}$-PSK modulation.

PIF-SLP algorithm can converge to optimal fast, significantly outperforming existing works.

II. System Model and Problem Formulation

A. System Model

We consider a downlink multi-user multi-input single-output (MU-MISO) system, where the base station (BS) equipped with N_t antennas provides service for K single-antenna users in the same time-frequency resource. The independent random data bits for each user are modulated to normalized data symbols. The data symbol vector $\tilde{\mathbf{s}} \triangleq [\tilde{s}_1, \cdots, \tilde{s}_K]^T \in \mathbb{C}^K$ contains the overall K data symbols in a symbol slot, which is mapped to the transmit signal $\tilde{\mathbf{x}} \triangleq [\tilde{x}_1, \cdots, \tilde{x}_{N_t}]^T \in \mathbb{C}^{N_t}$ at the BS via SLP. The received signal of user k in one symbol slot is given by

$$\tilde{y}_k = \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}} + \tilde{z}_k, \tag{1}$$

where $\tilde{\mathbf{h}}_k \in \mathbb{C}^{N_t}$ denotes the quasi-static Rayleigh flat-fading channel vector between BS and user k, and $\tilde{z}_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the complex-valued additive white Gaussian noise at user k. The channel matrix is denoted by $\tilde{\mathbf{H}} \triangleq [\tilde{\mathbf{h}}_1, \cdots, \tilde{\mathbf{h}}_K]^T \in \mathbb{C}^{K \times N_t}$. To focus on the precoding design, perfect CSI is assumed.

B. Constructive Interference

According to the concept of CI that was first introduced in [24], the constructive and destructive pattern of the noiseless received signal $\left\{\tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}}\right\}$ is jointly determined by CSI and data symbols. To predict the interference pattern and further exploit the known interference, CI precoding optimizes the transmit signal by judiciously utilizing both CSI and data symbols, such that all the multi-user interference can add up constructively at each receiver [14]. When interference exploitation is attained by CI precoding, the received instantaneous SINR at user k is given as

$$\mathrm{SINR}_k = \frac{|\tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}}|^2}{\sigma_k^2}. \tag{2}$$

It can be seen that there is no interference in the denominator. Because all signal is beneficial to correct decision, thus put in the numerator. Consequently, the instantaneous SINR above is equivalent to the conventional signal-to-noise ratio (SNR).

Geometrically, CI is obtained as long as the noiseless received signal of each user lies in the symbol-specified CI region in the complex plane, where the CI region refers to a polyhedron bounded by hyperplanes parallel to decision boundaries or Voronoi edges [12], [25], and the only vertex of one CI region is the SINR threshold-dependent nominal constellation symbol, as depicted in Fig. 1. For the sake of illustration, let $\tilde{s}_k$ be the symbol of interest of user k, which is an arbitrary constellation point drawn from a normalized $\mathcal{M}$-PSK constellation. We rotate $\tilde{s}_k$ to the positive real axis, thereby the rotated symbol is 1, which is corresponding to $\overrightarrow{OS}$ in Fig. 1. Other related signals are rotated by the same phase. Consequently, the received noiseless signal of user k, $\tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}}$, turns out to $\frac{\tilde{\mathbf{h}}_k^T}{\tilde{s}_k} \tilde{\mathbf{x}}$, which is denoted by $\overrightarrow{OB}$ in Fig. 1. For a given instantaneous SINR threshold γ_k for user k, the nominal constellation point is equivalent to $\sqrt{\gamma_k}\sigma_k \tilde{s}_k$ according to (2). We introduce $\overrightarrow{OA}$ as the rotated nominal constellation point, which is also the only vertex of the interested CI region. When $\overrightarrow{OB}$ is located in the depicted CI region, then the received signal is pushed away from decision boundaries and the instantaneous SINR is guaranteed to be no less than the prescribed threshold γ_k, thus constructive to correct decision. One of the criteria that specifies the location of $\overrightarrow{OB}$ in the CI region is $\left|\overrightarrow{CD}\right| \geq \left|\overrightarrow{CB}\right|$. The corresponding explicit mathematical formulation of CI constraints for $\mathcal{M}$-PSK signaling can be written as

$$\Re\{\hat{\mathbf{h}}_k^T \tilde{\mathbf{x}}\} - \frac{|\Im\{\hat{\mathbf{h}}_k^T \tilde{\mathbf{x}}\}|}{\tan\frac{\pi}{\mathcal{M}}} \geq \sqrt{\gamma_k}\sigma_k, \forall k, \tag{3}$$

where $\hat{\mathbf{h}}_k^T \triangleq \frac{\tilde{\mathbf{h}}_k^T}{\tilde{s}_k}$, γ_k denotes the pre-defined instantaneous SINR threshold for user k. It is worth noting that the CI constraint for each user already incorporates the SINR constraint.[1]

C. Problem Formulation

Throughout this paper, we are interested in minimizing the total transmit power subject to CI constraints, which is known as the PM-SLP problem. This optimization problem has the following mathematical form:

$$\begin{aligned} \min_{\tilde{\mathbf{x}}} \quad & \|\tilde{\mathbf{x}}\|^2 \\ \text{s.t. } & \Re\{\hat{\mathbf{h}}_k^T \tilde{\mathbf{x}}\} - \frac{|\Im\{\hat{\mathbf{h}}_k^T \tilde{\mathbf{x}}\}|}{\tan\frac{\pi}{\mathcal{M}}} \geq \sqrt{\gamma_k}\sigma_k, \forall k. \end{aligned} \tag{4}$$

[1]The CI constraints can be readily extended to multi-level modulation, such as QAM, by employing the symbol-scaling metric [26].

The quadratic objective function and linear constraints indicate that this problem is convex, and hence can be handled via off-the-shelf solvers. Unfortunately, most generic solvers, e.g., SeDuMi and SDPT3, are based on the high-complexity interior-point method (IPM). To alleviate the computational burden, efficient algorithms based on gradient projection method [12], suboptimal closed-form solution [15], and improved suboptimal closed-form solution [16] were proposed. Existing works, however, focus on centralized iterative algorithms and ignore the separable nature of the PM SLP problem. Exploiting such separability, we propose the PIF-SLP algorithm in the next section based on the PJ-ADMM framework.

III. Proposed PIF-SLP Algorithm

In this section, before elaborating on the proposed PIF-SLP algorithm, we prove the underlying separability of the PM-SLP by reformulating the original problem (4) into its separable equivalent. In addition, the computational complexity of our algorithm is analyzed at the end of this section.

A. Problem Reformulation

The real-valued equivalent of (4) can be written as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|^2 \\ \text{s.t.}\ & \mathbf{N}\mathbf{S}_k\mathbf{H}_k\mathbf{x} \succeq \sqrt{\gamma_k}\sigma_k\mathbf{1},\ \forall k, \end{aligned} \tag{5}$$

where

$$\mathbf{x} \triangleq \begin{bmatrix} \Re\{\tilde{\mathbf{x}}\} \\ \Im\{\tilde{\mathbf{x}}\} \end{bmatrix} \in \mathbb{R}^{2N_t}, \mathbf{N} \triangleq \begin{bmatrix} 1 & -\frac{1}{\tan\frac{\pi}{M}} \\ 1 & \frac{1}{\tan\frac{\pi}{M}} \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

$$\mathbf{S}_k \triangleq \begin{bmatrix} \Re\left\{\frac{1}{\tilde{s}_k}\right\} & -\Im\left\{\frac{1}{\tilde{s}_k}\right\} \\ \Im\left\{\frac{1}{\tilde{s}_k}\right\} & \Re\left\{\frac{1}{\tilde{s}_k}\right\} \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

$$\mathbf{H}_k \triangleq \begin{bmatrix} \Re\{\tilde{\mathbf{h}}_k^T\} & -\Im\{\tilde{\mathbf{h}}_k^T\} \\ \Im\{\tilde{\mathbf{h}}_k^T\} & \Re\{\tilde{\mathbf{h}}_k^T\} \end{bmatrix} \in \mathbb{R}^{2\times 2N_t}.$$

We further introduce $\bar{\mathbf{A}}_k \triangleq \mathbf{N}\mathbf{S}_k\mathbf{H}_k$, and $\mathbf{b}_k \triangleq \sqrt{\gamma_k}\sigma_k\mathbf{1}$. Accordingly, the CI constraints become

$$\bar{\mathbf{A}}_k\mathbf{x} \succeq \mathbf{b}_k,\ \forall k. \tag{6}$$

Stacking the CI constraints, the compact formulation can be written as

$$\mathbf{A}\mathbf{x} \succeq \mathbf{b}, \tag{7}$$

where $\mathbf{A} \triangleq \left[\bar{\mathbf{A}}_1^T, \cdots, \bar{\mathbf{A}}_K^T\right]^T \in \mathbb{R}^{2K\times 2N_t}$, $\mathbf{b} \triangleq \left[\mathbf{b}_1^T, \cdots, \mathbf{b}_K^T\right]^T \in \mathbb{R}^{2K}$. We can identify that the left-hand side of (7) can be expressed as a linear combination of the columns of $\mathbf{A}$, i.e., $\sum_{i=1}^{2N_t}\mathbf{a}_i x_i$, where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$, x_i is the i-th entry of $\mathbf{x}$. Accordingly, (5) can be rearranged as

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 \\ \text{s.t.}\ & \sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i \succeq \mathbf{b}, \end{aligned} \tag{8}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$ with $\sum_{i=1}^{N} n_i = 2N_t$ is the i-th block of $\mathbf{x}$, composed of the adjacent and/or disadjacent elements of $\mathbf{x}$, and $\mathbf{A}_i \in \mathbb{R}^{2K\times n_i}$ is the i-th column block of $\mathbf{A}$, each column of which is uniquely taken from the columns of $\mathbf{A}$. Mathematically, for the adjacent case, $\mathbf{x} = \left[\mathbf{x}_1^T, \cdots, \mathbf{x}_N^T\right]^T$, $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_N]$, while for the disadjacent case, $\mathbf{x}_i = \mathbf{E}_i^T\mathbf{x}$, $\mathbf{A}_i = \mathbf{A}\mathbf{E}_i$, where $\mathbf{E}_i \in \mathbb{R}^{2N_t\times n_i}$, and each column of $\{\mathbf{E}_i\}$ is uniquely picked from the columns of the $2N_t \times 2N_t$ identity matrix.

With such formulation, (8) is partitioned into N blocks, here we do not confine the number of blocks, so long as N is a positive integer not greater than $2N_t$.

B. PIF-SLP Algorithm

We reformulate (8) by introducing a slack variable vector $\mathbf{c} \in \mathbb{R}_+^{2K}$ to replace the original inequality constraints with corresponding equality constraints and nonnegativity constraints as follows:

$$\begin{aligned} \min_{\mathbf{x}_i,\mathbf{c}} \quad & \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 \\ \text{s.t.}\ & \sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i = \mathbf{b} + \mathbf{c}, \\ & \mathbf{c} \succeq 0. \end{aligned} \tag{9}$$

Since the feasible region of the slack variable $\mathbf{c}$ is $\mathbb{R}_+^{2K}$, by introducing an indicator function, the nonnegativity constraints can be incorporated into the objective function:

$$\begin{aligned} \min_{\mathbf{x}_i,\mathbf{c}} \quad & \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 + \mathcal{I}_{\mathbb{R}_+^{2K}}(\mathbf{c}) \\ \text{s.t.}\ & -\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \mathbf{c} = \mathbf{0}, \end{aligned} \tag{10}$$

where $\mathcal{I}_{\mathbb{R}_+^{2K}}$ is the indicator function of $\mathbb{R}_+^{2K}$ given by

$$\mathcal{I}_{\mathbb{R}_+^{2K}}(\mathbf{c}) = \begin{cases} 0, & \text{if } \mathbf{c} \in \mathbb{R}_+^{2K}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{11}$$

The augmented Lagrangian function of (10) is defined as (14) on the top of the next page.

The augmented Lagrangian function of (8) can be obtained by minimizing (14) with respect to $\mathbf{c}$, i.e.,

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\mathbf{c}} \bar{\mathcal{L}}_\rho(\mathbf{x}, \mathbf{c}, \boldsymbol{\lambda}). \tag{13}$$

The solution of the minimization with respect to $\mathbf{c}$ can be written as

$$\mathbf{c} = \arg\min_{\mathbf{c}\in\mathbb{R}_+^{2K}} \frac{\rho}{2}\left\| -\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \mathbf{c} + \frac{\boldsymbol{\lambda}}{\rho}\right\|^2, \tag{14}$$

which is equivalent to projecting the vector $\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t - \mathbf{b} - \frac{\boldsymbol{\lambda}^t}{\rho}$ onto $\mathbb{R}_+^{2K}$, denoted by $P_{\mathbb{R}_+^{2K}}\left(\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t - \mathbf{b} - \frac{\boldsymbol{\lambda}^t}{\rho}\right)$. Its closed-form solution is given by

$$\mathbf{c} = \max\left\{\mathbf{0}, \sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i - \mathbf{b} - \frac{\boldsymbol{\lambda}}{\rho}\right\}, \tag{15}$$

$$\bar{\mathcal{L}}_\rho(\mathbf{x},\mathbf{c},\boldsymbol{\lambda}) = \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 + I_{\mathbb{R}_+^{2K}}(\mathbf{c}) + \boldsymbol{\lambda}^T\left(-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \mathbf{c}\right) + \frac{\rho}{2}\left\|-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \mathbf{c}\right\|^2$$
$$= \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 + I_{\mathbb{R}_+^{2K}}(\mathbf{c}) + \frac{\rho}{2}\left\|-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \mathbf{c} + \frac{\boldsymbol{\lambda}}{\rho}\right\|^2 - \frac{1}{2\rho}\|\boldsymbol{\lambda}\|^2 \quad (14)$$

where $\max\{\cdot\}$ represents the elementwise maximum. Based on the above, we can write (13) as

$$\mathcal{L}_\rho(\mathbf{x},\boldsymbol{\lambda}) = \sum_{i=1}^{N}\|\mathbf{x}_i\|^2 - \frac{1}{2\rho}\|\boldsymbol{\lambda}\|^2 + \frac{\rho}{2}\left\|\max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i + \mathbf{b} + \frac{\boldsymbol{\lambda}}{\rho}, \mathbf{0}\right\}\right\|^2. \quad (16)$$

By adopting PJ-ADMM [27], each of the N blocks of transmit signal is updated alternately in parallel as follows:

$$\mathbf{x}_i^{t+1} = \arg\min_{\mathbf{x}_i}\mathcal{L}_\rho\left(\mathbf{x}_{\neq i}^t, \mathbf{x}_i, \boldsymbol{\lambda}^t\right) + \frac{1}{2}\left\|\mathbf{x}_i - \mathbf{x}_i^t\right\|_{\mathbf{P}_i}^2, \forall i, \quad (17)$$

where $\mathbf{P}_i$ is a symmetric and positive semi-definite matrix and $\|\mathbf{x}_i\|_{\mathbf{P}_i}^2 \triangleq \mathbf{x}_i^T\mathbf{P}_i\mathbf{x}_i$.

The above update of $\mathbf{x}_i^{t+1}$ is equivalent to solving an unconstrained quadratic programming problem (18) as shown on the top of the next page, whose optimal solution can be obtained by setting the gradient of the objective function with respect to $\mathbf{x}_i$ to zero, i.e.,

$$2\mathbf{x}_i + \mathbf{P}_i\left(\mathbf{x}_i - \mathbf{x}_i^t\right) + \rho\mathbf{A}_i^T\mathbf{A}_i\mathbf{x}_i - \rho\mathbf{A}_i^T\left(\mathbf{A}_i\mathbf{x}_i^t + \max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho}, \mathbf{0}\right\}\right) = 0, \forall i. \quad (19)$$

After some calculations, the closed-form solution for minimizing $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i^{t+1} = \left(2\mathbf{I} + \rho\mathbf{A}_i^T\mathbf{A}_i + \mathbf{P}_i\right)^{-1}\left[\mathbf{P}_i\mathbf{x}_i^t + \rho\mathbf{A}_i^T\left(\mathbf{A}_i\mathbf{x}_i^t + \max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho}, \mathbf{0}\right\}\right)\right], \forall i. \quad (20)$$

Note that when we take $N = 2N_t$, i.e., the transmit signal vector $\mathbf{x}$ is decomposed into $2N_t$ scalars, $\mathbf{A}_i$ reduces to a column vector $\mathbf{a}_i$, and $\mathbf{P}_i$ reduces to a scalar p_i, then the iteration of the transmit signal can be carried out via $2N_t$ parallel and distributed scalar operations, i.e.,

$$x_i^{t+1} = \frac{p_i x_i^t + \rho\mathbf{a}_i^T\left(\mathbf{a}_i x_i^t + \max\left\{-\sum_{i=1}^{2N_t}\mathbf{a}_i x_i^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho}, \mathbf{0}\right\}\right)}{2 + \rho\mathbf{a}_i^T\mathbf{a}_i + p_i}, \forall i. \quad (21)$$

If we group the real and imaginary parts of the same antenna's transmit signal into one block, the transmit signal vector will be decomposed into N_t blocks. $\mathbf{A}_i \in \mathbb{R}^{2K\times 2}$ is a matrix with orthogonal columns, which implies that the corresponding $\mathbf{A}_i^T\mathbf{A}_i$ is a 2×2 diagonal matrix with identical non-zero elements. Therefore, if we take $\mathbf{P}_i$ as a diagonal matrix too, then the matrix inverse operation during the update of $\mathbf{x}_i$ will be replaced by taking the reciprocals of the two entries in the main diagonal, with reduced complexity. Hence the $\mathbf{x}_i$-updates can be reformulated as

$$\mathbf{x}_i^{t+1} = \left[\mathbf{P}_i\mathbf{x}_i^t + \rho\mathbf{A}_i^T\left(\mathbf{A}_i\mathbf{x}_i^t + \max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho}, \mathbf{0}\right\}\right)\right] \oslash \mathbf{W}, \forall i, \quad (22)$$

where $\oslash$ denotes the element-wise division, $\mathbf{W} \triangleq diag\left(2\mathbf{I} + \rho\mathbf{A}_i^T\mathbf{A}_i + \mathbf{P}_i\right)$.

To further reduce the complexity by circumventing matrix inversion, we set the $\mathbf{P}_i$ as $\mathbf{P}_i = \tau_i\mathbf{I} - \rho\mathbf{A}_i^T\mathbf{A}_i\mathbf{x}_i$. Subsequently, the parallel inverse-free update of $\mathbf{x}_i$ turns to

$$\mathbf{x}_i^{t+1} = \frac{1}{2+\tau_i}\left(\tau_i\mathbf{x}_i^t + \rho\mathbf{A}_i^T\max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho}, \mathbf{0}\right\}\right), \forall i. \quad (23)$$

The Lagrangian multiplier is updated via the gradient iteration given by

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t + \beta\rho\max\left\{-\sum_{i=1}^{N}\mathbf{A}_i\mathbf{x}_i^{t+1} + \mathbf{b}, -\frac{\boldsymbol{\lambda}^t}{\rho}\right\}, \quad (24)$$

where $\beta > 0$ is a damping parameter. Consequently, we arrive at a PIF-SLP algorithm, which is summarized in Algorithm 1.

C. Computational Complexity Analysis

The computational overhead of the proposed PIF-SLP algorithm is assessed by accounting for the required float-point operations, i.e., flops. The PIF-SLP summarized in Algorithm 1 updates two variables alternately, i.e., the first step updates N blocks of the transmit signal $\{\mathbf{x}_i\}$ in parallel, and the second step updates the Lagrangian multiplier $\boldsymbol{\lambda}$. To simplify the analysis, it is assumed that each block of transmit signal has the same dimension, i.e., $2N_t/N$. As a result, the update of the transmit signal (23) requires $\mathcal{O}(2TK)+\mathcal{O}(2T(2K+1)N_t/N)$ flops per block, where T denotes the number of iterations. On the assumption that given $\{\mathbf{A}_i\mathbf{x}_i\}$ rather than $\{\mathbf{x}_i\}$, the update of the Lagrangian multiplier (24) requires $\mathcal{O}(2TK)$ flops.

$$\mathbf{x}_i^{t+1} = \arg\min_{\mathbf{x}_i} \|\mathbf{x}_i\|^2 + \frac{1}{2}\left\|\mathbf{x}_i - \mathbf{x}_i^t\right\|_{\mathbf{P}_i}^2 + \frac{\rho}{2}\left\|-\mathbf{A}_i\mathbf{x}_i + \mathbf{A}_i\mathbf{x}_i^t - \sum_{j=1}^{N}\mathbf{A}_j\mathbf{x}_j^t + \mathbf{b} + \frac{\boldsymbol{\lambda}^t}{\rho} + \max\{\mathbf{0}, \sum_{j=1}^{N}\mathbf{A}_j\mathbf{x}_j^t - \mathbf{b} - \frac{\boldsymbol{\lambda}^t}{\rho}\}\right\|^2, \forall i, \tag{18}$$

Algorithm 1 PIF-SLP for the PM-SLP problem (8)

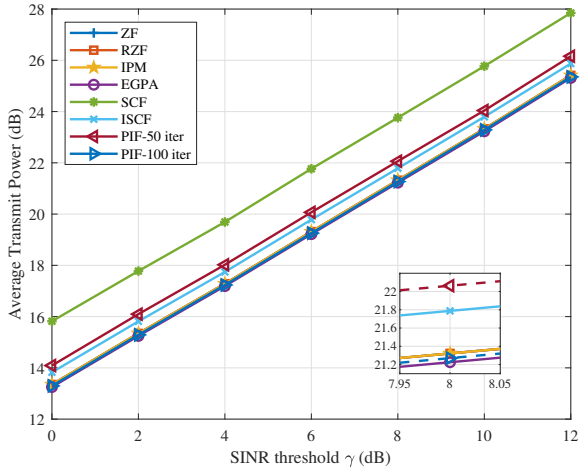
Input: $\mathbf{A}$, $\mathbf{b}$, ρ , τ_i, β
Output: $\mathbf{x}$

1: Initialize $\mathbf{x}_i^0$ $(i = 1, \cdots, N)$, and $\boldsymbol{\lambda}^0$;
2: Set $t \leftarrow 0$;
3: **repeat**
4: Update $\mathbf{x}_i^{t+1}$ for $i = 1, \cdots, N$ in parallel by:
5: **for** $i = 1, \cdots, N$ **do**
6: Update $\mathbf{x}_i^{t+1}$ by (23);
7: **end for**
8: Update $\boldsymbol{\lambda}^{t+1}$ by (24);
9: Set $t \leftarrow t + 1$;
10: **until** Convergence.

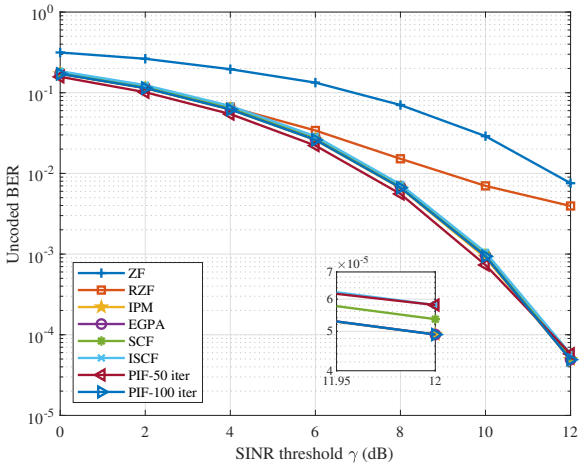
IV. Numerical Results

This section evaluates and compares the performance of the proposed PIF-SLP algorithm via Monte Carlo simulations. We assume each user has unit noise variance and an equal instantaneous SINR threshold, i.e., $\sigma_k^2 = \sigma^2 = 1$, $\gamma_k = \gamma, \forall k$. QPSK modulation is employed throughout the simulations. A downlink massive MU-MISO system with 128 transmit antennas to serve 112 single-antenna users is considered. The transmit signal vector is partitioned into 64 blocks, namely, $N = 64$, with 4 elements in each block. We choose $\tau_i = 0.2\rho\left(\frac{N}{2-\beta} - 1\right)\|\mathbf{A}_i\|^2$. The penalty parameter ρ is set to 0.06; the damping parameter β is set to 1. For a comprehensive comparison, we consider the followings: the conventional ZF and RZF schemes with symbol-level power normalization, the IPM for PM-SLP implemented by CVX [28], the efficient gradient projection algorithm (EGPA) for PM-SLP [12], the suboptimal closed-form (SCF) solution for PM-SLP [15], and the improved suboptimal closed-form (ISCF) solution for PM-SLP [16].

Fig. 2a depicts the average transmit power of the compared schemes with the SINR threshold for the considered system setting, where the transmit power of ZF and RZF is normalized by that of the IPM. It is observed that the transmit power of the proposed PIF-SLP algorithm approaches those of the IPM from low to high. Since we initialize the transmit signal as a zero vector. Specifically, the early termination of the proposed PIF-SLP algorithm at 50 iterations leads to a suboptimal solution of nearly 0.7 dB transmit power gap. When the number of iterations reaches 100, the proposed algorithm can almost provide optimal transmit power.



(a) Average transmit power v.s. SINR threshold



(b) Uncoded BER v.s. SINR threshold

Fig. 2. Transmit power and uncoded BER performance of different schemes in different SINR thresholds, QPSK, $N_t = 128$, $K = 112$, $N = 64$.

Fig. 2b shows the uncoded BER performance of the proposed PIF-SLP algorithm compared to other schemes at various SINR thresholds for the same system setting. At the interference-limited medium-to-high SINR threshold region, the PM-SLP schemes implemented by the proposed PIF-SLP algorithm, IPM, SCF, ISCF, and EGPA all achieve lower uncoded BER over the ZF and RZF schemes. The performance of the proposed algorithm increases stably with the number of iterations, providing a performance-complexity trade-off. With sufficient iterations, the uncoded BER performance of the proposed algorithm can match that of the IPM.

TABLE I
AVERAGE EXECUTION TIME IN SEC. OVER 2000 RANDOM CHANNEL REALIZATIONS, QPSK, $N_t = 128$, $K = 112$, $N = 64$.

	γ = 4 dB	γ = 8 dB	γ = 12 dB
ZF	0.0034	0.0033	0.0033
RZF	0.0027	0.0028	0.0028
IPM	0.4002	0.3962	0.3920
EGPA	0.2248	0.2312	0.2280
SCF	0.0209	0.0209	0.0207
ISCF	0.0218	0.0217	0.0217
PIF-50 iter	0.0104	0.0105	0.0104
PIF-100 iter	0.0193	0.0194	0.0192

Table I presents the average execution time per channel realization of the considered schemes in three different SINR thresholds, where the parameters are the same as those in Fig. 2. It should be noted that implementing the parallel approach in physical parallel computing processors is beyond the range of this paper. Thus the execution time for the PIF-SLP algorithm is the total time required for MATLAB simulation, which is an overestimate. We can observe that the proposed PIF-SLP algorithm is the most efficient SLP algorithm compared to the IPM and EGPA. The average execution time of the PIF-SLP is almost proportional to the number of iterations. When the number of iterations of the PIF-SLP algorithm increases to 100, the PIF-SLP, SCF, and ISCF have comparable average execution times.

V. CONCLUSION

In this paper, a parallelizable first-order fast algorithm for CI-based SLP is proposed for a massive MU-MISO downlink system based on PJ-ADMM. We reformulate the canonical PM-SLP optimization problem into separable equality-constrained optimization, which is further decomposed into multiple parallel subproblems. The second-order term in each subproblem is substracted by the PJ-ADMM framework, providing an inversion-free solution. Numerical results demonstrate that the proposed algorithm can provide optimal performance and show the superiority of the proposed algorithm in terms of computational efficiency over state-of-the-art works.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[2] L. Liu and W. Yu, "Massive connectivity with massive mimo—part ii: Achievable rate characterization," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2947–2959, 2018.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[4] C. Peel, B. Hochwald, and A. 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, 2005.

[5] M. Joham, K. Kusume, M. Gzara, W. Utschick, and J. Nossek, "Transmit wiener filter for the downlink of tddds-cdma systems," in *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications,*, vol. 1, 2002, pp. 9–13 vol.1.

[6] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *1999 IEEE 49th Vehicular Technology Conference (Cat. No.99CH36363)*, vol. 1, 1999, pp. 851–856 vol.1.

[7] D. Palomar, J. Cioffi, and M. Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, 2003.

[8] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual sinr constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, 2004.

[9] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, 2006.

[10] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[11] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[12] 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, 2015.

[13] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the miso downlink channel," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2239–2252, 2015.

[14] 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 Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 796–839, 2020.

[15] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Power minimizer symbol-level precoding: A closed-form suboptimal solution," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1730–1734, 2018.

[16] ——, "An approximate solution for symbol-level multiuser precoding using support recovery," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.

[17] 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.

[18] 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.

[19] A. Li, C. Shen, X. Liao, C. Masouros, and A. L. Swindlehurst, "Practical interference exploitation precoding without symbol-by-symbol optimization: A block-level approach," *arXiv preprint arXiv:2202.09830*, 2022.

[20] ——, "Block-level interference exploitation precoding without symbol-by-symbol optimization," *arXiv preprint arXiv:2203.12502*, 2022.

[21] 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.*, pp. 1–1, 2022.

[22] Z. Lei, X. Liao, Z. Gao, and A. Li, "CI-NN: A model-driven deep learning-based constructive interference precoding scheme," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1896–1900, 2021.

[23] A. Mohammad, C. Masouros, and Y. Andreopoulos, "An unsupervised deep unfolding framework for robust symbol level precoding," *arXiv preprint arXiv:2111.08129*, 2021.

[24] C. Masouros and E. Alsusa, "A novel transmitter-based selective-precoding technique for ds/cdma systems," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 637–640, 2007.

[25] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, 2018.

[26] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Vector perturbation based on symbol scaling for limited feedback miso downlinks," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, 2014.

[27] W. Deng, M.-J. Lai, Z. Peng, and W. Yin, "Parallel multi-block admm with o (1/k) convergence," *J. Sci. Comput.*, vol. 71, no. 2, pp. 712–736, 2017.

[28] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.