

An Approximate Solution for Symbol-Level Multiuser Precoding Using Support Recovery

Alireza Haqiqatnejad, Farbod Kayhan, and Björn Ottersten

Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg

email: {alireza.haqiqatnejad, farbod.kayhan, bjorn.ottersten}@uni.lu

Abstract

In this paper, we propose a low-complexity method to approximately solve the SINR-constrained optimization problem of symbol-level precoding (SLP). First, assuming a generic modulation scheme, the precoding optimization problem is recast as a standard non-negative least squares (NNLS). Then, we improve an existing closed-form SLP (CF-SLP) scheme using the conditions for nearly perfect recovery of the optimal solution support, followed by solving a reduced system of linear equations. We show through simulation results that in comparison with the CF-SLP method, the improved approximate solution of this paper, referred to as ICF-SLP, significantly enhances the performance with a negligible increase in complexity. We also provide comparisons with a fast-converging iterative NNLS algorithm, where it is shown that the ICF-SLP method is comparable in performance to the iterative algorithm with a limited maximum number of iterations. Analytic discussions on the complexities of different methods are provided, verifying the computational efficiency of the proposed method. Our results further indicate that the ICF-SLP scheme performs quite close to the optimal SLP, particularly in the large system regime.

Index Terms

Downlink MU-MIMO, NNLS optimization, SINR-constrained power minimization, symbol-level precoding.

I. INTRODUCTION

In wireless multiuser multi-input multi-output (MU-MIMO) broadcast channels, precoding techniques can be employed in order to mitigate the channel-induced multiuser interference (MUI) via spatially pre-processing the users' data stream prior to transmission. This pre-processing, in the optimal case, is shown to achieve the capacity of the MU-MIMO broadcast channel [1]. Beyond simple linear precoding schemes, such as (regularized) zero-forcing (ZF) [2], in a practical scenario the precoding design usually aims at optimizing a certain objective function subject to some given system/user requirements; this kind of design is often called objective-oriented precoding optimization [3]. Within a wide variety of objective-oriented design criteria, two closely-related formulations are frequently addressed, namely, signal-to-interference-plus-noise ratio (SINR)-constrained power minimization [4]–[6], and the max-min SINR with power constraints [6], [7], where “power” may refer to either *total* or *per-antenna* transmit power.

From a different point of view, multiuser precoding schemes can be classified broadly into two groups, namely, block-level (conventional) and symbol-level techniques. The conventional precoding typically exploits the channel state information (CSI) to mitigate the MUI, regardless of the instantaneous users' data symbols; see e.g. [4]. The precoder then may be redesigned according to the channel coherence time. On the other hand, symbol-level precoding (SLP) takes advantage of the readily-available data information (DI) by converting the instantaneous MUI into a constructive signal component, lying onto the so-called constructive interference (CI) regions [8], [9]. The symbol-level design, therefore, requires to be specifically optimized for every instantaneous realization of the users' symbols. In delay-sensitive wireless applications, online precoding computation may suffer from high computational complexity of the symbol-level design. Rather, an offline computation also leads to an unfavorable computation cost for high-order modulation schemes even with moderate number of users [10], [11]. Nonetheless, the considerable performance improvement offered by a symbol-level precoder is motivating to find a more practical solution with a reasonable complexity.

Recently, a promising effort has been made towards low complexity (sub-optimal) solutions for various types of the SLP design problem. The authors in [12] propose an iterative method with a closed-form update equation for the max-min SINR SLP, where the algorithm is shown to converge to the optimal solution in a few iterations. In [13], a closed-form sub-optimal solution is obtained for the SINR-constrained power minimization SLP using the Karush-Kuhn-Tucker (KKT) optimality conditions. In another recent work [14], the SINR-constrained power minimization SLP is addressed with strict phase constraints on the received signals, and a low complexity approximate method is suggested for this particular case. However, the major drawback of the two latter methods is poor performance of the approximate solution for large numbers of transmit antennas and users.

In this paper, we revisit the SINR-constrained power minimization SLP problem assuming a generic modulation scheme with distance-preserving CI regions (DPCIR) (Section II). The original formulation can be transformed into an equivalent non-negative least squares (NNLS) problem (Section III). The NNLS representation enables us to derive a low-complexity approximate solution in a systematic way (Section IV). This solution, which improves the method presented in [13], simply applies a validation step

The authors are supported by the Luxembourg National Research Fund (FNR) under CORE Junior project: C16/IS/11332341 Enhanced Signal Space opTimization for satellite comMunication Systems (ESSTIMS).

before calculating the final solution. Despite a slight increase in complexity, the new method shows noticeable performance gains. In particular, unlike [13], the gap to the optimal SLP remains almost steady with enlarging the system. It is further shown that the new method can be used as an alternative to (even fast-converging) NNLS algorithms, especially when complexity is a practical design limitation.

Notations: We use uppercase and lowercase bold-faced letters to denote matrices and vectors, respectively. The sets of real and complex numbers are represented by \mathbb{R} and \mathbb{C} . For a matrix \mathbf{A} , $\mathcal{R}(\mathbf{A})$ represents the column space of \mathbf{A} . $\text{diag}(\cdot)$, or $\text{blkdiag}(\cdot)$, represents a square (block) matrix having main-diagonal (block) entries and zero off-diagonals. For a set S , $|S|$ denotes the cardinality of S . Given two vectors \mathbf{x} and \mathbf{y} with equal dimensions, $\mathbf{x} \succeq \mathbf{y}$ (or $\mathbf{x} \succ \mathbf{y}$) denotes the entrywise inequality. $\|\cdot\|_2$ represent the vector Euclidean norm. \mathbf{I} and $\mathbf{0}$ respectively stand for the identity matrix and the zero matrix (or the zero vector, depending on the context) of appropriate dimensions. The operator \otimes stands for the Kronecker product.

II. SYSTEM MODEL AND CI CONSTRAINTS

We consider an MU-MIMO broadcast channel in which a common transmitter (e.g., a base station), equipped with N antennas, serves K single-antenna users by sending independent data streams, where $K \leq N$. We denote by row vectors $\mathbf{h}_k \in \mathbb{C}^{1 \times N}$, $k = 1, \dots, K$, the instantaneous (frequency-flat) fading channels of the transmit/receive antenna pairs. Focusing on a specific symbol instant, in the downlink transmission, independent data symbols $\{s_k\}_{k=1}^K$ are intended for different users, where the symbol s_k corresponds to the k th user.

The set of desired symbols for all K users needs to be mapped to N transmit antennas, yielding the transmit signal $\mathbf{u} = [u_1, \dots, u_N]^T \in \mathbb{C}^{N \times 1}$. This mapping is done by means of a multiuser precoding module. In this paper, we adopt a symbol-level precoding (SLP) scheme. Thereby, the optimal transmit vector \mathbf{u} is directly obtained as a result of an objective-oriented precoding optimization on a symbol-level basis. At the receiver of the k th user, the observed signal can be expressed as

$$r_k = \mathbf{h}_k \mathbf{u} + z_k, \quad k = 1, \dots, K, \quad (1)$$

where z_k represents the additive circularly symmetric complex Gaussian noise distributed as $z_k \sim \mathcal{CN}(0, \sigma_k^2)$. The k -th user may use the maximum-likelihood (ML) single-user detector to optimally detect its desired symbol s_k ; nevertheless, the structure of the receiver is independent of the precoder design. In the rest, we adopt the equivalent real-valued notations

$$\tilde{\mathbf{u}} = \begin{bmatrix} \text{Re}\{\mathbf{u}\} \\ \text{Im}\{\mathbf{u}\} \end{bmatrix}, \mathbf{H}_k = \begin{bmatrix} \text{Re}\{\mathbf{h}_k\} & -\text{Im}\{\mathbf{h}_k\} \\ \text{Im}\{\mathbf{h}_k\} & \text{Re}\{\mathbf{h}_k\} \end{bmatrix}, \mathbf{s}_k = \begin{bmatrix} \text{Re}\{s_k\} \\ \text{Im}\{s_k\} \end{bmatrix},$$

where $\tilde{\mathbf{u}} \in \mathbb{R}^{2N \times 1}$, and $\mathbf{H}_k \in \mathbb{R}^{2 \times 2N}$ and $\mathbf{s}_k \in \mathbb{R}^{2 \times 1}$ for all $k = 1, \dots, K$. Clearly, $\mathbf{H}_k \tilde{\mathbf{u}} = [\text{Re}\{\mathbf{h}_k \mathbf{u}\}, \text{Im}\{\mathbf{h}_k \mathbf{u}\}]^T$.

To exploit the DI in a symbol-level precoded broadcast, one needs to design the transmit signal such that each (noise-free) received signal $\mathbf{H}_k \tilde{\mathbf{u}}$ is observed within a pre-defined region corresponding to the intended symbol, called constructive interference region (CIR). The CIRs, which are modulation-specific regions, have been defined in several ways in the literature; see, e.g., [8], [9], [15]. As mentioned earlier, we focus on the so-called DPCIRs [15], which are presented in a generic form that is applicable to any given (two-dimensional) modulation scheme.

For the sake of simplicity of analysis, and without loss of generality, we assume that identical modulation schemes are employed for all K users. The associated symbol constellation is represented by $X = \{\mathbf{x}_m : \mathbf{x}_m \in \mathbb{R}^{2 \times 1}\}_{m=1}^M$, where X is an equiprobable set with unit average power. We denote by $\text{bd}(X)$ and $\text{int}(X)$, respectively, the sets of boundary and interior points of X . It has been shown in [11] that any $\mathbf{x} \in \mathbb{R}^{2 \times 1}$ belonging to the DPCIR of \mathbf{x}_m satisfies

$$\mathbf{A}_m (\mathbf{x} - \mathbf{x}_m) \succeq \mathbf{0}, \quad \text{if } \mathbf{x}_m \in \text{bd}(X), \quad (2)$$

or $\mathbf{A}_m (\mathbf{x} - \mathbf{x}_m) = \mathbf{0}$ otherwise, where $\mathbf{A}_m = [\mathbf{a}_{m,1}, \mathbf{a}_{m,2}]^T = [\mathbf{x}_m - \mathbf{x}_{m,1}, \mathbf{x}_m - \mathbf{x}_{m,2}]^T \in \mathbb{R}^{2 \times 2}$ contains the normal vectors of distance-preserving boundaries, with $\mathbf{x}_{m,1}$ and $\mathbf{x}_{m,2}$ denoting two (specific) neighboring constellation points of \mathbf{x}_i . Let $\boldsymbol{\delta}_m \in \mathbb{R}_+^{2 \times 1}$ be a non-negative vector, then the representation in (2) can be equally expressed by

$$\mathbf{A}_m (\mathbf{x} - \mathbf{x}_m) = \boldsymbol{\delta}_m, \quad \text{where } \begin{cases} \boldsymbol{\delta}_m \succeq \mathbf{0}, & \mathbf{x}_m \in \text{bd}(X), \\ \boldsymbol{\delta}_m = \mathbf{0}, & \mathbf{x}_m \in \text{int}(X), \end{cases} \quad (3)$$

For a detailed discussion on the characteristics of DPCIRs, the interested readers are kindly referred to [11].

III. SINR-CONSTRAINED POWER MINIMIZATION SLP

In this section, we overview the instantaneous (per-symbol) power minimization problem constrained by CIRs and given SINR requirements γ_k , $k = 1, \dots, K$. The users' intended symbols $\{\mathbf{s}_k\}_{k=1}^K$ are taken from the set of points $\{\mathbf{x}_m\}_{m=1}^M$ in X . We denote by m_k the index of the constellation point that corresponds to \mathbf{s}_k , i.e., $\mathbf{s}_k = \mathbf{x}_{m_k}$ where $\mathbf{x}_{m_k} \in X$ and $m_k \in \{1, \dots, M\}$. By assuming DPCIRs, the convex representation in (3) can be used to imply the CI constraint in the optimization problem. By

substituting $\mathbf{H}_k \tilde{\mathbf{u}}$ for \mathbf{x} and replacing the scaled symbol $\sigma_k \sqrt{\gamma_k} \mathbf{x}_{m_k}$, the joint CI/SINR constraint for the k -th user is expressed by

$$\mathbf{A}_{m_k} (\mathbf{H}_k \tilde{\mathbf{u}} - \sigma_k \sqrt{\gamma_k} \mathbf{x}_{m_k}) = \boldsymbol{\delta}_{m_k}, \quad \boldsymbol{\delta}_{m_k} \succeq \mathbf{0}, \quad (4)$$

where $\boldsymbol{\delta}_{m_k} = \mathbf{0}$ is imposed for $\mathbf{x}_{m_k} \in \text{int}(X)$. Let \mathbf{W} be a square binary weighting matrix defined as

$$\mathbf{W} \triangleq \text{diag}(w_{m_1}, \dots, w_{m_K}) \otimes \mathbf{I}_2, \quad w_{m_k} = \begin{cases} 1, & \mathbf{x}_{m_k} \in \text{bd}(X), \\ 0, & \mathbf{x}_{m_k} \in \text{int}(X). \end{cases} \quad (5)$$

By stacking the constraints in (4) for all $k \in \{1, \dots, K\}$ into a compact matrix form, we have

$$\mathbf{A}(\tilde{\mathbf{H}}\tilde{\mathbf{u}} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}^{1/2}\tilde{\mathbf{x}}) = \mathbf{W}\boldsymbol{\delta}, \quad \boldsymbol{\delta} \succeq \mathbf{0}, \quad (6)$$

where $\tilde{\mathbf{H}} \triangleq [\mathbf{H}_1^T, \dots, \mathbf{H}_K^T]^T$, $\mathbf{A} \triangleq \text{blkdiag}(\mathbf{A}_{m_1}, \dots, \mathbf{A}_{m_K})$, $\boldsymbol{\Sigma} \triangleq \text{diag}(\sigma_1, \dots, \sigma_K) \otimes \mathbf{I}_2$, $\boldsymbol{\Gamma} \triangleq \text{diag}(\gamma_1, \dots, \gamma_K)^T \otimes \mathbf{I}_2$, $\tilde{\mathbf{x}} \triangleq [\mathbf{x}_{m_1}, \dots, \mathbf{x}_{m_K}]^T$, $\boldsymbol{\delta} \triangleq [\boldsymbol{\delta}_{m_1}, \dots, \boldsymbol{\delta}_{m_K}]^T$, and $(\cdot)^{1/2}$ denotes the matrix square root. Then, the optimal symbol-level precoded transmit vector can be obtained by the following lemma [11].

Lemma 1. *The minimum-norm vector satisfying the DPCIR constraint of (6) is given by*

$$\tilde{\mathbf{u}}^* = \tilde{\mathbf{H}}^\dagger \left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{1/2}\tilde{\mathbf{x}} + \mathbf{A}^{-1}\mathbf{W}\boldsymbol{\delta}^* \right), \quad (7)$$

where $\boldsymbol{\delta}^*$ is the optimal solution to the following non-negative least squares (NNLS) problem

$$\min_{\boldsymbol{\delta} \succeq \mathbf{0}} \|\tilde{\mathbf{H}}^\dagger \boldsymbol{\Sigma}\boldsymbol{\Gamma}^{1/2}\tilde{\mathbf{x}} + \tilde{\mathbf{H}}^\dagger \mathbf{A}^{-1}\mathbf{W}\boldsymbol{\delta}\|^2. \quad (8)$$

It follows from Lemma 1 that the design problem of interest can be tackled through solving the NNLS optimization in (9). Furthermore, denoting $\mathbf{B} \triangleq -\tilde{\mathbf{H}}^\dagger \mathbf{A}^{-1}\mathbf{W}$ and $\mathbf{y} \triangleq \tilde{\mathbf{H}}^\dagger \boldsymbol{\Sigma}\boldsymbol{\Gamma}^{1/2}\tilde{\mathbf{x}}$, the NNLS problem (8) can be written in the standard form as

$$\min_{\boldsymbol{\delta} \succeq \mathbf{0}} \|\mathbf{y} - \mathbf{B}\boldsymbol{\delta}\|^2, \quad (9)$$

The NNLS problem, unlike its unconstrained counterpart, does not in general admit a closed-form solution due to the non-negativity constraints. Various efficient algorithms to solve an NNLS can be found in the literature on iterative optimization, such as the well-known active set based method proposed by Lawson and Hanson [16], the fast NNLS algorithm (FNNLS) [17], and those based on projected/proximal gradient method [18]–[20]. However, an NNLS algorithm, in the best known case, requires tens of iterations to converge. For instance, the accelerated gradient method have a linear convergence rate of $\mathcal{O}(n^{-2})$, where n is the number of iterations. With a convex objective function, this translates to a worst-case complexity bound of $\mathcal{O}(\epsilon^{-1/2})$ to reach an ϵ -optimal solution. As an illustrative example, using the accelerated projected gradient descent, it takes nearly 100 iterations to have a residual of 10^{-3} with respect to the optimum. In a symbol-level design application, this process needs to be done as repeatedly as either the frame length or the total number of possible symbol realizations for K users, i.e., M^K . This motivates us to still be looking for a more computationally efficient, though possibly approximate, solution for the SLP design problem.

IV. LOW-COMPLEXITY SOLUTION FOR NNLS-BASED SLP

The main goal of this section is to obtain a low-complexity solution for the NNLS design formulation of SLP in (9). We first proceed by reviewing some basic mathematical analysis on the NNLS problem.

Let $\boldsymbol{\delta}^* = [\delta_1^*, \dots, \delta_{2K}^*]^T$ denote the minimizer corresponding to the standard NNLS formulation in (9). We refer to the set of indices j for which $\delta_j^* > 0$ as the support of $\boldsymbol{\delta}^*$, or the optimal support, denoted by $\Lambda^* = \{j : \delta_j^* > 0\}$. Given the optimal support Λ^* , the minimizer of (9) can be simply computed by $(\mathbf{B}_{\Lambda^*})^\dagger \mathbf{y}$ with appropriate zero-padding, where \mathbf{B}_{Λ^*} denotes the matrix composed of those columns of \mathbf{B} associated with the indices in Λ^* . In other words, finding Λ^* is as complex as solving (9) for the optimal solution. Therefore, one may attempt to solve (9) equivalently by perfectly identifying Λ^* . This is in fact the underlying idea of active set methods, where at each iteration some constraints are set to be active (i.e., zero-valued in our context), while the other constraints are used in the update equation. However, here we are interested in having an estimate of Λ^* , say $\hat{\Lambda}$, obtained in a non-iterative manner. This allows us to derive an approximate solution $\hat{\boldsymbol{\delta}}$ in an explicit form. Our proposed method is essentially based on the following lemma which states sufficient conditions for nearly perfect recovery of the optimal support [21].

Lemma 2. *Let Λ be a subset of column indices of the matrix \mathbf{B} with $|\Lambda| \leq 2K$, and the columns associated with the indices in Λ are linearly independent. Let $\boldsymbol{\delta}^* \succeq \mathbf{0}$ be the minimizer of $\|\mathbf{y} - \mathbf{B}\boldsymbol{\delta}\|^2$. Then, Λ coincides with the support of $\boldsymbol{\delta}^*$ if*

$$\text{C1: } \mathbf{B}_\Lambda^\dagger \mathbf{y} \succ \mathbf{0}, \quad \text{and} \quad \text{C2: } \mathbf{y}^T \mathbf{P}_\Lambda^\perp \mathbf{b}_i < 0, \quad \forall i \in \Lambda^c,$$

where \mathbf{P}_Λ^\perp is the projector onto the orthogonal complement of $\mathcal{R}(\mathbf{B}_\Lambda)$, \mathbf{b}_i denotes the i th column of \mathbf{B}_Λ , and $\Lambda^c = \{1, \dots, 2K\} - \Lambda$.

Algorithm 1 APGD algorithm for the SLP NNLS problem (9)

1: **input** $\mathbf{B}, \mathbf{y}, n_{\max}$
2: **initialize** $\boldsymbol{\vartheta}_0 = \boldsymbol{\delta}_0 \in \mathbb{R}_+^{2K \times 1}, \mathbf{Q} = \mathbf{I} - \frac{\mathbf{B}^T \mathbf{B}}{\|\mathbf{B}^T \mathbf{B}\|_F}, \boldsymbol{\phi} = \frac{\mathbf{B}^T \mathbf{y}}{\|\mathbf{B}^T \mathbf{B}\|_F}$
3: **set** $\eta = \frac{1 - \sqrt{\kappa}}{1 + \sqrt{\kappa}}, \kappa = \frac{\sigma_{\max}(\mathbf{B})}{\sigma_{\min}(\mathbf{B})}, n = 0$
4: **while** $n < n_{\max}$ **do**
5: $n = n + 1$
6: $\boldsymbol{\delta}_n = \max\{\mathbf{Q}\boldsymbol{\vartheta}_{n-1} + \boldsymbol{\phi}, \mathbf{0}\}$
7: $\boldsymbol{\vartheta}_n = \boldsymbol{\delta}_n + \eta(\boldsymbol{\delta}_n - \boldsymbol{\delta}_{n-1})$
8: **end while**

note: $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ respectively denote the maximum and minimum singular values of an input matrix.

Based on Lemma 2, both the conditions C1 and C2 together are sufficient for a candidate support Λ to be optimal. In fact, C1 measures if the resultant solution satisfies the positivity constraint (notice that the constraint cannot be satisfied with equality due to the definition of support), while the projection in C2 can be read as the deviation of \mathbf{y} from the column space of \mathbf{B}_Λ . In other words, C1 is required to validate the columns already indexed in Λ , whereas C2 assesses the possibility of including any of the columns belonging to Λ^c . Armed with these two conditions, we are ready to approximately solve the NNLS problem in (9), as will be explained in the sequel.

A. The Proposed Approximate Solution

First, we exploit the projection condition C2 to produce a rough estimate of Λ^* . Let

$$d_i \triangleq \mathbf{y}^T \mathbf{P}_\Lambda^\perp \mathbf{b}_i, \quad i = 1, \dots, 2K,$$

where

$$\mathbf{P}_\Lambda^\perp = \mathbf{I} - \mathbf{B}_\Lambda \left(\mathbf{B}_\Lambda^T \mathbf{B}_\Lambda \right)^{-1} \mathbf{B}_\Lambda^T.$$

Treating the entire columns of \mathbf{B} as candidate columns to be indexed in Λ , we assume $\Lambda^c = \{1, \dots, 2K\}$, yielding $\mathbf{P}_\Lambda^\perp = \mathbf{I}$. Hence,

$$d_i = \mathbf{y}^T \mathbf{b}_i, \quad i = 1, \dots, 2K. \quad (10)$$

With the inner products in (10), we define $\hat{\Lambda} \triangleq \{i : d_i > 0\}$ with $|\hat{\Lambda}| = L_1$, which builds our initial estimate of Λ^* . Notice that the conditions in (10) are similarly implied from the KKT optimality conditions, as discussed in [13]. Next, we validate this estimate by excluding those columns in $\hat{\Lambda}$ that result in negative elements for $\boldsymbol{\delta}$, i.e.,

$$\hat{\Lambda} \triangleq \left\{ l : l \in \hat{\Lambda}, [(\mathbf{B}_{\hat{\Lambda}})^\dagger \mathbf{y}]_l > 0 \right\}, \quad (11)$$

where $[\cdot]_l$ denotes the l th element of an input vector. Clearly, we have $|\hat{\Lambda}| \triangleq L_2 \leq L_1$, which reduces the possibility of having negative elements in the final solution as a result of the additional validation step in (11). Our simulations indicate that in most cases $\hat{\Lambda}$ gives a more accurate estimate of the optimal support Λ^* , compared to that given by $\hat{\Lambda}$, as we will see in the next section. Notice, however, that the positivity constraints may still be violated even after the validation step in (11) since the remaining set of columns in $\hat{\Lambda}$ does not necessarily guarantee that $(\mathbf{B}_{\hat{\Lambda}})^\dagger \mathbf{y} \succ \mathbf{0}$. Therefore, one needs to ignore all the negative elements in the final solution, if any. More precisely, due to the fact that $\mathcal{R}(\mathbf{B}_{\hat{\Lambda}}) \subseteq \mathcal{R}(\mathbf{B}_{\hat{\Lambda}})$, perfect recovery of the optimal support is possible only if $\Lambda^* \subseteq \hat{\Lambda}$. In such case, we obtain $(\mathbf{B}_{\hat{\Lambda}})^\dagger \mathbf{y} \succ \mathbf{0}$ and $\hat{\Lambda}$ is the optimal support. Consequently, the approximate solution $\hat{\boldsymbol{\delta}} = [\hat{\delta}_1, \dots, \hat{\delta}_{2K}]^T$ can be represented as a zero-padded version of the vector $(\mathbf{B}_{\hat{\Lambda}})^\dagger \mathbf{y}$, i.e.,

$$\hat{\delta}_l = \max \left\{ [(\mathbf{B}_{\hat{\Lambda}})^\dagger \mathbf{y}]_l, 0 \right\}, \quad l \in \hat{\Lambda}, \quad (12)$$

and $\hat{\delta}_l = 0$ otherwise, for $l = 1, \dots, 2K$. Having an explicit expression for $\hat{\boldsymbol{\delta}}$, the corresponding transmit vector is readily computable by replacing $\boldsymbol{\delta}$ in (7).

B. Computational Complexity Analysis

We compare the computational complexity of the proposed method with an iterative NNLS algorithm. As our benchmark for comparison, we consider the accelerated projected gradient descent (APGD) algorithm [18]. The pseudocode of APGD to (approximately) solve the NNLS (9) via a limited number of iterations is given in Algorithm 1. We evaluate the worst-case complexity in terms of the number of arithmetic operations. For an iterative method, this can be interpreted as the required number of operations until the stopping condition is met.

The main loop of APGD is preceded by an initialization step performing two matrix multiplications and one singular value decomposition (SVD) with complexity orders of K^2N , KN and K^3 , respectively. Within the main loop, the per-iteration complexity is dominated by a matrix multiplication of order K^2 . To be more accurate, the complexity of APGD depends also on the convergence specifications, e.g., the condition number of \mathbf{B} ; however, we consider only those complexity terms directly relating to the problem size. On the other hand, the dominant arithmetic operations in (10), (11) and (12), i.e., $2K$ vector multiplications and two matrix pseudo-inversions, result in computation costs of order KN and $N(L_1^2 + L_2^2)$, respectively, for the proposed method. Remark that the closed-form solution in [13] can be implemented in an equivalent way using (10) and (12); therefore we assess the complexity of [13] based on the method of this paper.

TABLE I
COMPLEXITIES OF DIFFERENT NNLS-BASED SLP SOLUTIONS.

Solution Method	Complexity Order
APGD algorithm [18]	$K^2 \cdot \mathcal{O}(K + N) + \mathcal{O}(K^2) \epsilon^{-1/2}$
Closed-form SLP [13]	$N \cdot \mathcal{O}(K + L_2^2)$
Improved closed-form SLP	$N \cdot \mathcal{O}(K + L_1^2 + L_2^2)$

In Table I, we summarize the dominating complexity orders of different methods, where that of the APGD corresponds to an ϵ -optimal solution. Due to the sparsity-promoting nature of the NNLS problem [22], in practice we have $L_2 \leq L_1 \ll 2K$. Based on this observation and the results in Table I, we conclude that both closed-form methods (potentially) decrease the computation cost of the precoding design. In fact, even the complexity of the initialization step in APGD (without any iterations) is higher than the two closed-form methods.

V. SIMULATION RESULTS

In this section, we provide some simulation results to evaluate and compare the performances of various approaches to solve the SINR-constrained SLP problem. Our simulation setup is as follows. We consider a downlink MU-MIMO system with N transmit antennas and K (single-antenna) users, where $N/K \triangleq \beta$. For all users $k \in \{1, \dots, K\}$, we assume unit noise variances $\sigma_k^2 = 1$ and equal target SINRs $\gamma_k \triangleq \gamma$. The users' channel vectors $\{\mathbf{h}_k\}_{k=1}^K$ are independently generated following the standard circularly symmetric complex Gaussian distribution, i.e., $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The maximum number of iterations, n_{\max} , for the APGD algorithm is set to be 25. The results are all averaged over 10^3 channel coherence blocks each of length 10^3 symbols. Later on in this section, we refer to the SLP methods of interest as:

- *ZF-SLP*: symbol-level ZF, assuming $\boldsymbol{\delta} = \mathbf{0}$ in (9).
- *Optimal SLP*: optimal solution to (9).
- *NNLS-SLP (APGD)*: solving (9) via APGD algorithm.
- *CF-SLP*: closed-form approximate SLP solution of [13].
- *ICF-SLP*: Improved CF-SLP proposed in this paper.

In Fig. 1, the total transmit powers obtained from various precoding schemes are plotted as a function of target SINR, where three different modulations are assessed in the whole range of depicted SINR. The results correspond to a fully-loaded system with $N = K$. It can be seen that the ICF-SLP method improves the accuracy of the approximate solution by up to 3 dB, compared to its naive counterpart CF-SLP. Further, ICF-SLP outperforms the NNLS-SLP method via APGD with $n_{\max} = 25$. The observations show that both the methods have nearly the same complexity in this range of K . The promising fact about Fig. 1 is that ICF-SLP performs well close to the optimal SLP with a far less complexity, as we will see next.

In another set of simulations for an under-loaded system with $\beta = 6/5$, we evaluate the performance/complexity of different solution approaches to the SLP problem (9). The results are presented in Fig. 2 as a function of the number of users K . The optimal SLP solution is obtained using the *lsqnonneg* function of MATLAB, which uses the Lawson and Hanson active set method to solve NNLS. As it can be seen, the resulting performance of CF-SLP noticeably degrades with increasing K , whereas the proposed ICF-SLP shows a decreasing trend in transmit power (as that of the optimal SLP) for large system dimensions. The optimality gap of ICF-SLP with $K = 100$ is just 0.15 dBW. This improvement becomes of great significance when we consider also the time complexities of the two solutions; see Table I. Therefore, the time complexity results in Fig. 2 are consistent with the analytical evaluations reported in Table I. This can be further verified through comparing ICF-SLP and the APGD-based NNLS-SLP method. The latter method has a higher complexity growth rate, which is theoretically proportional to $\mathcal{O}(K^2N)$ in the limiting case. This might suggest a performance-complexity tradeoff. However, notice that with $n_{\max} = 25$, the dominating complexity order of the APGD algorithm in the large system limit stems from the initialization step, which is higher than the entire computation cost of ICF-SLP.

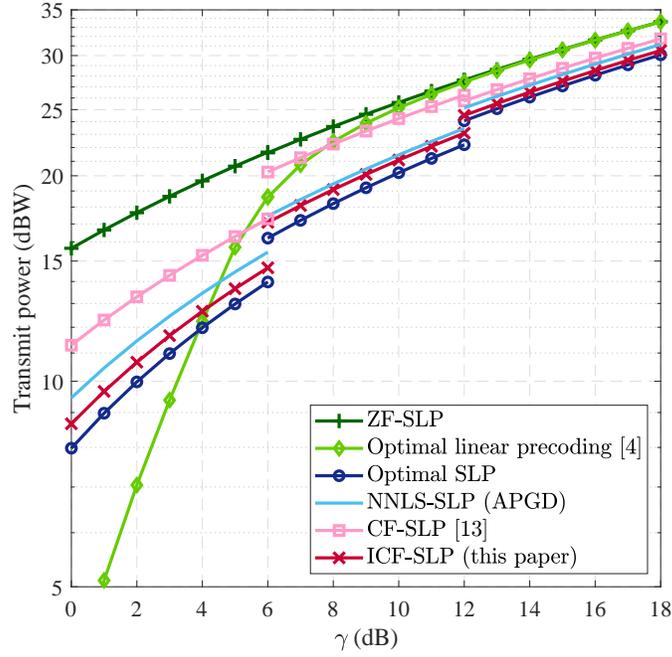


Fig. 1. Transmit power versus target SINR with $N = K = 8$. Three different modulations, namely QPSK, 8PSK and 16QAM, are respectively used in 0-6, 6-12 and 12-18 dB SINR ranges.

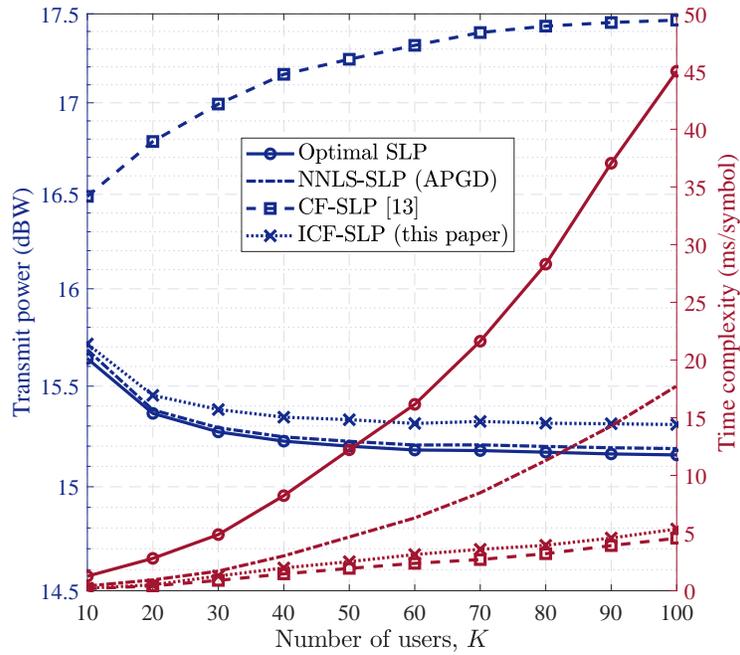


Fig. 2. Transmit power and time complexity versus number of users, $\beta = 6/5$. Same line types and markers as those in the legend refer to the right axes but with a different color.

VI. CONCLUSION

We proposed a low-complexity method to approximately solve the SLP power minimization problem with SINR constraints. Due to the required per-symbol computation, solving the SLP optimization problem for the exact solution may lead to an impractical transmitter complexity. To address this issue, we exploited the structure of an equivalent NNLS formulation of the original problem. We modified an existing approximate solution by applying a computationally efficient validation step before calculating the final solution. Based on our simulation results, this modification considerably reduces the loss with respect to the optimal solution, particularly in the large system regime. Further, the new approximate solution is shown to be comparable with the solution obtained from an iterative NNLS algorithm, from both performance and complexity points of view. It is, however, concluded that as far as a low-complexity implementation of SLP with a close-to-optimal performance is of concern, the proposed method provides a more efficient solution.

REFERENCES

- [1] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [2] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
- [3] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.
- [4] M. Bengtsson and B. Ottersten, *Handbook of Antennas in Wireless Communications*, 2001, ch. Optimal and suboptimal transmit beamforming.
- [5] 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, Jan. 2004.
- [6] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.
- [7] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.
- [8] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, Jul. 2015.
- [9] 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, May 2015.
- [10] —, "Symbol-level multiuser MISO precoding for multi-level adaptive modulation," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 8, pp. 5511–5524, Aug. 2017.
- [11] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Symbol-level precoding design based on distance preserving constructive interference regions," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5817–5832, Nov. 2018.
- [12] A. Li and C. Masouros, "Interference exploitation precoding made practical: Optimal closed-form solutions for PSK modulations," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 11, pp. 7661–7676, 2018.
- [13] 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, Nov. 2018.
- [14] J. Krivochiza, A. Kalantari, S. Chatzinotas, and B. Ottersten, "Low complexity symbol-level design for linear precoding systems," in *Symposium on Information Theory and Signal Processing in the Benelux*, Mar. 2016.
- [15] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, Apr. 2018.
- [16] C. Lawson and R. Hanson, *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 1995.
- [17] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 11, no. 5, pp. 393–401, 1997.
- [18] R. A. Polyak, "Projected gradient method for non-negative least square," *Contemp Math*, vol. 636, pp. 167–179, 2015.
- [19] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $o(1/k^2)$," in *Dokl. Akad. Nauk SSSR*, vol. 269, 1983, pp. 543–547.
- [20] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [21] Y. Itoh, M. F. Duarte, and M. Parente, "Perfect recovery conditions for non-negative sparse modeling," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 69–80, Jan. 2017.
- [22] M. Slawski and M. Hein, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization," *Electronic J. Statistics*, vol. 7, pp. 7661–7676, 2013.