Recovery Conditions of Sparse Signals Using Orthogonal Least Squares-Type Algorithms

Liyang Lu ⓘ, *Graduate Student Member, IEEE*, Wenbo Xu ⓘ, *Member, IEEE*, Yue Wang ⓘ, *Senior Member, IEEE*, and Zhi Tian ⓘ, *Fellow, IEEE*

Abstract—Orthogonal least squares (OLS)-type algorithms are efficient in reconstructing sparse signals, which include the well-known OLS, multiple OLS (MOLS) and block OLS (BOLS). In this paper, we first investigate the noiseless exact recovery conditions of these algorithms. Specifically, based on mutual incoherence property (MIP), we provide theoretical analysis of OLS and MOLS to ensure that the correct nonzero support can be selected during the iterative procedure. Nevertheless, theoretical analysis for BOLS utilizes the block-MIP to deal with the block sparsity. Furthermore, the noiseless MIP-based analyses are extended to the noisy scenario. Our results indicate that for K-sparse signals, when MIP or SNR satisfies certain conditions, OLS and MOLS obtain reliable reconstruction in at most K iterations, while BOLS succeeds in at most (K/d) iterations where d is the block length. It is shown that our derived theoretical results improve the existing ones, which are verified by simulation tests.

Index Terms—Block sparsity, compressed sensing, mutual incoherence property (MIP), orthogonal least squares (OLS), signal recovery.

I. INTRODUCTION

IN COMPRESSED sensing (CS) [1], [2], [3], the main task is to recover a high-dimensional sparse signal from low-dimensional compressed measurements, which has a broad impact in different fields [4], [5]. Specifically, it seeks to accurately recover a K-sparse signal $\mathbf{x} \in \mathcal{R}^N$ with K nonzero elements ($K \ll N$) from its linear measurement vector $\mathbf{y} = \mathbf{D}\mathbf{x}$ [6], [7], [8], where the measurement matrix $\mathbf{D}$ satisfies restricted isometry property (RIP) [9], [10], [11], [12]. Exploiting the signal sparsity, a large number of reconstruction algorithms have been developed to exactly recover the original sparse signals under different conditions on $\mathbf{D}$ [13], [14], [15], [16], [17], [18].

At present, there are two categories of sparse signal recovery approaches: one exploits convex optimization techniques [19],

[20] and the other is based on greedy matching pursuit [21], [22], [23], [24], [25], [26]. Optimization-based algorithms, such as the well-known basis pursuit (BP) [19], enjoy satisfactory performance, at the cost of high computational complexity that may not be amenable to practical implementation. On the other hand, greedy algorithms, such as matching pursuit (MP) and orthogonal MP (OMP) [27], [28], admit simple and fast implementations, and hence have been widely used in practice. Another popular family of greedy algorithms is the orthogonal least squares (OLS) type, which is shown to have preferable convergence property [29]. This paper focuses on the family of OLS-type algorithms for sparse signal recovery, because of their computational simplicity and competitive performance.

OLS is an iterative CS algorithm that picks out the support of the sparse signal by updating only one index to the list per iteration. Then, several variants are developed to enhance the sparse recovery performance. In [16], multiple OLS (MOLS) is proposed which selects multiple indices per iteration. In [30], a block version of the OLS algorithm, named BOLS, is proposed to utilize the block structure of the underlying signals. It works well for practical scenarios where the nonzero support of the sparse signal appears in clusters rather than spreading out randomly [31], [32].

A fundamental problem in CS reconstruction is to characterize the recovery capabilities of these algorithms, where RIP is usually used as a representative metric. The authors in [15], [33] exploit RIP to analyze the isometry constant of OLS for exactly recovering a K-sparse signal in K iterations. In [16], the authors utilize RIP to obtain the recovery conditions for MOLS and show its performance in noisy case. Furthermore, [34] uses similar methodology to analyze MOLS and proposes a less restrictive performance guarantee condition.

Unfortunately, to obtain the RIP constant of a given measurement matrix is NP-hard. By contrary, a computationally friendly approach is the mutual incoherence property (MIP) [13], [35], [36]. This technique has been utilized to analyze the performance guarantee of OMP-type algorithms in the last decade. Compared with the OMP-type, the OLS-type algorithms are more reliable and the performance is less dependent on the amplitude distribution of nonzero entries [37]. However, the analysis of OLS-type algorithms is more involved, because their atom selection rules are more complicated than those of the OMP-type algorithms. As a result, the study on OLS-type algorithms with MIP tool is seldom in current literature. Though the authors in [37] present the MIP-based conditions of OLS,

the results only apply for decaying signals. To better understand their recovery capabilities for general sparse signals and show comparison with OMP-type algorithms [13], [14], it is necessary to investigate the MIP-based performance for the OLS-type algorithms.

Most work on MIP-based analysis is devoted to the recovery of noiseless signals [13], [14]. Nevertheless, noise is always present in practical applications [37], [38]. A fundamental question is thus at what signal-to-noise ratio (SNR) levels the OLS-type algorithms realize reliable recovery. Since the MIP metric is powerful in assessing recovery performance, MIP analysis can shed light on the target SNR levels for ensuring signal reconstruction at high accuracy.

Based on the above observations, this paper focuses on the recovery conditions for OLS-type algorithms, both in the noiseless and the noisy scenarios. Note that the mapping normalization factor in OLS-type algorithms makes the analysis greatly different from that of the OMP-type in [13], [14], which poses major challenges [15], [29]. We shall show that the MIP technique is effective to bound this mapping normalization factor. Another major difference from the study in [13], [14] is that we consider the noise effect. We develop the MIP and SNR-based recovery conditions for correct atom selection when additive noise is present. It is shown that the OLS-type algorithms can choose all the correct atoms from the measurement matrix under these conditions and thus guarantee their performance. The contributions of this paper are summarized as below.

1) For the noiseless case, we develop the MIP-based asymptotic exact recovery conditions (ERCs) of the OLS-type algorithms. The derived conditions imply higher reconstructible sparsity levels than those in [13], [14]. Our analytical results reveal that the OLS-type algorithms can recover any K-sparse or block k-sparse signal within K or k iterations under the derived conditions, respectively.

2) In the noisy case, the recovery conditions are developed for OLS-type algorithms by extending the noiseless results. The proposed noisy recovery conditions imply lower signal power required for reliable recovery compared to the known bound in [38]. These results unveil that if the nonzero coefficients of a sparse signal are larger than the derived bounds specified by the MIP and noise variance, the OLS-type algorithms can achieve reliable recovery.

3) Validating simulations are performed to confirm the derived theoretical results. It is shown that the theoretical guarantees in this paper lead to better bounds compared with the existing ones. Empirical simulations indicate that under the derived conditions, the OLS-type algorithms can reliably reconstruct the original signals.

The rest of this paper is organized as follows. In Section II, we introduce notations, CS background and OLS-type algorithms. In Section III, we study the noiseless recovery conditions for OLS-type algorithms. In Section IV, the noisy recovery conditions for OLS-type algorithms are derived. In Section V, simulation tests are presented, followed by conclusions in Section VI.

II. PRELIMINARIES

A. Notations

We briefly summarize the notations in this paper. Vectors are denoted by boldface lowercase letters, e.g., $\mathbf{x}$, and matrices are denoted by boldface uppercase letters, e.g., $\mathbf{D}$. The i-th element of $\mathbf{x}$ is denoted as $\mathbf{x}_i$. The element in the i-th row and j-th column of matrix $\mathbf{D}$ is represented by $\mathbf{D}_{ij}$, and $\mathbf{D}_i$ is the i-th column of $\mathbf{D}$. $\mathbf{D_S}$ is a sub-matrix of $\mathbf{D}$ that contains columns indexed by $\mathbf{S}$. $\mathbf{D} \backslash \mathbf{D_S}$ is the residual matrix after all the columns indexed by $\mathbf{S}$ are removed from $\mathbf{D}$. $\mathbf{D}^T$ represents the transpose of the matrix $\mathbf{D}$. The ℓ_0-norm, ℓ_1-norm, ℓ_2-norm, ℓ_∞-norm and $\ell_{2,\infty}$-norm of $\mathbf{x}$ are represented by $||\mathbf{x}||_0, ||\mathbf{x}||_1, ||\mathbf{x}||_2, ||\mathbf{x}||_\infty$ and $||\mathbf{x}||_{2,\infty}$, respectively. The spectral norm of a matrix $\mathbf{D}$ is denoted by $\rho(\mathbf{D}) = \sqrt{\lambda_{\max}(\mathbf{D}^T\mathbf{D})}$, where $\lambda_{\max}(\cdot)$ is the largest eigenvalue of its objective matrix. The set that includes the indices of the nonzero elements in $\mathbf{x}$ is represented by $\text{supp}(\mathbf{x})$, and $\text{span}(\mathbf{D_S})$ is the span of columns in $\mathbf{D_S}$. The diagonalization of $\mathbf{x}$ is denoted by $\text{diag}(\mathbf{x})$. $|\mathbf{S}|$ stands for the cardinality of a set $\mathbf{S}$ and $|c|$ means the absolute value of a constant c. $< \cdot, \cdot >$ represents the inner product operation. The coherence, block-coherence and sub-coherence of a specific matrix are represented by μ, μ_B and ν, respectively. K denotes the overall sparsity of a sparse signal. For a block sparse signal, k and d represent its block sparsity and block length, respectively. $\mathbf{I}$ stands for the identity matrix. If $\mathbf{D_S}$ has the full column rank, $\mathbf{D_S}^\dagger = (\mathbf{D_S}^T\mathbf{D_S})^{-1}\mathbf{D_S}^T$ represents the pseudoinverse of $\mathbf{D_S}$. $\mathbf{P_S} = \mathbf{D_S}\mathbf{D_S}^\dagger$ denotes the projection onto $\text{span}(\mathbf{D_S})$, and $\mathbf{P_S}^\perp = \mathbf{I} - \mathbf{P_S}$ is the projection onto the orthogonal complement of $\text{span}(\mathbf{D_S})$. Throughout the paper, the columns in the measurement matrix are normalized to have unit ℓ_2-norm.

B. Compressed Sensing and Sparse Signal Models

The basic model of CS is given as follows:

$$\mathbf{y} = \mathbf{Dx}, \tag{1}$$

where $\mathbf{y} \in \mathcal{R}^M$ is the low-dimensional measurement vector, $\mathbf{D} \in \mathcal{R}^{M \times N}$ is the measurement matrix and $\mathbf{x} \in \mathcal{R}^N$ is the signal vector with $N > M$. Generally, infinitely many choices of $\mathbf{x}$ satisfy (1) for a given $\mathbf{y}$ because (1) is underdetermined. Therefore, in order to guarantee uniqueness of $\mathbf{x}$, CS considers the case of sparse signals, i.e., there are only a few nonzero elements in $\mathbf{x}$ relative to its dimension. The problem of obtaining the sparse solution to (1) can be expressed as

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{Dx}. \tag{2}$$

This ℓ_0-norm minimization problem is non-convex and NP-hard. Fortunately, the problem can be relaxed by its convex surrogate ℓ_1 optimization [2],

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{Dx}. \tag{3}$$

Many methods mentioned in the first section can be applied to solve (3).

When $\mathbf{x}$ is block sparse, block CS is proposed in [14], [39], [40], [41] to recover the signal. Letting d denote the block length,

Algorithm 1: Orthogonal Least Squares.

Input: D, y, total sparsity level K and residual tolerant ζ
Output: $x \in \mathcal{R}^N$, $S \subseteq \{1, 2, \ldots, N\}$
1: **Initialization** : $l = 0$, $r^0 = y$, $S^0 = \emptyset$, $x^0 = 0$
2: **while** $l < K$ and $||r^l||_2 > \zeta$ **do**
3:     Set $i^{l+1} = \underset{j \in \{1,\ldots,N\} \backslash S^l}{\arg\min} ||P_{S^l \cup \{j\}}^{\perp} y||_2^2$
4:     Augment $S^{l+1} = S^l \cup \{i^{l+1}\}$
5:     Estimate $x^{l+1} = \underset{x: \text{supp}(x) = S^{l+1}}{\arg\min} ||y - Dx||_2^2$
6:     Update $r^{l+1} = y - Dx^{l+1}$
7:     $l = l + 1$
8: **end while**
9: **return** $S = S^l$ and $x = x^l$

the block sparse signal x is defined as

$$x = [\underbrace{x_1 \cdots x_d}_{x^T[1]} \underbrace{x_{d+1} \cdots x_{2d}}_{x^T[2]} \cdots \underbrace{x_{N-d+1} \cdots x_N}_{x^T[N_B]}]^T, \quad (4)$$

where $N = N_B d$ and $x[i] \in \mathcal{R}^d$ is the i-th block of x ($i \in \{1, 2, \ldots, N_B\}$). A signal x is called block k-sparse if x has k nonzero ℓ_2-norm blocks. The measurement matrix can be rewritten as a concatenation of N_B column blocks, i.e.,

$$D = [\underbrace{D_1 \cdots D_d}_{D[1]} \underbrace{D_{d+1} \cdots D_{2d}}_{D[2]} \cdots \underbrace{D_{N-d+1} \cdots D_N}_{D[N_B]}], \quad (5)$$

where $D[i] \in \mathcal{R}^{M \times d}$ is the i-th block of D. The next subsection describes the OLS-type algorithms for reconstructing the sparse x from y in (1).

C. OLS-Type Algorithms

1) OLS Algorithm: The standard OLS algorithm is given in Algorithm 1. To facilitate the theoretical derivations in ensuing sections, we rewrite Step 3 in Algorithm 1 in an alternative form:

$$i^{l+1} = \underset{j \in \{1,\ldots,N\} \backslash S^l}{\arg\max} \left| \left\langle \frac{P_{S^l}^{\perp} D_j}{||P_{S^l}^{\perp} D_j||_2}, r^l \right\rangle \right|$$

$$= \underset{j \in \{1,\ldots,N\} \backslash S^l}{\arg\max} \left| \left\langle \frac{D_j}{||P_{S^l}^{\perp} D_j||_2}, r^l \right\rangle \right|. \quad (6)$$

The derivation details can be found in [16], [42]. This formula also offers a more intuitive explanation of OLS. Specifically, it selects a candidate column that is most closely related with the current residual after the column is orthogonally projected onto the subspace span(D_{S^l}).

2) MOLS Algorithm: In the l-th iteration, the support selection criteria of MOLS algorithm [16] is given as follows:

$$Q_M^{l+1} = \underset{Q_M: |Q_M| = L}{\arg\min} \sum_{j_M \in Q_M} ||P_{S^l \cup \{j_M\}}^{\perp} y||_2^2. \quad (7)$$

Owning to the selection of multiple candidates per iteration, MOLS converges faster than OLS.

3) BOLS Algorithm: When x is block sparse, [30] proposes a tailored BOLS algorithm, which seeks a block candidate that makes the most significant reduction in the residual power. The support selection criteria of BOLS is given by

$$i_B^{l+1} = \underset{j_B \in \{1,\ldots,N_B\} \backslash S_B^l}{\arg\min} ||P_{S^l \cup \{(j_B-1)d+1,\ldots,j_B d\}}^{\perp} y||_2^2, \quad (8)$$

where the entries in S_B^l correspond to the block indices of S^l. All these algorithms follow the same procedure, except that the support selection criteria are different.

III. EXACT RECOVERY CONDITIONS

A. ERCs for OLS and MOLS Algorithms

In the absence of block sparsity, we provide the ERCs for OLS and MOLS. Relevant concepts are defined below.

Definition 1 (Matrix coherence [13], [14]): The coherence of a matrix D, which represents the similarity of its elements, is defined as

$$\mu = \max_{i,j \neq i} | < D_i, D_j > |. \quad (9)$$

Definition 2 (Mixed norm [13], [14]): Given a matrix $D \in \mathcal{R}^{M \times N}$, $||D||_{1,1}$ equals the maximum absolute column sum of D, i.e.,

$$||D||_{1,1} = \max_j \sum_{i=1}^{M} |D_{ij}|, \quad (10)$$

and $||D||_{\infty,\infty}$ represents the maximum absolute row sum of D, i.e.,

$$||D||_{\infty,\infty} = \max_i \sum_{j=1}^{N} |D_{ij}|. \quad (11)$$

Let $M = md$ and $N = N_B d$,

$$\rho_c(D) = \max_j \sum_i \rho(D[i,j]) \quad (12)$$

and

$$\rho_r(D) = \max_i \sum_j \rho(D[i,j]), \quad (13)$$

where m is an integer and $D[i,j]$ is the (i,j)-th block of D.

Without loss of generality, suppose that the first K elements of x are nonzero. With the set of the selected indices, i.e., $S^l = \{K - l + 1, \ldots, K\}$, in the $(l+1)$-th iteration, we define

$$Q_0 = [D_1, D_2, \ldots, D_K] \in \mathcal{R}^{M \times K}, \quad (14)$$

$$Q_{0\backslash S} = Q_0 \backslash Q_{S^l} \in \mathcal{R}^{M \times (K-|S^l|)} \quad (15)$$

and

$$\overline{Q}_0 = D \backslash Q_0 \in \mathcal{R}^{M \times (N-K)}. \quad (16)$$

$R_{0\backslash S}$ and $\overline{R}_0$ corresponding to $Q_{0\backslash S}$ and $\overline{Q}_0$ are defined by

$$R_{0\backslash S} = f(Q_{0\backslash S}) = \begin{bmatrix} \frac{1}{||P_{S^l}^{\perp} D_1||_2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{||P_{S^l}^{\perp} D_{K-l}||_2} \end{bmatrix} \quad (17)$$

and $\overline{\mathbf{R}}_0 = f(\overline{\mathbf{Q}}_0)$. Using these arguments, we present the ERCs for OLS and MOLS in the following theorems.

Theorem 1: Assume that OLS or MOLS has chosen l correct atoms after l iterations. A sufficient condition for OLS to select one correct atom or MOLS to select at least one correct atom in the $(l+1)$-th iteration is

$$||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1} < 1. \tag{18}$$

Proof: See Appendix A. ∎

Then, we present the MIP-based sufficient condition to ensure that Theorem 1 can be established in each iteration. To obtain this sufficient condition, we give the subsequent lemmas. The first one raises the bounds on the mapping normalization factor, i.e., $||\mathbf{P}_{\mathbf{S}^l}^\perp \mathbf{D}_i||_2$, in terms of the MIP. The second one presents the probability of a random event.

Lemma 1: Suppose $\mu < \frac{1}{K-1}$, then $\frac{1}{\sqrt{\mathcal{T}}} \le ||\mathbf{P}_{\mathbf{S}^l}^\perp \mathbf{D}_i||_2 \le 1$ for $i \in \{1, 2, \dots, N\}\backslash\mathbf{S}^l$, where $\mathcal{T} = \left(1 - \frac{(1+(K-1)\mu)K\mu^2}{(1-(K-1)\mu)^2}\right)^{-1}$.

Proof: See Appendix B. ∎

Notably, under the same assumption made as in Lemma 1, i.e., $\mu < \frac{1}{K-1}$, the following bounds can be obtained by direct calculations using the Lemmas 2 and 5 in [38]:

$$\sqrt{1 - K\mu} \le ||\mathbf{P}_{\mathbf{S}^l}^\perp \mathbf{D}_i||_2 \le 1. \tag{19}$$

It is worthy of noting that $\frac{1}{\sqrt{\mathcal{T}}} \ge \sqrt{1 - K\mu}$ since $\mu < \frac{1}{K-1}$, which means Lemma 1 improves the result (19).

Lemma 2: For a random matrix $\mathbf{B} \in \mathcal{R}^{M \times K}$ $(K \ge 2)$ whose entries are independently and identically distributed as $\mathcal{N}(0, \frac{1}{M})$, the following probability holds:

$$P\left\{||\mathbf{B}^T\mathbf{B} - \mathbf{I}||_{1,1} \le \frac{(K-\mathcal{T})\tau}{2}\right\} \ge 1 - \frac{Ke^{-\frac{K-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(K-1)}}, \tag{20}$$

where τ is a positive constant and $\mathcal{C} = \frac{M(K-\mathcal{T})^2\tau^2}{4(K-1)^2} - 1$.

Proof: See Appendix C. ∎

Remark 1: Suppose that the matrix $\mathbf{B}$ in Lemma 2 is a sub matrix of the Gaussian matrix $\mathbf{H} \in \mathcal{R}^{M \times N}$ $(N > M)$. Since $\mu < \frac{1}{K-1}$ and $\mathcal{T}$, which is given in Lemma 1, is monotonically increasing with respect to μ, then $\mathcal{T}$ is limited by an upper bound. When K is fixed, we obtain

$$\lim_{M \to \infty} \frac{M(K-\mathcal{T})^2\tau^2}{4(K-1)^2} - 1 \to \infty. \tag{21}$$

Therefore,

$$\lim_{M \to \infty} 1 - \frac{Ke^{-\frac{K-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(K-1)}} = 1. \tag{22}$$

Theorem 2: Let μ be the matrix coherence of the measurement matrix $\mathbf{D}$. The ERC (18) is satisfied if

$$K < \left(\sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}} - \frac{\beta}{3\alpha}\right), \tag{23}$$

where $q = \frac{27\alpha^2\delta - 9\alpha\beta\gamma + 2\beta^3}{27\alpha^3}$, $p = \frac{3\alpha\gamma - \beta^2}{3\alpha^2}$, $\alpha = -\frac{\mu^4}{2} + \frac{3}{2}\mu^3$, $\beta = \frac{\mu^4}{2} - 3\mu^3 - 4\mu^2$, $\gamma = \frac{3\mu^3}{2} + 7\mu^2 + \frac{7}{2}\mu$, $\delta = -\frac{1}{2}\mu^3 - 2\mu^2 - \frac{5}{2}\mu - 1$ and $\Delta = (\frac{q}{2})^2 + (\frac{p}{3})^3$.

Proof: See Appendix D. ∎

(a) Results of $\hat{\mathcal{Q}}/\mathcal{Q}$ and $\hat{\mathcal{P}}/\mathcal{P}$.   (b) Results of $\mathcal{P}^3/\mathcal{Q}^2$ and $\hat{\mathcal{P}}^3/\hat{\mathcal{Q}}^2$.
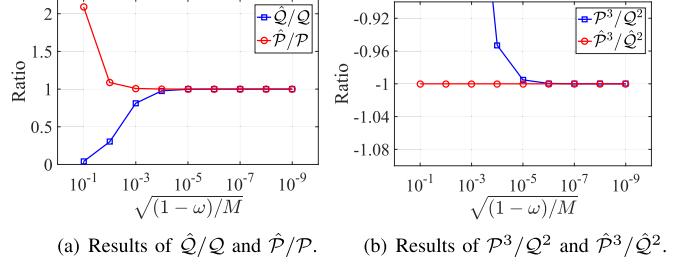
Fig. 1. Ratios of different parameters with the decrease of $\sqrt{(1-\omega)/M}$.

Remark 2: Denote the right-side of (23) as $K^*(\mu)$. The upper bound of $K^*(\mu)$ is $K^*\left(\sqrt{\frac{N-M}{M(N-1)}}\right)$, where M and N are the dimensions of the measurement matrix. Moreover, we have the following observations. 1) When M is fixed, the upper bound of $K^*(\mu)$ degrades with the increase of N. 2) When N is fixed, the upper bound of $K^*(\mu)$ is improved with the increase of M. 3) The upper bound is directly related to the number of measurements M and the compression ratio $\omega = \frac{M}{N}$. That is, when $N \gg 0$ and $M \gg 0$, it holds that

$$K^*(\mu) \le \frac{2}{3}\left(\sqrt{\frac{M}{1-\omega}} + 1\right). \tag{24}$$

Proof: It is known that $\mu \ge \sqrt{\frac{N-M}{M(N-1)}}$ [43]. Meanwhile, the partial derivative of $K^*(\mu)$ with respect to μ satisfies $\nabla K^*(\mu) < 0$. Therefore, $K^*(\mu) \le K^*\left(\sqrt{\frac{N-M}{M(N-1)}}\right)$.

When M is fixed, $\sqrt{\frac{N-M}{M(N-1)}} = \frac{1}{\sqrt{M}}\sqrt{1 - \frac{M-1}{N-1}}$ is improved with the increase of N, leading to the degradation of $K^*\left(\sqrt{\frac{N-M}{M(N-1)}}\right)$. When N is fixed, $\sqrt{\frac{N-M}{M(N-1)}} = \sqrt{\frac{\frac{N}{M}-1}{N-1}}$ decreases with the increase of M, which causes the improvement of $K^*\left(\sqrt{\frac{N-M}{M(N-1)}}\right)$.

Moreover, in asymptotic case, we have

$$\lim_{M/N=\omega; M,N\to\infty} K^*\left(\sqrt{\frac{N-M}{M(N-1)}}\right) = K^*\left(\sqrt{\frac{1-\frac{M}{N}}{M(1-\frac{1}{N})}}\right)$$

$$= K^*\left(\sqrt{\frac{1-\omega}{M}}\right). \tag{25}$$

Rewrite $K^*\left(\sqrt{\frac{1-\omega}{M}}\right)$ as follows:

$$K^*\left(\sqrt{\frac{1-\omega}{M}}\right) = \sqrt[3]{-\mathcal{Q} + \sqrt{\mathcal{Q}^2 + \mathcal{P}^3}}$$

$$+ \sqrt[3]{-\mathcal{Q} - \sqrt{\mathcal{Q}^2 + \mathcal{P}^3}} + \mathcal{F}, \tag{26}$$

where $\mathcal{Q} = \frac{q}{2}$, $\mathcal{P} = \frac{p}{3}$ and $\mathcal{F} = -\frac{\beta}{3\alpha}$ by replacing μ as $\sqrt{\frac{1-\omega}{M}}$ in Theorem 2.

Meanwhile, the magnitudes of $\mathcal{Q}$ and $\mathcal{P}$ are mainly determined by their highest order terms, which can be verified by the following simulations. We denote the highest order terms in $\mathcal{Q}$ and $\mathcal{P}$ as $\hat{\mathcal{Q}}$ and $\hat{\mathcal{P}}$ respectively. The results are shown in Fig. 1. When M is large enough, the ratio $\frac{\mathcal{P}^3}{\mathcal{Q}^2}$ converges

to -1, indicating that $\mathcal{Q}^2$ and $\mathcal{P}^3$ are on the same order and can be well approximated by their highest order terms $\hat{\mathcal{Q}}^2$ and $\hat{\mathcal{P}}^3$ respectively. Furthermore, when M is large enough, $\hat{\mathcal{Q}}/\mathcal{Q}$ and $\hat{\mathcal{P}}/\mathcal{P}$ are equal to 1, indicating that $\mathcal{Q}$ and $\mathcal{P}$ can be well approximated by their highest terms, i.e., $\hat{\mathcal{Q}}$ and $\hat{\mathcal{P}}$. In a word,

$$\begin{cases} \lim_{M\to\infty} \mathcal{P}^3/\mathcal{Q}^2 = \lim_{M\to\infty} \hat{\mathcal{P}}^3/\hat{\mathcal{Q}}^2 = -1, \\ \lim_{M\to\infty} \hat{\mathcal{Q}}/\mathcal{Q} = 1, \\ \lim_{M\to\infty} \hat{\mathcal{P}}/\mathcal{P} = 1. \end{cases} \tag{27}$$

Based on the above arguments, we obtain

$$\lim_{M\to\infty} K^*\left(\sqrt{\frac{1-\omega}{M}}\right)$$

$$= \sqrt[3]{-\frac{1}{729}\left(\frac{M}{1-\omega}\right)^{\frac{3}{2}} + \sqrt{\left(\left(\frac{1}{729}\right)^2 + \left(-\frac{1}{81}\right)^3\right)\left(\frac{M}{1-\omega}\right)^3}}$$

$$+ \sqrt[3]{-\frac{1}{729}\left(\frac{M}{1-\omega}\right)^{\frac{3}{2}} - \sqrt{\left(\left(\frac{1}{729}\right)^2 + \left(-\frac{1}{81}\right)^3\right)\left(\frac{M}{1-\omega}\right)^3}}$$

$$+ \frac{8}{9}\sqrt{\left(\frac{M}{1-\omega}\right)} + \frac{2}{3} = \frac{2}{3}\left(\sqrt{\frac{M}{1-\omega}} + 1\right). \tag{28}$$

∎

In practice, a simple checking operation can be made to check if the real sparsity K satisfies $K < K^*(\mu)$. If so, the reliable recovery is possible; otherwise the recovery cannot be guaranteed. K can be either a prior information, or a predefined approximate value from experience, or an estimated value calculated based on the measurement vector and the measurement matrix [44], [45]. For all the cases, our theoretical results provide a way to verify if a measurement matrix satisfies the exact recovery condition via the computationally affordable MIP. Even when the sparsity is unknown in advance, our theoretical results still play as a guideline as preliminary judgments about whether the recovery can be successful based on the estimated value, which is meaningful for subsequent decision-making.

Remark 3: Similar to the derivation in Remark 2, we have

$$\lim_{\mu\to 0} K^*(\mu) = \frac{2}{3}\left(\frac{1}{\mu} + 1\right). \tag{29}$$

It is known that the MIP-based sufficient condition for the ERC of OMP derived by Tropp [13] is

$$K < \frac{1}{2}\left(\frac{1}{\mu} + 1\right). \tag{30}$$

Therefore, the sufficient condition for the ERC of OLS and MOLS based on (29), i.e.,

$$K < \frac{2}{3}\left(\frac{1}{\mu} + 1\right) \tag{31}$$

improves the result (30) in the case of low coherence, which indicates OLS and MOLS algorithms can reconstruct more exactly than OMP algorithm from the perspective of reconstructible sparsity level.

It is proved in [37] that the optimal value of MIP-based sufficient condition for decaying signals is $K < \frac{1}{\mu}$. Our result in Remark 3 indicates that $K < \frac{2}{3}\left(\frac{1}{\mu} + 1\right)$ is sufficient for general sparse signals, which means that (31) is at most $\frac{2}{3}$ times worse than the optimal value. However, our analysis applies for general sparse signals rather than for only decaying sparse signals. Moreover, Remark 3 shows that our derived result improves that in [13]. That is to say, our result exhibits stronger generality and superiority.

To elucidate the role of our theoretical analysis in selecting measurement matrix, consider a practical sparse recovery scenario, where the overall sparsity and the length of a sparse signal are $K = 4$ and $N = 8192$, respectively. We generate a random Gaussian matrix with $M = 2048$, i.e., the compression rate is $\frac{1}{4}$, which is a widely used setting for compressed measurement matrix in the literature of CS with coherence μ roughly equal to 0.109 [2]. Our theoretical result in Remark 3 indicates that the upper bound of the reconstructible sparsity satisfies $K^*(\mu) \approx 6.78 > 4$. It means that this Gaussian matrix satisfies the exact recovery condition of OLS and MOLS. That is, the OLS and MOLS algorithms can perform exact recovery with this measurement matrix. In this sense, our theoretical results offer a theoretical guarantee on the effectiveness of this measurement matrix, which guides the actual implementation of the recovery process. It is worth noting that our results offer sufficient conditions, which is a common practice in the CS literature. Not surprisingly, the OLS and MOLS algorithms may still be able to perform exact recovery even when such conditions are not met.

B. ERCs for BOLS Algorithm Exploiting Block Structure

The main difference of the analysis between BOLS and the aforementioned two algorithms is that we consider the block structure within the sparse signals. To do so, we adopt two independent concepts of coherence: block-coherence that indicates the inter-block global coherence properties of the measurement matrix $\mathbf{D}$, and sub-coherence that captures the intra-block local coherence properties of $\mathbf{D}$.

Definition 3 (Block-coherence [14]): The block-coherence of $\mathbf{D}$ is defined as

$$\mu_B = \max_{i,j\neq i} \frac{\rho(\mathbf{M}[i,j])}{d}, \tag{32}$$

where $\mathbf{M}[i,j] = \mathbf{D}^T[i]\mathbf{D}[j]$. It is similar to (9) by replacing columns $\mathbf{D}_i$ of $\mathbf{D}$ by its sub-blocks $\mathbf{D}[i]$.

Definition 4 (Sub-coherence [14]): The sub-coherence of $\mathbf{D}$ is defined as

$$\nu = \max_l \max_{i,j\neq i} | < \mathbf{D}_i, \mathbf{D}_j > |, \quad \mathbf{D}_i, \mathbf{D}_j \in \mathbf{D}[l]. \tag{33}$$

Based on aforementioned definitions, we derive the following Lemmas.

Lemma 3: Let k be the block sparsity level, d be the block length, μ be the matrix coherence and ν be the sub-coherence. Suppose $\mu < \frac{1}{kd-1}$, then $\frac{1}{\sqrt{\mathcal{T}_B}} \leq ||\mathbf{P}_{\mathbf{S}^l}^{\perp}\mathbf{D}_i||_2 \leq 1$ for $i \in \{1, 2, \ldots, N\}\backslash\mathbf{S}^l$, where $\mathcal{T}_B = \left(1-\left(\frac{\sqrt{1+(kd-1)\mu}\sqrt{kd\mu^2}}{1-(d-1)\nu-(k-1)d\mu}\right)^2\right)^{-1}$.

Proof: See Appendix E. ∎

Lemma 4: For a random matrix $\mathbf{B} \in \mathcal{R}^{M \times kd}$ $(k \geq 2)$ whose entries are independently and identically distributed as $\mathcal{N}(0, \frac{1}{M})$, the following probability holds:

$$P\left\{\rho_c(\mathbf{B}^T\mathbf{B} - \mathbf{I}) \leq \frac{(k-\mathcal{T}_B)d\tau}{2}\right\} \geq 1 - \frac{ke^{-\frac{kd-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(kd-1)}},$$
(34)

where τ is a positive constant, d is the block length and $\mathcal{C} = \frac{M(k-\mathcal{T}_B)^2 d^2\tau^2}{4(kd-1)^2} - 1$.

The proof is omitted since it is similar to that of Lemma 2.

Remark 4: Similar to Remark 1, we have the following probability:

$$\lim_{M\to\infty} 1 - \frac{ke^{-\frac{kd-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(kd-1)}} = 1.$$
(35)

Intuitively, we give a geometric interpretation of BOLS in terms of orthogonal projections as shown in Proposition 1 below to facilitate the following derivation.

Proposition 1: In the $(l+1)$-th iteration, the BOLS algorithm identifies a block support index:

$$i_B^{l+1} = \underset{j_B \in \{1,..,N_B\}\backslash\mathbf{S}_B^l}{\arg\max} \left(\frac{|<\mathbf{D}_{(j_B-1)d+1}, \mathbf{r}^l>|}{||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_{(j_B-1)d+1}||_2}\right)^2$$

$$+ \sum_{j=(j_B-1)d+2}^{j_Bd} \left(\frac{|<\mathbf{D}_j, \mathbf{r}^l>|}{||\mathbf{P}_{\mathbf{S}^l\cup\{(j_B-1)d+1,...,j-1\}}^\perp\mathbf{D}_j||_2}\right)^2.$$
(36)

Proof: See Appendix F. ∎

The ERC and the MIP-based sufficient condition for BOLS are given in Theorem 3 and Theorem 4.

Theorem 3: Assume that BOLS has chosen l correct blocks after l iterations. A sufficient condition for BOLS to select one correct block in the $(l+1)$-th iteration is

$$\rho_c((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)) < 1,$$
(37)

where $\mathbf{R}_{0\backslash\mathbf{S}}$ and $\overline{\mathbf{R}}_0$ are given by

$$\mathbf{R}_{0\backslash\mathbf{S}} = \text{diag}\left(\frac{1}{||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_1||_2}, \frac{1}{||\mathbf{P}_{\mathbf{S}^l\cup\{1\}}^\perp\mathbf{D}_2||_2}, \ldots,\right.$$

$$\left.\frac{1}{||\mathbf{P}_{\mathbf{S}^l\cup\{1,...,d-1\}}^\perp\mathbf{D}_d||_2}, \cdots\right)$$
(38)

and

$$\overline{\mathbf{R}}_0 = \text{diag}\left(\frac{1}{||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_{kd+1}||_2}, \frac{1}{||\mathbf{P}_{\mathbf{S}^l\cup\{kd+1\}}^\perp\mathbf{D}_{kd+2}||_2}, \ldots,\right.$$

$$\left.\frac{1}{||\mathbf{P}_{\mathbf{S}^l\cup\{(2k-1)d+1,...,2kd-1\}}^\perp\mathbf{D}_{2kd}||_2}, \cdots\right),$$
(39)

if $\{1, \ldots, kd\}$ are support indices in $\mathbf{Q}_0$ and $\{1, \ldots, d\} \not\subset \mathbf{S}^l$.

Proof: See Appendix G. ∎

Theorem 4: Let μ be the matrix coherence, μ_B be the block-coherence and ν be the sub-coherence of the measurement matrix $\mathbf{D}$. The ERC (37) is satisfied if

$$kd < \left(\sqrt[3]{-\frac{q_B}{2} + \sqrt{\Delta_B}} + \sqrt[3]{-\frac{q_B}{2} - \sqrt{\Delta_B}} - \frac{\beta_B}{3\alpha_B}\right),$$
(40)

where $q_B = \frac{27\alpha_B^2\delta_B - 9\alpha_B\beta_B\gamma_B + 2\beta_B^3}{27\alpha_B^3}$, $p_B = \frac{3\alpha_B\gamma_B - \beta_B^2}{3\alpha_B^2}$, $\alpha_B = -\mu_B\mu^3 + 3\mu_B\mu^2$, $\beta_B = (2+\mu_B)\mu^3 - (d\mu_B + \mu_B + 2)\mu^2 + 6\mu_B\mathcal{G}\mu$, $\gamma_B = -2\mu^3 + 2\mu^2 - (2d\mu_B\mathcal{G} + 4\mathcal{G})\mu + 3\mu_B\mathcal{G}^2$, $\delta_B = -d\mu_B\mathcal{G}^2 - 2\mathcal{G}^2$, $\mathcal{G} = (d-1)\nu - 1 - d\mu$ and $\Delta_B = (\frac{q_B}{2})^2 + (\frac{p_B}{3})^3$.

Proof: See Appendix H. ∎

Remark 5: Theorem 4 applies to OLS when $d = 1$, which is identical to Theorem 2.

While the main result in Theorem 4 appears to be cumbersome, it accurately quantifies the performance bounds of BOLS and offers adequate intuitions for a given problem setting. Along this line, Remark 6 arises.

Remark 6: To clearly observe the relationship between the upper bound of sparsity and matrix coherence, and the relationship between the upper bound of sparsity and block-coherence, let us consider the special case of a block structure matrix, that is, the sub-coherence satisfies that $\nu = 0$. Denote the right-side of (40) as $K_B^*(\mu, \mu_B)$. The upper bound of $K_B^*(\mu, \mu_B)$ is $K_B^*\left(\sqrt{\frac{N-M}{M(N-1)}}, \frac{1}{d}\sqrt{\frac{N-M}{M(N-1)}}\right)$. Depending on the assumptions adopted, the corresponding analytical results for the upper bound are given as follows. 1) When M is fixed, the upper bound of $K_B^*(\mu, \mu_B)$ decreases with the increase of N. 2) When N is fixed, the upper bound of $K_B^*(\mu, \mu_B)$ is improved with the increase of M. 3) Letting $\omega = \frac{M}{N}$, and for $M \gg 0$ and $N \gg 0$, the upper bound of the theoretical threshold $K_B^*(\mu, \mu_B)$ satisfies

$$K_B^*(\mu, \mu_B) \leq \mathcal{W}(d)\sqrt{\frac{M}{1-\omega}} + \frac{5}{9}d + \frac{1}{9},$$
(41)

where

$$\mathcal{W}(d) = \sqrt[3]{-\mathcal{U} + \sqrt{\mathcal{U}^2 + \mathcal{V}^3}}$$

$$+ \sqrt[3]{-\mathcal{U} - \sqrt{\mathcal{U}^2 + \mathcal{V}^3}} + \frac{2}{9}d + \frac{2}{3},$$
(42)

$\mathcal{U} = -\frac{8}{729}d^3 + \frac{4}{81}d^2 - \frac{2}{27}d + \frac{1}{27}$, and $\mathcal{V} = -\frac{4}{81}d^2 + \frac{4}{27}d - \frac{1}{9}$.

Proof: By direct calculations, the theoretical threshold $K_B^*(\mu, \mu_B)$ is improved with the decrease of μ or μ_B, i.e., $\nabla K_B^*(\mu) < 0$ and $\nabla K_B^*(\mu_B) < 0$. In order to obtain the upper bound of $K_B^*(\mu, \mu_B)$, it is necessary to derive the minimum value of μ_B.

For a given matrix $\mathbf{Q} \in \mathcal{R}^{d \times d}$,

$$\rho(\mathbf{Q}) \geq ||\mathbf{Q}||_{\max},$$
(43)

where $|| \cdot ||_{\max}$ represents the largest element of its objective. Therefore, the block-coherence of a given matrix $\mathbf{D}$ satisfies

$$\mu_B \geq \frac{\mu}{d} \geq \frac{1}{d}\sqrt{\frac{N-M}{M(N-1)}},$$
(44)

Finally, the proof of Remark 6 can be completed by using the similar proof of Remark 2. ∎

In general, the derived upper bound is improved with the increase of d and M. On the one hand, the results mean that

the performance of the algorithm generally increases with the increase of the number of measurements. On the other hand, the larger d makes the block structure more powerful, leading to better performance of BOLS.

Remark 7: Define the right-side of (40) as $K_B^*(\mu, \mu_B, \nu)$. In asymptotic case, we have

$$\lim_{\mu_B, \nu \to 0; \, \mu \to \mu_B} K_B^*(\mu, \mu_B, \nu) = \frac{2}{3}\left(\frac{1}{\mu_B} + \frac{7}{6}d - \frac{1}{6}\right). \quad (45)$$

The well-known MIP-based sufficient condition for the ERC of BOMP by using orthogonal block uncertainty relations in [14] is

$$kd < \frac{1}{2}\left(\frac{1}{\mu_B} + d\right). \quad (46)$$

Since $d \geq 1$ and $\mu_B \in [0, 1]$, the sufficient condition based on (45), i.e.,

$$kd < \frac{2}{3}\left(\frac{1}{\mu_B} + \frac{7}{6}d - \frac{1}{6}\right), \quad (47)$$

is improved compared with (46). This result implies that BOLS performs better than BOMP from the perspective of reconstructible sparsity level.

IV. NOISY RECOVERY CONDITIONS

In the noisy case, to guarantee reliable recovery, the SNR should be sufficiently high for correct atom selection. Meanwhile, the level of MIP needs to be small, otherwise the atoms cannot be separated well. Considering these issues, this section presents the recovery conditions in the noisy case for the OLS-type algorithms based on the theoretical results for the noiseless case in Theorem 2 and Theorem 4.

The noisy form of the CS model (1) is given by

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \boldsymbol{\epsilon}, \quad (48)$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_M)$ is the Gaussian noise. Note that the residual in the l-th iteration is

$$\mathbf{r}^l = (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{y} = (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}\mathbf{x} + (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\boldsymbol{\epsilon}. \quad (49)$$

Define $\mathbf{s}^l = (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}\mathbf{x}$ and $\mathbf{n}^l = (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\boldsymbol{\epsilon}$ as the signal and noise parts of the residual, respectively. Then the analysis of recovery conditions for OLS-type algorithms under the noisy model (48) is investigated in the following subsections.

A. Noisy Recovery Conditions for OLS and MOLS Algorithms

To obtain the recovery conditions for the OLS and MOLS algorithms in the noisy case, we first present the following lemma.

Lemma 5: If the noiseless condition in (23) holds, then

$$\|(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)\|_{1,1} \leq \frac{2K\mathcal{T}\mu}{2 - (K - \mathcal{T})\mu} < 1. \quad (50)$$

The proof of Lemma 5 follows directly from the proof of Theorem 2 in Appendix D. Then, based on Lemma 5, the following theorem holds.

Theorem 5: Suppose that the condition in (23) holds and the remaining nonzero vector $\mathbf{x}_{0\backslash\mathbf{S}}$ in the $(l + 1)$-th iteration satisfies

$$\|\mathbf{x}_{0\backslash\mathbf{S}}\|_2 > \frac{2\sqrt{K - l}(2 - (K - \mathcal{T})\mu)\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu)(1 - (K - 1)\mu)}. \quad (51)$$

Then, in the noisy case, the OLS algorithm chooses one correct atom and the MOLS algorithm chooses at least one correct atom in the $(l + 1)$-th iteration with probability at least $1 - \frac{1}{M}$.

Proof: See Appendix I. ∎

Based on Theorem 5, we have the subsequent corollary.

Corollary 1: Suppose that the condition in (23) holds and all the nonzero entries $\mathbf{x}_i$ satisfy

$$|\mathbf{x}_i| > \frac{2(2 - (K - \mathcal{T})\mu)\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu)(1 - (K - 1)\mu)}. \quad (52)$$

Then, in the noisy case, the OLS and MOLS algorithms choose the true support set with probability at least $1 - \frac{1}{M}$.

Theorem 5 and Corollary 1 indicate that if the nonzero entries are large enough, the OLS and MOLS algorithms choose the correct atoms in the noisy case.

Remark 8: For MOLS algorithm, Theorem 5 and Corollary 1 present the strongest sufficient condition, i.e., they hold for all $L \in [1, \min\{K, \frac{M}{K}\}]$ [16]. Since $\max_L(\cdot) \geq g_L\|\cdot\|_\infty$, where $g_L \in [1, L]$ denotes a nonlinear parameter with respect to L, the conditions (51) in Theorem 5, and (52) in Corollary 1 are relaxed to

$$\|\mathbf{x}_{0\backslash\mathbf{S}}\|_2$$
$$> \frac{(1 + \frac{g_L}{L})\sqrt{K - l}(2 - (K - \mathcal{T})\mu)\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu)(1 - (K - 1)\mu)} \quad (53)$$

and

$$|\mathbf{x}_i| > \frac{(1 + \frac{g_L}{L})(2 - (K - \mathcal{T})\mu)\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu)(1 - (K - 1)\mu)}, \quad (54)$$

respectively. If the nonlinear parameter g_L is equal to 1, (53) and (54) correspond to the easiest-satisfied conditions. In this case, a larger L leads to a smaller lower bound of $\|\mathbf{x}_{0\backslash\mathbf{S}}\|_2$ and $|\mathbf{x}_i|$ required for reliable recovery.

B. Noisy Recovery Conditions for BOLS Algorithm

Similarly, to derive the recovery conditions for the BOLS algorithm in the noisy case, we give the following lemmas.

Lemma 6: If the noiseless condition in (40) holds, then

$$\rho_c((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)) < \frac{2\mathcal{T}_B kd\mu_B}{2 - (k - \mathcal{T}_B)d\mu_B} < 1. \quad (55)$$

Lemma 7: The Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_M)$ satisfies

$$\mathrm{P}\left\{\|\boldsymbol{\epsilon}\|_{2,\infty} \leq \sqrt{d}\sigma\sqrt{M + 2\sqrt{M \log M}}\right\} \geq 1 - \frac{1}{M}, \quad (56)$$

where d is the block length.

The proof of Lemma 7 can be extended from that of Lemma 5.1 in [46]. Based on Lemmas 6 and 7, we obtain the following Theorem 6.

(a) The bounds versus μ.   (b) The bounds versus k.
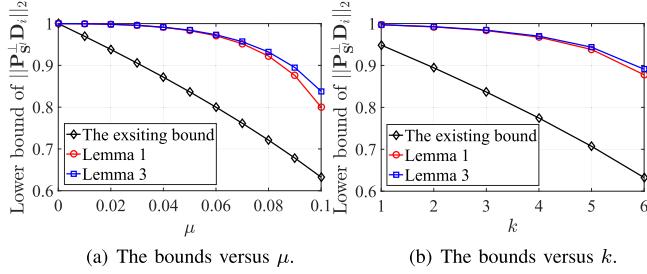
Fig. 2. Comparisons of lower bounds of $||\mathbf{P}_{\mathbf{S}^l}^{\perp}\mathbf{D}_i||_2$ with (a) $K = 6, k = 3$, $d = 2$ and $\nu = \frac{1}{2}\mu$; (b) $K = kd, d = 2, \mu = 0.05$ and $\nu = \frac{1}{2}\mu$.

Theorem 6: Suppose that the condition in (40) holds and the remaining nonzero vector $\mathbf{x}_{0\setminus\mathbf{S}}$ in the $(l+1)$-th iteration satisfies

$$||\mathbf{x}_{0\setminus\mathbf{S}}||_2$$
$$> \frac{2\sqrt{k-l}(2 - (k - \mathcal{T}_B)d\mu_B)\sqrt{d}\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (k - \mathcal{T}_B)d\mu_B - 2\mathcal{T}_B kd\mu_B)(1 - (kd - 1)\mu)}. \tag{57}$$

Then, in the noisy case, the BOLS algorithm chooses one correct block in the $(l+1)$-th iteration with probability at least $1 - \frac{1}{M}$.

Proof: See Appendix J. ∎

The following corollary is derived from Theorem 6.

Corollary 2: Suppose that the condition in (40) holds and all the nonzero blocks $\mathbf{x}[i]$ satisfy

$$||\mathbf{x}[i]||_2 > \frac{2(2 - (k - \mathcal{T}_B)d\mu_B)\sqrt{d}\sigma\sqrt{M + 2\sqrt{M \log M}}}{(2 - (k - \mathcal{T}_B)d\mu_B - 2\mathcal{T}_B kd\mu_B)(1 - (kd - 1)\mu)}. \tag{58}$$

Then, in the noisy case, the BOLS algorithm chooses the true support set with probability at least $1 - \frac{1}{M}$.

V. SIMULATION TESTS

In this section, we present simulation tests to illustrate our theoretical results shown in Section III and Section IV, and compare them with the existing ones.

A. Simulation Tests for Noiseless Recovery Conditions

1) Comparisons of the Lower Bounds of $||\mathbf{P}_{\mathbf{S}^l}^{\perp}\mathbf{D}_i||_2$: In this subsection, we exploit simulations to illustrate Lemma 1, Lemma 3 and compare them with (19) [38].

The simulation parameters satisfy $\mu(K - 1) < 1$ and $\mu(kd - 1) < 1$, which are given in Lemma 1, Lemma 3 and the precondition of (19). Fig. 2 shows that our derived results are much tighter than the existing bound which indicates that the follow-up theoretical results based on these bounds, such as Theorems 2 and 4, are much sharper than the one using (19). Meanwhile, it is observed that the lower bounds using sub-coherence in Lemma 1 is much tighter than that just using conventional matrix coherence. This means the characteristics of block structure plays a role in improving the boundary.

*2) Comparisons of the Sufficient Conditions for ERCs:* The simulations of Theorems 2 and 4 are conducted. We compare
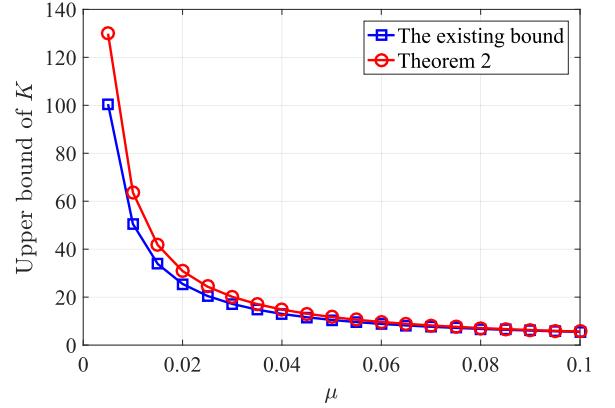
Fig. 3. Comparisons of upper bounds of K as a function of μ.
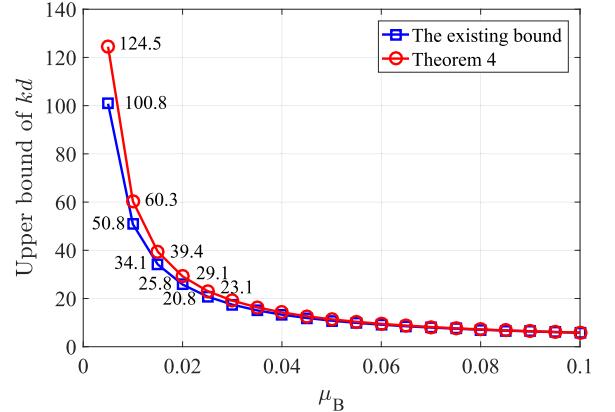
Fig. 4. Comparisons of upper bounds of kd as a function of μ_B with $\mu = 1.2\mu_B, \nu = 0.8\mu_B$ and $d = 4$.
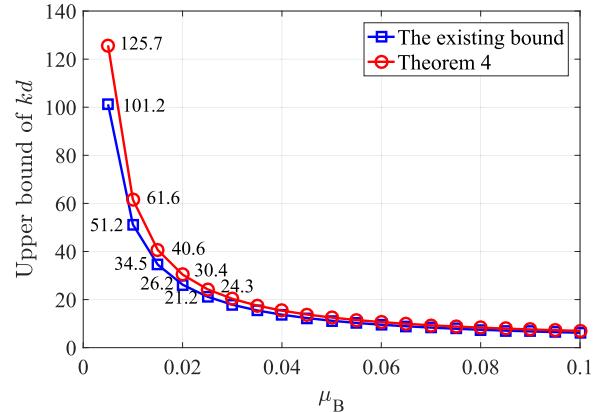
Fig. 5. Comparisons of upper bounds of kd as a function of μ_B with $\mu = 1.2\mu_B, \nu = 0.8\mu_B$ and $d = 8$.

Theorem 2 with Theorem 3.5 in [13], and Theorem 4 with Theorem 3 in [14] respectively.

Figs. 3–5 respectively present the existing bounds and our derived bounds for OLS, MOLS, BOLS with $d = 4$ and BOLS with $d = 8$. The results show that the upper bounds of K and kd in Theorems 2 and 4 are higher than the existing bounds under the same parameter settings, resulting in more reliable reconstructible sparsity levels of OLS-type algorithms than those of OMP-type algorithms. They also show that for the block sparse
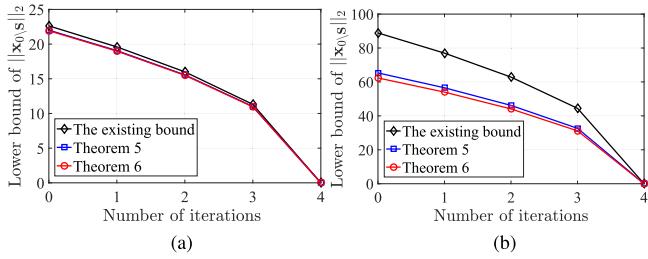
Fig. 6. Comparisons of lower bounds of $||\mathbf{x}_{0\setminus\mathbf{S}}||_2$ as a function of l with $\mu_B = 0.025$, $\nu = \mu_B$, $k = 4$, $\sigma = 0.1$, $M = 512$ and (a) $d = 2$; (b) $d = 4$.
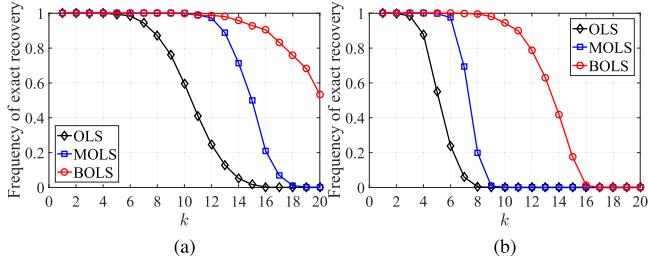
Fig. 7. Performance of OLS, MOLS and BOLS with $M = 128$, $N = 256$, (a) $d = 4$; (b) $d = 8$.

signals, the upper bounds of kd are improved with the increase of d.

B. Simulation Tests for Noisy Recovery Conditions

In this subsection, we perform simulations to illustrate Theorem 5, Theorem 6, and compare them with the existing bound (Eqn. 10 [38]).

For fair comparison, μ and K in Theorem 5 and the existing bound are replaced by μ_B and $k \times d$ respectively. Then, the lower bounds of $||\mathbf{x}_{0\setminus\mathbf{S}}||_2$ as a function of the number of iteration l are presented in Fig. 6. It shows that the lower bounds in Theorems 5 and 6 are smaller than the existing bound for an arbitrary iteration number, which indicates that the conditions required by OLS-type algorithms for reliable reconstruction are more relaxed than that of OMP. Moreover, it implies that BOLS performs better than OLS and MOLS.

C. Empirical Simulation Tests for the OLS-Type Algorithms

This subsection presents empirical tests on the performance of OLS-type algorithms. For each problem setting, we perform 1,000 independent tests for each algorithm to calculate the frequency of exact recovery. We adopt a Gaussian random matrix $\mathbf{D} \in \mathcal{R}^{M \times N}$ ($M = 128$ and $N = 256$) as our measurement matrix with each entry being independently and identically distributed as $\mathcal{N}\left(0, \frac{1}{M}\right)$. For each value of k, we generate a block k-sparse signal, whose support is selected randomly and the nonzero elements independently follow standard Gaussian distribution, i.e., $\mathcal{N}(0, 1)$. The recovery is successful if the reconstructed vector is within a certain small Euclidean distance, which is set as 10^{-6}, of the original vector.

As shown in Fig. 7, BOLS outperforms other greedy algorithms, which means that the utilization of the block property exhibits competitive reconstruction performance.

TABLE I
UPPER BOUNDS OF RECONSTRUCTIBLE SPARSITY LEVELS WITH THE SETTINGS IN FIG. 7

	Theoretical condition	Simulation result
OLS	$K < 8.2 \Rightarrow K \leq 8$	$K \leq 12$
MOLS	$K < 8.2 \Rightarrow K \leq 8$	$K \leq 32$
BOLS ($d = 4$)	$kd < 18.1 \Rightarrow kd \leq 18$	$kd \leq 40$
BOLS ($d = 8$)	$kd < 21.2 \Rightarrow kd \leq 21$	$kd \leq 48$
$K < 1/\mu + 1$	$K < 12.3$	$-$

When the block length becomes larger, the performance of BOLS is also improved. In Table I, we calculate the upper bounds of the reconstructible sparsity levels in Remarks 3 and 7 for OLS-type algorithms. The coherence parameters are set to be their lower bounds, i.e., $\mu = \frac{1}{\sqrt{M}}$ and $\mu_B = \frac{1}{\sqrt{dM}}$ as given in Section III-A [14]. It can be seen that the bound $K < \frac{1}{\mu} + 1$ is $K < 12.3$ in the simulation, which is more relaxed than the sufficient conditions for OLS and MOLS, i.e., $K < 8.2$. This means that the Neumann series is convergent in the proof procedures. Evidently, the theoretical thresholds are more pessimistic than the empirical results. However, this negativity is moderated compared with the results in [14]. The theoretical thresholds indicate that the frequencies of exact recovery are close to 100% for OLS and MOLS with $K = 8$, for BOLS ($d = 4$) with $kd = 18$ and for BOLS ($d = 8$) with $kd = 21$. The empirical results show that the frequencies of exact recovery close to 100% for OLS, MOLS, BOLS ($d = 4$) and BOLS ($d = 8$) up to $K = 12$, $K = 32$, $kd = 40$ and $kd = 48$ respectively.

VI. CONCLUSION

In this paper, we study the OLS-type algorithms in reconstructing sparse signals in both noiseless and noisy cases. We exploit MIP and asymptotic analysis to analyze exact recovery conditions of OLS and MOLS for general sparse signals. Analytical results for BOLS are derived by exploiting the notions of block-coherence and sub-coherence. The various notions of coherence, describing the similarity among matrix columns, are simpler and more intuitive than RIP. Our derived performance guarantees reveal that the OLS-type algorithms perform reliable recovery if the sparsity level is lower than the MIP-based conditions in noiseless case, or if the power of nonzero coefficients is larger than the conditions related to both MIP and noise variance in noisy case. Due to the sharpness of the mapping normalization factor bound, the developed theoretical results improve the existing ones.

Note that our derived upper bounds of reconstructible sparsity themselves do not rely on any prior sparsity information. Based on these bounds, one future work is to design new blind recovery algorithms when the real sparsity value is not available.

APPENDIX A
PROOF OF THEOREM 1

Proof: We first use mathematical induction to prove that Theorem 1 is suitable for OLS. Assume that OLS has selected l atoms from $\mathbf{Q}_0$ after l steps ($l \in \{1, \ldots, K - 1\}$). Then, we

derive a condition to guarantee that the next selected entry is also correct.

Since the expression $||(\mathbf{DR})^T\mathbf{r}^l||_\infty$ gives the largest magnitude attained among the inner products, where $\mathbf{R} = f(\mathbf{D})$. In consequence, to see whether the next chosen index corresponds to an atom in $\mathbf{Q}_{0\backslash\mathbf{S}}$, we need to examine whether the following quotient holds:

$$Z(\mathbf{r}^l, \mathbf{S}^l) = \frac{||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T\mathbf{r}^l||_\infty}{||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty} < 1. \quad (59)$$

Following Tropp's [13] and Soussen's [29] analysis, we know that $\mathbf{r}^l \in \text{span}(\mathbf{Q}_{0\backslash\mathbf{S}}) = \text{span}(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})$. Therefore,

$$((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger)^T(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l = \mathbf{r}^l. \quad (60)$$

Substituting (60) into (59) yields

$$Z(\mathbf{r}^l, \mathbf{S}^l)$$
$$= \frac{||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger)^T(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty}{||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty}$$
$$\leq ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger)^T||_{\infty,\infty}$$
$$= ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1}. \quad (61)$$

Then, if the following condition holds, i.e.,

$$Z(\mathbf{r}^l, \mathbf{S}^l) \leq ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1} < 1, \quad (62)$$

OLS selects a correct atom from $\mathbf{Q}_{0\backslash\mathbf{S}}$.

Next, we prove that the theorem is also valid for MOLS. Omit some similar proof to that of OLS. For MOLS's selecting at least one correct index [34], it needs to examine whether the following quotient holds:

$$Z_M(\mathbf{r}^l, \mathbf{S}^l) = \frac{||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T\mathbf{r}^l||_{L,1}}{||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty} < L, \quad (63)$$

where $||\cdot||_{L,1}$ calculates the absolute sum of L largest elements of its objective. Then, we have

$$Z_M(\mathbf{r}^l, \mathbf{S}^l)$$
$$\leq \frac{L \times ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T\mathbf{r}^l||_\infty}{||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty}$$
$$= L \times \frac{||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger)^T(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty}{||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{r}^l||_\infty}$$
$$\leq L \times ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1} < L, \quad (64)$$

and thus the proof of the theorem is completed. ∎

APPENDIX B
PROOF OF LEMMA 1

Proof: Note that for $i \in \{1, 2, \ldots, N\}\backslash\mathbf{S}^l$, $\mathbf{D}_i = \mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i + \mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_i$. Then, $||\mathbf{D}_i||_2^2 = ||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2^2 + ||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_i||_2^2 = 1$ for normalized matrix $\mathbf{D}$. By rewriting $\mathbf{P}_{\mathbf{S}^l}$ and using submultiplicativity of matrix norm, we have

$$||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2 = ||\mathbf{D}_{\mathbf{S}^l}(\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l})^{-1}\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2$$
$$\leq \rho(\mathbf{D}_{\mathbf{S}^l})\rho((\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l})^{-1})||\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2. \quad (65)$$

By exploiting Lemma 2 in [38],

$$\rho(\mathbf{D}_{\mathbf{S}^l}) = \sqrt{\lambda_{\max}(\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l})} \leq \sqrt{1 + (K-1)\mu}. \quad (66)$$

Next, we derive the upper bound of the second term in the last inequality of (65) and we express $\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l}$ as $\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l} = \mathbf{I} + \mathbf{A}$, where $\mathbf{A}_{ii} = 0$ for all i. Then, by applying Gerš gorin's disc theorem, we obtain

$$\rho(\mathbf{A}) \leq (l-1)\mu \leq (K-1)\mu < 1. \quad (67)$$

Using Neumann series expansion (Corollary 5.6.16 [47]), we have $(\mathbf{I} + \mathbf{A})^{-1} = \sum_{k=0}^\infty(-\mathbf{A})^k$. Therefore,

$$\rho((\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}_l})^{-1}) = \rho\left(\sum_{k=0}^\infty(-\mathbf{A})^k\right) \leq \sum_{k=0}^\infty\left(\rho(\mathbf{A})^k\right)$$
$$= \frac{1}{1 - \rho(\mathbf{A})} \leq \frac{1}{1 - (K-1)\mu}. \quad (68)$$

Moreover, the upper bound of $||\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2$ is calculated by

$$||\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2 \leq \sqrt{l\mu^2} \leq \sqrt{K\mu^2}. \quad (69)$$

Combining (66), (68) and (69) yields

$$||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2 \leq \frac{\sqrt{1 + (K-1)\mu}\sqrt{K\mu^2}}{1 - (K-1)\mu}. \quad (70)$$

Finally, we obtain

$$1 \overset{(a)}{\geq} ||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_i||_2 = \sqrt{1 - ||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2^2}$$
$$\geq \sqrt{1 - \left(\frac{\sqrt{1 + (K-1)\mu}\sqrt{K\mu^2}}{1 - (K-1)\mu}\right)^2}. \quad (71)$$

The reason why (a) in (71) holds is that in the first iteration, $\mathbf{S}^0 = \emptyset$, and thus $||\mathbf{P}_{\mathbf{S}^0}^\perp\mathbf{D}_i||_2 = ||\mathbf{D}_i||_2 = 1$. For $l > 1$, $||\mathbf{P}_{\mathbf{S}^l}^\perp\mathbf{D}_i||_2 < ||\mathbf{D}_i||_2 = 1$. ∎

APPENDIX C
PROOF OF LEMMA 2

Proof: Define $\mathbf{a}^j = \mathbf{B}_{\bullet\backslash j}^T\mathbf{B}_j \in \mathcal{R}^{K-1}$ $(j \in \{1, 2, \ldots, K\})$, where $\mathbf{B}_{\bullet\backslash j} = \mathbf{B}\backslash\mathbf{B}_j$. Then, according to the central limit theorem, when M is sufficiently large, the entries of $\mathbf{a}^j$ are independently and identically distributed as $\mathcal{N}(0, \frac{1}{M})$. Meanwhile, the random variable $\mathcal{A}^j = M||\mathbf{a}^j||_2^2 \sim \chi_{K-1}^2$. Therefore,

$$P\left\{||\mathbf{a}^j||_1 \leq \frac{(K-\mathcal{T})\tau}{2}\right\}$$
$$\geq P\left\{\sqrt{K-1}||\mathbf{a}^j||_2 \leq \frac{(K-\mathcal{T})\tau}{2}\right\}$$
$$= P\left\{M||\mathbf{a}^j||_2^2 \leq \frac{M(K-\mathcal{T})^2\tau^2}{4(K-1)}\right\}. \quad (72)$$

Moreover, the Lemma 4 in [48] indicates that for any $\mathcal{C} > 0$,

$$P\{\mathcal{A}^j > (1+\mathcal{C})(K-1)\} \leq \frac{e^{-\frac{K-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(K-1)}}. \quad (73)$$

Hence,

$$
\mathrm{P} \left\{ ||\mathbf{B}^T\mathbf{B} - \mathbf{I}||_{1,1} \leq \frac{(K-\mathcal{T})\tau}{2} \right\}
$$

$$
= 1 - \mathrm{P} \left\{ \bigcup_{j=1}^{K} \left(||\mathbf{a}^j||_1 > \frac{(K-\mathcal{T})\tau}{2} \right) \right\}
$$

$$
\geq 1 - \sum_{j=1}^{K} \mathrm{P} \left\{ ||\mathbf{a}^j||_1 > \frac{(K-\mathcal{T})\tau}{2} \right\}
$$

$$
\geq 1 - \frac{Ke^{-\frac{K-1}{2}(\mathcal{C}-\log(1+\mathcal{C}))}}{\mathcal{C}\sqrt{\pi(K-1)}}, \tag{74}
$$

where $\mathcal{C} = \frac{M(K-\mathcal{T})^2\tau^2}{4(K-1)^2} - 1$. ∎

APPENDIX D
PROOF OF THEOREM 2

Proof: By exploiting submultiplicativity,

$$
||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1}
$$

$$
= \max_{\overline{\mathbf{G}}_i} ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger\overline{\mathbf{G}}_i||_1
$$

$$
= \max_{\overline{\mathbf{G}}_i} ||((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^{-1}(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\overline{\mathbf{G}}_i||_1
$$

$$
\leq ||((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^{-1}||_{1,1}
$$

$$
\times \max_{\overline{\mathbf{G}}_i} ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\overline{\mathbf{G}}_i||_1, \tag{75}
$$

where $\overline{\mathbf{G}}_i$ represents the i-th column of $\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0$. Now let us bound the term $||((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^{-1}||_{1,1}$ in the last inequality of (75). Let the off-diagonal part of $\mathbf{Q}_{0\backslash\mathbf{S}}^T\mathbf{Q}_{0\backslash\mathbf{S}}$ be $\mathbf{A}$, i.e., $\mathbf{Q}_{0\backslash\mathbf{S}}^T\mathbf{Q}_{0\backslash\mathbf{S}} = \mathbf{I} + \mathbf{A}$. Then, according to Lemma 2, we have

$$
||\mathbf{A}||_{1,1} \leq \frac{(K-l-\mathcal{T})\mu}{2} \leq \frac{(K-\mathcal{T})\mu}{2}. \tag{76}
$$

Now the assumption (23) implies $\frac{(K-\mathcal{T})\mu}{2} < 1$. Then, by exploiting Neumann series expansion,

$$
||(\mathbf{Q}_{0\backslash\mathbf{S}}^T\mathbf{Q}_{0\backslash\mathbf{S}})^{-1}||_{1,1} = \left\| \sum_{k=0}^{\infty}(-\mathbf{A})^k \right\|_{1,1} \leq \frac{1}{1 - \frac{K\mu}{2} + \frac{\mathcal{T}\mu}{2}}. \tag{77}
$$

Observe that

$$
||((\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^{-1}||_{1,1}
$$

$$
\leq ||(\mathbf{R}_{0\backslash\mathbf{S}})^{-1}||_{1,1}||(\mathbf{Q}_{0\backslash\mathbf{S}}^T\mathbf{Q}_{0\backslash\mathbf{S}})^{-1}||_{1,1}||(\mathbf{R}_{0\backslash\mathbf{S}})^{-1}||_{1,1}
$$

$$
\leq ||(\mathbf{Q}_{0\backslash\mathbf{S}}^T\mathbf{Q}_{0\backslash\mathbf{S}})^{-1}||_{1,1} \leq \frac{1}{1 - \frac{K\mu}{2} + \frac{\mathcal{T}\mu}{2}}. \tag{78}
$$

On the other hand, by the definition of matrix coherence, it is straightforward to obtain

$$
\max_{\overline{\mathbf{G}}_i} ||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^T\overline{\mathbf{G}}_i||_1 \leq K\mathcal{T}\mu. \tag{79}
$$

Combining (78) and (79), we get

$$
||(\mathbf{Q}_{0\backslash\mathbf{S}}\mathbf{R}_{0\backslash\mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1} \leq \frac{2K\mathcal{T}\mu}{2 - (K-\mathcal{T})\mu} < 1. \tag{80}
$$

By writing out $\mathcal{T}$, we obtain the following cubic inequality,

$$
\alpha K^3 + \beta K^2 + \gamma K + \delta < 0, \tag{81}
$$

where α, β, γ and δ are given in the description of Theorem 2. Finally, by exploiting Cardano formula to solve the valid solution of the inequality (81) in real domain, the proof is completed. ∎

APPENDIX E
PROOF OF LEMMA 3

Proof: According to submultiplicativity of matrix norm,

$$
||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2 = ||\mathbf{D}_{\mathbf{S}^l}(\mathbf{D}_{\mathbf{S}^l}\mathbf{D}_{\mathbf{S}^l})^{-1}\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2
$$

$$
\leq \rho(\mathbf{D}_{\mathbf{S}^l})\rho((\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l})^{-1})||\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2. \tag{82}
$$

Similar to that in Appendix B, we have

$$
\rho(\mathbf{D}_{\mathbf{S}^l}) \leq \sqrt{1 + (kd-1)\mu} \tag{83}
$$

and

$$
||\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_i||_2 \leq \sqrt{l\mu^2} \leq \sqrt{kd\mu^2}. \tag{84}
$$

Then, it remains to derive the upper bound of the second term in the last inequality of (82) and we express $\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l}$ as $\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l} = \mathbf{I} + \mathbf{A}$, where $\mathbf{A}$ is an $ld \times ld$ matrix with blocks $\mathbf{A}[l, r]$ of size $d \times d$ such that $\mathbf{A}_{ii} = 0$ for all i because of the normalized matrix $\mathbf{D}$. Since $\mathbf{A}[l, r] = \mathbf{D}_{\mathbf{S}^l}^T[l]\mathbf{D}_{\mathbf{S}^l}[r]$ for all $l \neq r$, and $\mathbf{A}[r, r] = \mathbf{D}_{\mathbf{S}^l}^T[r]\mathbf{D}_{\mathbf{S}^l}[r] - \mathbf{I}$, we have

$$
\rho(\mathbf{A}) \leq (d-1)\nu + (l-2)d\mu \leq (d-1)\nu + (k-1)d\mu
$$

$$
\leq (kd-1)\mu < 1. \tag{85}
$$

Therefore, $(\mathbf{I} + \mathbf{A})^{-1} = \sum_{k=0}^{\infty}(-\mathbf{A})^k$ and we obtain

$$
\rho\left((\mathbf{D}_{\mathbf{S}^l}^T\mathbf{D}_{\mathbf{S}^l})^{-1}\right) = \rho\left(\sum_{k=0}^{\infty}(-\mathbf{A})^k\right) \leq \sum_{k=0}^{\infty}\left(\rho(\mathbf{A})^k\right)
$$

$$
= \frac{1}{1 - \rho(\mathbf{A})}
$$

$$
\leq \frac{1}{1 - (d-1)\nu - (k-1)d\mu}. \tag{86}
$$

Hence,

$$
||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2^2 \leq \frac{kd\mu^2 + kd(kd-1)\mu^3}{(1 - (d-1)\nu - (k-1)d\mu)^2}. \tag{87}
$$

Finally, we have

$$
1 \geq ||\mathbf{P}_{\mathbf{S}^l}^{\perp}\mathbf{D}_i||_2 = \sqrt{1 - ||\mathbf{P}_{\mathbf{S}^l}\mathbf{D}_i||_2^2}
$$

$$
\geq \sqrt{1 - \frac{kd\mu^2 + kd(kd-1)\mu^3}{(1 - (d-1)\nu - (k-1)d\mu)^2}}. \tag{88}
$$

∎

APPENDIX F
PROOF OF PROPOSITION 1

Proof: From the geometric properties of a right triangle, we get

$$||\mathbf{P}^{\perp}_{\mathbf{S}^l \cup \{(j_B-1)d+1,\ldots,j_B d\}} \mathbf{y}||^2_2$$
$$+ ||\mathbf{P}_{\mathbf{S}^l \cup \{(j_B-1)d+1,\ldots,j_B d\}} \mathbf{y}||^2_2 = ||\mathbf{y}||^2_2. \quad (89)$$

Therefore, support selection step in BOLS is altered to

$$i^{l+1}_B = \underset{j_B \in \{1,\ldots,N_B\} \setminus \mathbf{S}^l_B}{\arg\max} ||\mathbf{P}_{\mathbf{S}^l \cup \{(j_B-1)d+1,\ldots,j_B d\}} \mathbf{y}||^2_2. \quad (90)$$

Without loss of generality, we assume the i_B-th block contains the indices $\{1,\ldots,d\}$, where $d \geq 2$. Then, thanks to the decomposition characteristics [42] of $||\mathbf{P}_{\mathbf{S}^l \cup \{1,\ldots,d\}} \mathbf{y}||^2_2$, we have

$$||\mathbf{P}_{\mathbf{S}^l \cup \{1,\ldots,d\}} \mathbf{y}||^2_2$$

$$= ||\mathbf{P}_{\mathbf{S}^l \cup \{1,\ldots,d-1\}} \mathbf{y}||^2_2 + \left(\frac{|<\mathbf{D}_d, \mathbf{r}^l>|}{||\mathbf{P}^{\perp}_{\mathbf{S}^l \cup \{1,\ldots,d-1\}} \mathbf{D}_d||_2} \right)^2$$

$$\overset{(a)}{=} ||\mathbf{P}_{\mathbf{S}^l} \mathbf{y}||^2_2 + \left(\frac{|<\mathbf{D}_1, \mathbf{r}^l>|}{||\mathbf{P}^{\perp}_{\mathbf{S}^l} \mathbf{D}_1||_2} \right)^2$$

$$+ \sum_{j=2}^{d} \left(\frac{|<\mathbf{D}_j, \mathbf{r}^l>|}{||\mathbf{P}^{\perp}_{\mathbf{S}^l \cup \{1,\ldots,j-1\}} \mathbf{D}_j||_2} \right)^2, \quad (91)$$

where equation (a) is derived by writing out $||\mathbf{P}_{\mathbf{S}^l \cup \{1,\ldots,d-1\}} \mathbf{y}||^2_2$ and we obtain (36). ∎

APPENDIX G
PROOF OF THEOREM 3

Proof: Based on Proposition 1, the atom selection indicator of the BOLS algorithm is given by

$$||(\mathbf{D}_{\bullet \setminus \mathbf{S}} \mathbf{R}_{\bullet \setminus \mathbf{S}})^T \mathbf{r}^l||_{2,\infty}, \quad (92)$$

where $\mathbf{D}_{\bullet \setminus \mathbf{S}} = \mathbf{D} \setminus \mathbf{D}_{\mathbf{S}^l} \in \mathcal{R}^{M \times (N-ld)}$, and $\mathbf{R}_{\bullet \setminus \mathbf{S}} \in \mathcal{R}^{(N-ld) \times (N-ld)}$ is given as follows:

$$\mathbf{R}_{\bullet \setminus \mathbf{S}} = \begin{bmatrix} \mathbf{R}_{0 \setminus \mathbf{S}} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{R}}_0 \end{bmatrix}. \quad (93)$$

Similar to the assumption in Appendix A, to see whether the next chosen index corresponds to an atom in $\mathbf{Q}_{0 \setminus \mathbf{S}}$, we need to examine whether the following quotient holds:

$$Z(\mathbf{r}^l, \mathbf{S}^l) = \frac{||(\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0)^T \mathbf{r}^l||_{2,\infty}}{||(\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \mathbf{r}^l||_{2,\infty}} < 1. \quad (94)$$

Since $\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}} (\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger}$ is Hermitian, we have

$$\left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger} \right)^T (\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \mathbf{r}^l = \mathbf{r}^l. \quad (95)$$

Substituting (95) into (94) yields

$$Z(\mathbf{r}^l, \mathbf{S}^l) \leq \rho_r \left((\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0)^T \left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger} \right)^T \right)$$

$$= \rho_c ((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger} (\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0)). \quad (96)$$

The following proof of BOLS in selecting a new block index in each step is similar to that in Section V-A [14]. Finally, we complete the proof. ∎

APPENDIX H
PROOF OF THEOREM 4

Proof: By using submultiplicativity, we have

$$\rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger} (\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0) \right)$$

$$= \rho_c \left(\left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}} \right)^{-1} (\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0 \right)$$

$$\leq \rho_c \left(\left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}} \right)^{-1} \right)$$

$$\times \rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0 \right). \quad (97)$$

Denote the off-diagonal part of $\mathbf{Q}_{0 \setminus \mathbf{S}}^T \mathbf{Q}_{0 \setminus \mathbf{S}}$ as $\mathbf{A}$. Then, due to Lemma 4,

$$\rho_c(\mathbf{A}) \leq \frac{(k - l - \mathcal{T}_B) d \mu_B}{2} \leq \frac{(k - \mathcal{T}_B) d \mu_B}{2}. \quad (98)$$

The result (40) now indicates that $\frac{(k-\mathcal{T}_B) d \mu_B}{2} < 1$ and thus $\rho_c(\mathbf{A}) < 1$, which leading to the following Neumann series expansion,

$$\rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}}^T \mathbf{Q}_{0 \setminus \mathbf{S}})^{-1} \right) = \rho_c \left(\sum_{k=0}^{\infty} (-\mathbf{A})^k \right) \leq \sum_{k=0}^{\infty} \rho_c \left((-\mathbf{A})^k \right)$$

$$\leq \frac{1}{1 - \frac{(k-\mathcal{T}_B) d \mu_B}{2}}. \quad (99)$$

Therefore,

$$\rho_c \left(\left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T \mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}} \right)^{-1} \right)$$

$$\leq \rho_c \left(\mathbf{R}_{0 \setminus \mathbf{S}}^{-1} \right) \rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}}^T \mathbf{Q}_{0 \setminus \mathbf{S}})^{-1} \right) \rho_c \left((\mathbf{R}_{0 \setminus \mathbf{S}}^T)^{-1} \right)$$

$$\leq \rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}}^T \mathbf{Q}_{0 \setminus \mathbf{S}})^{-1} \right)$$

$$\leq \frac{2}{2 - (k - \mathcal{T}_B) d \mu_B}. \quad (100)$$

Note that

$$\rho_c \left((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^T (\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0) \right)$$

$$= \rho_c \left(\mathbf{R}_{0 \setminus \mathbf{S}}^T \mathbf{Q}_{0 \setminus \mathbf{S}}^T \overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0 \right)$$

$$\leq \rho_c \left(\mathbf{R}_{0 \setminus \mathbf{S}}^T \right) \rho_c \left(\mathbf{Q}_{0 \setminus \mathbf{S}}^T \overline{\mathbf{Q}}_0 \right) \rho_c \left(\overline{\mathbf{R}}_0 \right)$$

$$\leq \mathcal{T}_B k d \mu_B. \quad (101)$$

Combining (100) and (101), we have

$$\rho_c((\mathbf{Q}_{0 \setminus \mathbf{S}} \mathbf{R}_{0 \setminus \mathbf{S}})^{\dagger} (\overline{\mathbf{Q}}_0 \overline{\mathbf{R}}_0)) \leq \frac{2 \mathcal{T}_B k d \mu_B}{2 - (k - \mathcal{T}_B) d \mu_B}$$

$$< 1. \quad (102)$$

Simplify (102) and we get a cubic inequality with respect to kd,

$$\alpha_B(kd)^3 + \beta_B(kd)^2 + \gamma_B kd + \delta_B < 0, \qquad (103)$$

where α_B, β_B, γ_B and δ_B are given in the description of the theorem. Finally, the proof is completed by exploiting Cardano formula to calculate the solution of (103) in real domain. ∎

APPENDIX I
PROOF OF THEOREM 5

Proof: We first prove the theorem for OLS. It is clear that a sufficient condition for OLS to select a correct atom in the $(l+1)$-th iteration is

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{r}^l||_\infty > ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{r}^l||_\infty. \qquad (104)$$

By using the residual definition (49), a sufficient condition for (104) is

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty - ||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_\infty$$
$$> ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{s}^l||_\infty + ||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_\infty, \qquad (105)$$

which implies

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty - ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{s}^l||_\infty > 2||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_\infty. \qquad (106)$$

Following the Lemma 4 in [38] and (18), we have

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty - ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{s}^l||_\infty$$
$$\geq (1 - ||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)||_{1,1})||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty. \qquad (107)$$

Combining (106), (107) and Lemma 5, the condition

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty > \frac{2(2 - (K - \mathcal{T})\mu)||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_\infty}{2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu} \qquad (108)$$

ensures that OLS selects a correct atom in the $(l+1)$-th iteration. Note that

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_\infty$$
$$= ||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}_{0\backslash \mathbf{S}}\mathbf{x}_{0\backslash \mathbf{S}}||_\infty$$
$$\geq \frac{||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}_{0\backslash \mathbf{S}}\mathbf{x}_{0\backslash \mathbf{S}}||_2}{\sqrt{K - l}}$$
$$\geq \frac{(1 - (K - 1)\mu)||\mathbf{x}_{0\backslash \mathbf{S}}||_2}{\sqrt{K - l}}, \qquad (109)$$

where the last inequality follows from Lemmas 2 and 5 in [38]. Combining (108) and (109) yields

$$||\mathbf{x}_{0\backslash \mathbf{S}}||_2 > \frac{2\sqrt{K - l}(2 - (K - \mathcal{T})\mu)||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_\infty}{(2 - (K - \mathcal{T})\mu - 2K\mathcal{T}\mu)(1 - (K - 1)\mu)}. \qquad (110)$$

Then, using Lemma 5.1 in [46], we obtain that (51) guarantees OLS for selecting a correct atom in the $(l+1)$-th iteration with the probability at least $1 - \frac{1}{M}$.

For MOLS algorithm, the sufficient condition for selecting at least one correct atom is

$$L||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{r}^l||_\infty > ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{r}^l||_{L,1}$$
$$\geq L||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{r}^l||_\infty, \qquad (111)$$

where $||\cdot||_{L,1}$ is defined in Appendix A. Then, by eliminating L on both sides of the inequality (111), the proof is completed for MOLS by using the similar procedures as described above. ∎

APPENDIX J
PROOF OF THEOREM 6

Proof: A sufficient condition for BOLS to select a correct block in the $(l+1)$-th iteration is

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{r}^l||_{2,\infty} > ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{r}^l||_{2,\infty}. \qquad (112)$$

Similar to the proof in Appendix I,

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_{2,\infty} - ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{s}^l||_{2,\infty}$$
$$> 2||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_{2,\infty} \qquad (113)$$

is the sufficient condition for (112). By using Lemma 4 in [38] and (37), we have

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_{2,\infty} - ||(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)^T \mathbf{s}^l||_{2,\infty}$$
$$\geq (1 - \rho_c((\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^\dagger(\overline{\mathbf{Q}}_0\overline{\mathbf{R}}_0)))||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_{2,\infty}. \qquad (114)$$

Combining (113), (114) and Lemma 6, the condition

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_{2,\infty} > \frac{2(2 - (k - \mathcal{T}_B)d\mu_B)||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_{2,\infty}}{2 - (k - \mathcal{T}_B)d\mu_B - 2\mathcal{T}_B kd\mu_B} \qquad (115)$$

ensures that BOLS selects a correct block in the $(l+1)$-th iteration. Observe that

$$||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T \mathbf{s}^l||_{2,\infty}$$
$$= ||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}_{0\backslash \mathbf{S}}\mathbf{x}_{0\backslash \mathbf{S}}||_{2,\infty}$$
$$\geq \frac{||(\mathbf{Q}_{0\backslash \mathbf{S}}\mathbf{R}_{0\backslash \mathbf{S}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{S}^l})\mathbf{D}_{0\backslash \mathbf{S}}\mathbf{x}_{0\backslash \mathbf{S}}||_2}{\sqrt{k - l}}$$
$$\geq \frac{(1 - (kd - 1)\mu)||\mathbf{x}_{0\backslash \mathbf{S}}||_2}{\sqrt{k - l}}. \qquad (116)$$

Combining (115) and (116) yields

$$||\mathbf{x}_{0\backslash \mathbf{S}}||_2 > \frac{2\sqrt{k-l}(2 - (k - \mathcal{T}_B)d\mu_B)||(\mathbf{Q}\mathbf{R})^T \mathbf{n}^l||_{2,\infty}}{(2 - (k - \mathcal{T}_B)d\mu_B - 2\mathcal{T}_B kd\mu_B)(1 - (kd - 1)\mu)}. \qquad (117)$$
∎

Finally, based on Lemma 7, the proof is completed.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès et al., "Compressive sampling," in *Proc. Int. Congr. Mathematicians*, Madrid, Spain, 2006, pp. 1433–1452.

[3] M. Rani, S. Dhok, and R. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.

[4] Y. Wang and Z. Tian, "Big data in 5G," in *Encyclopedia of Wireless Networks*. New York, NY, USA: Springer, 2018. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-32903-1_58-1

[5] Y. Wang, Z. Tian, and X. Cheng, "Enabling technologies for spectrum and energy efficient NOMA-mmWave-maMIMO systems," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 53–59, Oct. 2020.

[6] L. Rencker, F. Bach, W. Wang, and M. D. Plumbley, "Sparse recovery and dictionary learning from nonlinear compressive measurements," *IEEE Trans. Signal Process.*, vol. 67, no. 21, pp. 5659–5670, Nov. 2019.

[7] J. Wen, Z. Zhou, Z. Liu, M.-J. Lai, and X. Tang, "Sharp sufficient conditions for stable recovery of block sparse signals by block orthogonal matching pursuit," *Appl. Comput. Harmon. Anal.*, vol. 47, no. 3, pp. 948–974, Nov. 2019.

[8] S. Daei, F. Haddadi, and A. Amini, "Exploiting prior information in block-sparse signals," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 5093–5102, Oct. 2019.

[9] Y. Wang, Z. Tian, C. Feng, S. Feng, and P. Zhang, "Performance analysis of generalized block diagonal structured random matrices in compressive sensing," in *Proc. Int. Symp. Commun. Inf. Technol.*, 2012, pp. 793–797.

[10] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[11] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9/10, pp. 589–592, May 2008.

[12] G. Li, J. Yan, and Y. Gu, "Information theoretic lower bound of restricted isometry property constant," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, U.K., 2019, pp. 5297–5301.

[13] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[14] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[15] J. Wen, J. Wang, and Q. Zhang, "Nearly optimal bounds for orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5347–5356, Oct. 2017.

[16] J. Wang and P. Li, "Recovery of sparse signals using multiple orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2049–2062, Apr. 2017.

[17] W. Xu, Q. Liu, Y. Wang, and X. Bian, "Performance limits of one-bit compressive classification," *Signal Process.*, vol. 178, Jan. 2021, Art. no. 107808.

[18] W. Xu, J. Lin, K. Niu, Z. He, and Y. Wang, "Performance analysis of partial support recovery and signal reconstruction of compressed sensing," *IET Signal Proc.*, vol. 8, no. 2, pp. 188–201, Apr. 2014.

[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Aug. 2001.

[20] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with prior information: Strategies, geometry, and bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4472–4496, Jul. 2017.

[21] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[22] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[23] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[24] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[25] Z. Li, W. Xu, Y. Tian, Y. Wang, and J. Lin, "Compressed sensing reconstruction algorithms with prior information: Logit weight simultaneous orthogonal matching pursuit," in *Proc. IEEE Veh. Technol. Conf.*, 2014, pp. 1–5.

[26] W. Xu, Z. Li, Y. Tian, Y. Wang, and J. Lin, "Perturbation analysis of simultaneous orthogonal matching pursuit," *Signal Process.*, vol. 116, pp. 91–100, Nov. 2015.

[27] J. Wen, R. Zhang, and W. Yu, "Signal-dependent performance analysis of orthogonal matching pursuit for exact sparse recovery," *IEEE Trans. Signal Process.*, vol. 68, pp. 5031–5046, 2020.

[28] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit: The Gaussian case," Caltech, Pasadena, CA, 2007, ACM Tech. Rep. 2007-01, 2007.

[29] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.

[30] A. Majumdar and R. K. Ward, "Fast group sparse classification," *Can. J. Elect. Comput. Eng.*, vol. 34, no. 4, pp. 136–144, 2009.

[31] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[32] Y. C. Eldar and M. Mishali, "Block sparsity and sampling over a union of subspaces," in *Proc. IEEE Int. Conf. Digit. Signal Process.*, Santorini, Greece, 2009, pp. 1–8.

[33] J. Kim, J. Wang, and B. Shim, "Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 69, pp. 1521–1536, Feb. 2021.

[34] J. Kim and B. Shim, "A near-optimal restricted isometry condition of multiple orthogonal least squares," *IEEE Access*, vol. 7, pp. 46822–46830, 2019.

[35] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 255–261, Jan. 2006.

[36] Y. C. Eldar, "Uncertainty relations for shift-invariant analog signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5742–5757, Dec. 2009.

[37] C. Herzet, A. Drmeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 459–470, Jan. 2016.

[38] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[39] M. Stojnic, "ℓ_2/ℓ_1-optimization in block-sparse compressed sensing and its strong thresholds," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 350–357, Apr. 2010.

[40] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.

[41] S. Jiang and Y. Gu, "Block-sparsity-induced adaptive filter for multi-clustering system identification," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5318–5330, Oct. 2015.

[42] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

[43] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.

[44] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2116–2125, Jun. 2012.

[45] Y. Luo, J. Dang, and Z. Song, "Optimal compressive spectrum sensing based on sparsity order estimation in wideband cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12094–12106, Dec. 2019.

[46] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, Jul. 2009.

[47] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[48] T. T. Cai, "On block thresholding in wavelet regression: Adaptivity, block size, and threshold level," *Statist. Sinica*, vol. 12, no. 4, pp. 1241–1273, Oct. 2002.

Liyang Lu (Graduate Student Member, IEEE) received the B.S. degree in communication engineering in 2017, from the Beijing University of Posts and Telecommunications, Beijing, China, where he is currently working toward the Ph.D. degree in information and communication engineering. His research interests include compressive sensing, cognitive radios, integrated sensing and communication, sparse representation-based classification and signal optimization.

Wenbo Xu (Member, IEEE) received the B.S. degree from the School of Information Engineering and the Ph.D. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2005 and 2010, respectively. Since 2010, she has been with BUPT, where she is currently the Professor with the School of Artificial Intelligence. Her research interests include sparse signal processing, machine learning, and signal processing in wireless networks.

Yue Wang (Senior Member, IEEE) received the Ph.D. degree in communication and information system from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently a Research Assistant Professor with Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA, where he was a Postdoctoral Researcher. Prior to that, he was a Senior Engineer with Huawei Technologies Company, Ltd., Shenzhen, China. From 2009 to 2011, he was the Visiting Ph.D. Student with Electrical and Computer Engineering Department, Michigan Technological University, Houghton, MI, USA. His general interests include signal processing, wireless communications, machine learning, and their applications in cyber physical systems. His current research focuses on compressive sensing, massive MIMO, millimeter-wave communications, cognitive radios, DoA estimation, high-dimensional data analysis, and distributed optimization and learning.

Zhi Tian (Fellow, IEEE) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 1994, and the M.S. and Ph.D. degrees from George Mason University, Fairfax, VA, USA, in 1998 and 2000, respectively. She is currently the Professor with the Electrical and Computer Engineering Department of George Mason University, Fairfax, VA, USA, since 2015. Prior to that, she was on the Faculty of Michigan Technological University, Houghton, MI, USA, from 2000 to 2014. She was the Program Director with the U.S. National Science Foundation from 2012 to 2014. Her general interests include the areas of statistical signal processing, wireless communications, and estimation and detection theory. Her research focuses on compressed sensing for random processes, statistical inference of network data, distributed network optimization and learning, and millimeter-wave communications. She was an IEEE Distinguished Lecturer for IEEE Communications Society and IEEE Vehicular Technology Society. She was the Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. She was the recipient of the IEEE Communications Society TCCN Publication Award in 2018. She was the Chair of the IEEE Signal Processing Society Big Data Special Interest Group, General Co-Chair of the 2016 IEEE GlobalSIP Conference, and is the Unclassified Technical Program Co-Chair of the 2022 IEEE MILCOM Conference. She was a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society (2019–2021).