Recovery of Sparse Signals Using Multiple Orthogonal Least Squares

Jian Wang, *Student Member, IEEE*, and Ping Li

***Abstract*—Sparse recovery aims to reconstruct sparse signals from compressed linear measurements. In this paper, we propose a sparse recovery algorithm called multiple orthogonal least squares (MOLS), which extends the well-known orthogonal least squares (OLS) algorithm by allowing multiple L indices to be selected per iteration. Owing to its ability to catch multiple support indices in each selection, MOLS often converges in fewer iterations and hence improves the computational efficiency over the conventional OLS algorithm. Theoretical analysis shows that MOLS ($L > 1$) performs exact recovery of K-sparse signals ($K > 1$) in at most K iterations, provided that the sensing matrix obeys the restricted isometry property with isometry constant $\delta_{LK} < \sqrt{L}/(\sqrt{K} + 2\sqrt{L})$. When $L = 1$, MOLS reduces to the conventional OLS algorithm and our analysis shows that exact recovery is guaranteed under $\delta_{K+1} < 1/(\sqrt{K} + 2)$. This condition is nearly optimal with respect to δ_{K+1} in the sense that, even with a small relaxation (e.g., $\delta_{K+1} = 1/\sqrt{K}$), exact recovery with OLS may not be guaranteed. The recovery performance of MOLS in the noisy scenario is also studied. It is shown that stable recovery of sparse signals can be achieved with the MOLS algorithm when the signal-to-noise ratio scales linearly with the sparsity level of input signals.**

***Index Terms*—Compressed sensing (CS), sparse recovery, orthogonal matching pursuit (OMP), orthogonal least squares (OLS), multiple orthogonal least squares (MOLS), restricted isometry property (RIP), signal-to-noise ratio (SNR).**

I. INTRODUCTION

IN RECENT years, sparse recovery has attracted much attention in applied mathematics, electrical engineering, and statistics [1]–[5]. The main task of sparse recovery is to recover a high dimensional K-sparse vector $\mathbf{x} \in \mathcal{R}^n$ ($\|\mathbf{x}\|_0 \leq K \ll n$) from a small number of linear measurements

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ ($m < n$) is often called the sensing matrix. Although the system is underdetermined, owing to the signal sparsity, $\mathbf{x}$ can be accurately recovered from the measurements $\mathbf{y}$ by solving an ℓ_0-minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x}. \tag{2}$$

This method, however, is known to be intractable due to the combinatorial search involved and therefore impractical for realistic applications. Thus, much attention has been focused on developing efficient algorithms for recovering the sparse signal. In general, the algorithms can be classified into two major categories: those using convex optimization techniques [1]–[6] and those based on greedy searching principles [7]–[15]. Other algorithms relying on nonconvex methods have also been proposed [16]–[20]. The optimization-based approaches replace the nonconvex ℓ_0-norm with its convex surrogate ℓ_1-norm, translating the combinatorial hard search into a computationally tractable problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{ subject to } \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x}. \tag{3}$$

This algorithm is known as basis pursuit (BP) [1]. It has been revealed that under appropriate constraints on the sensing matrix, BP yields exact recovery of the sparse signal. Here and throughout the paper, exact recovery means accurate reconstruction of all coefficients of the underlying signal.[1]

The second category of approaches for sparse recovery is greedy method, in which signal support is iteratively identified according to various greedy principles. Due to its computational simplicity and competitive performance, greedy method has gained considerable popularity in practical applications. Representative algorithms include matching pursuit (MP) [8], orthogonal matching pursuit (OMP) [7], [21]–[28] and orthogonal least squares (OLS) [9], [29]–[32]. In a nutshell, both OMP and OLS identify the support of the underlying sparse signal by adding one index to the list at a time, and estimate the sparse coefficients over the enlarged support. The main difference between OMP and OLS lies in the greedy rule of updating the support at each iteration. While OMP chooses a column that is most strongly correlated with the signal residual, OLS seeks a candidate that gives the most significant decrease in the residual power. It has been shown that OLS has better convergence property but is computationally more expensive than the OMP algorithm [31].

In this paper, with the aim of improving the recovery accuracy and also reducing the computational cost of OLS, we propose a new method called multiple orthogonal least squares (MOLS). Our method can be viewed as an extension of the OLS

[1]Note that our definition of exact recovery differs from that in some other literature (e.g., [21]) which refers to exact reconstruction of the support of $\mathbf{x}$.

Manuscript received July 9, 2016; revised November 4, 2016; accepted November 21, 2016. Date of publication December 14, 2016; date of current version February 7, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wenwu Wang. The research was supported primarily by NSF-III-1360971, NSF-Bigdata-1419210, ONRN00014-13-1-0764, and AFOSR-FA9550-13-1-0137. During the revision, the work of Jian Wang was also supported in part by NRF-2014R1A5A1011478.

J. Wang was with the Department of Statistics and Department of Computer Science, Rutgers University, Piscataway, NJ, 08854, USA. He is currently with the Department of Electrical & Computer Engineering, Seoul National University Korea (e-mail: wangjianeee@gmail.com).

P. Li is with the Department of Statistics and Department of Computer Science, Rutgers University, Piscataway, NJ, 08854, USA (e-mail: pingli@stat.rutgers.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2639467

TABLE I
THE MOLS ALGORITHM

Input	sensing matrix $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$, measurements vector $\mathbf{y} \in \mathcal{R}^m$, sparsity level K, residual tolerant ϵ, and selection parameter $L \leq \min\{K, \frac{m}{K}\}$.
Initialize	iteration count $k = 0$, estimated support $\mathcal{T}^0 = \emptyset$, and residual vector $\mathbf{r}^0 = \mathbf{y}$.
While	$\|\mathbf{r}^k\|_2 > \epsilon$ and $k < K$, **do**
	$k = k + 1$.
	Identify $\mathcal{S}^k = \underset{\mathcal{S}:\lvert\mathcal{S}\rvert=L}{\arg\min} \sum_{i \in \mathcal{S}} \|\mathbf{P}^{\perp}_{\mathcal{T}^{k-1}\cup\{i\}}\mathbf{y}\|_2^2$.
	Enlarge $\mathcal{T}^k = \mathcal{T}^{k-1} \cup \mathcal{S}^k$.
	Estimate $\mathbf{x}^k = \underset{\mathbf{u}:\mathrm{supp}(\mathbf{u})=\mathcal{T}^k}{\arg\min} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$.
	Update $\mathbf{r}^k = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^k$.
End	
Output	the estimated support $\hat{\mathcal{T}} = \underset{\mathcal{S}:\lvert\mathcal{S}\rvert=K}{\arg\min} \|\mathbf{x}^k - \mathbf{x}^k_{\mathcal{S}}\|_2$ and the estimated signal $\hat{\mathbf{x}}$ satisfying $\hat{\mathbf{x}}_{\Omega\setminus\hat{\mathcal{T}}} = \mathbf{0}$ and $\hat{\mathbf{x}}_{\hat{\mathcal{T}}} = \mathbf{\Phi}^{\dagger}_{\hat{\mathcal{T}}}\mathbf{y}$.

algorithm in that multiple indices are allowed to be chosen at a time. It is inspired by that some sub-optimal candidates in each identification of OLS are likely to be reliable and could be utilized to better reduce the power of signal residual at each iteration, thereby accelerating the convergence of the algorithm. The main steps of the MOLS algorithm are specified in Table I. Owing to selection of multiple "good" candidates in each time, MOLS often converges in fewer iterations and improves the computational efficiency over the conventional OLS algorithm.

Greedy methods with a similar flavor to MOLS in adding multiple indices per iteration include stagewise OMP (StOMP) [10], regularized OMP (ROMP) [11], and generalized OMP (gOMP) [12], [33] (also known as orthogonal super greedy algorithm (OSGA) [34]), etc. These algorithms identify candidates at each iteration according to correlations between columns of the sensing matrix and the residual vector. Specifically, StOMP picks indices whose magnitudes of correlation exceed a deliberately designed threshold. ROMP first chooses a set of K indices with strongest correlations and then narrows down the candidates to a subset based on a predefined regularization rule. The gOMP algorithm finds a fixed number of indices with strongest correlations in each selection. Other well-known greedy methods adopting a different strategy of adding as well as pruning indices from the list include compressive sampling matching pursuit (CoSaMP) [13] and subspace pursuit (SP) [14] and hard thresholding pursuit (HTP) [15], etc.

The contributions of this paper are summarized as follows.

i) We propose a new algorithm, referred to as MOLS, for solving sparse recovery problems. We analyze the MOLS algorithm using the restricted isometry property (RIP) introduced in the compressed sensing (CS) theory [6] (see Definition 1 below). Our analysis shows that MOLS ($L > 1$) exactly recovers any K-sparse signal ($K > 1$) within K iterations if the sensing matrix $\mathbf{\Phi}$ obeys the RIP with isometry constant

$$\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K} + 2\sqrt{L}}. \tag{4}$$

When $L = 1$, MOLS reduces to the conventional OLS algorithm. We establish the condition for OLS as

$$\delta_{K+1} < \frac{1}{\sqrt{K} + 2}. \tag{5}$$

This condition is nearly sharp in the sense that, even with a slight relaxation (e.g., relaxing to $\delta_{K+1} = \frac{1}{\sqrt{K}}$), exact recovery of K-sparse signals may not be ensured.

ii) We also analyze the recovery performance of MOLS in the presence of noise. Our result demonstrates that MOLS can achieve stable recovery of sparse signals when the signal-to-noise ratio (SNR) scales linearly with the sparsity level of input signals. In particular, for the case of OLS (i.e., when $L = 1$), we show that the linear scaling law of the SNR is necessary for exactly recovering the support of the underlying signals of interest. We stress that our analysis for MOLS is strongly motivated from [12], [27], [35], where similar results for the OMP and gOMP algorithms were established. The connections will be discussed in detail in Sections III and IV-A.

The rest of this paper is organized as follows: In Section II, we introduce notations, definitions, and lemmas that are used in this paper. In Section III, we give a useful observation regarding the identification step of MOLS and analyze the theoretical performance of MOLS in recovering sparse signals. In Section IV, we present proofs for our theoretical results. In Section V, we study the empirical performance of the MOLS algorithm and concluding remarks are given in Section VI.

II. PRELIMINARIES

A. Notations

We briefly summarize notations in this paper. Let $\Omega = \{1, \cdots, n\}$ and let $\mathcal{T} = \mathrm{supp}(\mathbf{x}) = \{i | i \in \Omega, x_i \neq 0\}$ denote the support of vector $\mathbf{x}$. For $\mathcal{S} \subseteq \Omega$, $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. $\mathcal{T} \setminus \mathcal{S}$ is the set of all elements contained in $\mathcal{T}$ but not in $\mathcal{S}$. $\mathbf{x}_{\mathcal{S}} \in \mathcal{R}^{|\mathcal{S}|}$ is the restriction of the vector $\mathbf{x}$ to the elements with indices in $\mathcal{S}$. Throughout the paper we assume that $\mathbf{\Phi}$ is normalized to have unit column norm (i.e., $\|\phi_i\|_2 = 1$ for $i = 1, \cdots, n$).[2] $\mathbf{\Phi}_{\mathcal{S}} \in \mathcal{R}^{m \times |\mathcal{S}|}$ is a submatrix of $\mathbf{\Phi}$ that only contains columns indexed by $\mathcal{S}$. If $\mathbf{\Phi}_{\mathcal{S}}$ has full column rank, then $\mathbf{\Phi}^{\dagger}_{\mathcal{S}} = (\mathbf{\Phi}'_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}})^{-1}\mathbf{\Phi}'_{\mathcal{S}}$ is the pseudoinverse of $\mathbf{\Phi}_{\mathcal{S}}$ where $\mathbf{\Phi}'_{\mathcal{S}}$ denotes the transpose of $\mathbf{\Phi}_{\mathcal{S}}$. $\mathrm{span}(\mathbf{\Phi}_{\mathcal{S}})$ is the span of columns in $\mathbf{\Phi}_{\mathcal{S}}$. $\mathbf{P}_{\mathcal{S}} = \mathbf{\Phi}_{\mathcal{S}}\mathbf{\Phi}^{\dagger}_{\mathcal{S}}$ is the projection onto $\mathrm{span}(\mathbf{\Phi}_{\mathcal{S}})$. $\mathbf{P}^{\perp}_{\mathcal{S}} = \mathbf{I} - \mathbf{P}_{\mathcal{S}}$ is the projection onto the orthogonal complement of $\mathrm{span}(\mathbf{\Phi}_{\mathcal{S}})$ where $\mathbf{I}$ is the identity matrix.

B. Definitions and Lemmas

Definition 1 (RIP [6]): A matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a constant $\delta \in (0, 1)$ such that

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \tag{6}$$

for all K-sparse vectors $\mathbf{x}$. In particular, the minimum of all constants δ satisfying (6) is called the isometry constant δ_K.

The following lemmas are useful for our analysis.

Lemma 1 (Monotonicity [6]): If a matrix satisfies the RIP of both orders K_1 and K_2 where $K_1 \leq K_2$, then $\delta_{K_1} \leq \delta_{K_2}$.

Lemma 2 (Consequences of RIP [13]): Let $\mathcal{S} \subseteq \Omega$. If $\delta_{|\mathcal{S}|} < 1$, then for any vector $\mathbf{u} \in \mathcal{R}^{|\mathcal{S}|}$,

$$(1 - \delta_{|\mathcal{S}|})\,\|\mathbf{u}\|_2 \leq \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}}\mathbf{u}\|_2 \leq (1 + \delta_{|\mathcal{S}|})\,\|\mathbf{u}\|_2\,,$$

$$(1 - \delta_{|\mathcal{S}|})\|(\mathbf{\Phi}'_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}})^{-1}\mathbf{u}\|_2 \leq \|\mathbf{u}\|_2 \leq (1 + \delta_{|\mathcal{S}|})\|(\mathbf{\Phi}'_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}})^{-1}\mathbf{u}\|_2.$$

[2] [36] showed that the behavior of OLS is unchanged whether columns of $\mathbf{\Phi}$ are normalized or not. As MOLS is a direct extension of OLS, one can verify that the normalization does not matter either for the behavior of MOLS.

Lemma 3 (Lemma 5 in [37]): Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \Omega$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$. Then the eigenvalues of $\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{P}^{\perp}_{\mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1}$ satisfy

$$\lambda_{\min}(\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{P}^{\perp}_{\mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1}) \geq \lambda_{\min}(\mathbf{\Phi}'_{\mathcal{S}_1 \cup \mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1 \cup \mathcal{S}_2}),$$
$$\lambda_{\max}(\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{P}^{\perp}_{\mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1}) \leq \lambda_{\max}(\mathbf{\Phi}'_{\mathcal{S}_1 \cup \mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1 \cup \mathcal{S}_2}).$$

Lemma 4 (Proposition 3.1 in [13]): Let $\mathcal{S} \subset \Omega$. If $\delta_{|\mathcal{S}|} < 1$, then for any vector $\mathbf{u} \in \mathcal{R}^m$, $\|\mathbf{\Phi}'_{\mathcal{S}} \mathbf{u}\|_2 \leq \sqrt{1 + \delta_{|\mathcal{S}|}} \|\mathbf{u}\|_2$.

Lemma 5 (Lemma 2.1 in [5]): Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \Omega$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$. If $\delta_{|\mathcal{S}_1| + |\mathcal{S}_2|} < 1$, then $\|\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{\Phi} \mathbf{v}\|_2 \leq \delta_{|\mathcal{S}_1| + |\mathcal{S}_2|} \|\mathbf{v}\|_2$ holds for any vector $\mathbf{v} \in \mathcal{R}^n$ supported on $\mathcal{S}_2$.

Lemma 6: Let $\mathcal{S} \subseteq \Omega$. If $\delta_{|\mathcal{S}|} < 1$, then for any $\mathbf{u} \in \mathcal{R}^{|\mathcal{S}|}$,

$$\sqrt{1 - \delta_{|\mathcal{S}|}} \|(\mathbf{\Phi}^{\dagger}_{\mathcal{S}})' \mathbf{u}\|_2 \leq \|\mathbf{u}\|_2 \leq \sqrt{1 + \delta_{|\mathcal{S}|}} \|(\mathbf{\Phi}^{\dagger}_{\mathcal{S}})' \mathbf{u}\|_2. \quad (7)$$

The upper bound in (7) has appeared in [13, Proposition 3.1] and we give a proof for the lower bound in Appendix A.

III. Sparse Recovery with MOLS

A. Observation

We begin with an important observation on the identification step of MOLS. As shown in Table I, at the $(k+1)$-th iteration $(k \geq 0)$, MOLS adds to $\mathcal{T}^k$ a set of L indices,

$$\mathcal{S}^{k+1} = \arg \min_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i \in \mathcal{S}} \|\mathbf{P}^{\perp}_{\mathcal{T}^k \cup \{i\}} \mathbf{y}\|_2^2. \quad (8)$$

Intuitively, a straightforward implementation requires to sort all elements in $\{\|\mathbf{P}^{\perp}_{\mathcal{T}^k \cup \{i\}} \mathbf{y}\|_2^2\}_{i \in \Omega \setminus \mathcal{T}^k}$ and then find the smallest L ones (and their corresponding indices). This implementation, however, is computationally expensive as it requires to construct $n - Lk$ different orthogonal projections (i.e., $\mathbf{P}^{\perp}_{\mathcal{T}^k \cup \{i\}}, \forall i \in \Omega \setminus \mathcal{T}^k$). It is thus desirable to find a cost-effective alternative to (8) for the identification step of MOLS.

Interestingly, the following proposition illustrates that (8) can be substantially simplified. It is inspired by the technical report of Blumensath and Davies [36], which gave a geometric interpretation of OLS in terms of orthogonal projections.

Proposition 1: At the $(k+1)$-th iteration, the MOLS algorithm identifies a set of L indices:

$$\mathcal{S}^{k+1} = \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i \in \mathcal{S}} \frac{|\langle \phi_i, \mathbf{r}^k \rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2} \quad (9)$$

$$= \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i \in \mathcal{S}} \left| \left\langle \frac{\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2}, \mathbf{r}^k \right\rangle \right|. \quad (10)$$

One can interpret from (9) that to identify $\mathcal{S}^{k+1}$, it suffices to find the largest L values in $\{\frac{|\langle \phi_i, \mathbf{r}^k \rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2}\}_{i \in \Omega \setminus \mathcal{T}^k}$, which is much simpler than (8), as it involves only one projection operator (i.e., $\mathbf{P}^{\perp}_{\mathcal{T}^k}$). Indeed, by numerical experiments, we have confirmed that the simplification offers massive reduction in the computational cost. The proof of this proposition is essentially identical to some analysis in [29], which is related to OLS, but with extension to the case of selecting multiple indices per iteration. This extension is crucial in that not only does it enable a low-complexity implementation for MOLS, it also plays a key role in the analysis of MOLS in Section IV. We thus include the proof (in Appendix B) for completeness.

Following the arguments in [31], [36], we give a geometric interpretation of the selection rule in MOLS: the columns of sensing matrix are projected onto the subspace that is orthogonal to the span of active columns, and the L normalized projected columns that are best correlated with the residual vector are selected. Thus, the greedy selection rule in MOLS can also be viewed as an extension of the gOMP rule.

B. Main Results

In this subsection, we study the recovery performance of MOLS under the RIP framework. We first provide the condition of MOLS for exact recovery of sparse signals.

Theorem 1: Let $\mathbf{x} \in \mathcal{R}^n$ be any K-sparse signal and let $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ be the sensing matrix with unit ℓ_2-norm columns. Also, let $L \leq \min\{K, \frac{m}{K}\}$ be the number of indices selected at each iteration of MOLS. Then if $\mathbf{\Phi}$ satisfies the RIP with

$$\begin{cases} \delta_{LK} < \frac{\sqrt{L}}{\sqrt{K} + 2\sqrt{L}}, & L > 1, \\ \delta_{K+1} < \frac{1}{\sqrt{K} + 2}, & L = 1, \end{cases} \quad (11)$$

MOLS exactly recovers $\mathbf{x}$ from $\mathbf{y} = \mathbf{\Phi x}$ in at most K steps.

Note that MOLS reduces to the conventional OLS algorithm when $L = 1$. Theorem 1 suggests that OLS recovers any K-sparse signal in exact K iterations under $\delta_{K+1} < \frac{1}{\sqrt{K}+2}$. Similar results have also been established for the OMP algorithm. For example, it has been shown in [24], [25] that $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$ is sufficient for OMP to exactly recover K-sparse signals. The condition is recently improved to $\delta_{K+1} < \frac{\sqrt{4K+1}-1}{2K}$ [38] (by utilizing techniques developed in [39]) and further to $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ [41].

It is worth mentioning that there exist K-sparse signals and measurement matrices having columns of equal length and satisfying $\delta_{K+1} = \frac{1}{\sqrt{K}}$, for which OMP makes a wrong selection at the first iteration and thus fails to recover the signals in K iterations (see, e.g., [25, Example 1]). Since OLS coincides with OMP in the first iteration when columns of $\mathbf{\Phi}$ have the same length [31], this counterexample immediately applies to OLS, which therefore implies that

$$\delta_{K+1} < \frac{1}{\sqrt{K}} \quad (12)$$

is also necessary for OLS. The claim that $\delta_{K+1} < \frac{1}{\sqrt{K}}$ being necessary for exact recovery with K iterations of OLS has also been proved in [42, Lemma 1]. As $\frac{1}{\sqrt{K}+2} \approx \frac{1}{\sqrt{K}}$ for large K, the proposed bound $\delta_{K+1} < \frac{1}{\sqrt{K}+2}$ is nearly sharp. We would like to mention that although $\delta_{K+1} < \frac{1}{\sqrt{K}}$ places a fundamental limit on exact recovery with OLS in K iterations, if OLS is allowed to perform more than K steps, there is still significant room to improve the condition. This case has been studied thoroughly in [30] under the name pure OMP (POMP).

Next, we consider the model where the measurements are contaminated with noise as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}, \quad (13)$$

where $\mathbf{v}$ is the noise. (13) is more applicable and includes the noise-free model (1) as a special case when $\mathbf{v} = \mathbf{0}$. Note that when $\mathbf{v} \neq \mathbf{0}$, accurate recovery of all coefficients of $\mathbf{x}$ is impossible. Thus we use the ℓ_2-norm distortion ($\|\mathbf{x} - \hat{\mathbf{x}}\|_2$) as a performance measure and will derive an upper bound for it.

As detailed in Table I, MOLS runs until $\|\mathbf{r}^k\|_2 \le \epsilon$ or $k = K$. When the iteration terminates, if the number of selected indices exceeds K, these candidates are pruned to a subset of K indices (i.e., $\hat{\mathcal{T}}$) according to K most significant coordinates in $\mathbf{x}^k$. Finally, a standard least squares (LS) procedure is solved on $\hat{\mathcal{T}}$, yielding a K-sparse signal estimate (i.e., $\hat{\mathbf{x}}$). This step is also known as debiasing, which is helpful for lowering the reconstruction error. The following theorem provides an upper bound on $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$ when the algorithm is terminated by $\|\mathbf{r}^k\|_2 \le \epsilon$ where $k \le K$.

Theorem 2: Consider the measurement model in (13). If MOLS satisfies $\|\mathbf{r}^{\bar{k}}\|_2 \le \epsilon$ for some $\bar{k} \le K$ and $\mathbf{\Phi}$ obeys the RIP of order $\max\{L\bar{k} + K, 2K\}$, then the output $\hat{\mathbf{x}}$ satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \le \frac{2\epsilon\sqrt{1+\delta_{2K}} + 2(\sqrt{1+\delta_{2K}} + \sqrt{1-\delta_{L\bar{k}+K}})\|\mathbf{v}\|_2}{\sqrt{(1-\delta_{L\bar{k}+K})(1-\delta_{2K})}}. \quad (14)$$

In particular, when $\epsilon = \|\mathbf{v}\|_2$, $\hat{\mathbf{x}}$ satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \le \frac{(4\sqrt{1+\delta_{2K}} + 2\sqrt{1-\delta_{L\bar{k}+K}})\|\mathbf{v}\|_2}{\sqrt{(1-\delta_{L\bar{k}+K})(1-\delta_{2K})}}. \quad (15)$$

Proof: See Appendix C. ∎

From Theorem 2, we can see that when $\|\mathbf{v}\|_2 = 0$, $\epsilon = 0$ directly implies $\mathbf{x} = \hat{\mathbf{x}}$, Thus the stopping rule of MOLS in the noise-free case can simply be $\|\mathbf{r}^k\|_2 = 0$; furthermore, the pruning operation in Table I may become unnecessary for this case because the current signal estimate is already K-sparse, although more indices may have been selected. To be specific, suppose $\|\mathbf{r}^{\bar{k}}\|_2 = 0$ for some $\bar{k} \le K$. Then by Table I we have $\mathbf{y} = \mathbf{\Phi}\mathbf{x}^{\bar{k}}$, which implies that $\mathbf{\Phi}(\mathbf{x} - \mathbf{x}^{\bar{k}}) = \mathbf{0}$, where $\mathbf{x} - \mathbf{x}^{\bar{k}}$ is an $(L\bar{k} + K)$-sparse vector. Since $\delta_{L\bar{k}+K} < 1$ ensures arbitrary $L\bar{k} + K$ columns of $\mathbf{\Phi}$ to be linearly independent, by basic linear algebra theory we have $\mathbf{x}^{\bar{k}} = \mathbf{x}$. On the other hand, since MOLS iterates at most K times, if $\|\mathbf{r}^k\|_2 \ne 0$ for $k = 1, \cdots, K$ we can infer that $\mathbf{x}^K \ne \mathbf{x}$, which also implies that the algorithm fails to catch all support indices.

While the residual tolerance ϵ of MOLS can be set to zero for the noise-free scenario, such is not the case when noise is present. When $\|\mathbf{v}\|_2 \ne 0$, ϵ should be chosen properly to avoid late or early termination of the algorithm. Note that late termination makes some portion of the noise be treated as signal (over-fitting), while early termination often leads to signal that is not fully reconstructed (under-fitting). In both cases, the reconstruction quality is degraded. Such issue is typically among sparse recovery algorithms. To achieve proper termination under noise, current methods generally assume that the noise is either bounded or Gaussian with known variance and then use information about the noise to establish a stopping criterion (see, e.g., [4], [43]). It is also commonly assumed that knowledge about the signal sparsity is known in prior in the algorithm design/analysis [13]–[15]. In practice, neither of the two information is strictly easier to determine than the other. If one of them is available, it might be possible to estimate the other via, e.g., cross-validation [44], [45].

In Theorem 2, we have analyzed the case where $\|\mathbf{r}^k\|_2 \le \epsilon$ is satisfied when $k \le K$. We now consider the remaining case that $\|\mathbf{r}^k\|_2 \le \epsilon$ is not met for $k \le K$. Clearly if ϵ is set too small, then the residual tolerance may be difficult to meet by increasing the number of iterations, which therefore indicates the need of setting a maximum iteration number for MOLS (in order to prevent over-fitting and save computations), In this paper, we let MOLS execute at most K iterations. This, however, raises the question of whether a reliable recovery can indeed be achieved if the algorithm is forced to stop after K iterations. The following theorem aims to answer this question; to this end, we build a condition under which MOLS successfully selects all support indices by iterating at most K times. In our analysis, we parameterize the dependence on noise $\mathbf{v}$ and signal $\mathbf{x}$ with two quantities: i) the SNR and ii) the minimum-to-average ratio (MAR) [46], which are defined, respectively, as below,

$$snr := \frac{\|\mathbf{\Phi}\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2} \text{ and } \kappa := \frac{\min_{j \in \mathcal{T}} |x_j|}{\|\mathbf{x}\|_2/\sqrt{K}}. \quad (16)$$

Theorem 3: Consider the measurement model in (13) and let $L \le \min\{K, \frac{m}{K}\}$ be the number of indices selected at each iteration of MOLS. Suppose $\|\mathbf{r}^k\|_2 \le \epsilon$ is not met for $k \le K$. Then if the sensing matrix $\mathbf{\Phi}$ has unit ℓ_2-norm columns satisfying (11) and the SNR satisfies

$$\begin{cases} \sqrt{snr} \ge \frac{3(1+\delta_{K+1})}{\kappa(1-(\sqrt{K}+2)\delta_{K+1})}\sqrt{K}, & L = 1, \\ \sqrt{snr} \ge \frac{(2\sqrt{L}+1)(1+\delta_{LK})}{\kappa(\sqrt{L}-(\sqrt{K}+2\sqrt{L})\delta_{LK})}\sqrt{K}, & L > 1, \end{cases} \quad (17)$$

MOLS chooses all support indices in at most K iterations.

One can interpret from Theorem 3 that MOLS catches all support indices of $\mathbf{x}$ in at most K iterations when the SNR scales linearly with the sparsity K. In particular, for the special case of $L = 1$, the algorithm returns the true support of $\mathbf{x}$ (i.e., $\hat{\mathcal{T}} = \mathcal{T}^K = \mathcal{T}$). We would like to point out that the SNR being proportional to K is also necessary for exactly recovering $\mathcal{T}$ with OLS. Indeed, there exist a sensing matrix $\mathbf{\Phi}$ satisfying (11) and a K-sparse signal, for which the OLS algorithm fails to recover the support of the signal under

$$snr \ge K. \quad (18)$$

Specifically, OLS is not guaranteed to make a correct selection at the first iteration. See [27, Theorem 3.2] for more details.[3] Moreover, we would like to mention that the factor κ in (17) may be close to zero when there are strong variations of magnitudes in $\mathbf{x}_{\mathcal{T}}$. In the case, (17) would hardly be satisfied, which in turn implies that it may be difficult to identify those nonzero elements whose magnitudes are very small.

The result in Theorem 3 is closely related to the results in [38], [47]. There, the researchers considered the OMP algorithm with data-driven stopping rules (i.e., residual based stopping rules), and established conditions for exact support recovery that depend on the minimum magnitude of nonzero elements of input signals. It can be shown that the results of [38], [47] essentially require a same scaling law of the SNR as the result in Theorem 3.

Our proof of Theorem 3 extends the proof technique in [12, Theorem 3.4 and 3.5] by considering the measurement noise. We mention that [12, Theorem 4.2] also provided a noisy case analysis based on ℓ_2-norm distortion of the signal recovery with gOMP, but the corresponding result is inferior. Indeed, the result in [12] suggested a recovery distortion upper bounded by $\mathcal{O}(\sqrt{K})\|\mathbf{v}\|_2$; whereas, the following theorem shows that (under similar RIP conditions to [12]) the recovery distortion with MOLS is at most proportional to the noise power.

[3] [27, Theorem 3.2] concerns the first step of OMP, but it immediately applies to OLS, since OLS coincides with OMP in the first step [31].

Corollary 1: Consider the measurement model in (13) and suppose $\|\mathbf{r}^k\|_2 \leq \epsilon$ is not met for $k \leq K$. Then if the sensing matrix $\mathbf{\Phi}$ satisfies the RIP of order LK and the SNR satisfies (17), the output $\hat{\mathbf{x}}$ of MOLS satisfies

$$\begin{cases} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{\|\mathbf{v}\|_2}{\sqrt{1-\delta_K}}, & L = 1, \\ \|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \left(1 + \sqrt{\frac{1+\delta_{2K}}{1-\delta_{LK}}}\right) \frac{2\|\mathbf{v}\|_2}{\sqrt{1-\delta_{2K}}}, & L > 1. \end{cases} \quad (19)$$

Proof: See Appendix D. ∎

We can gain good insights by studying the rate of convergence of MOLS. The next theorem demonstrates that in the noise-free case, the residual power of MOLS decays exponentially with the number of iterations.[4] The proof is left to Appendix E. It is essentially inspired by [48], where exponential convergence was established for MP and OMP.

Theorem 4: Consider the measurement model in (1). Let $\mathbf{\Phi}$ be the sensing matrix with unit ℓ_2-norm columns. If $\mathbf{\Phi}$ satisfies the RIP of order $Lk + K$ and $\delta_{Lk+1} < \frac{\sqrt{5}-1}{2}$, then the residual of MOLS satisfies

$$\|\mathbf{r}^{k+1}\|_2^2 \leq (\alpha(k, L))^{k+1} \|\mathbf{y}\|_2^2, \quad 0 \leq k < K, \quad (20)$$

where

$$\alpha(k, L) := 1 - \frac{L(1 - \delta_{Lk+K})^2}{K(1 + \delta_L)(1 + \delta_{Lk+K})} \left(1 - \frac{\delta_{Lk+1}^2}{1 - \delta_{LK}}\right).$$

IV. PROOFS

A. Main Idea

This section is dedicated to the proofs of our main results in Section III, We focus on the proof for Theorem 3, as Theorem 1 deduces from Theorem 3.

The proof of Theorem 3 follows along a similar line as the analysis in [27, Theorem 3.1], in which the condition for exact support recovery with OMP under measurement noise was established in terms of the SNR and MAR. Specifically, the proof works by mathematical induction. For convenience of stating the results, we say that MOLS makes a success at an iteration if it selects at least one correct index at the iteration. We first derive a condition for success at the first iteration. Then we assume that MOLS has been successful in previous k $(1 \leq k < K)$ iterations and derive a condition for it to make a success also at the $(k+1)$-th iteration. Clearly under the two conditions, MOLS succeeds in every iteration and hence will select all true indices in at most K iterations.

The proof of Theorem 3 owes a lot to the work by Li *et al.* [35]. In particular, some ideas for proving Proposition 2 are inspired from [35, Eq. (25), (26)], which offered similar bounds for the gOMP algorithm. Note that MOLS differs with gOMP in the identification step. In Proposition 1, the difference is precisely characterized by the denominator term (i.e., $\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2$) on the right-hand side of (10), for which there is no counterpart in gOMP. In fact, the existence of the denominator term results in a looser bound of v_L for the MOLS algorithm (see (29) in Proposition 2) and also makes the subsequent proof for Theorem 3 more complex than the gOMP case [35]. Moreover, we would like to mention that in the proof of Proposition 2, the way we lower bound $\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2$ is motivated from [30, Theorem 2].

B. Proof of Theorem 3

1) Success at the first iteration: From (9), MOLS selects at the first iteration the index set

$$\begin{aligned} \mathcal{T}^1 &= \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i \in \mathcal{S}} \frac{|\langle \phi_i, \mathbf{r}^0 \rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^0} \phi_i\|_2} \\ &\overset{(a)}{=} \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}} \mathbf{y}\|_1 = \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}} \mathbf{y}\|_2, \quad (21) \end{aligned}$$

where (a) is because $\mathcal{T}^0 = \emptyset$ so that $\|\mathbf{P}^{\perp}_{\mathcal{T}^0} \phi_i\|_2 = \|\phi_i\|_2 = 1$.

By noting that $L \leq K$,[5]

$$\begin{aligned} \|\mathbf{\Phi}'_{\mathcal{T}^1} \mathbf{y}\|_2 &\geq \sqrt{\frac{L}{K}} \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{y}\|_2 \overset{(13)}{=} \sqrt{\frac{L}{K}} \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{\Phi} \mathbf{x} + \mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \\ &\overset{(a)}{\geq} \sqrt{\frac{L}{K}} \left(\|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}}\|_2 - \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2\right) \\ &\overset{\text{Lemma 2,4}}{\geq} \sqrt{\frac{L}{K}} \Big((1-\delta_K)\|\mathbf{x}\|_2 - \sqrt{1+\delta_K}\|\mathbf{v}\|_2\Big), \quad (22) \end{aligned}$$

where (a) is from the triangle inequality.

On the other hand, if no correct index is chosen at the first iteration (i.e., $\mathcal{T}^1 \cap \mathcal{T} = \emptyset$), then

$$\begin{aligned} \|\mathbf{\Phi}'_{\mathcal{T}^1} \mathbf{y}\|_2 &= \|\mathbf{\Phi}'_{\mathcal{T}^1} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}} + \mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \\ &\leq \|\mathbf{\Phi}'_{\mathcal{T}^1} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}}\|_2 + \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \\ &\overset{\text{Lemma 5,4}}{\leq} \delta_{K+L} \|\mathbf{x}\|_2 + \sqrt{1+\delta_K}\|\mathbf{v}\|_2. \end{aligned}$$

This, however, contradicts (22) if

$$\begin{aligned} &\delta_{K+L}\|\mathbf{x}\|_2 + \sqrt{1+\delta_K}\|\mathbf{v}\|_2 \\ &< \sqrt{\frac{L}{K}} \Big((1-\delta_K)\|\mathbf{x}\|_2 - \sqrt{1+\delta_K}\|\mathbf{v}\|_2\Big), \quad (23) \end{aligned}$$

Since $\delta_K \leq \delta_{K+L}$ by monotonicity of the isometry constant, one can show that (23) holds true whenever

$$\left((1-\delta_{K+L})\sqrt{\frac{L}{K}} - \delta_{K+L}\right) \frac{\|\mathbf{x}\|_2}{\|\mathbf{v}\|_2} > \left(1 + \sqrt{\frac{L}{K}}\right)\sqrt{1+\delta_{K+L}}. \quad (24)$$

Furthermore, note that

$$\frac{\|\mathbf{x}\|_2}{\|\mathbf{v}\|_2} \overset{\text{RIP}}{\geq} \frac{\|\mathbf{\Phi}\mathbf{x}\|_2}{\sqrt{1+\delta_K}\|\mathbf{v}\|_2} \overset{(16)}{=} \frac{\sqrt{snr}}{\sqrt{1+\delta_K}} \overset{\text{Lemma 1}}{\geq} \frac{\sqrt{snr}}{\sqrt{1+\delta_{K+L}}}. \quad (25)$$

Hence, (24) is guaranteed by

$$\left((1-\delta_{K+L})\sqrt{\frac{L}{K}} - \delta_{K+L}\right)\sqrt{snr} > \left(1 + \sqrt{\frac{L}{K}}\right)(1+\delta_{K+L}),$$

which, under $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$, can be rewritten as

$$\sqrt{snr} > \frac{(1+\delta_{K+L})(\sqrt{K}+\sqrt{L})}{\sqrt{L} - (\sqrt{K}+\sqrt{L})\delta_{K+L}}. \quad (26)$$

[4]Note that the proof for the convergence property of MOLS is valid for the noise-free case only. Whether similar result can be established for the noisy case is an interesting open question.

[5]The average of L largest elements in $\{\langle \phi_i, \mathbf{y} \rangle^2\}_{i \in \Omega}$ must be no less than that of any other subset of $\{\langle \phi_i, \mathbf{y} \rangle^2\}_{i \in \Omega}$ whose cardinality is no less than L.

Therefore, under $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$ and (26), at least one correct index is chosen at the first iteration of MOLS.

2) Success at the $(k+1)$-th iteration: We assume that MOLS selects at least one correct index at each of the previous k $(1 \le k < K)$ iterations and denote by ℓ the number of correct indices in $\mathcal{T}^k$. Then,

$$\ell = |\mathcal{T} \cap \mathcal{T}^k| \ge k. \tag{27}$$

Also, we assume that $\mathcal{T}^k$ does not contain all correct indices $(\ell < K)$. Under these assumptions, we derive a condition that ensures MOLS to select at least one correct index at the $(k+1)$-th iteration. We first introduce two quantities that are useful for stating results. Let u_1 denote the largest value of $\frac{|\langle \phi_i, \mathbf{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2}$, $i \in \mathcal{T}$ and let v_L denote the L-th largest value of $\frac{|\langle \phi_i, \mathbf{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2}$, $i \in \Omega \setminus (\mathcal{T} \cup \mathcal{T}^k)$. It is clear that if

$$u_1 > v_L,$$

u_1 belongs to the set of L largest elements amongst all elements in $\{\frac{|\langle \phi_i, \mathbf{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2}\}_{i\in\Omega\setminus\mathcal{T}^k}$. Then it follows from (9) that at least one correct index (i.e., the one corresponding to u_1) will be selected at the $(k+1)$-th iteration. The following results are a lower bound for u_1 and an upper bound for v_L.

Proposition 2: Let $\mathbf{\Phi}$ have normalized columns and satisfy the RIP with $\delta_{LK} < \frac{\sqrt{5}-1}{2}$. Then, we have

$$u_1 \ge \frac{(1-\delta_{K+Lk-\ell})\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 - \sqrt{1+\delta_{K+Lk-\ell}}\|\mathbf{v}\|_2}{\sqrt{K-\ell}}, \tag{28}$$

$$v_L \le \frac{1}{\sqrt{L}}\left(\left(\delta_{L+K-\ell} + \frac{\delta_{L+Lk}\delta_{Lk+K-\ell}}{1-\delta_{Lk}}\right)\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 + \sqrt{1+\delta_{L+Lk}}\|\mathbf{v}\|_2\right)\left(1 + \frac{\delta^2_{Lk+1}}{1-\delta_{Lk}-\delta^2_{Lk+1}}\right)^{1/2}. \tag{29}$$

The proof is given in Appendix F. Since $1 \le k \le \ell < K$ and $1 \le L \le K$, we obtain from Lemma 1 that

$$\begin{aligned} K-\ell < LK &\Rightarrow \delta_{K-\ell} \le \delta_{LK}, \\ Lk+K-\ell \le LK &\Rightarrow \delta_{Lk+K-\ell} \le \delta_{LK}, \\ Lk < LK &\Rightarrow \delta_{LK} \le \delta_{LK}, \\ Lk+1 \le LK &\Rightarrow \delta_{Lk+1} \le \delta_{LK}, \\ L+Lk \le LK &\Rightarrow \delta_{L+Lk} \le \delta_{LK}. \end{aligned} \tag{30}$$

From (29) and (30), we have

$$\begin{aligned} v_L \le\ & \frac{1}{\sqrt{L}}\left(1 + \frac{\delta^2_{LK}}{1-\delta_{LK}-\delta^2_{LK}}\right)^{1/2} \\ & \times\left(\left(\delta_{LK} + \frac{\delta^2_{LK}}{1-\delta_{LK}}\right)\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 + \sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2\right) \\ =\ & \frac{1}{\sqrt{L}}\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta^2_{LK}}\right)^{1/2}\left(\frac{\delta_{LK}\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2}{1-\delta_{LK}} + \sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2\right). \end{aligned} \tag{31}$$

Also, using (28) and (30), we have

$$u_1 \ge \frac{1}{\sqrt{K-\ell}}\left((1-\delta_{LK})\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 - \sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2\right). \tag{32}$$

Thus, by (31) and (32), $u_1 > v_L$ is guaranteed if

$$\begin{aligned} & \frac{1}{\sqrt{K-\ell}}\left((1-\delta_{LK})\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 - \sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2\right) \\ & > \frac{1}{\sqrt{L}}\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta^2_{LK}}\right)^{1/2}\left(\frac{\delta_{LK}\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2}{1-\delta_{LK}} + \sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2\right). \end{aligned} \tag{33}$$

One can show that when $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ is satisfied, (33) holds true under (see Appendix G)

$$\sqrt{snr} \ge \frac{(2\sqrt{L}+1)(1+\delta_{LK})}{\kappa(\sqrt{L}-(\sqrt{K}+2\sqrt{L})\delta_{LK})}\sqrt{K}, \tag{34}$$

Therefore, under $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ and (34), MOLS selects at least one correct index at the $(k+1)$-th iteration.

3) Overall condition: Thus far we have obtained conditions $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$ and (26) for success of MOLS at the first iteration and conditions $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ and (34) for success of the general iteration. We now combine them to get an overall condition ensuring the success of MOLS in every iteration. The following two cases need to be considered.

i) $L \ge 2$: Since $K \ge L$ (see Table I), $L \ge 2$ implies $\delta_{LK} \ge \delta_{K+L}$. By noting that $\frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}} > \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$, $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ is more restrictive than $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$. Also, note that (34) is more restrictive than (26). Therefore, under $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ and (34), MOLS ensures selection of all support indices in at most K steps.

ii) $L = 1$: In this case, MOLS reduces to OLS and conditions $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$ and $\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}$ become, respectively: $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$ and $\delta_K < \frac{1}{\sqrt{K}+2}$, both of which are ensured by $\delta_{K+1} < \frac{1}{\sqrt{K}+2}$. Moreover, since $\delta_{K+1} \ge \delta_K$, both (26) and (34) hold true under

$$\sqrt{snr} \ge \frac{3(1+\delta_{K+1})}{\kappa(1-(\sqrt{K}+2)\delta_{K+1})}\sqrt{K}. \tag{35}$$

Therefore, $\delta_{K+1} < \frac{1}{\sqrt{K}+2}$ and (35) ensure selection of all support indices in K iterations of OLS.

The proof is now complete.

V. Empirical Results

A. Experimental Setup

We empirically study the performance of MOLS in recovering sparse signals for both noise-free and noisy scenarios. In the noise-free case, we adopt the testing strategy in [12], [14], which measures the performance of recovery algorithms by testing their empirical frequency of exact reconstruction of sparse signals. In the noisy case, we use the mean square error (MSE) as a metric to evaluate the recovery performance. In each trial, we construct an $m \times n$ matrix (where $m = 128$ and $n = 256$)

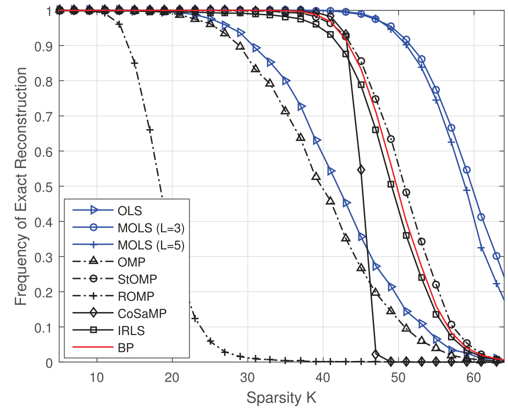
with each element an independent and identically distributed (i.i.d.) draw of a Gaussian distribution, with zero mean and $\frac{1}{m}$ variance. For each value of K in $\{5, \cdots, 64\}$, we generate a K-sparse signal of size $n \times 1$, whose support is chosen uniformly at random and nonzero elements are 1) drawn independently from a standard Gaussian distribution, or 2) chosen randomly from the set $\{\pm 1\}$. We refer to the two types of signals as the sparse Gaussian signal and the sparse 2-ary pulse amplitude modulation (2-PAM) signal, respectively. For comparative purposes, our simulation includes the following recovery approaches:[6]

1) OMP and OLS;
2) MOLS (https://sites.google.com/site/jianwanghomepage);
3) StOMP (http://sparselab.stanford.edu/);
4) ROMP (http://www.cmc.edu/pages/faculty/DNeedell);
5) CoSaMP (http://www.cmc.edu/pages/faculty/DNeedell);
6) BP (or BPDN for the noisy case) (http://cvxr.com/cvx/);
7) Iterative reweighted LS (IRLS);
8) Linear minimum MSE (LMMSE) estimator.

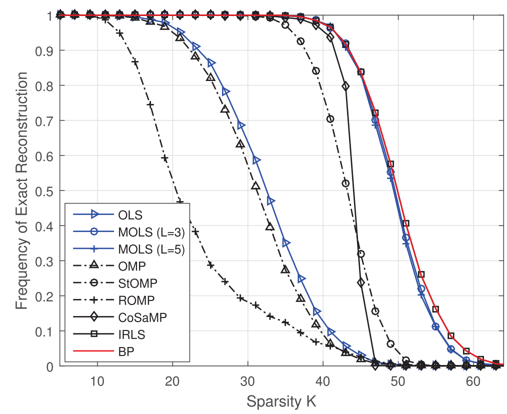
B. The Noise-Free Case

In the noise-free case, we perform 2,000 independent trials for each recovery approach and plot the empirical frequency of exact reconstruction as a function of the sparsity level. By comparing the maximal sparsity level, i.e., the so called *critical sparsity* [14], of sparse signals at which exact reconstruction is always ensured, recovery accuracy of different algorithms can be compared empirically. As shown in Fig. 1, for both sparse Gaussian and sparse 2-PAM signals, MOLS outperforms other greedy approaches with respect to the critical sparsity. Even when compared to the BP and IRLS methods, MOLS still exhibits competitive reconstruction performance. For the Gaussian case, the critical sparsity of MOLS is 43, which is higher than that of BP and IRLS, while for the 2-PAM case, MOLS, BP and IRLS have almost identical critical sparsity (around 37). A key reason for the superior performance of MOLS is that MOLS actually estimates a support of greater size (up to KL), which is thus more likely to contain all support indices. Note that as long as all support indices are included, those selected incorrect indices do not affect the recovery, because the LS projection yields exact recovery of $\mathbf{x}$:[7] Fig. 2 depicts the support detection performance of MOLS in recovering sparse signals. To measure the support detection performance, we adopt the false alarm ratio expressed by $|\hat{\mathcal{T}} \backslash \mathcal{T}|/K$, which is equivalent to the miss-detection ratio since $|\hat{\mathcal{T}}| = |\mathcal{T}| = K$ (and thus $|\mathcal{T} \backslash \hat{\mathcal{T}}| = |\hat{\mathcal{T}} \backslash \mathcal{T}|$). We observe that the false alarm ratio goes up as the sparse level increases. In particular, MOLS performs better in recovering sparse Gaussian signals than sparse 2-PAM signals, which matches the results in Fig. 1.

In Fig. 4, we plot the running time and the number of iterations for exact reconstruction of K-sparse Gaussian and 2-PAM signals as a function of K. The running time is measured using the MATLAB program under the 28-core 64-bit processor, 256 Gb RAM, and Windows Server 2012 R2 environments.



(a) Sparse Gaussian signal



(b) Sparse 2-PAM signal

Fig. 1. Frequency of exact recovery of sparse signals as a function of K.

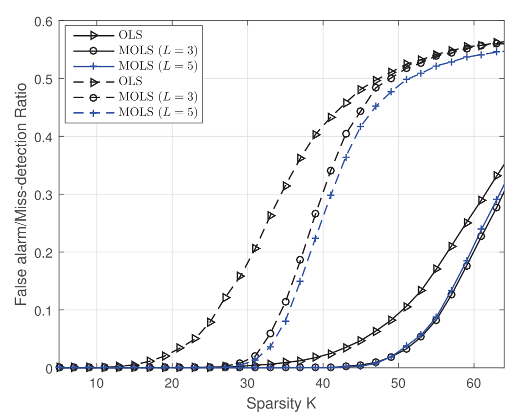


Fig. 2. False alarm/Miss-detection ration for recovering sparse signals as a function of K, where the solid and dash lines correspond to sparse Gaussian and sparse 2-PAM signals, respectively.

[6]In our test, both OMP and OLS run K iterations before stopping. For MOLS, since $L \leq K$, we choose $L = 3, 5$ in our simulation. Interested readers may try other options. We suggest to choose L to be small integers and have empirically confirmed that choices of $L = 2, 3, 4, 5$ generally lead to similar recovery performance. Moreover, ϵ is set to be equal to the noise power ($\epsilon = \|\mathbf{v}\|_2$) when $\mathbf{v} \neq \mathbf{0}$ and zero when $\mathbf{v} = \mathbf{0}$. StOMP has two thresholding strategies: false alarm control (FAC) and false discovery control (FDC) [10]. We exclusively use the FAC strategy since it outperforms FDC. For CoSaMP, we set the maximal iteration number to 50 to avoid repeated iterations. In addition, we implement IRLS (with $p = 1$) as featured in [17], and terminate BPDN when the ℓ_2-norm of residual falls below the noise power, as in [43].

[7]Suppose $\mathcal{T} \subseteq \mathcal{T}^{\bar{k}}$ for some $\bar{k} \leq K$. Since $L \leq \min\{K, \frac{m}{K}\}$, the number of selected columns satisfies $L\bar{k} \leq m$. For random matrices, those selected columns are linearly independent with probability one, which allows us to recover $\mathbf{x}$ via an LS procedure: $\mathbf{x}^{\bar{k}}_{\mathcal{T}^{\bar{k}}} = \arg\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{\Phi}_{\mathcal{T}^{\bar{k}}}\mathbf{u}\|_2 = \mathbf{\Phi}^{\dagger}_{\mathcal{T}^{\bar{k}}}\mathbf{y} = \mathbf{\Phi}^{\dagger}_{\mathcal{T}^{\bar{k}}}\mathbf{\Phi}_{\mathcal{T}^{\bar{k}}}\mathbf{x}_{\mathcal{T}^{\bar{k}}} = \mathbf{x}_{\mathcal{T}^{\bar{k}}}$, That is, $\mathbf{x}^{\bar{k}} = \mathbf{x}$. We note that the LS procedure is also involved in other OMP-type algorithms under test. We all use matrix left division of Matlab to solve them in the simulation.

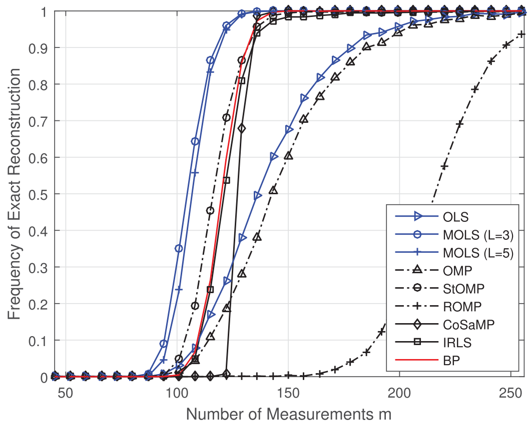
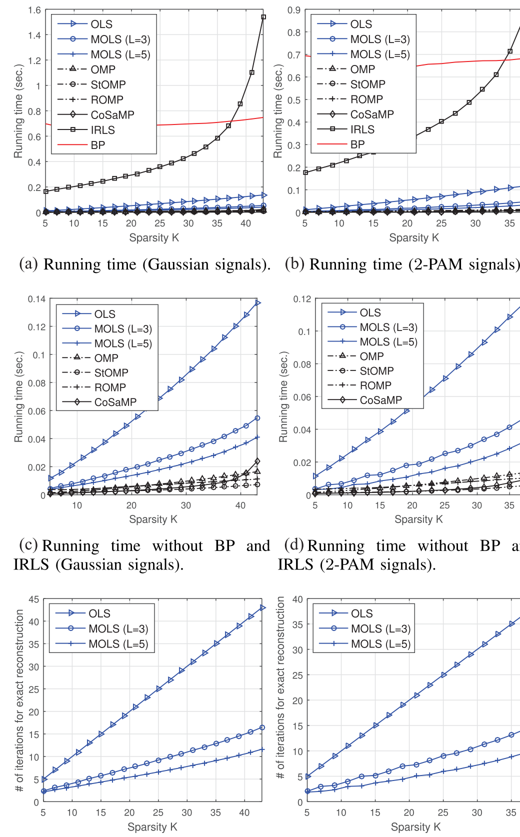


Fig. 3. Frequency of exact recovery of sparse signals as a function of m.

Overall, we observe that for both sparse Gaussian and 2-PAM cases, the running time of BP and IRLS is longer than that of OMP, CoSaMP, StOMP and MOLS. In particular, the running time of BP is more than one order of magnitude higher than the rest of algorithms require. This is because the complexity of BP is quadratic in the number of measurements ($\mathcal{O}(m^2 n^{3/2})$) [49], while that of greedy algorithms is $\mathcal{O}(Kmn)$ [10]–[13], [32].[8] Moreover, the running time of MOLS is two to three times as much as that of OMP. We also observe that the number of iterations of MOLS for exact reconstruction is smaller than that of OLS. The associated running time of MOLS is also less than that of OLS. This is because MOLS can select more than one support index at a time so that it may take much fewer than K iterations to catch all support indices (or to satisfy $\|\mathbf{r}^k\|_2 = 0$). It should be noted that in some cases, MOLS may be required to run K iterations before stopping. For example, if MOLS selects only one correct indices per time, then it would need K iterations to catch all support indices. Even MOLS may fail to select all support indices in K iterations for some sparsity region (see, e.g., when $K \geq 43$ in Fig. 1(a)). For those cases, however, the complexity and running time of MOLS are higher than OLS, because at the k-th iteration ($k = 1, \cdots, K$) MOLS actually solves a linear system of size Lk, whereas OLS solves a linear system of size k. Nevertheless, these situations do not occur often in the exact recovery sparsity region. Indeed, as observed in Fig. 4, in which the result is averaged over 2,000 independent trials, MOLS has shorter running time over OLS.

In Fig. 3, by varying the number of measurements m, we plot empirical frequency of exact reconstruction of K-sparse Gaussian signals as a function of m. We let sparsity $K = 45$, for which none of the reconstruction methods in Fig. 1(a) are guaranteed to perform exact recovery. Overall, we observe that the performance comparison is similar to Fig. 1(a) in that MOLS performs the best and OLS, OMP and ROMP perform worse than other methods. Moreover, for all recovery methods under test, the frequency of exact reconstruction improves as m increases. In particular, MOLS roughly requires $m \geq 135$ to ensure exact recovery of sparse signals, while BP, CoSaMP and StOMP always succeed when $m \geq 150$.

[8]Similar to OMP, ROMP and StOMP, the MOLS algorithm involves LS projections over a sequentially enlarged support list, which essentially allows for computational savings through recursively using the QR factorization of previously selected columns. See [12, Appendix F] for more details on recursive use of QR factorization.



(a) Running time (Gaussian signals). (b) Running time (2-PAM signals).

(c) Running time without BP and IRLS (Gaussian signals). (d) Running time without BP and IRLS (2-PAM signals).

(e) # of iterations (Gaussian signals). (f) # of iterations (2-PAM signals).

Fig. 4. Running time and number of iterations for exact reconstruction of K-sparse Gaussian and 2-PAM signals.

C. The Noisy Case

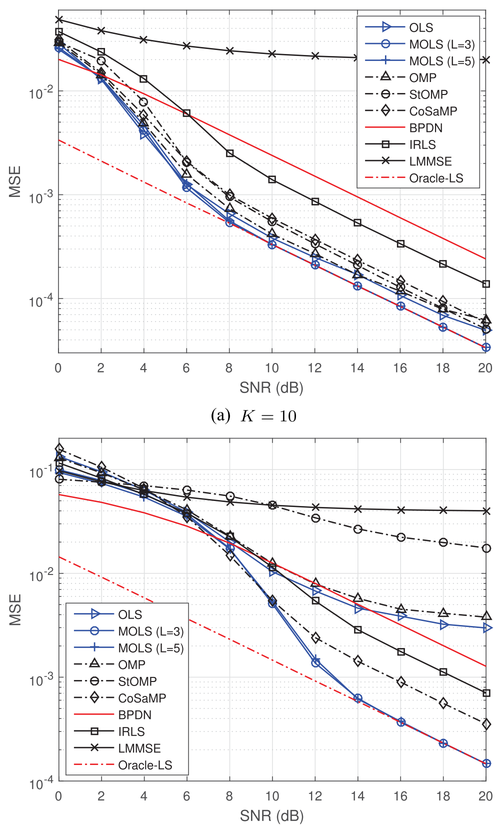
In the noisy case, we empirically compare MSE performance of each recovery method. The MSE is defined as

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2, \tag{36}$$

where $\hat{x}_i$ is the estimate of x_i. For each simulation point of the algorithm, we perform 2,000 independent trials. In Fig. 5, we plot the MSE performance for each recovery method as a function of SNR (in dB) (i.e., $\mathrm{SNR} := 10 \log_{10} snr$). In this case, the system model is expressed as $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$ where $\mathbf{v}$ is the noise vector whose elements are generated from Gaussian distribution $\mathcal{N}(0, \frac{K}{m} 10^{-\frac{\mathrm{SNR}}{10}})$.[9] The benchmark performance of Oracle least squares estimator (Oracle-LS), the best possible estimation having prior knowledge on the support of input signals, is plotted as well.

In general, we observe that for all reconstruction methods, the MSE performance improves with the SNR. For the whole SNR region under test, the MSE performance of MOLS is very competitive. In particular, in the high SNR region the MSE of MOLS matches with that of the Oracle-LS estimator, because MOLS detects all support indices of sparse signals successfully in that SNR region. We should mention that the actual recovery

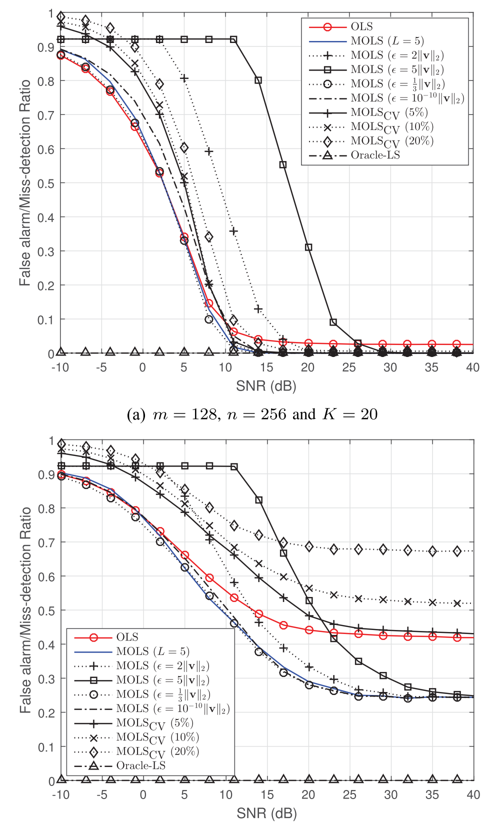
[9]Note that $\mathbb{E}|(\mathbf{\Phi x})_i|^2 = \frac{K}{m}$, as each component of $\mathbf{\Phi}$ has power $\frac{1}{m}$ and signal $\mathbf{x}$ is K-sparse with nonzero elements drawn i.i.d. from $\mathcal{N}(0,1)$. From the definition of SNR, we have $\mathbb{E}|v_i|^2 = \mathbb{E}|(\mathbf{\Phi x})_i|^2 \cdot 10^{-\frac{\mathrm{SNR}}{10}} = \frac{K}{m} 10^{-\frac{\mathrm{SNR}}{10}}$.



(a) $K = 10$

(b) $K = 20$

Fig. 5. MSE for recovering sparse 2-PAM signals as a function of SNR.



(a) $m = 128$, $n = 256$ and $K = 20$

(b) $m = 80$, $n = 256$ and $K = 20$

Fig. 6. False alarm/Miss-detection ratio for recovering sparse 2-PAM signals as a function of SNR where the tasted SNR $\in \{-10, -8, \cdots, 40\}$ (in dB).

error of MOLS may be much smaller than indicated in Theorem 2 and 3. Consider MOLS ($L = 5$) for example. When SNR $= 20$ dB, $K = 20$, and $v_j \sim \mathcal{N}(0, \frac{K}{m} 10^{-\frac{\text{SNR}}{10}})$, we have $\mathbb{E}\|\mathbf{v}\|_2 = (K \cdot 10^{-\frac{\text{SNR}}{10}})^{1/2} = \frac{\sqrt{5}}{5}$. Thus, by assuming small isometry constants we obtain, respectively, from Theorem 2 and 3 that $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{4\sqrt{5}}{5}$ and $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{6\sqrt{5}}{5}$.[10] Whereas, the ℓ_2-norm of the actual recovery error of MOLS is $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 = (n \cdot \text{MSE})^{1/2} \approx 0.2$ (Fig. 5(b)), which is much smaller. The gap between the theoretical and empirical results is perhaps due to i) that our analysis is based on the RIP and hence is essentially a worst-case-analysis and ii) that some inequalities (relaxations) in our analysis may be loose.

D. Discussions

As in [4], [43], knowledge about noise (typically the noise power) is necessary to set the residual tolerance ϵ for MOLS. In practice, however, this knowledge is unknown and may be hard to estimate accurately. If the noise power is under- or over-estimated, then ϵ may be set improperly and consequently, the algorithm may have either an early or late termination. In Fig. 6, by varying ϵ, we empirically study the effect of early/late termination on the support detection performance (i.e., the false alarm/miss-detection ratio) of MOLS. Specifically, we consider the MOLS with $\epsilon = \eta\|\mathbf{v}\|_2$ where $\eta \in \{10^{-10}, \frac{1}{3}, 2, 3\}$. Note that $\eta = 2$ and $\eta = 3$ represent early termination situations, while $\eta = 10^{-10}$ and $\eta = \frac{1}{3}$ lead to late termination. In particular, $\eta = 10^{-10}$ essentially represents the case of $\epsilon = 0$, in which case the algorithm always runs K iterations to stop because the criterion $\|\mathbf{r}^k\|_2 \leq \epsilon$ cannot be met in advance.

We also provide an alternative way to establish the stopping rule when information about noise is completely unavailable. To be specific, instead of setting the residual tolerance, we employ cross-validation (CV) [44] to predict a reasonable termination. In CV, measurements are divided into two distinct sets: a training/estimation set and a test/CV set. The signal estimate resulting in the minimal CV error indicates a proper termination. See [44, Section 3.2] for more details. In our simulation, we use MOLS$_{\text{CV}}$ ($\#\%$) to mean MOLS adopting $\#$ percent of measurements for CV.

In Fig. 6(a), we plot the miss-detection/false alarm ratio as a function of the SNR where the tested SNR region is $\{-10, -8, \cdots, 40\}$ (in dB). We observe that MOLS is very sensitive to early termination. Indeed, the support detection performance degrades significantly when ϵ is set to be twice or thrice the noise power. In contrast, late termination results in only moderate performance degradation. This is because MOLS employs a pruning step (followed by a subsequent step of debiasing), which can largely reduce the potential over-fitting issue caused by selection of incorrect indices. We also observe that MOLS offers superior accuracy in the high SNR region but performs worse than OLS in the low SNR region. The inferior performance of MOLS to OLS in low SNR is mainly due to selection of too many incorrect indices before stopping. Those incorrect indices are supposed to be finally pruned; however,

[10] We can verify condition (17) in Theorem 3. Specifically, since 2-PAM signals have $\kappa = 1$ and snr $= 10^{\frac{\text{SNR}}{10}} = 100$, when $\mathbf{\Phi}$ has small isometry constants, (17) roughly becomes $\sqrt{100} \geq \frac{\sqrt{5}+1}{\sqrt{5}}\sqrt{20}$, which is true.

accuracy of the pruning step is also limited by the high-level noise (i.e., low SNR). In addition, we see that the CV-aided MOLS is effective in the high SNR region. In particular, MOLS_{CV} (5%) and MOLS_{CV} (10%) exhibit promising detection performance that gradually approaches the ideal case as the SNR goes high. Fig. 6(b) depicts the support detection performance of MOLS for correlated sensing matrix (where only 80 random measurements are used). We observe that for the whole SNR range under test, MOLS_{CV} does not perform well, owing to lack of measurements; whereas, using a relatively small ϵ (i.e., letting the algorithm iterate relatively more steps) seems to be a good choice.

VI. Conclusion

In this paper, we have studied a sparse recovery algorithm called MOLS, which extends the conventional OLS algorithm by allowing multiple candidates entering the list in each selection. Our method is motivated by the fact that "sub-optimal" candidates in each of the OLS identification may also be reliable and can be selected to accelerate the convergence of the algorithm. We have demonstrated by RIP analysis that MOLS performs exact recovery of any K-sparse signal in at most K iterations when the isometry constant scales inversely with $\sqrt{K}$. For the special case of MOLS when $L = 1$ (i.e., the OLS case), we have shown that any K-sparse signal can be exactly recovered in K iterations under $\delta_{K+1} < \frac{1}{\sqrt{K+2}}$, which is a nearly optimal condition for the OLS algorithm. We have also studied the MOLS algorithm in the noisy scenario. Our result showed that stable recovery of sparse signals can be achieved with MOLS when the SNR has a linear scaling in the sparsity level of signals to be recovered. In particular, for the case of OLS, there exists a counterexample illustrating that the linear scaling law for the SNR is actually necessary [27].

It is interesting to note the gap between conditions of OLS and OMP (i.e., $\delta_{K+1} < \frac{1}{\sqrt{K+2}}$ and $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ [41]) in the noise-free case. This gap is due to the difference in their identification rules. As indicated in (9), OLS has an extra denominator term (i.e., $\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \phi_i\|_2$) compared to the OMP algorithm. While this term does not affect the first iteration of OLS because $\|\mathbf{P}^{\perp}_{\mathcal{T}^0} \phi_i\|_2 = \|\phi_i\|_2 = 1$ (so that OLS coincides with OMP for this step), it does make a difference in subsequent iterations and ultimately leads to an overall condition for OLS that is more restrictive than the OMP algorithm. Whether it is possible to bridge this gap (i.e., to show that condition $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ [41] also works for the OLS algorithm) is an interesting open question.

Appendix A
Proof of Lemma 6

Proof: We focus on the proof for the lower bound in (7). Since $\delta_{|\mathcal{S}|} < 1$, $\mathbf{\Phi}_{\mathcal{S}}$ has full column rank. Suppose that $\mathbf{\Phi}_{\mathcal{S}}$ has singular value decomposition $\mathbf{\Phi}_{\mathcal{S}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$. Then from the definition of RIP, the minimum diagonal entry of $\mathbf{\Sigma}$ satisfies $\sigma_{\min} \geq \sqrt{1 - \delta_{|\mathcal{S}|}}$. Note that

$$\begin{aligned}(\mathbf{\Phi}^{\dagger}_{\mathcal{S}})' &= ((\mathbf{\Phi}'_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}})^{-1}\mathbf{\Phi}'_{\mathcal{S}})' \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}'((\mathbf{U}\mathbf{\Sigma}\mathbf{V}')'\mathbf{U}\mathbf{\Sigma}\mathbf{V}')^{-1} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}', \quad \text{(A.1)}\end{aligned}$$

where $\mathbf{\Sigma}^{-1}$ is the diagonal matrix formed by replacing nonzero diagonal entries of $\mathbf{\Sigma}$ by their reciprocal. Hence, all singular values of $(\mathbf{\Phi}^{\dagger}_{\mathcal{S}})'$ are upper bounded by $\frac{1}{\sigma_{\min}} \leq \frac{1}{\sqrt{1-\delta_{|\mathcal{S}|}}}$. ■

Appendix B
Proof of Proposition 1

Proof: Since $\mathbf{P}^{\perp}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}$ and $\mathbf{P}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}$ are orthogonal with each other, we have $\|\mathbf{P}^{\perp}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}\|_2^2 = \|\mathbf{y}\|_2^2 - \|\mathbf{P}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}\|_2^2$, and hence (8) is equivalent to

$$\mathcal{S}^{k+1} = \arg \max_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i \in \mathcal{S}} \|\mathbf{P}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}\|_2^2. \quad \text{(B.1)}$$

It is well-known that $\|\mathbf{P}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}\|_2^2$ can be decomposed as follows (see, e.g., [29], [30], [36]):

$$\|\mathbf{P}_{\mathcal{T}^k \cup \{i\}}\mathbf{y}\|_2^2 = \|\mathbf{P}_{\mathcal{T}^k}\mathbf{y}\|_2^2 + \left(\frac{|\langle \phi_i, \mathbf{r}^k \rangle|}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2}\right)^2. \quad \text{(B.2)}$$

By relating (B.1) and (B.2), we obtain (9).

Furthermore, noting from Table I that

$$\mathbf{r}^k = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^k = \mathbf{y} - \mathbf{\Phi}_{\mathcal{T}^k}\mathbf{\Phi}^{\dagger}_{\mathcal{T}^k}\mathbf{y} = \mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{y} \quad \text{(B.3)}$$

where

$$\mathbf{P}^{\perp}_{\mathcal{T}^k} = (\mathbf{P}^{\perp}_{\mathcal{T}^k})^2 = (\mathbf{P}^{\perp}_{\mathcal{T}^k})', \quad \text{(B.4)}$$

if we write $|\langle \phi_i, \mathbf{r}^k \rangle| = |\phi'_i(\mathbf{P}^{\perp}_{\mathcal{T}^k})'\mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{y}| = |\langle \mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i, \mathbf{r}^k \rangle|$ in (9), we obtain (10). ■

Appendix C
Proof of Theorem 2

Proof: Let $\mathbf{z}^{\bar{k}}$ denote the best K-sparse approximation of $\mathbf{x}^{\bar{k}}$. Recalling from Table I that $\hat{\mathcal{T}} = \arg\min_{\mathcal{S}:|\mathcal{S}|=K} \|\mathbf{x}^{\bar{k}} - \mathbf{x}^{\bar{k}}_{\mathcal{S}}\|_2$, we have $\mathbf{z}^{\bar{k}}_{\hat{\mathcal{T}}} = \mathbf{x}^{\bar{k}}_{\hat{\mathcal{T}}}$ and $\mathbf{z}^{\bar{k}}_{\Omega \setminus \hat{\mathcal{T}}} = \mathbf{0}$.

Observe that

$$\begin{aligned}\|\mathbf{z}^{\bar{k}} - \mathbf{x}\|_2 &= \|\mathbf{z}^{\bar{k}} - \mathbf{x}^{\bar{k}} + \mathbf{x}^{\bar{k}} - \mathbf{x}\|_2 \\ &\overset{(a)}{\leq} \|\mathbf{z}^{\bar{k}} - \mathbf{x}^{\bar{k}}\|_2 + \|\mathbf{x}^{\bar{k}} - \mathbf{x}\|_2 \\ &\overset{(b)}{\leq} 2\|\mathbf{x}^{\bar{k}} - \mathbf{x}\|_2 \overset{\text{RIP}}{\leq} \frac{2\|\mathbf{\Phi}(\mathbf{x}^{\bar{k}} - \mathbf{x})\|_2}{\sqrt{1 - \delta_{L\bar{k}+K}}} \\ &\overset{(c)}{\leq} \frac{2(\|\mathbf{r}^{\bar{k}}\|_2 + \|\mathbf{v}\|_2)}{\sqrt{1 - \delta_{L\bar{k}+K}}} \leq \frac{2(\epsilon + \|\mathbf{v}\|_2)}{\sqrt{1 - \delta_{L\bar{k}+K}}}, \quad \text{(C.1)}\end{aligned}$$

where (a) is from the triangle inequality, (b) is because $\mathbf{z}$ is the best K-sparse approximation to $\mathbf{x}^{\bar{k}}$ and hence is a better approximation than $\mathbf{x}$, and (c) is due to the fact that $\mathbf{r}^{\bar{k}} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{\bar{k}} = \mathbf{\Phi}(\mathbf{x}^{\bar{k}} - \mathbf{x}) + \mathbf{v}$.

On the other hand, since $\mathbf{z}^{\bar{k}}$ and $\mathbf{x}$ are both K-sparse,

$$\begin{aligned}
\|\mathbf{z}^{\bar{k}} - \mathbf{x}\|_2 &\overset{\text{RIP}}{\geq} \frac{\|\mathbf{\Phi}(\mathbf{z}^{\bar{k}} - \mathbf{x})\|_2}{\sqrt{1+\delta_{2K}}} = \frac{\|\mathbf{\Phi}\mathbf{z}^{\bar{k}} - \mathbf{y} + \mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}} \\
&\overset{(a)}{\geq} \frac{\|\mathbf{\Phi}\mathbf{z}^{\bar{k}} - \mathbf{y}\|_2 - \|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}} \\
&\overset{(b)}{\geq} \frac{\|\mathbf{\Phi}\hat{\mathbf{x}} - \mathbf{y}\|_2 - \|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}} \\
&= \frac{\|\mathbf{\Phi}(\hat{\mathbf{x}} - \mathbf{x}) - \mathbf{v}\|_2 - \|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}} \\
&\overset{(c)}{\geq} \frac{\|\mathbf{\Phi}(\hat{\mathbf{x}} - \mathbf{x})\|_2 - 2\|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}} \\
&\overset{\text{RIP}}{\geq} \frac{\sqrt{1-\delta_{2K}}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 - 2\|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}}, \qquad \text{(C.2)}
\end{aligned}$$

where (a) and (c) are from the triangle inequality and (b) is because $\mathbf{z}^{\bar{k}}$ is supported on $\hat{\mathcal{T}}$ and $\hat{\mathbf{x}}_{\hat{\mathcal{T}}} = \mathbf{\Phi}^{\dagger}_{\hat{\mathcal{T}}}\mathbf{y} = \arg\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{\Phi}_{\hat{\mathcal{T}}}\mathbf{u}\|_2$ (see Table I). Combining (C.1) and (C.2) yileds the desired result. ■

Appendix D
Proof of Corollary 1

Proof: Let $\bar{k}$ $(\leq K)$ be the number of actually performed iterations of MOLS. We first consider the case $L = 1$. In this case, $\bar{k} = K$. By Theorem 3 we have $\hat{\mathcal{T}} = \mathcal{T}^K = \mathcal{T}$ and $\hat{\mathbf{x}} = \mathbf{x}^K$ (where $\hat{\mathbf{x}}_{\mathcal{T}} = \mathbf{\Phi}^{\dagger}_{\mathcal{T}}\mathbf{y}$ and $\hat{\mathbf{x}}_{\Omega\setminus\mathcal{T}} = \mathbf{0}$). Hence,

$$\begin{aligned}
\|\hat{\mathbf{x}} - \mathbf{x}\|_2 &= \|\mathbf{\Phi}^{\dagger}_{\mathcal{T}}\mathbf{y} - \mathbf{x}_{\mathcal{T}}\|_2 = \|\mathbf{\Phi}^{\dagger}_{\mathcal{T}}\mathbf{v}\|_2 \\
&\overset{\text{RIP}}{\leq} \frac{\|\mathbf{\Phi}_{\mathcal{T}}\mathbf{\Phi}^{\dagger}_{\mathcal{T}}\mathbf{v}\|_2}{\sqrt{1-\delta_K}} = \frac{\|\mathbf{P}_{\mathcal{T}}\mathbf{v}\|_2}{\sqrt{1-\delta_K}} \\
&\leq \frac{\|\mathbf{v}\|_2}{\sqrt{1-\delta_K}}. \qquad \text{(D.1)}
\end{aligned}$$

Next, we prove the case of $L > 1$. In this case, by Theorem 3 we have $\mathcal{T}^K \supseteq \mathcal{T}$. Consider the best K-sparse approximation $\mathbf{z}^K$ of $\mathbf{x}^K$ and observer that

$$\begin{aligned}
\|\mathbf{z}^K - \mathbf{x}\|_2 &= \|\mathbf{z}^K - \mathbf{x}^K + \mathbf{x}^K - \mathbf{x}\|_2 \\
&\overset{(a)}{\leq} \|\mathbf{z}^K - \mathbf{x}^K\|_2 + \|\mathbf{x}^K - \mathbf{x}\|_2 \overset{(b)}{\leq} 2\|\mathbf{x}^K - \mathbf{x}\|_2 \\
&\overset{(c)}{=} 2\|\mathbf{\Phi}^{\dagger}_{\mathcal{T}^K}\mathbf{y} - \mathbf{x}_{\mathcal{T}^K}\|_2 \overset{(d)}{=} 2\|\mathbf{\Phi}^{\dagger}_{\mathcal{T}^K}\mathbf{v}\|_2 \\
&\overset{\text{RIP}}{\leq} \frac{2\|\mathbf{\Phi}_{\mathcal{T}^K}\mathbf{\Phi}^{\dagger}_{\mathcal{T}^K}\mathbf{v}\|_2}{\sqrt{1-\delta_{|\mathcal{T}^K|}}} = \frac{2\|\mathbf{P}_{\mathcal{T}^K}\mathbf{v}\|_2}{\sqrt{1-\delta_{LK}}} \\
&\leq \frac{2\|\mathbf{v}\|_2}{\sqrt{1-\delta_{LK}}}, \qquad \text{(D.2)}
\end{aligned}$$

where (a) uses the triangle inequality, (b) is because $\mathbf{z}^K$ is the best K-sparse approximation to $\mathbf{x}^K$ and hence is a better approximation than $\mathbf{x}$ (note that both $\mathbf{z}^K$ and $\mathbf{x}$ are K-sparse), (c) is because $(\mathbf{x}^K)_{\mathcal{T}^K} = \mathbf{\Phi}^{\dagger}_{\mathcal{T}^K}\mathbf{y}$, $(\mathbf{x}^K)_{\Omega\setminus\mathcal{T}^K} = \mathbf{0}$ and $\mathcal{T}^K \supseteq \mathcal{T}$, and (d) is from $\mathbf{y} = \mathbf{\Phi}_{\mathcal{T}}\mathbf{x}_{\mathcal{T}} + \mathbf{v} = \mathbf{\Phi}_{\mathcal{T}^K}\mathbf{x}_{\mathcal{T}^K} + \mathbf{v}$.

On the other hand, by following (C.2) with $\bar{k} = K$, we have

$$\|\mathbf{z}^K - \mathbf{x}\|_2 \geq \frac{\sqrt{1-\delta_{2K}}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 - 2\|\mathbf{v}\|_2}{\sqrt{1+\delta_{2K}}}. \qquad \text{(D.3)}$$

Combining (D.2) and (D.3) yields the desired result. ■

Appendix E
Proof of Theorem 4

We shall prove Theorem 4 in two steps. First, we show that the residual power difference of MOLS satisfies

$$\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 \geq \frac{1 - \delta_{LK} - \delta_{Lk+1}^2}{(1+\delta_L)(1-\delta_{LK})} \max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2, \qquad \text{(E.1)}$$

which is essentially an extension of the OLS result in [30, Theorem 2]. In the second step, we show that

$$\max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2 \geq \frac{L(1-\delta_{Lk+K})^2}{K(1+\delta_{Lk+K})}\|\mathbf{r}^k\|_2^2. \qquad \text{(E.2)}$$

The theorem is established by combining (E.1) and (E.2).

1) *Proof of (E.1)*: First, for any integer $0 \leq k < K$,

$$\begin{aligned}
\mathbf{r}^k - \mathbf{r}^{k+1} &\overset{\text{(B.3)}}{=} \mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{y} - \mathbf{P}^{\perp}_{\mathcal{T}^{k+1}}\mathbf{y} \\
&\overset{(a)}{=} (\mathbf{P}_{\mathcal{T}^{k+1}} - \mathbf{P}_{\mathcal{T}^{k+1}}\mathbf{P}_{\mathcal{T}^k})\mathbf{y} \\
&= \mathbf{P}_{\mathcal{T}^{k+1}}(\mathbf{y} - \mathbf{P}_{\mathcal{T}^k}\mathbf{y}) = \mathbf{P}_{\mathcal{T}^{k+1}}\mathbf{r}^k, \qquad \text{(E.3)}
\end{aligned}$$

where (a) is because $\text{span}(\mathbf{\Phi}_{\mathcal{T}^k}) \subseteq \text{span}(\mathbf{\Phi}_{\mathcal{T}^{k+1}})$ so that $\mathbf{P}_{\mathcal{T}^k}\mathbf{y} = \mathbf{P}_{\mathcal{T}^{k+1}}(\mathbf{P}_{\mathcal{T}^k}\mathbf{y})$. Since $\mathcal{T}^{k+1} \supseteq \mathcal{S}^{k+1}$, we have $\text{span}(\mathbf{\Phi}_{\mathcal{T}^{k+1}}) \supseteq \text{span}(\mathbf{\Phi}_{\mathcal{S}^{k+1}})$ and hence

$$\|\mathbf{r}^k - \mathbf{r}^{k+1}\|_2^2 \overset{\text{(E.3)}}{=} \|\mathbf{P}_{\mathcal{T}^{k+1}}\mathbf{r}^k\|_2^2 \geq \|\mathbf{P}_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2.$$

Further, since $\|\mathbf{r}^k - \mathbf{r}^{k+1}\|_2^2 = \|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2$,

$$\begin{aligned}
\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 &\geq \|\mathbf{P}_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2 = \|\mathbf{P}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2 \\
&= \|(\mathbf{\Phi}^{\dagger}_{\mathcal{S}^{k+1}})'\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2 \\
&\overset{\text{Lemma 6}}{\geq} \frac{\|\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2}{1+\delta_{|\mathcal{S}^{k+1}|}} = \frac{\|\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2}{1+\delta_L}, \qquad \text{(E.4)}
\end{aligned}$$

Next, we build a lower bound for the term $\|\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2$ in (E.4). Denote $S^* = \arg\max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2$. Then,

$$\begin{aligned}
\sum_{i\in\mathcal{S}^{k+1}} \frac{|\langle\phi_i, \mathbf{r}^k\rangle|^2}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2^2} &\overset{(9)}{=} \max_{\mathcal{S}:|\mathcal{S}|=L} \sum_{i\in\mathcal{S}} \frac{|\langle\phi_i, \mathbf{r}^k\rangle|^2}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2^2} \geq \sum_{i\in\mathcal{S}^*} \frac{|\langle\phi_i, \mathbf{r}^k\rangle|^2}{\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2^2} \\
&\overset{(a)}{\geq} \sum_{i\in\mathcal{S}^*} |\langle\phi_i, \mathbf{r}^k\rangle|^2 = \max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2, \qquad \text{(E.5)}
\end{aligned}$$

where (a) holds because $\|\mathbf{P}^{\perp}_{\mathcal{T}^k}\phi_i\|_2 \leq \|\phi_i\|_2 = 1$.

On the other hand, since $\delta_{Lk+1} < \frac{\sqrt{5}-1}{2}$,

$$\begin{aligned}\sum_{i\in\mathcal{S}^{k+1}} \frac{|\langle\phi_i, \mathbf{r}^k\rangle|^2}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2^2} &\le \frac{\sum_{i\in\mathcal{S}^{k+1}} |\langle\phi_i, \mathbf{r}^k\rangle|^2}{\min_{i\in\mathcal{S}^{k+1}} \|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2^2} \\ &= \frac{\sum_{i\in\mathcal{S}^{k+1}} |\langle\phi_i, \mathbf{r}^k\rangle|^2}{1-\max_{i\in\mathcal{S}^{k+1}} \|\mathbf{P}_{\mathcal{T}^k}\phi_i\|_2^2} \\ &\overset{(a)}{\le} \left(1-\frac{\delta^2_{Lk+1}}{1-\delta_{LK}}\right)^{-1} \|\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2. \quad \text{(E.6)}\end{aligned}$$

where (a) is because for any $i \notin \mathcal{T}^k$,

$$\begin{aligned}\|\mathbf{P}_{\mathcal{T}^k}\phi_i\|_2^2 &= \|\mathbf{P}'_{\mathcal{T}^k}\phi_i\|_2^2 = \|(\mathbf{\Phi}^\dagger_{\mathcal{T}^k})'\mathbf{\Phi}'_{\mathcal{T}^k}\phi_i\|_2^2 \\ &\overset{\text{Lemma 6}}{\le} \frac{\|\mathbf{\Phi}'_{\mathcal{T}^k}\phi_i\|_2^2}{1-\delta_{LK}} \overset{\text{Lemma 5}}{\le} \frac{\delta^2_{Lk+1}}{1-\delta_{LK}}. \quad \text{(E.7)}\end{aligned}$$

Combining (E.5) and (E.6) yields

$$\|\mathbf{\Phi}'_{\mathcal{S}^{k+1}}\mathbf{r}^k\|_2^2 \ge \left(1-\frac{\delta^2_{Lk+1}}{1-\delta_{LK}}\right) \max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2. \quad \text{(E.8)}$$

Finally, using (E.4) and (E.8), we obtain (E.1).

2) Proof of (E.2): Since $L \le K$,

$$\begin{aligned}\max_{\mathcal{S}:|\mathcal{S}|=L} \|\mathbf{\Phi}'_{\mathcal{S}}\mathbf{r}^k\|_2^2 &\ge \frac{L}{K}\|\mathbf{\Phi}'_{\mathcal{T}}\mathbf{r}^k\|_2^2 \overset{(a)}{=} \frac{L}{K}\|\mathbf{\Phi}'_{\mathcal{T}\cup\mathcal{T}^k}\mathbf{r}^k\|_2^2 \\ &= \frac{L}{K}\|\mathbf{\Phi}'_{\mathcal{T}\cup\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\cup\mathcal{T}^k}(\mathbf{x}-\mathbf{x}^k)_{\mathcal{T}\cup\mathcal{T}^k}\|_2^2 \\ &\overset{\text{RIP}}{\ge} \frac{L}{K}(1-\delta_{Lk+K})^2\|(\mathbf{x}-\mathbf{x}^k)_{\mathcal{T}\cup\mathcal{T}^k}\|_2^2 \\ &\overset{\text{RIP}}{\ge} \frac{L(1-\delta_{Lk+K})^2}{K(1+\delta_{Lk+K})}\|\mathbf{\Phi}_{\mathcal{T}\cup\mathcal{T}^k}(\mathbf{x}-\mathbf{x}^k)_{\mathcal{T}\cup\mathcal{T}^k}\|_2^2 \\ &= \frac{L(1-\delta_{Lk+K})^2}{K(1+\delta_{Lk+K})}\|\mathbf{r}^k\|_2^2, \quad \text{(E.9)}\end{aligned}$$

where (a) is because $\mathbf{\Phi}'_{\mathcal{T}^k}\mathbf{r}^k \overset{\text{(B.3)}}{=} \mathbf{\Phi}'_{\mathcal{T}^k}(\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{y}) \overset{\text{(B.4)}}{=} \mathbf{\Phi}'_{\mathcal{T}^k}(\mathbf{P}^\perp_{\mathcal{T}^k})'\mathbf{y} = (\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}^k})'\mathbf{y} = \mathbf{0}$. From (E.1) and (E.9),

$$\begin{aligned}\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 &\ge \frac{L(1-\delta_{Lk+K})^2}{K(1+\delta_L)(1+\delta_{Lk+K})}\left(1-\frac{\delta^2_{Lk+1}}{1-\delta_{LK}}\right) \\ &\quad\times \|\mathbf{r}^k\|_2^2,\end{aligned}$$

which implies that

$$\|\mathbf{r}^{k+1}\|_2^2 \le \alpha(k,L)\|\mathbf{r}^k\|_2^2, \quad \text{(E.10)}$$

where

$$\alpha(k,L) := 1 - \frac{L(1-\delta_{Lk+K})^2}{K(1+\delta_L)(1+\delta_{Lk+K})}\left(1-\frac{\delta^2_{Lk+1}}{1-\delta_{LK}}\right).$$

Since δ_{Lk+1}, δ_{Lk+1}, and δ_{Lk+K} are increasing monotone functions of k (by Lemma 1), $\alpha(k,L)$ is increasing with k. Repeating (E.10) we obtain

$$\|\mathbf{r}^{k+1}\|_2^2 \le \prod_{i=0}^{k}\alpha(i,L)\|\mathbf{r}^0\|_2^2 \le (\alpha(k,L))^{k+1}\|\mathbf{y}\|_2^2. \quad \text{(E.11)}$$

Appendix F
Proof of Proposition 2

A. Proof of (28)

Since u_1 is the largest value of $\left\{\frac{|\langle\phi_i,\mathbf{r}^k\rangle|}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2}\right\}_{i\in\mathcal{T}\setminus\mathcal{T}^k}$,

$$\begin{aligned}u_1 &\overset{(a)}{\ge} \frac{1}{\sqrt{|\mathcal{T}\setminus\mathcal{T}^k|}}\left(\sum_{i\in\mathcal{T}\setminus\mathcal{T}^k}\frac{\langle\phi_i,\mathbf{r}^k\rangle^2}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2^2}\right)^{1/2} \\ &\overset{(b)}{\ge} \frac{1}{\sqrt{K-\ell}}\sqrt{\sum_{i\in\mathcal{T}\setminus\mathcal{T}^k}\langle\phi_i,\mathbf{r}^k\rangle^2} = \frac{\|\mathbf{\Phi}'_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{r}^k\|_2}{\sqrt{K-\ell}} \quad \text{(F.1)}\end{aligned}$$

where (a) is due to Cauchy-Schwarz inequality and (b) is because $\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2 \le \|\phi_i\|_2 = 1$.

To lower bound $\|\mathbf{\Phi}'_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{r}^k\|_2$ in (F.1), we apply the improved result of [12, Lemma 3.7] recently obtained by Li *et al.* [35]. Specifically, [35, Eq. (25)] implies that

$$\begin{aligned}&\|\mathbf{\Phi}'_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{r}^k\|_2 \\ &\ge (1-\delta_{K+Lk-\ell})\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 - \sqrt{1+\delta_{K+Lk-\ell}}\,\|\mathbf{v}\|_2. \quad \text{(F.2)}\end{aligned}$$

Using (F.1) and (F.2), we obtain (28).

B. Proof of (29)

First, let $\mathcal{F}$ be the index set corresponding to L largest elements in $\left\{\frac{|\langle\phi_i,\mathbf{r}^k\rangle|}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2}\right\}_{i\in\Omega\setminus(\mathcal{T}\cup\mathcal{T}^k)}$. Since v_L is the L-th largest value in $\left\{\frac{|\langle\phi_i,\mathbf{r}^k\rangle|}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2}\right\}_{i\in\mathcal{F}}$,

$$\begin{aligned}v_L &\le \frac{1}{\sqrt{L}}\left(\sum_{i\in\mathcal{F}}\frac{|\langle\phi_i,\mathbf{r}^k\rangle|^2}{\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2^2}\right)^{1/2} \le \frac{1}{\sqrt{L}}\left(\frac{\sum_{i\in\mathcal{F}}|\langle\phi_i,\mathbf{r}^k\rangle|^2}{\min_{i\in\mathcal{F}}\|\mathbf{P}^\perp_{\mathcal{T}^k}\phi_i\|_2^2}\right)^{1/2} \\ &= \frac{1}{\sqrt{L}}\left(\frac{\sum_{i\in\mathcal{F}}|\langle\phi_i,\mathbf{r}^k\rangle|^2}{1-\max_{i\in\mathcal{F}}\|\mathbf{P}_{\mathcal{T}^k}\phi_i\|_2^2}\right)^{1/2} \\ &\overset{\text{(E.7)}}{\le} \left(1-\frac{\delta^2_{Lk+1}}{1-\delta_{LK}}\right)^{-1/2}\frac{\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{r}^k\|_2}{\sqrt{L}}, \quad \text{(F.3)}\end{aligned}$$

where in the last inequality we have used the fact that $1 \le k < K$ and $1 \le L \le K$ so that $\delta_{Lk+1} \overset{\text{Lemma 1}}{\le} \delta_{LK} < \frac{\sqrt{5}-1}{2}$.

Next, we find an upper bound for $\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{r}^k\|_2$ in (F.3). Since

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{r}^k\|_2 &= \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^\perp_{\mathcal{T}^k}(\mathbf{\Phi}\mathbf{x}+\mathbf{v})\|_2 \\ &\le \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{\Phi}\mathbf{x}\|_2 + \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{v}\|_2 \\ &\le \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 + \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{v}\|_2 \\ &\quad + \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2, \quad \text{(F.4)}\end{aligned}$$

we upper bound $\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2$, $\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^\perp_{\mathcal{T}^k}\mathbf{v}\|_2$, and $\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2$, respectively.

Since $\mathcal{F}$ and $\mathcal{T}\setminus\mathcal{T}^k$ are disjoint ($\mathcal{F}\cap(\mathcal{T}\setminus\mathcal{T}^k)=\emptyset$), and also noting that $\mathcal{T}\cap\mathcal{T}^k=\ell$ by hypothesis in (27), we have $|\mathcal{F}|+|\mathcal{T}\setminus\mathcal{T}^k| = L+K-\ell$. Hence, by Lemma 5, we have

$$\left\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 \le \delta_{L+K-\ell}\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2. \quad \text{(F.5)}$$

Following the argument in [35, Eq. (26)], we can upper bound $\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{v}\|_2$ as

$$\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{v}\|_2 \leq \sqrt{1+\delta_{L+Lk}}\|\mathbf{v}\|_2. \tag{F.6}$$

Moreover, since $\mathcal{F}\cap\mathcal{T}^k=\emptyset$ and $|\mathcal{F}|+|\mathcal{T}^k|=L+Lk$,

$$\begin{aligned}
&\left\|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{P}_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 \\
&= \|\mathbf{\Phi}'_{\mathcal{F}}\mathbf{\Phi}_{\mathcal{T}^k}(\mathbf{\Phi}^{\dagger}_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k})\|_2 \\
&\leq \delta_{L+Lk}\|\mathbf{\Phi}^{\dagger}_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 \\
&= \delta_{L+Lk}\|(\mathbf{\Phi}'_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}^k})^{-1}\mathbf{\Phi}'_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 \\
&\overset{\text{Lemma 2}}{\leq} \frac{\delta_{L+Lk}}{1-\delta_{LK}}\left\|\mathbf{\Phi}'_{\mathcal{T}^k}\mathbf{\Phi}_{\mathcal{T}\setminus\mathcal{T}^k}\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2 \\
&\overset{\text{Lemma 5}}{\leq} \frac{\delta_{L+Lk}\delta_{Lk+K-\ell}}{1-\delta_{LK}}\left\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\right\|_2.
\end{aligned} \tag{F.7}$$

where in the last inequality we have used the fact that $|\mathcal{T}^k\cup(\mathcal{T}\setminus\mathcal{T}^k)|=Lk+K-\ell$. (Note that $\mathcal{T}^k$ and $\mathcal{T}\setminus\mathcal{T}^k$ are disjoint and $|\mathcal{T}\setminus\mathcal{T}^k|=K-\ell$.)

Finally, using (F.3), (F.4), (F.5), (F.6) and (F.7) yields (29).

Appendix G
Proof of (34)

Proof: Rearranging the terms in (33) we obtain

$$\begin{aligned}
&\sqrt{\frac{L}{K-\ell}}(1-\delta_{LK})-\frac{\delta_{LK}}{1-\delta_{LK}}\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta_{LK}^2}\right)^{1/2} \\
&> \left(\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta_{LK}^2}\right)^{1/2}+\sqrt{\frac{L}{K-\ell}}\right)\frac{\sqrt{1+\delta_{LK}}\|\mathbf{v}\|_2}{\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2}.
\end{aligned} \tag{G.1}$$

In the following, we will show that (G.1) is guaranteed by (34).

First, since

$$\begin{aligned}
\frac{\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2}{\|\mathbf{v}\|_2} &\geq \frac{\sqrt{|\mathcal{T}\setminus\mathcal{T}^k|}\min_{j\in\mathcal{T}}|x_j|}{\|\mathbf{v}\|_2} \overset{(16)}{=} \frac{\kappa\sqrt{K-\ell}\|\mathbf{x}\|_2}{\sqrt{K}\|\mathbf{v}\|_2} \\
&\overset{\text{RIP}}{\geq} \frac{\kappa\sqrt{K-\ell}\,\|\mathbf{\Phi x}\|_2}{\sqrt{K(1+\delta_{LK})}\|\mathbf{v}\|_2} \overset{(16)}{=} \frac{\kappa\sqrt{(K-\ell)snr}}{\sqrt{K(1+\delta_{LK})}},
\end{aligned}$$

it is clear that (G.1) holds true under

$$\begin{aligned}
&\sqrt{\frac{L}{K-\ell}}(1-\delta_{LK})-\frac{\delta_{LK}}{1-\delta_{LK}}\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta_{LK}^2}\right)^{1/2} \\
&> \left(\left(\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta_{LK}^2}\right)^{1/2}+\sqrt{\frac{L}{K-\ell}}\right)\frac{\sqrt{K}(1+\delta_{LK})}{\kappa\sqrt{(K-\ell)snr}}.
\end{aligned} \tag{G.2}$$

For notational simplicity, denote $\beta := (\frac{1-\delta_{LK}}{1-\delta_{LK}-\delta_{LK}^2})^{1/2}$, $\gamma := \sqrt{\frac{L}{K-\ell}}$, $\delta := \delta_{LK}$, and $\tau := \frac{\sqrt{K}}{\kappa\sqrt{(K-\ell)snr}}$. Then, we can rewrite (G.2) as $\gamma(1-\delta)-\frac{\delta\beta}{1-\delta} > (\beta+\gamma)(1+\delta)\tau$. That is,

$$\gamma\left(1-\delta-(1+\delta)\tau\right) > \beta\left(\frac{\delta}{1-\delta}+(1+\delta)\tau\right). \tag{G.3}$$

Next, we build an upper bound for β in (G.3). Define

$$f(\delta)=(1-\delta)^{3/2}(1-\delta-\delta^2)^{1/2} \text{ and } g(\delta)=1-2\delta.$$

Since $L\leq K$ and hence $\delta<\frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}}\leq\frac{1}{3}<\frac{\sqrt{5}-1}{2}$, one can show that $f(\delta)>g(\delta)$, which immediately implies that

$$\left(\frac{1-\delta}{1-\delta-\delta^2}\right)^{1/2} < \frac{(1-\delta)^2}{1-2\delta}. \tag{G.4}$$

That is, $\beta<\frac{(1-\delta)^2}{1-2\delta}$. Thus the right-hand side of (G.3) satisfies

$$\begin{aligned}
\beta\left(\frac{\delta}{1-\delta}+(1+\delta)\tau\right) &< \frac{(1-\delta)(\delta+(1-\delta^2)\tau)}{1-2\delta} \\
&\overset{(a)}{\leq} \frac{(1-\delta)(\delta+(1+\delta)\tau)}{1-2\delta},
\end{aligned} \tag{G.5}$$

where (a) is because $1-\delta^2\leq 1+\delta$.

By (G.3) and (G.5), we directly have that (G.1) holds if $\gamma\left(1-\delta-(1+\delta)\tau\right) > \frac{(1-\delta)(\delta+(1+\delta)\tau)}{1-2\delta}$. Equivalently,

$$\delta < \frac{1}{1+\tau}\left(\frac{\gamma}{u+\gamma}-\tau\right) \text{ where } u := \frac{1-\delta}{1-2\delta}. \tag{G.6}$$

We now simplify (G.6). Observe that

$$\begin{aligned}
\delta < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}} &\overset{(27)}{\Rightarrow} \delta < \frac{\sqrt{L}}{\sqrt{K-\ell}+2\sqrt{L}} \\
&\Leftrightarrow \delta < \frac{\gamma}{1+2\gamma} \Leftrightarrow u < 1+\gamma \\
&\Leftrightarrow \frac{\gamma}{u+\gamma} > \frac{\gamma}{1+2\gamma}.
\end{aligned} \tag{G.7}$$

From (G.6) and (G.7), we further have that (G.1) holds true whenever

$$\delta < \frac{1}{1+\tau}\left(\frac{\gamma}{1+2\gamma}-\tau\right) = \frac{1}{1+\tau}\left(\frac{\sqrt{L}}{\sqrt{K-\ell}+2\sqrt{L}}-\tau\right). \tag{G.8}$$

Since $\delta < \frac{\sqrt{L}}{\sqrt{K}+2\sqrt{L}} \leq \frac{\sqrt{L}}{\sqrt{K-\ell}+2\sqrt{L}}$, (G.8) can be rewritten as

$$\sqrt{snr} > \frac{(1+\delta)(\sqrt{K-\ell}+2\sqrt{L})}{\kappa(\sqrt{L}-(\sqrt{K-\ell}+2\sqrt{L})\delta)\sqrt{K-\ell}}\sqrt{K}. \tag{G.9}$$

Finally, since $1\leq K-\ell<K$, (G.9) is guaranteed by (34) and this completes the proof. ■

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, 2001.

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

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[5] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, nos. 9/10, pp. 589–592, 2008.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Syst. Comput.*, Nov. 1993, vol. 1, pp. 40–44.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[9] 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.

[10] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[11] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[12] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[13] 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, Mar. 2009.

[14] 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.

[15] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.

[16] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[17] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3869–3872.

[18] S. Foucart and M Lai, "Sparsest solutions of underdetermined linear systems via q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.

[19] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[20] L. Chen and Y. Gu, "The convergence guarantees of a non-convex approach for sparse recovery using regularized least squares," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 3350–3354.

[21] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[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] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.

[24] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit algorithm," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, Jun. 2012.

[25] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, Sep. 2012.

[26] J. Wen, X. Zhu, and D. Li, "Improved bounds on restricted isometry constant for orthogonal matching pursuit," *Electron. Lett.*, vol. 49, no. 23, pp. 1487–1489, 2013.

[27] J. Wang, "Support recovery with orthogonal matching pursuit in the presence of noise," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5868–5877, Nov. 2015.

[28] J. Wang and B. Shim, "Exact recovery of sparse signals using orthogonal matching pursuit: How many iterations do we need?" *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4194–4202, Aug. 2016.

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

[30] S. Foucart, "Stability and robustness of weak orthogonal matching pursuits," in *Recent Advances in Harmonic Analysis and Applications*. Berlin, Germany: Springer, 2013, pp. 395–405.

[31] 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.

[32] C. Maung and H. Schweitzer, "Improved greedy algorithms for sparse approximation of a matrix in terms of another matrix," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 769–780, Mar. 2015.

[33] J. Wang, S. Kwon, P. Li, and B. Shim, "Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1076–1089, Feb. 2016.

[34] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.

[35] B. Li, Y. Shen, Z. Wu, and J. Li, "Sufficient conditions for generalized orthogonal matching pursuit in noisy case," *Signal Process.*, vol. 108, pp. 111–123, 2015.

[36] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," Univ. Southampton, Southampton, U.K., *Tech. Rep.*, 2007. [Online]. Available: http://www.personal.soton.ac.uk/tb1m08/papers/BDOMPvsOLS07.pdf, 2007

[37] 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.

[38] L. Chang and J. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5702–5715, Sep. 2014.

[39] L. Chang and J. Wu, "Achievable angles between two compressed sparse vectors under norm/distance constraints imposed by the restricted isometry property: A plane geometry approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2059–2081, Apr. 2013.

[40] P. Zhao, J. Yang, T. Zhang, and P. Li, "Adaptive stochastic alternating direction method of multipliers," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 69–77.

[41] Q. Mo, "A sharp restricted isometry constant bound of orthogonal matching pursuit," arXiv:1501.01708, 2015.

[42] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[43] E. Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[44] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Proc. IEEE Workshop Statist. Signal Process.*, 2007, pp. 299–303.

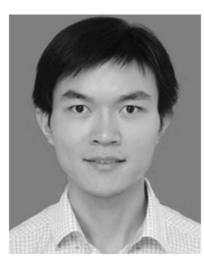
[45] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, Dec. 2009.

[46] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit: A brownian motion analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1010–1021, Mar. 2012.

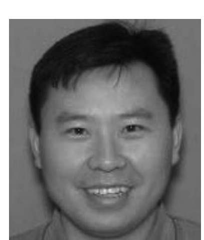
[47] R. Wu, W. Huang, and D. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 403–406, Apr. 2013.

[48] 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.

[49] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, Philadelphia, PA, USA: SIAM, 1994.



Jian Wang (S'11) received the B.S. degree in material engineering and the M.S. degree in information and communication engineering from Harbin Institute of Technology, Harbin, China, and the Ph.D. degree in electrical and computer engineering from Korea University, Seoul, South Korea, in 2006, 2009, and 2013, respectively. After graduation, he held positions of Postdoctoral Research Associate in the Deparment of Statistics and Department of Computer Science, Rutgers University, New Brunswick, NJ, USA, and the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. He is currently a Research Assistant Professor in the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea. His research interests include sparse and low-rank recovery, lattice, wireless sensor networks, signal processing in wireless communications, and statistical learning.



Ping Li received his Ph.D. in Statistics from Stanford University, where he also earned two Master's degrees in Computer Science and Electric Engineering. He also completed graduate degrees from the University of Washington (Seattle). Ping Li is a recipient of the Air Force Office of Scientific Research Young Investigator Award (AFOSR-YIP) and a receipt of the Office of Naval Research Young Investigator Award (ONR-YIP). Ping Li (with coauthors) won the Best Paper Award in NIPS 2014, the Best Paper Award in ASONAM 2014, and the Best Student Paper Award in KDD 2006.