Recovery of Sparse Signals Using Multiple Orthogonal Least Squares

Jian Wang, Ping Li
Department of Statistics and Biostatistics
Department of Computer Science
Rutgers University
Piscataway, New Jersey 08854, USA
Email: \{jwang, pingli\}@stat.rutgers.edu

Abstract

We study the problem of sparse recovery from compressed measurements. This problem has generated a great deal of interest in recent years. To recover the sparse signal, we propose a new method called multiple orthogonal least squares (MOLS), which extends the well-known orthogonal least squares (OLS) algorithm by choosing multiple $L$ indices per iteration. Due to the inclusion of multiple support indices in each selection, the MOLS algorithm converges in much fewer iterations and hence improves the computational efficiency over the OLS algorithm. Theoretical analysis shows that MOLS performs the exact recovery of any $K$-sparse signal within $K$ iterations if the measurement matrix satisfies the restricted isometry property (RIP) with isometry constant $\delta_{LK} \leq \sqrt{\frac{L}{K+\sqrt{ML}}}$. Empirical experiments demonstrate that MOLS is very competitive in recovering sparse signals compared to the state of the art recovery algorithms.

Index Terms

Compressed sensing (CS), orthogonal matching pursuit (OMP), orthogonal least squares (OLS), restricted isometry property (RIP)
Recovery of Sparse Signals Using Multiple Orthogonal Least Squares

I. INTRODUCTION

In recent years, sparse recovery has attracted much attention in signal processing, image processing, and computer science [1]–[4]. The main task of sparse recovery is to recover a high dimensional $K$-sparse vector $x \in \mathbb{R}^n$ ($\|x\|_0 \leq K$) from a small number of linear measurements

$$y = \Phi x,$$

where $\Phi \in \mathbb{R}^{m \times n}$ ($m \ll n$) is often called the measurement matrix. Although the system is underdetermined, owing to the signal sparsity, $x$ can be accurately recovered from the measurements $y$ by solving an $\ell_0$-minimization problem:

$$\min_{x} \|x\|_0 \quad \text{subject to} \quad y = \Phi x.$$  \hfill (2)

This method, however, is known to be intractable due to the combinatorial search involved and therefore impractical for realistic applications.

Much recent effort has been devoted to develop efficient algorithms for recovering the sparse signal. In general, the algorithms can be classified into two major categorizes: those using convex optimization techniques and those relying on greedy searching principles. The optimization-based approaches replace the nonconvex $\ell_0$-norm with its convex relaxation $\ell_1$-norm, translating the combinatorial hard search into a computationally tractable problem:

$$\min_{x} \|x\|_1 \quad \text{subject to} \quad y = \Phi x.$$  \hfill (3)

The algorithm is known as basis pursuit (BP) [5]. It has been revealed that under appropriate constraints on the measurement matrix, BP yields exact recovery of the sparse signal.

The second category of approaches for sparse recovery are greedy algorithms, in which signal support is iteratively constructed according to various greedy principles. Greedy algorithms have gained considerable popularity due to their computational simplicity and competitive performance. Representative algorithms include matching pursuit (MP) [6], orthogonal matching pursuit (OMP) [7]–[12] and orthogonal least squares (OLS) [13]–[16]. Both OMP and OLS update the
support of the underlying sparse signal by adding one index 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. While OMP finds a column that is most strongly correlated with the signal residual, OLS seeks to maximally reduce the residual energy with an enlarged support set. It has been shown that OLS has better convergence property but is computationally more expensive than OMP [16].

In this paper, with the aim of improving the recovery accuracy and reducing the computational cost of OLS, we propose a new method called multiple orthogonal least squares (MOLS), which can be viewed as an extension of OLS in the sense that multiple indices are allowed to be chosen at a time. Our method is inspired by the fact that those sub-optimal candidates in each of the OLS identification are likely to be reliable, and hence can be utilized to better reduce the energy of signal residual at each iteration. The main steps of the MOLS algorithm are specified in Table 1. Owing to the selection of multiple “good” candidates in each time, MOLS converges in much fewer iterations in the end and thus reduces the computational cost over the conventional OLS algorithm.

Greedy methods with a similar flavor to MOLS in adding multiple indices per iteration include stagewise OMP (StOMP) [17], regularized OMP (ROMP) [18], and generalized OMP (gOMP) [19] (also known as orthogonal super greedy algorithm (OSGA) [20]), etc. These algorithms identify candidates at each iteration according to the correlations between columns of the measurement 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 greedy methods adopting a different strategy of adding as well as pruning indices from the list include compressive sampling matching pursuit (CoSaMP) [21] and subspace pursuit (SP) [22] and hard thresholding pursuit (HTP) [23].

The main contributions of this paper are summarized as follows.

• We propose a new algorithm, referred to as MOLS, for recovering sparse signals from compressed measurements. We analyze the MOLS algorithm using the restricted isometry property (RIP). A measurement matrix Φ is said to satisfy the RIP of order K if there
**TABLE I**

**THE MOLS ALGORITHM**

| Input | measurement matrix $\Phi \in \mathbb{R}^{m \times n}$, measurements vector $y \in \mathbb{R}^m$, sparsity level $K$, and selection parameter $L \leq K$. |
|-------|----------------------------------------------------------------------------------------------------------------------------------|
| Initialize | iteration count $k = 0$, estimated support $\mathcal{T}^0 = \emptyset$, and residual vector $r^0 = y$. |
| While | $\|r^k\|_2 \geq \epsilon$ and $k < \min\{K, \lfloor \frac{m}{L} \rfloor\}$, do |
| | $k = k + 1$. |
| | Identify $\mathcal{S}^k = \arg\min_{\mathcal{S} \subseteq [L]} \sum_{i \in \mathcal{S}} \| \mathbf{P}_{\mathcal{T}^k \cup \{i\}} \mathbf{y} \|_2^2$. |
| | Enlarge $\mathcal{T}^k = \mathcal{T}^{k-1} \cup \mathcal{S}^k$. |
| | Estimate $\mathbf{x}^k = \arg\min_{u : \text{supp}(u) = \mathcal{T}^k} \| y - \Phi u \|_2$. |
| | Update $r^k = y - \Phi \mathbf{x}^k$. |
| End | |
| Output | the estimated support $\hat{T} = \arg\min_{\mathcal{S} \subseteq [K]} \| \mathbf{x}^\mathcal{S} - \hat{x} \|_2$ and $K$-sparse signal $\hat{x}$ satisfying $\hat{x}_{\mathcal{H} \setminus \hat{T}} = \mathbf{0}$ and $\hat{x}_{\hat{T}} = \mathbf{x}^\mathcal{S}$. |

exists a constant $\delta \in [0, 1)$ such that \[ (4) \]
\[
\sqrt{1 - \delta} \| x \|_2 \leq \| \Phi x \|_2 \leq \sqrt{1 + \delta} \| x \|_2
\]
for all $K$-sparse vectors $x$. In particular, the minimum of all constants $\delta$ satisfying \[ (4) \] is the isometry constant $\delta_K$. Our result shows that when $L > 1$, the MOLS algorithm exactly recovers any $K$-sparse signal in $K$ iterations if the measurement matrix $\Phi$ obeys the RIP with isometry constant
\[
\delta_{LK} \leq \frac{\sqrt{L}}{\sqrt{K} + \sqrt{5L}}.
\]

- For the special case when $L = 1$, MOLS reduces to the conventional OLS algorithm and the recovery condition is given by
\[
\delta_{K+1} < \frac{1}{\sqrt{K + 1}}.
\]
We show that the condition in (6) is nearly optimal for OLS in the sense that, even with a slight relaxation of the condition (e.g., to $\delta_{K+1} = \frac{1}{\sqrt{K}}$), the exact recovery of OLS may fail.

- We conduct experiments to test the effectiveness of the MOLS algorithm. Our empirical results demonstrate that MOLS (with $L > 1$) converges in much fewer iterations and has lower computational cost than the OLS algorithm, while exhibiting better recovery accuracy. The empirical results also show that MOLS is very competitive in recovering sparse signals when compared to the state of the art sparse recovery algorithms.

The rest of this paper is organized as follows: In Section II, we introduce notations and lemmas that are used in the paper. In Section III, we give useful observations regarding MOLS. In Section IV, we analyze the recovery condition of MOLS. In Section V, we empirically study the recovery performance of the MOLS algorithm. Concluding remarks are given in Section VI.

II. PRELIMINARIES

A. Notations

We first briefly summarize notations used in this paper. Let $\mathcal{H} = \{1, 2, \ldots, n\}$ and $\mathcal{T} = \text{supp}(x) = \{i| i \in \mathcal{H}, x_i \neq 0\}$ denote the support of vector $x$. For $S \subseteq \mathcal{H}$, $|S|$ is the cardinality of $S$. $\mathcal{T} \setminus S$ is the set of all elements contained in $\mathcal{T}$ but not in $S$. $x_S \in \mathbb{R}^{|S|}$ is a restriction of the vector $x$ to the elements with indices in $S$. For mathematical convenience, we assume that $\Phi$ has unit $\ell_2$-norm columns throughout this paper. $\Phi_S \in \mathbb{R}^{m \times |S|}$ is a submatrix of $\Phi$ that only contains columns indexed by $S$. If $\Phi_S$ is full column rank, then $\Phi^\dagger_S = (\Phi_S'\Phi_S)^{-1}\Phi_S'$ is the pseudoinverse of $\Phi_S$. Span($\Phi_S$) represents the span of columns in $\Phi_S$. $P_S = \Phi_S\Phi^\dagger_S$ stands for the projection onto Span($\Phi_S$). $P^\perp_S = I - P_S$ is the projection onto the orthogonal complement of Span($\Phi_S$), where $I$ denotes the identity matrix.

B. Lemmas

The following lemmas are useful for our analysis.

Lemma 1 (Lemma 3 in [24]): If a measurement matrix satisfies the RIP of both orders $K_1$ and $K_2$ where $K_1 \leq K_2$, then $\delta_{K_1} \leq \delta_{K_2}$. This property is often referred to as the monotonicity of the isometry constant.
Lemma 2 (Consequences of RIP [21], [25]): Let $S \subseteq H$. If $\delta_{|S|} < 1$ then for any $u \in \mathcal{R}^{|S|}$,
\[
(1 - \delta_{|S|}) \|u\|_2 \leq \|\Phi'_S \Phi_S u\|_2 \leq (1 + \delta_{|S|}) \|u\|_2,
\]
\[
\frac{1}{1 + \delta_{|S|}} \|u\|_2 \leq \|(\Phi'_S \Phi_S)^{-1} u\|_2 \leq \frac{1}{1 - \delta_{|S|}} \|u\|_2.
\]

Lemma 3 (Lemma 2.1 in [26]): Let $S_1, S_2 \subseteq H$ and $S_1 \cap S_2 = \emptyset$. If $\delta_{|S_1|+|S_2|} < 1$, then
\[
\|\Phi'_{S_1} \Phi v\|_2 \leq \delta_{|S_1|+|S_2|} \|v\|_2
\]
holds for any vector $v \in \mathcal{R}^n$ supported on $S_2$.

Lemma 4: Let $S \subseteq H$. If $\delta_{|S|} < 1$ then for any $u \in \mathcal{R}^{|S|}$,
\[
\frac{1}{\sqrt{1 + \delta_{|S|}}} \|u\|_2 \leq \|(\Phi'_S)^* u\|_2 \leq \frac{1}{\sqrt{1 - \delta_{|S|}}} \|u\|_2.
\]

Proof: See Appendix A.

III. SIMPLIFICATION OF MOLS IDENTIFICATION

Let us begin with an interesting (and important) observation regarding 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,
\[
S^{k+1} = \arg \min_{S : |S| = L} \sum_{i \in S} \|P_{\mathcal{T}^k \cup \{i\}}^\perp y\|_2^2. \tag{7}
\]
Intuitively, a straightforward implementation of (7) requires to sort all elements in $\{\|P_{\mathcal{T}^k \cup \{i\}}^\perp y\|_2^2\}_{i \in H \setminus \mathcal{T}^k}$ and find the $L$ smallest ones (and their corresponding indices). This implementation, however, is computationally expensive as it requires to construct a number of $n - Lk$ different orthogonal projections (i.e., $P_{\mathcal{T}^k \cup \{i\}}^\perp$, $\forall i \in H \setminus \mathcal{T}^k$). Therefore, it is highly desirable to find a cost-effective alternative to (7) for the MOLS identification.

Interestingly, the following proposition illustrates that (7) can be substantially simplified.

Proposition 1: At the $(k+1)$-th iteration, the MOLS algorithm identifies a set of $L$ indices:
\[
S^{k+1} = \arg \max_{S : |S| = L} \sum_{i \in S} \frac{|\langle \phi_i, \mathbf{r}^k \rangle|}{\|P_{\mathcal{T}^k} \phi_i\|_2}, \tag{8}
\]
\[
= \arg \max_{S : |S| = L} \sum_{i \in S} \frac{\langle P_{\mathcal{T}^k} \phi_i, \mathbf{r}^k \rangle}{\|P_{\mathcal{T}^k} \phi_i\|_2}. \tag{9}
\]
**Proof:** Since $P_{\hat{\tau}_k}\cup\{i\}^\perp$ and $P_{\tau_k}\cup\{i\}^\perp$ are orthogonal,

$$\|P_{\hat{\tau}_k}\cup\{i\}^\perp y\|_2^2 = \|y\|_2^2 - \|P_{\tau_k}\cup\{i\}^\perp y\|_2^2,$$

and hence (7) is equivalent to

$$S^{k+1} = \arg\max_{S:|S|=L} \sum_{i\in S} \|P_{\tau_k}\cup\{i\}^\perp y\|_2^2; \quad (10)$$

By noting that $P_{\tau_k}\cup\{i\}$ can be decomposed as (see Appendix B)

$$P_{\tau_k}\cup\{i\} = P_{\tau_k} + P_{\hat{\tau}_k} \phi_i (\phi_i^\perp P_{\hat{\tau}_k} \phi_i)^{-1} \phi_i^\perp P_{\hat{\tau}_k}, \quad (11)$$

we have

$$\|P_{\tau_k}\cup\{i\}^\perp y\|_2^2 = \|P_{\tau_k} y + P_{\hat{\tau}_k} \phi_i (\phi_i^\perp P_{\hat{\tau}_k} \phi_i)^{-1} \phi_i^\perp P_{\hat{\tau}_k} y\|_2^2$$

$$(a) = \|P_{\tau_k} y\|_2^2 + \|P_{\hat{\tau}_k} \phi_i (\phi_i^\perp P_{\hat{\tau}_k} \phi_i)^{-1} \phi_i^\perp P_{\hat{\tau}_k} y\|_2^2$$

$$(b) = \|P_{\tau_k} y\|_2^2 + \frac{|\phi_i^\perp P_{\hat{\tau}_k} y|^2}{\|P_{\tau_k} \phi_i\|^2} \|P_{\tau_k} \phi_i\|^2$$

$$(c) = \|P_{\tau_k} y\|_2^2 + \frac{|\phi_i^\perp P_{\hat{\tau}_k} y|^2}{\|P_{\tau_k} \phi_i\|^2}$$

$$(d) = \|P_{\tau_k} y\|_2^2 + \left(\frac{|\phi_i^\perp P_{\hat{\tau}_k} y|}{\|P_{\tau_k} \phi_i\|}\right)^2, \quad (12)$$

where (a) is because $P_{\tau_k} y$ and $P_{\hat{\tau}_k} \phi_i (\phi_i^\perp P_{\hat{\tau}_k} \phi_i)^{-1} \phi_i^\perp P_{\hat{\tau}_k} y$ are orthogonal, (b) follows from that fact that $\phi_i^\perp P_{\hat{\tau}_k} y$ and $\phi_i^\perp P_{\hat{\tau}_k} \phi_i$ are scalars, (c) is from $P_{\tau_k} = (P_{\tau_k})^2$ and $P_{\hat{\tau}_k} = (P_{\hat{\tau}_k})^\perp$, and hence

$$|\phi_i^\perp P_{\hat{\tau}_k} \phi_i| = |\phi_i^\perp (P_{\hat{\tau}_k})^\perp P_{\hat{\tau}_k} \phi_i| = \|P_{\tau_k} \phi_i\|^2,$$

and (d) is due to $r^k = P_{\hat{\tau}_k} y$.

By relating (10) and (12), we have

$$S^{k+1} = \arg\max_{S:|S|=L} \sum_{i\in S} \frac{|\langle \phi_i, r^k \rangle|}{\|P_{\tau_k} \phi_i\|_2}. \quad (11)$$

Furthermore, if we write

$$|\langle \phi_i, r^k \rangle| = |\phi_i^\perp (P_{\hat{\tau}_k})^\perp P_{\hat{\tau}_k} y| = \|P_{\tau_k} \phi_i, r^k\|_2,$$
then \((8)\) becomes
\[
S^{k+1} = \arg \max_{S : |S| = L} \sum_{i \in S} \left| \frac{\langle \phi_i, r^k \rangle}{\|P_{T_k}^\perp \phi_i\|_2} \right|.
\]
This completes the proof.

Interestingly, we can interpret from \((8)\) that to identify \(S^{k+1}\), it suffices to find the \(L\) largest values in \(\left\{ \frac{|\langle \phi_i, r^k \rangle|}{\|P_{T_k}^\perp \phi_i\|_2} \right\}_{i \in H \setminus T^k}\), which is much simpler than \((7)\) as it involves only one projection operator (i.e., \(P_{T_k}^\perp\)). Indeed, by numerical experiments, we have confirmed that the simplification offers large reduction in computational cost. The result \((8)\) will also play an important role in our analysis.

Another interesting point we would like to mention is that the result \((9)\) provides a geometric interpretation of the selection rule in MOLS: the columns of measurement matrix are projected onto the subspace that is orthogonal to the span of the active columns, and the \(L\) normalized projected columns that are best correlated with the residual vector are selected.

IV. SIGNAL RECOVERY USING MOLS

A. Exact recovery condition

In this subsection, we study the recovery condition of MOLS which guarantees the selection of all support indices within \(K\) iterations. For the convenience of stating the results, we say MOLS makes a success in an iteration if it selects at least one correct index at that iteration. Clearly if MOLS makes a success in each iteration, then it is guaranteed to select all support indices within \(K\) iterations.

- **Success at the first iteration**

  Observe from \((8)\) that when \(k = 0\) (i.e., at the first iteration), we have \(\|P_{T_k}^\perp \phi_i\|_2 = \|\phi_i\|_2 = 1\), and MOLS selects the set

  \[
  T^1 = S^1 = \arg \max_{S : |S| = L} \sum_{i \in S} \left| \frac{\langle \phi_i, r^k \rangle}{\|P_{T_k}^\perp \phi_i\|_2} \right|
  = \arg \max_{S : |S| = L} \sum_{i \in S} \|\phi_i, r^k\|_2
  = \arg \max_{S : |S| = L} \sum_{i \in S} \|\Phi_{Sy}\|_2.
  \]  

  (13)
That is,
\[ \| \Phi'_T y \|_2 \geq \max_{S : |S| = L} \| \Phi'_S y \|_2. \] (14)

By noting that \( L \leq K \), we know
\[
\| \Phi'_T y \|_2 \geq \sqrt{\frac{L}{K}} \| \Phi'_T y \|_2 \\
\geq \sqrt{\frac{L}{K}} \| \Phi'_T \Phi_T x_T \|_2 \\
\geq \sqrt{\frac{L}{K}} (1 - \delta_K) \| x \|_2. \] (15)

On the other hand, if no correct index is chosen at the first iteration (i.e., \( T^1 \cap T = \emptyset \)), then
\[
\| \Phi'_T y \|_2 = \| \Phi'_T \Phi_T x_T \|_2 \\
\leq (a) \delta_{K+L} \| x \|_2, \] (16)

where (a) follows from Lemma 3. This, however, contradicts (15) if
\[ \delta_{K+L} < \sqrt{\frac{L}{K}} (1 - \delta_K), \] (17)

or
\[ \delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K} + \sqrt{L}}. \] (18)

Therefore, under (18), at least one correct index is chosen at the first iteration of MOLS.

**Success at the** \((k + 1)\)-**th iteration** \((k > 0)\)

Assume that MOLS selects at least one correct index in each of the previous \(k \geq 1\) iterations and denote by \( \ell \) the number of correct indices in \( S^k \). Then,
\[ \ell = |T \cap T^k| \geq k. \] (19)

Also, assume that \( T^k \) does not contain all correct indices \((\ell < K)\). Under these assumptions, we will build a condition to ensure that MOLS selects at least one correct index at the \((k + 1)\)-th iteration.
For convenience, we introduce two additional notations. Let $u_1$ denote the largest value of $\frac{|\langle \phi_i, r_k \rangle|}{\| P_{T_k} \phi_i \|_2}$, $i \in T \setminus T^k$. Also, let $v_L$ denote the $L$-th largest value of $\frac{|\langle \phi_i, r_k \rangle|}{\| P_{T_k} \phi_i \|_2}$, $i \in H \setminus (T \cup T^k)$. Then it follows from (8) that as long as $u_1 > v_L$, (20)

$u_1$ is contained in the top-$L$ among all values in $\left\{ \frac{|\langle \phi_i, r_k \rangle|}{\| P_{T_k} \phi_i \|_2} \right\}_{i \in H \setminus T^k}$, and hence at least one correct index (i.e., the one corresponding to $u_1$) is selected at the $(k + 1)$-th iteration of MOLS. The following proposition gives the lower bound of $u_1$ and the upper bound of $v_L$.

**Proposition 2:**

$$ u_1 \geq \left( 1 - \frac{\delta_{K-\ell}}{1 - \delta_{Lk}} \right) \frac{\| x_{T \setminus T^k} \|_2}{\sqrt{K - \ell}}, $$

$$ v_L \leq \left( 1 + \frac{\delta_{Lk+1}^2}{1 - \delta_{Lk} - \delta_{Lk+1}^2} \right)^{1/2} \times \left( \delta_{L+K-\ell} + \frac{\delta_{L+K-\ell} \delta_{Lk} \delta_{Lk+K-\ell}}{1 - \delta_{Lk}} \right) \frac{\| x_{T \setminus T^k} \|_2}{\sqrt{L}}. $$

(22)

**Proof:** See Appendix C.

By noting that $1 \leq k \leq \ell < K$ and $1 \leq L \leq K$ and also using the monotonicity of the restricted isometry constant in Lemma 1, we have

$$ K - \ell < LK \Rightarrow \delta_{K-\ell} < \delta_{LK}, $$

$$ Lk + K - \ell \leq LK \Rightarrow \delta_{Lk+K-\ell} \leq \delta_{LK}, $$

$$ Lk < LK \Rightarrow \delta_{Lk} < \delta_{LK}, $$

$$ Lk + 1 \leq LK \Rightarrow \delta_{Lk+1} \leq \delta_{LK}, $$

$$ L + Lk \leq LK \Rightarrow \delta_{L+Lk} \leq \delta_{LK}. $$

(23)
From (22) and (24), we have
\[
v_L < \left( 1 + \frac{\delta_{LK}^2}{1 - \delta_{LK} - \delta_{LK}^2} \right)^{1/2} \times \left( \delta_{LK} + \frac{\delta_{LK}^2}{1 - \delta_{LK}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{L}} \times \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{L}}.
\]
\[
= \left( \frac{\delta_{LK}}{(1 - \delta_{LK} - \delta_{LK}^2)^{1/2} (1 - \delta_{LK})^{1/2}} \right) \times \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{L}}.
\]
(24)

Also, from (21) and (24), we have
\[
\begin{aligned}
\quad u_1 &= \left( 1 - \delta_{LK} - \frac{\delta_{LK}^2}{1 - \delta_{LK}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{K - \ell}} \\
&= \left( \frac{1 - 2\delta_{LK}}{1 - \delta_{LK}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{K - \ell}} \\
&\geq \left( \frac{1 - 2\delta_{LK}}{1 - \delta_{LK}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{K}},
\end{aligned}
\]
(25)

where (a) follows from \( \sqrt{K - \ell} < \sqrt{K} \). Using (24) and (25), we obtain the sufficient condition of \( u_1 > v_L \) as
\[
\begin{aligned}
\left( \frac{1 - 2\delta_{LK}}{1 - \delta_{LK}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{K}} \\
\geq \left( \frac{\delta_{LK}}{(1 - \delta_{LK} - \delta_{LK}^2)^{1/2} (1 - \delta_{LK})^{1/2}} \right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{L}}.
\end{aligned}
\]
(26)

which is true under (see Appendix D)
\[
\delta_{LK} \leq \frac{\sqrt{L}}{\sqrt{K + \sqrt{5L}}},
\]
(27)

Therefore, under (27), MOLS selects at least one correct index at the \((k + 1)\)-th iteration.

So far, we have obtained condition (18) for guaranteeing the success of MOLS at the first iteration and condition (27) for the success of the succeeding iterations. We now combine them to get a sufficient condition of MOLS ensuring the exact recovery of all \(K\)-sparse signals within \(K\) iterations.
Theorem 5 (Recovery Guarantee for MOLS): The MOLS algorithm exactly recovers any $K$-sparse ($K \geq L$) vector $x$ from the measurements $y = \Phi x$ within $K$ iterations if the measurement matrix $\Phi$ satisfies the RIP with isometry constant

$$\begin{cases} 
\delta_{LK} < \frac{\sqrt{L}}{\sqrt{K} + \sqrt{5L}}, & L > 1, \\
\delta_{K+1} < \frac{1}{\sqrt{K} + 1}, & L = 1. 
\end{cases} \tag{28}$$

Proof: Our goal is to prove an exact recovery condition for the MOLS algorithm. To the end, we first derive a condition ensuring the selection of all support indices in $K$ iterations of MOLS, and then show that when all support indices are chosen, the recovery of sparse signal is exact.

Clearly, the condition that guarantees MOLS to select all support indices is determined by the more strict condition between (18) and (27). We consider the following two cases.

- $L = 1$: In this case, the condition (18) for the success of the first iteration becomes

$$\delta_{K+1} < \frac{1}{\sqrt{K} + 1}. \tag{29}$$

In fact, (29) also guarantees the success of the oncoming iterations, and hence becomes the overall condition for this case. We justify this by mathematical induction on $k$. Under the hypothesis that the former $k$ ($< K$) iterations are successful (i.e., that the algorithm selects one correct index in each of the $k$ iterations), we have $T^k \subset T$ and hence the residual $r^k = y - \Phi x^k = \Phi(x - x^k)$ can be viewed as the measurements of $K$-sparse vector $x - x^k$ (which is supported on $T \cup T^k = T$) using the measurement matrix $\Phi$. Therefore, (29) guarantees a success at the $(k + 1)$-th iteration of algorithm. In summary, (29) will ensure the selection of all support indices in $K$ iterations.

- $L \geq 2$: Since $\delta_{LK} \geq \delta_{K+L}$ and

$$\frac{\sqrt{L}}{\sqrt{K} + \sqrt{L}} > \frac{\sqrt{L}}{\sqrt{K} + \sqrt{5L}}, \tag{30}$$

(27) is more strict than (18) and becomes the overall condition for this case.
When all support indices are selected by MOLS (i.e., $\mathcal{T} \subseteq \mathcal{T}^{\bar{k}}$ where $\bar{k}$ $(\leq K)$ denotes the number of actually performed iterations),

$$x_{\mathcal{T}^{\bar{k}}}^k = \arg \min_u \| y - \Phi_{\mathcal{T}^{\bar{k}}} u \|_2$$

$$= \Phi_{\mathcal{T}^{\bar{k}}}^\dagger y$$

$$= \Phi_{\mathcal{T}^{\bar{k}}}^\dagger \Phi x$$

$$= \Phi_{\mathcal{T}^{\bar{k}}}^\dagger \Phi_{\mathcal{T}^{\bar{k}}} x_{\mathcal{T}^{\bar{k}}}$$

$$= x_{\mathcal{T}^{\bar{k}}}.$$ \hspace{1cm} (31)

As a result, the residual vector becomes zero ($r_{\hat{k}} = y - \Phi x_{\hat{k}} = 0$). The algorithm terminates and returns the exact recovery of the sparse signal ($\hat{x} = x$). \hfill \blacksquare

B. Optimality of recovery condition for OLS

When $L = 1$, MOLS reduces to the conventional OLS algorithm. Theorem 5 suggests that under $\delta_{K+1} < 1/\sqrt{K+1}$, OLS recovers any $K$-sparse signal in $K$ iterations. In fact, this condition is not only sufficient but also nearly necessary for OLS in the sense that, even with a slight relaxation on the isometry constant (e.g., relaxing to $\delta_{K+1} = 1/\sqrt{K}$), the recovery may fail. We justify this argument with the following example. Let

$$\Phi = \begin{bmatrix} 1 & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ 0 & \frac{\sqrt{14}}{4} & -\frac{4\sqrt{7\sqrt{14}}}{28} \\ 0 & 0 & -\sqrt{\frac{10-\sqrt{2}}{14}} \end{bmatrix}$$

and

$$x = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$  

Then,

$$\Phi'\Phi = \begin{bmatrix} 1 & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ -\frac{1}{2\sqrt{2}} & 1 & -\frac{1}{2\sqrt{2}} \\ -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 1 \end{bmatrix}. \hspace{1cm} (32)$$

$^1$Since $\bar{k} \leq \min\{K, \lfloor \frac{m}{L} \rfloor\}$ by Table I, the number of totally selected indices (within $K$ iterations of MOLS) does not exceed $m$, which ensures that the sparse signal can be recovered through an LS projection.
One can show that the eigenvalues of $\Phi' \Phi$ are \( \lambda_1(\Phi' \Phi) = \lambda_2(\Phi' \Phi) = 1 + \frac{1}{2\sqrt{2}} \) and \( \lambda_3(\Phi' \Phi) = 1 - \frac{1}{\sqrt{2}} \). Hence, $\Phi$ satisfies the RIP with

\[
\delta_3 = \max \{ \lambda_{\text{max}}(\Phi' \Phi) - 1, 1 - \lambda_{\text{min}}(\Phi' \Phi) \}
\]

\[
= \max \left\{ \frac{1}{2\sqrt{2}}, \frac{1}{\sqrt{2}} \right\}
\]

\[
= \frac{1}{\sqrt{2}}.
\]

(33)

However, in this case OLS identifies an incorrect index

\[
t^1 = \arg \min_{i \in \{1, 2, 3\}} \| P_i^\perp y \|_2 = 3
\]

at the first iteration, and hence the recovery fails.

C. The relationship between MOLS and gOMP

We can gain good insights by studying the relationship between MOLS and the generalized OMP (gOMP) \cite{19}. gOMP differs from MOLS in that, at each iteration, gOMP finds \( L \) indices with strongest correlations to the residual vector. Interestingly, it can be shown that MOLS belongs to the family of “weak gOMP” (also known as WOSGA \cite{20}), which includes gOMP as a special case.

By definition, the weak gOMP algorithm with parameter \( \mu^{k+1} \in (0, 1] \) identifies the set \( S^{k+1} \) of \( L \) indices at the \( (k+1) \)-th iteration \( (k \geq 0) \) such that

\[
\| \Phi_{\bar{S}^{k+1}}^\prime r^k \|_2 \geq \mu^{k+1} \max_{S : |S| = L} \| \Phi_{S}^\prime r^k \|_2.
\]

(35)

Clearly it embraces the gOMP algorithm as a special case when \( \mu^{k+1} = 1 \).

**Theorem 6:** MOLS belongs to the family of weak gOMP algorithms with parameter

\[
\mu^{k+1} = \sqrt{1 - \frac{\delta^2_k}{1 - \delta_k}}.
\]

**Proof:** Denote by \( \bar{S}^k \) the set of \( L \) indices selected at the \( (k+1) \)-th iteration such that

\[
\bar{S}^{k+1} = \arg \max_{S : |S| = L} \| \Phi_{S}^\prime r^k \|_2.
\]

(36)

Note that in order to recover a \( K \)-sparse signal exactly with \( K \) steps of OLS, any selection of incorrect index will not be allowed.
From (8),

\[
\sum_{i \in S^k} \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2} \geq \sum_{i \in S^k} \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2} \\
\geq \sum_{i \in S^k} |\langle \phi_i, r^k \rangle|^2 \\
= \max_{S: |S| = L} \| \Phi_S r^k \|_2^2
\]  

(37)

where (37) holds because \( \phi_i \) has unit \( \ell_2 \)-norm and hence \( \| P_{T_k} \phi_i \|_2 \leq 1 \).

On the other hand,

\[
\sum_{i \in S^k} \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2} \\
\leq \left( \frac{1}{\min_{i \in S^k} \| P_{T_k} \phi_i \|_2} \right) \sum_{i \in S^k} |\langle \phi_i, r^k \rangle|^2 \\
\leq \left( \frac{1}{\| \phi_i \|_2^2 - \max_{i \in S^k} \| P_{T_k} \phi_i \|_2} \right) \sum_{i \in S^k} |\langle \phi_i, r^k \rangle|^2 \\
= \left( \frac{1}{1 - \max_{i \in S^k} \| (\Phi_{T_k})' \Phi_{T_k} \phi_i \|_2} \right) \| \Phi_{S^k} r^k \|_2^2 \\
\leq \left( \frac{1}{1 - \frac{\delta_{L_k}^2 + 1}{1 - \delta_{L_k}}} \right) \| \Phi_{S^k} r^k \|_2^2
\]  

(38)

where (38) is due to the fact that, for any \( i \in S^k \),

\[
\| (\Phi_{T_k})' \Phi_{T_k} \phi_i \|_2^2 \overset{(a)}{\leq} \frac{\delta_{L_k}^2 + 1}{1 - \delta_{L_k}} \| \Phi_{T_k} \phi_i \|_2^2 \\
\overset{(b)}{\leq} \frac{\delta_{L_k}^2}{1 - \delta_{L_k}} \| \phi_i \|_2^2 \\
= \frac{\delta_{L_k}^2 + 1}{1 - \delta_{L_k}} \| \phi_i \|_2^2
\]  

(39)

where (a) and (b) follow from Lemma 4 and 3, respectively.

By combining (38) and (37), we obtain

\[
\| \Phi_{S^k} r^k \|_2^2 \geq \left( 1 - \frac{\delta_{L_k}^2 + 1}{1 - \delta_{L_k}} \right) \max_{S: |S| = L} \| \Phi_{S} r^k \|_2^2.
\]  

(40)

This completes the proof.
V. EMPirical Results

In this section, we empirically study the performance of the MOLS algorithm for recovering sparse signals. We adopt the testing strategy in [19], [22], [27] which measures the performance of recovery algorithms by testing their empirical frequency of exact reconstruction of sparse signals. In each trial, we construct an $m \times n$ matrix (where $m = 128$ and $n = 256$) with entries drawn independently from a Gaussian distribution with zero mean and $\frac{1}{m}$ variance. For each value of $K$ in $\{5, \ldots, 64\}$, we generate a $K$-sparse signal of $n \times 1$ 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 $\{\pm 1, \pm 3\}$. We refer to the two types of signals as the sparse Gaussian signal and the sparse 4-ary pulse amplitude modulation (4-PAM) signal, respectively.

We should mention that reconstructing sparse 4-PAM signals is a particularly challenging case for OMP and OLS. For comparative purposes, we consider the following recovery approaches in our simulation:

1) OLS and MOLS.
2) OMP and gOMP (http://isl.korea.ac.kr/paper/gOMP.zip),
3) StOMP (http://sparaselab.stanford.edu/),
4) ROMP (http://www.cmc.edu/pages/faculty/DNeedell),
5) CoSaMP (http://www.cmc.edu/pages/faculty/DNeedell),
6) BP by linear programming (LP) (http://cvxr.com/cvx/).

For each recovery approach, we perform 10,000 independent trials and plot the empirical frequency of exact reconstruction as a function of the sparsity level $K$. By comparing the maximal sparsity level, i.e., the so called critical sparsity [22], of the sparse signals at which the exact reconstruction always ensured, recovery accuracy of different algorithms can be compared empirically. As shown in Fig. 1, for both sparse Gaussian and sparse 4-PAM signals, the MOLS algorithm outperforms other greedy approaches with respect to the critical sparsity. Even when compared to the BP method, the MOLS algorithm still exhibits very competitive reconstruction performance. For the Gaussian case, MOLS has much higher critical sparsity than BP, while for

3Note that for both gOMP and MOLS, the selection parameter should obey $L \leq K$. Hence, we choose $L = 3, 5$ in our simulation. Moreover, StOMP has two thresholding strategies: false alarm control (FAC) and false discovery control (FDC). We exclusively use FAC, since the FAC outperforms FDC.
Fig. 1. Frequency of exact reconstruction of $K$-sparse signals.

4-PAM case, MOLS and BP have almost identical performance.

In Fig. 2 we plot the running time and the number of iterations of MOLS and OLS for the exact reconstruction of $K$-sparse Gaussian and 4-PAM signals as a function of $K$. The running
time is measured using the MATLAB program under the 16-core 64-bit processor, 256 Gb RAM, and Window 8 environments. We observe that for both sparse Gaussian and 4-PAM cases, the number of iterations of MOLS for exact reconstruction is much smaller than that of the OLS algorithm, and accordingly, the associated running time of MOLS is much less than that of OLS.
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 inspired by the fact that sub-optimal candidates in the OLS identification are likely to be reliable and can be utilized to better reduce the residual energy at each time, thereby accelerating the convergence of the algorithm. We have demonstrated by RIP analysis that MOLS performs the exact recovery of any $K$-sparse signal within $K$ iterations if $\delta_{LK} \leq \sqrt{\frac{L}{K+\sqrt{5L}}}$. In particular, as a special case of MOLS when $L = 1$, the OLS algorithm exactly recovers any $K$-sparse signal in $K$ iterations under $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$, which coincides with the best-so-far recovery condition of OMP [12] and is proved to be nearly optimal for the OLS recovery. In addition, we have shown by empirical experiments that the MOLS algorithm has faster convergence and lower computational cost than the conventional OLS algorithm, while exhibiting improved recovery accuracy. Our empirical results have also shown that MOLS achieves very promising performance in recovering sparse signals when compared to the state of the art reconvey algorithms.

Acknowledgement

This work is supported in part by ONR-N00014-13-1-0764, NSF-III-1360971, AFOSR-FA9550-13-1-0137, and NSF-Bigdata-1419210.

Appendix A

Proof of Lemma 4

Proof: Since $\delta_{|S|} < 1$, $\Phi_S$ has full column rank. Suppose that $\Phi_S$ has singular value decomposition $\Phi_S = U \Sigma V^\prime$. Then from the definition of RIP, the maximum and minimum diagonal entries of $\Sigma$ (denoted by $\sigma_{max}$ and $\sigma_{min}$, respectively) satisfy

$$\sigma_{max} \leq \sqrt{1 + \delta_{|S|}} \quad \text{and} \quad \sigma_{min} \geq \sqrt{1 - \delta_{|S|}}.$$  

Note that

$$\left(\Phi_S^\dagger\right)^\prime = ((\Phi_S^\prime \Phi_S)^{-1} \Phi_S^\prime)^\prime = U \Sigma V^\prime ((U \Sigma V^\prime)^\prime (U \Sigma V^\prime)^{-1} = U \Sigma^{-1} V^\prime.$$  

(41)
where \( \Sigma^{-1} \) is the diagonal matrix formed by replacing every (non-zero) diagonal entry of \( \Sigma \) by its reciprocal. Therefore, all singular values of \( (\Phi_S^\dagger)’ \) lie between \( \frac{1}{\sigma_{\text{max}}} \) and \( \frac{1}{\sigma_{\text{min}}} \). This together with (41) implies that all singular values of \( (\Phi^\dagger)’ \) lie between \( \frac{1}{\sqrt{1+\delta|S|}} \) and \( \frac{1}{\sqrt{1-\delta|S|}} \), which completes the proof.

**APPENDIX B**

**PROOF OF (11)**

*Proof:* Since \( P_{T_k} + P_{T_k}^\perp = I \),

\[
P_{T_k \cup \{i\}} = P_{T_k} P_{T_k \cup \{i\}} + P_{T_k}^\perp P_{T_k \cup \{i\}}
\]

\[
= P_{T_k} + P_{T_k}^\perp P_{T_k \cup \{i\}}
\]

\[
= P_{T_k} + P_{T_k}^\perp [\Phi_{T_k}, \phi_i]
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}^{-1}
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}
\]

\[
= P_{T_k} + [0 P_{T_k}^\perp \phi_i]
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}
\begin{bmatrix}
M_1 & M_2 \\
M_3 & M_4
\end{bmatrix}
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}
\]

\[
= (a) P_{T_k} + [0 P_{T_k}^\perp \phi_i]
\begin{bmatrix}
M_1 & M_2 \\
M_3 & M_4
\end{bmatrix}
\begin{bmatrix}
\Phi_{T_k}' \\
\phi_i'
\end{bmatrix}
\]

(42)

where \((a)\) is from the partitioned inverse formula and

\[
M_1 = (\Phi_{T_k}' P_i^\perp \Phi_{T_k})^{-1},
\]

\[
M_1 = -(\Phi_{T_k}' P_i^\perp \Phi_{T_k})^{-1} \Phi_{T_k}' \phi_i (\phi_i')^{-1},
\]

\[
M_1 = (\Phi_{T_k}' P_i^\perp \Phi_{T_k})^{-1},
\]

\[
M_1 = (\Phi_{T_k}' P_i^\perp \Phi_{T_k})^{-1}.
\]

(43)

After some manipulations, we have

\[
P_{T_k \cup \{i\}}
\]

\[
= P_{T_k} + [0 P_{T_k}^\perp \phi_i (\phi_i' P_{T_k}^\perp \phi_i)^{-1} \phi_i' P_{T_k}^\perp]
\]

\[
= P_{T_k} + P_{T_k}^\perp \phi_i (\phi_i' P_{T_k}^\perp \phi_i)^{-1} \phi_i' P_{T_k}^\perp.
\]

(44)
APPENDIX C

PROOF OF PROPOSITION 21

Proof:

• Proof of (21)

Since $u_1$ is the largest value of \( \frac{|\langle \phi_i, r^k \rangle|}{\|P^\perp_{T_k} \phi_i\|_2} \), $i \in T \setminus T^k$, it is clear that

\[
    u_1 \geq \frac{1}{\sqrt{K - \ell}} \sum_{i \in T \setminus T^k} \frac{|\langle \phi_i, r^k \rangle|}{\|P^\perp_{T_k} \phi_i\|_2}
\]

\[
    \geq \frac{1}{\sqrt{K - \ell}} \sum_{i \in T \setminus T^k} |\langle \phi_i, r^k \rangle|
\]

\[
    = \frac{1}{\sqrt{K - \ell}} \|\Phi'_{T \setminus T^k} r^k\|_2
\]

\[
    \geq \frac{1}{\sqrt{K - \ell}} \|\Phi'_{T \setminus T^k} P^\perp_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2
\]

\[
    = \frac{1}{\sqrt{K - \ell}} \left( \|\Phi'_{T \setminus T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k} - \Phi'_{T \setminus T^k} P_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2 \right)
\]

\[
    \geq (a) \frac{1}{\sqrt{K - \ell}} \left( \|\Phi'_{T \setminus T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2 - \|\Phi'_{T \setminus T^k} P_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2 \right),
\]

where (a) uses the triangle inequality. Since $|T \setminus T^k| = K - \ell$,

\[
    \|\Phi'_{T \setminus T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2 \geq (1 - \delta_{K-\ell}) \|x_{T \setminus T^k}\|_2,
\]

and

\[
    \|\Phi'_{T \setminus T^k} P_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2
\]

\[
    = \|\Phi'_{T \setminus T^k} \Phi_{T \setminus T^k} (\Phi'_{T^k} \Phi_{T^k})^{-1} \Phi'_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2
\]

\[
    \leq (a) \delta_{L_k + K - \ell} \|\Phi'_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k}\|_2
\]

\[
    \leq (b) \frac{\delta_{L_k + K - \ell}}{1 - \delta_{L_k}} \|x_{T \setminus T^k}\|_2
\]

\[
    \leq \frac{\delta_{L_k + K - \ell}}{1 - \delta_{L_k}} \|x_{T \setminus T^k}\|_2,
\]

(45)
where (a) and (b) follow from Lemma 3 and 2, respectively (\(T^k\) and \(T \setminus T^k\) are disjoint sets and \(|T^k \cup (T \setminus T^k)| = Lk + K - \ell\)).

Finally, by combining (45), (46), and (47), we obtain

\[
u_1 \geq \left(1 - \delta_{K-\ell} - \frac{\delta_{Lk+K-\ell}}{1 - \delta_{Lk}}\right) \frac{\|x_{T \setminus T^k}\|_2}{\sqrt{K - \ell}}. \tag{48}
\]

- **Proof of (22)**

Let \(\mathcal{F}\) be the index set corresponding to \(L\) largest elements in \(\left\{ \frac{|\langle \phi_i, r^k \rangle|}{\|\mathbf{P}_{T^k \phi_i} \|_2} \right\}_{i \in H \setminus (T \cup T)}\). Then,

\[
\left(\sum_{i \in \mathcal{F}} \frac{|\langle \phi_i, r^k \rangle|^2}{\|\mathbf{P}_{T^k \phi_i} \|_2^2}\right)^{1/2} = \left(\frac{1}{\min_{i \in \mathcal{F}} \|\mathbf{P}_{T^k \phi_i} \|_2} \sum_{i \in \mathcal{F}} |\langle \phi_i, r^k \rangle|^2\right)^{1/2} \leq \left(\frac{1}{1 - \max_{i \in \mathcal{F}} \|\mathbf{P}_{T^k \phi_i} \|_2 \sum_{i \in \mathcal{F}} |\langle \phi_i, r^k \rangle|^2\right)^{1/2} = \left(\frac{1}{1 - \max_{i \in \mathcal{F}} \|\mathbf{P}_{T^k \phi_i} \|_2} \frac{\|\mathbf{\Phi}^T_{\mathcal{F}^T} r^k\|_2}{\|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T^k \phi_i} \|_2}\right)^{1/2} \leq (a) \left(\frac{1}{1 - \frac{\delta_{Lk+1}}{1 - \delta_{Lk}}}\right)^{1/2} \|\mathbf{\Phi}^T_{\mathcal{F}^T} r^k\|_2 \leq (b) \left(\frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}}\right)^{1/2} \|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2 \leq \left(\frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}}\right)^{1/2} \left(\|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2 - \mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2\right) \leq \left(\frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}}\right)^{1/2} \left(\|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2 + \|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2\right) \leq \left(\frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}}\right)^{1/2} \left(\|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2 + \|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2\right) \leq \left(\frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}}\right)^{1/2} \left(\|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2 + \|\mathbf{\Phi}^T_{\mathcal{F}^T} \mathbf{\Phi}_{T \setminus T^k} x_{T \setminus T^k}\|_2\right) \tag{49}
\]
where (a) is because for any \( i \in F \),
\[
\| (\Phi_{T_k}^\dagger)^T \Phi_{T_k} \phi_i \|_2^2 \leq \frac{1}{1 - \delta_{Lk}} \| \Phi_{T_k}^\dagger \phi_i \|_2^2 \\
\leq \frac{\delta_{Lk+1}^2}{1 - \delta_{Lk}} \| \phi_i \|_2^2 \\
= \frac{\delta_{Lk+1}^2}{1 - \delta_{Lk}},
\]
and (b) from \( r^k = y - \Phi x^k = y - \Phi_{T_k} \Phi_{T_k}^\dagger y = P_{T_k} y = P_{T_k} \Phi_{T_k} \Phi_{T_k}^\dagger y \).

Since \( F \) and \( T \setminus T^k \) are disjoint (i.e., \( F \cap (T \setminus T^k) = \emptyset \)) and \(|F| + |T \setminus T^k| = L + K - \ell \)
(note that \( T \cap T^k = \ell \) by hypothesis). Using this together with Lemma 3, we have
\[
\| \Phi_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \|_2 \leq \delta_{L+K-\ell} \| x_{T \setminus T^k} \|_2 .
\]
Moreover, since \( F \cap T^k = \emptyset \) and \(|F| + |T^k| = L + K \),
\[
\| \Phi_{T_k} P_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \|_2 \\
\leq \delta_{L+K} \| \Phi_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \|_2 \\
= \delta_{L+K} \| (\Phi_{T_k}^\dagger \Phi_{T_k}^\dagger)^{-1} \Phi_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \|_2 \\
\leq \frac{\delta_{L+K}}{1 - \delta_{Lk}} \| \Phi_{T_k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \|_2 \\
\leq \frac{\delta_{L+K} \delta_{L+K+K-\ell}}{1 - \delta_{Lk}} \| x_{T \setminus T^k} \|_2 ,
\]
where (a) is from Lemma 2 and (b) is due to Lemma 3 (since \( T^k \) and \( T \setminus T^k \) are disjoint
and \(|T \setminus T^k| = K - \ell, |T^k \cup (T \setminus T^k)| = Lk + K - \ell \)).

Invoking (51) and (52) into (49), we have
\[
\left( \sum_{i \in F} \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2} \right)^{1/2} \leq \left( \frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}^2} \right)^{1/2} \\
\times \left( \delta_{L+K-\ell} + \frac{\delta_{L+K} \delta_{L+K+K-\ell}}{1 - \delta_{Lk}} \right) \| x_{T \setminus T^k} \|_2 .
\]
On the other hand, by noting that \( v_L \) is the \( L \)-th largest value of \( \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2}, i \in F \), we obtain
\[
\left( \sum_{i \in F} \frac{|\langle \phi_i, r^k \rangle|^2}{\| P_{T_k} \phi_i \|_2^2} \right)^{1/2} \geq \sqrt{L} v_L .
\]
From (53) and (54), we have

\[
v_L \leq \left( \frac{1 - \delta_{Lk}}{1 - \delta_{Lk} - \delta_{Lk+1}^2} \right)^{1/2} \times \left( \delta_{L+K-\ell} + \frac{\delta_{L+Lk} \delta_{Lk+K-\ell}}{1 - \delta_{Lk}} \right) \frac{\|X_T \setminus T^k\|_2}{\sqrt{L}}.
\]

(55)

\textbf{APPENDIX D}

\textbf{PROOF OF (27)}

\textit{Proof:} Observe that (26) is equivalent to

\[
\sqrt{\frac{K}{L}} \leq \frac{(1 - 2\delta_{LK})(1 - \delta_{LK} - \delta_{Lk}^2)^{1/2}}{\delta_{LK}(1 - \delta_{LK})^{1/2}}.
\]

(56)

Let

\[
f(\delta_{LK}) = \frac{(1 - 2\delta_{LK})(1 - \delta_{LK} - \delta_{Lk}^2)^{1/2}}{\delta_{LK}(1 - \delta_{LK})^{1/2}}
\]

and

\[
g(\delta_{LK}) = \frac{1}{\delta_{LK}} - \sqrt{5}.
\]

Then one can check that \( \forall \delta_{LK} \in (0, \frac{1}{2}) \), \( f(\delta_{LK}) > g(\delta_{LK}) \). Hence, (56) is guaranteed under

\[
\sqrt{\frac{K}{L}} \leq \frac{1}{\delta_{LK}} - \sqrt{5}.
\]

(57)

Equivalently,

\[
\delta_{LK} \leq \frac{\sqrt{L}}{\sqrt{K} + \sqrt{5L}}.
\]

(58)

\textbf{REFERENCES}

[1] D. L. Donoho and P. B. Stark, “Uncertainty principles and signal recovery,” SIAM Journal on Applied Mathematics, vol. 49, no. 3, pp. 906–931, 1989.

[2] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” IEEE Trans. Inform. Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM review, pp. 129–159, 2001.
[6] 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.
[7] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers. IEEE, Nov. Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.
[8] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” IEEE Trans. Inform. Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[9] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inform. Theory, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[10] M. A. Davenport and M. B. Wakin, “Analysis of Orthogonal Matching Pursuit using the restricted isometry property,” IEEE Trans. Inform. Theory, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.
[11] T. Zhang, “Sparse recovery with orthogonal matching pursuit under rIP,” IEEE Trans. Inform. Theory, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.
[12] 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.
[13] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” International Journal of control, vol. 50, no. 5, pp. 1873–1896, 1989.
[14] L. Rebollo-Neira and D. Lowe, “Optimized orthogonal matching pursuit approach,” IEEE Signal Processing Letters, vol. 9, no. 4, pp. 137–140, Apr. 2002.
[15] S. Foucart, “Stability and robustness of weak orthogonal matching pursuits,” in Recent Advances in Harmonic Analysis and Applications, pp. 395–405. Springer, 2013.
[16] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, “Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares,” IEEE Trans. Inform. Theory, vol. 59, no. 5, pp. 3158–3174, May 2013.
[17] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” IEEE Trans. Inform. Theory, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.
[18] 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.
[19] J. Wang, S. Kwon, and B. Shim, “Generalized orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
[20] E. Liu and V. N. Temlyakov, “The orthogonal super greedy algorithm and applications in compressed sensing,” IEEE Trans. Inform. Theory, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.
[21] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 301–321, Mar. 2009.
[22] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” IEEE Trans. Inform. Theory, vol. 55, no. 5, pp. 2230–2249, May. 2009.
[23] S. Foucart, “Hard thresholding pursuit: an algorithm for compressive sensing,” SIAM Journal on Numerical Analysis, vol. 49, no. 6, pp. 2543–2563, 2011.
[24] E. J. Candès and T. Tao, “Decoding by linear programming,” IEEE Trans. Inform. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[25] S. Kwon, J. Wang, and B. Shim, “Multipath matching pursuit,” IEEE Trans. Inform. Theory, vol. 60, no. 5, pp. 2986–3001, May 2014.

[26] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9-10, pp. 589–592, 2008.

[27] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, “Error correction via linear programming,” in IEEE Symposium on Foundations of Computer Science (FOCS), 2005, pp. 668–681.