An Optimal Restricted Isometry Condition for Exact Sparse Recovery with Orthogonal Least Squares

Junhan Kim, Jian Wang and Byonghyo Shim

Abstract—The orthogonal least squares (OLS) algorithm is popularly used in sparse recovery, subset selection, and function approximation. In this paper, we analyze the performance guarantee of OLS. Specifically, we show that if a sampling matrix $\Phi$ has unit $\ell_2$-norm columns and satisfies the restricted isometry property (RIP) of order $K + 1$ with

$$\delta_{K+1} < C_K$$

then OLS exactly recovers any $K$-sparse vector $x$ from its measurements $y = \Phi x$ in $K$ iterations. Furthermore, we show that the proposed guarantee is optimal in the sense that OLS may fail the recovery under $\delta_{K+1} \geq C_K$. Additionally, we show that if the columns of a sampling matrix are $\ell_2$-normalized, then the proposed condition is also an optimal recovery guarantee for the orthogonal matching pursuit (OMP) algorithm. Also, we establish a recovery guarantee of OLS in the more general case where a sampling matrix might not have unit $\ell_2$-norm columns. Moreover, we analyze the performance of OLS in the noisy case. Our result demonstrates that under a suitable constraint on the minimum magnitude of nonzero elements in an input signal, the proposed RIP condition ensures OLS to identify the support exactly.

Index Terms—Orthogonal least squares (OLS), orthogonal matching pursuit (OMP), restricted isometry property (RIP), sparse recovery

I. INTRODUCTION

ORTHOGONAL least squares (OLS) is a popular algorithm for sparse recovery, subset selection, and function approximation [1], [2]. The main goal of the OLS algorithm is to reconstruct a high dimensional $K$-sparse vector $x \in \mathbb{R}^n$ ($\|x\|_0 \leq K \ll n$) from a small number of linear measurements

$$y = \Phi x,$$  \hspace{1cm} (1)

where $\Phi \in \mathbb{R}^{m \times n}$ ($m \ll n$) is the sampling (measurement) matrix. In order to solve the problem, OLS identifies the support (the index set of nonzero entries) of $x$ sequentially in a way to minimize the residual power. To be specific, in each iteration, OLS identifies the column of $\Phi$ that leads to the most significant reduction in the $\ell_2$-norm of a residual

and then adds the index of the column to the estimated support. The vestige of indices in the enlarged support is then removed from $y$, generating an updated residual for the upcoming iteration (see Table I).

As a framework to analyze the performance of the OLS algorithm, the restricted isometry property (RIP) has been widely employed [3]–[8]. A matrix $\Phi$ is said to satisfy the RIP of order $K$ if there exists a constant $\delta \in [0, 1)$ such that

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2$$  \hspace{1cm} (2)

for any $K$-sparse vector $x$ [9]. In particular, the minimum of $\delta$ satisfying (2) is called the RIP constant and denoted by $\delta_K$. In [3], it has been shown that OLS guarantees the exact recovery of any $K$-sparse vector in $K$ iterations if a sampling matrix has unit $\ell_2$-norm columns and obeys the RIP with

$$\delta_{K+1} < \frac{1}{\sqrt{K + 2}}.$$  \hspace{1cm} (3)

This condition has been improved to [6]

$$\delta_{K+1} < \frac{1}{\sqrt{K + 1}}.$$  \hspace{1cm} (4)

On the other hand, it has been reported in [6] that when $K = 2$, there exists a counterexample for which OLS fails.
the reconstruction under
\[ \delta_{K+1} = \frac{1}{\sqrt{K + \frac{4}{3}}}. \]

Based on the counterexample, it has been conjectured that for all of \( K \), a recovery guarantee of OLS cannot be weaker than
\[ \delta_{K+1} \leq \frac{1}{\sqrt{K + \frac{4}{3}}}. \quad (5) \]

An aim of this paper is to bridge the gap between (4) and (5). Towards this end, we first put forth an improved performance guarantee of OLS. Specifically, we show that OLS exactly recovers any \( K \)-sparse vector in \( K \) iterations, provided that a sampling matrix has unit \( \ell_2 \)-norm columns and satisfies the RIP of order \( K + 1 \) with (Theorem 1)

\[ \delta_{K+1} < C_K = \begin{cases} \frac{1}{\sqrt{K}}, & K = 1, \\ \frac{1}{\sqrt{K+4}}, & K = 2, \\ \frac{1}{\sqrt{K+6}}, & K = 3, \\ \frac{1}{\sqrt{K+8}}, & K \geq 4. \end{cases} \quad (6) \]

The significance of our result lies in the fact that it not only outperforms the existing result in (4), but also states an optimal recovery guarantee of OLS. By the optimality, we mean that if the condition is violated, then there exists a counterexample for which OLS fails the recovery. In fact, for any positive integer \( K \) and constant
\[ \delta^* \in [C_K, 1), \]
there always exist an \( \ell_2 \)-normalized matrix \( \Phi \) with \( \delta_{K+1} = \delta^* \) and a \( K \)-sparse vector \( x \) such that \( x \) cannot be recovered from \( y = \Phi x \) by OLS (Theorem 2).

We note that the aforementioned results are based on the assumption that the columns of a sampling matrix are \( \ell_2 \)-normalized. This assumption might not be satisfied in some applications (e.g., \( \Phi \) is Gaussian random [9]). For this case, we establish a condition of general (not necessarily \( \ell_2 \)-normalized) sampling matrices for the success\(^1\) of OLS (Theorem 3). By comparing our result with existing ones, we show that our result outperforms the existing one by a large margin.

We also provide a performance guarantee of OLS in the noisy scenario. Specifically, we study a sufficient condition under which OLS identifies the support of \( x \) accurately in the presence of measurement noise. Our result states that under a suitable condition on the minimum magnitude of nonzero elements in \( x \), (6) ensures OLS to reconstruct the support of \( x \) accurately (Theorem 4). We show that the proposed sufficient condition is better (more relaxed) than existing ones.

Finally, we extend our analyses for orthogonal matching pursuit (OMP) [10], [11], which is the most famous sparse recovery algorithm. Our results demonstrate that if an \( \ell_2 \)-normalized sampling matrix is employed, then (6) is also an optimal performance guarantee of OMP in the noiseless scenario (Theorems 5 and 6). We also show that under a proper constraint on the minimum magnitude of nonzero entries in \( x \), the proposed RIP condition (6) ensures OMP to identify the support of \( x \) exactly in the noisy scenario (Theorem 7). Through these results, we also demonstrate that the RIP-based performance guarantee of OMP can be improved by employing an \( \ell_2 \)-normalized sampling matrix.

II. Preliminaries

We first summarize the notations used in this paper.

- \( \Omega = \{1, \ldots, n\} \);
- \( T = \text{supp}(x) = \{ i \in \Omega : x_i \neq 0 \} \) is the support of \( x \);
- For any \( J \subseteq \Omega \), \( |J| \) is the cardinality of \( J \) and \( T \setminus J = \{ i : i \in T, i \notin J \} \);
- \( x_J \in \mathbb{R}^{|J|} \) is the restriction of \( x \in \mathbb{R}^n \) to the elements indexed by \( J \);
- \( \phi_i \in \mathbb{R}^m \) is the \( i \)-th column of \( \Phi \in \mathbb{R}^{m \times n} \);
- \( \Phi_j \in \mathbb{R}^{m \times |J|} \) is the submatrix of \( \Phi \) with the columns indexed by \( J \);
- \( \sigma_{\min}(\Phi_j) \) is the smallest singular value of \( \Phi_j \);
- \( 0_{d_1 \times d_2} \) and \( 1_{d_1 \times d_2} \) are \( (d_1 \times d_2) \)-dimensional matrices with entries being zeros and ones, respectively;
- \( \text{Id}_d \) is the \( d \)-dimensional identity matrix;
- \( P_J = \Phi_J \Phi_J^\dagger \) and \( P_J^* = \text{Id}_m - P_J \) are the orthogonal projections onto \( \text{span}(\Phi_J) \) and its orthogonal complement, respectively.

We next give some lemmas useful in our analysis. The first lemma is about the monotonicity of the RIP constant.

Lemma 1 ([9], [13, Lemma 1]). If a matrix \( \Phi \in \mathbb{R}^{m \times n} \) satisfies the RIP of orders \( K_1 \) and \( K_2 \) \((K_1 \leq K_2)\), then \( \delta_{K_1} \leq \delta_{K_2} \).

The second lemma is often called the modified RIP of a projected matrix.

Lemma 2 ([14, Lemma 1]). Let \( T, J \subseteq \Omega \). If a matrix \( \Phi \in \mathbb{R}^{m \times n} \) satisfies the RIP of order \( |T \cup J| \), then for any \( x \in \mathbb{R}^n \) with \( \text{supp}(x) = T \),
\[ (1 - \delta_{|T \cup J|}) \| x_{T \setminus J} \|_2^2 \leq \| P_{J} \Phi x \|_2^2 \leq (1 + \delta_{|T \cup J|}) \| x_{T \setminus J} \|_2^2. \]

The third lemma gives an equivalent form to the identification rule of OLS in Table I.

Lemma 3 ([2, Theorem 1]). Consider the system model in (1). Let \( T^k \) and \( x^k \) be the estimated support and the residual generated in the \( k \)-th iteration of the OLS algorithm, respectively. Then the index \( k+1 \) chosen in the \((k+1)\)-th iteration of OLS satisfies
\[ t_{k+1} = \arg \max_j \| (t^k, \phi_j) \|_2. \quad (7) \]
One can deduce from (7) that OLS picks a support index in the $(k+1)$-th iteration (i.e., $k^+T \in T)$ if and only if

$$\max_{j \in \Omega \cap T} \frac{|\langle \Phi_j, \phi^j \rangle |}{\|P_{T^k}^+ \Phi_j\|_2} > \max_{j \in \Omega \cap (T^k \cap T^k)} \frac{|\langle \Phi_j, \phi^j \rangle |}{\|P_{T^k}^+ \Phi_j\|_2}.$$  \hspace{1cm} (8)

To examine if OLS is successful in the $(k+1)$-th iteration, therefore, it suffices to check whether (8) holds.

The next lemma plays a key role in bounding the right-hand side of (8).

**Lemma 4.** Consider the system model in (1) where $\Phi$ has unit $\ell_2$-norm columns. Let $T^k$ be a subset of $T = \text{supp}(x)$ and $r^k = P_{T^k}^+ \Phi x$. If $\Phi$ obeys the RIP of order $K + 1$ with

$$\delta_{K+1} \leq \frac{1}{2},$$  \hspace{1cm} (9)

then

$$\max_{j \in \Omega \cap T} \frac{|\langle \Phi_j, \phi^j \rangle |}{\|P_{T^k}^+ \Phi_j\|_2} \leq \frac{\delta_{K+1}\|\phi^k\|_2^2}{\|X_{T^k \cap T^k}\|_2}.$$  \hspace{1cm} (10)

**Proof.** Let $j \in \Omega \cap T$ and $t = \frac{|\langle \Phi_j, \phi^j \rangle |}{\|P_{T^k}^+ \Phi_j\|_2}$. Then, what we need to show is

$$t \leq \delta_{K+1} \frac{\|r^k\|_2^2}{2}$$  \hspace{1cm} (11)

under (9). Let $\Psi = P_{T^k}^+ [\Phi_{T^k \cap T^k}]$ and

$$\mathbf{u} = \left[ \begin{array}{c} \frac{X_{T^k \cap T^k} \Phi}{\|P_{T^k} \Phi\|_2} \\ \frac{\text{sgn}(\phi^j \Phi^j)}{\|P_{T^k}^+ \Phi_j\|_2} \end{array} \right], \quad \mathbf{v} = \left[ \begin{array}{c} X_{T^k \cap T^k} \Phi \\ \frac{\text{sgn}(\phi^j \Phi^j)}{\|P_{T^k}^+ \Phi_j\|_2} \end{array} \right],$$  \hspace{1cm} (12)

where \(\text{sgn}(\cdot)\) is the signum function. Noting that $r^k = P_{T^k}^+ \Phi r^k x_{T^k \cap T^k}$ and $P_{T^k}^+ = (P_{T^k}^+)^2$, we have

$$\|\Psi \mathbf{u}\|_2^2 = \|r^k - \text{sgn}(\phi^j \Phi^j) x_{T^k \cap T^k}\|_2^2 \cdot \frac{P_{T^k}^+ \Phi_j}{\|P_{T^k}^+ \Phi_j\|_2} = \|r^k\|_2^2 - 2t + \|x_{T^k \cap T^k}\|_2^2$$  \hspace{1cm} (13)

and

$$\|\Psi \mathbf{v}\|_2^2 = \|r^k + \text{sgn}(\phi^j \Phi^j) x_{T^k \cap T^k}\|_2^2 \cdot \frac{P_{T^k}^+ \Phi_j}{\|P_{T^k}^+ \Phi_j\|_2} = \|r^k\|_2^2 + 2t + \|x_{T^k \cap T^k}\|_2^2.$$  \hspace{1cm} (14)

Also, from Lemma 2, we have

$$\|\Psi \mathbf{u}\|_2^2 \geq (1 - \delta_{K+1}) \left( 1 + \frac{1}{\|P_{T^k} \Phi\|_2} \right) \|x_{T^k \cap T^k}\|_2^2$$  \hspace{1cm} (15)

and

$$\|\Psi \mathbf{v}\|_2^2 \leq (1 + \delta_{K+1}) \left( 1 + \frac{1}{\|P_{T^k} \Phi\|_2} \right) \|x_{T^k \cap T^k}\|_2^2.$$  \hspace{1cm} (16)

By combining (13)-(16), we obtain

$$\|r^k\|_2^2 - 2t \geq (1 - (1 + \|P_{T^k}^+ \Phi_j\|_2^2) \delta_{K+1}) \|x_{T^k \cap T^k}\|_2^2 \frac{\|r^k\|_2^2}{\|P_{T^k}^+ \Phi_j\|_2^2}$$  \hspace{1cm} (17)

and

$$\|r^k\|_2^2 + 2t \leq (1 + (1 + \|P_{T^k}^+ \Phi_j\|_2^2) \delta_{K+1}) \|x_{T^k \cap T^k}\|_2^2 \frac{\|r^k\|_2^2}{\|P_{T^k}^+ \Phi_j\|_2^2}.$$  \hspace{1cm} (18)

Note that since $\Phi$ has unit $\ell_2$-norm columns, $\|P_{T^k}^+ \Phi_j\|_2^2 \leq \|\Phi_j\|_2^2 = 1$ and thus we have

$$1 - (1 + \|P_{T^k}^+ \Phi_j\|_2^2) \delta_{K+1} \geq 1 - 2\delta_{K+1} \geq 0,$$

where (a) is from (9). Then, by combining (17) and (18), we have

$$\|r^k\|_2^2 - 2t \geq (\|r^k\|_2^2 + 2t) \{1 - (1 + \|P_{T^k}^+ \Phi_j\|_2^2) \delta_{K+1}\},$$

which is equivalent to

$$2t \geq (1 + \|P_{T^k}^+ \Phi_j\|_2^2) \delta_{K+1} \|r^k\|_2^2.$$

Finally, by exploiting $\|P_{T^k}^+ \Phi_j\|_2^2 \leq 1$, we obtain the desired result (11). \(\square\)

**Remark 1.** The bound in (10) is tight in the sense that the equality of (10) is attainable. To see this, consider the following example:

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1_{K \times 1} \end{bmatrix} \text{ and } \Phi = \begin{bmatrix} \sqrt{\frac{1}{K}} \mathbf{1}_{K \times 1} \\ \mathbf{I}_K \end{bmatrix}.$$  \hspace{2cm} (11)

We now take a look at the left- and right-hand sides of (10) when $k = 0$. Since $y = \Phi x = \mathbf{x}$, the left-hand side of (10) is

$$\max_{j \in \Omega \cap T} \frac{|\langle \phi^j, \phi \rangle |}{\|P_{T^k}^+ \Phi_j\|_2} = |\langle \mathbf{y}, \phi_1 \rangle| = 1.$$  \hspace{2cm} (12)

Furthermore, since the RIP constant $\delta_{K+1}$ of $\Phi$ is $\delta_{K+1} = \frac{1}{\sqrt{K}}$ (see [19, p.3655]), the right-hand side of (10) is also

$$\frac{\|r^k\|_2^2}{\|x_{T^k \cap T^k}\|_2^2} = \delta_{K+1} \|\mathbf{y}\|_2^2 = 1.$$  \hspace{2cm} (13)

Therefore, the bound in (10) is tight.

**Remark 2.** Lemma 4 is motivated by [6, Lemma 4], where the inequality

$$\max_{j \in \Omega \cap T} |\langle \phi^j, \phi \rangle | \leq \frac{\|r^k\|_2^2}{\sqrt{\alpha} \|x_{T^k \cap T^k}\|_2}$$  \hspace{2cm} (14)

was established under

$$\delta_{K+1} \leq \frac{1}{\sqrt{\alpha} + 1}.$$  \hspace{2cm} (15)

We note that our result in Lemma 4 outperforms this result in the following aspects:

i) Under the same RIP condition ($\delta_{K+1} \leq \frac{1}{\sqrt{\alpha} + 1}$), a tighter upper bound of $\max_{j \in \Omega \cap T} |\langle \phi^j, \phi \rangle |$ can be established using Lemma 4. By applying [6, Lemma 4] with $\alpha = 3$, we have

$$\max_{j \in \Omega \cap T} |\langle \phi^j, \phi \rangle | \leq \frac{\|r^k\|_2^2}{\sqrt{3} \|x_{T^k \cap T^k}\|_2}$$  \hspace{2cm} (16)

and

$$\max_{j \in \Omega \cap T} |\langle \phi^j, \phi \rangle | \leq \frac{\|r^k\|_2^2}{\sqrt{\alpha} \|x_{T^k \cap T^k}\|_2}$$  \hspace{2cm} (17)
under $\delta_{K+1} \leq \frac{1}{2}$. Under the same condition on $\delta_{K+1}$, it can be deduced from Lemma 4 that

$$
\max_{j \in \Omega \setminus T} \frac{|\langle \mathbf{r}^k, \phi_j \rangle |}{\| \mathbf{P}^k_{T^k} \phi_j \|_2} \leq \frac{\delta_{K+1} \| \mathbf{r}^k \|_2^2}{\| x_{T^k \setminus T^k} \|_2} \leq \frac{\| \mathbf{r}^k \|_2^2}{2\| x_{T^k \setminus T^k} \|_2^2},
$$

where (a) is because $\| \mathbf{P}^k_{T^k} \phi_j \|_2 \leq \| \phi_j \|_2 = 1$ for each of $j \in \Omega \setminus T$. Clearly, the bound in (22) is tighter than that in (21) by the factor of $\frac{\sqrt{2}}{\sqrt{3}}$.

ii) In [6], by putting $\alpha = \| T \setminus T^k / \| \mathbf{P}^k_{T^k} \phi_j \|_2$ into (19), the inequality

$$
\max_{j \in \Omega \setminus T} \frac{|\langle \mathbf{r}^k, \phi_j \rangle |}{\| \mathbf{P}^k_{T^k} \phi_j \|_2} < \frac{\| \mathbf{r}^k \|_2^2}{\sqrt{T \setminus T^k / \| x_{T \setminus T^k} \|_2}}
$$

is established under

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

We mention that the inequality (23) can be established under a weaker RIP condition using Lemma 4. Specifically, when $K \geq 4$, (23) can be obtained from Lemma 4 under

$$
\delta_{K+1} < \frac{1}{\sqrt{K}}
$$

because

$$
\max_{j \in \Omega \setminus T} \frac{|\langle \mathbf{r}^k, \phi_j \rangle |}{\| \mathbf{P}^k_{T^k} \phi_j \|_2} \leq \frac{\delta_{K+1} \| \mathbf{r}^k \|_2^2}{\| x_{T \setminus T^k} \|_2} \leq \frac{\| \mathbf{r}^k \|_2^2}{\sqrt{T \setminus T^k / \| x_{T \setminus T^k} \|_2}} \leq \frac{\| \mathbf{r}^k \|_2^2}{\sqrt{T \setminus T^k / \| x_{T \setminus T^k} \|_2}}.
$$

Clearly, (25) is less restrictive than (24) in all sparsity region.

III. OPTIMAL PERFORMANCE GUARANTEE OF OLS

In this section, we present an optimal RIP condition for the success of the OLS algorithm when an $\ell_2$-normalized sampling matrix is employed. First, we present a sufficient condition ensuring the success of OLS.

**Theorem 1.** Let $\Phi \in \mathbb{R}^{m \times n}$ be a sampling matrix having unit $\ell_2$-norm columns. If $\Phi$ satisfies the RIP of order $K+1$ with

$$
\delta_{K+1} < C_K = \begin{cases} 
\frac{1}{\sqrt{K}}, & K = 1, \\
\frac{1}{\sqrt{K^2 + 1}} & K = 2, \\
\frac{1}{\sqrt{K + 1}} & K = 3, \\
\frac{1}{\sqrt{K}} & K \geq 4,
\end{cases}
$$

then the OLS algorithm exactly reconstructs any $K$-sparse vector $\mathbf{x} \in \mathbb{R}^n$ from its samples $\mathbf{y} = \Phi \mathbf{x}$ in $K$ iterations.

**Proof.** We show that the OLS algorithm picks a support index in each iteration. In other words, we show that $T^k \subset T$ for all $k \in \{0, \ldots, K\}$. In doing so, we have $T^K = T$, and thus OLS can recover $\mathbf{x}$ accurately:

$$
(x^K)_{T^k} = A^\dagger_{T^k} y = A^\dagger_{T^k} A_{T^k} x_T = A^\dagger_{T} A_{T} x_T = x_T. \tag{27}
$$

First, we consider the case for $k = 0$. This case is trivial since

$$
T^0 = \emptyset \subset T.
$$

Next, we assume that $T^k \subset T$ for some integer $k (0 \leq k < K)$ and then show that OLS picks a support index in the $(k+1)$-th iteration. As mentioned, $T^{k+1} \in T$ if and only if the condition (8) holds. Since the left-hand side of (8) satisfies [6, Proposition 2]

$$
\max_{j \in \Omega \setminus T} \frac{|\langle \mathbf{r}^k, \phi_j \rangle |}{\| \mathbf{P}^k_{T^k} \phi_j \|_2} \geq \frac{\| \mathbf{r}^k \|_2^2}{\sqrt{K - k} \| x_{T \setminus T^k} \|_2},
$$

it suffices to show that

$$
\max_{j \in \Omega \setminus T} \frac{|\langle \mathbf{r}^k, \phi_j \rangle |}{\| \mathbf{P}^k_{T^k} \phi_j \|_2} < \frac{\| \mathbf{r}^k \|_2^2}{\sqrt{K - k} \| x_{T \setminus T^k} \|_2}. \tag{28}
$$

Towards this end, we consider three cases: (i) $K = 1$, (ii) $K \in \{2, 3\}$, and (iii) $K \geq 4$.

(i) $K = 1$

In this case, $k = 0$ and thus we need to show that

$$
\max_{j \in \Omega \setminus T} |\langle \mathbf{y}, \phi_j \rangle | < \frac{\| \mathbf{y} \|_2^2}{\| x_T \|_2}. \tag{29}
$$

Without loss of generality, we assume that

$$
\mathbf{x} = \begin{bmatrix} c \\
0_{(n-1) \times 1}
\end{bmatrix}
$$

for some $c \in \mathbb{R} \setminus \{0\}$. Then, $\mathbf{y} = c \phi_1$ and thus the right-hand side of (29) is simply $|c|$. As a result, it suffices to show that

$$
\max_{j \in \Omega \setminus T} |\langle \mathbf{y}, \phi_j \rangle | < |c|. \tag{30}
$$

Let $j \in \Omega \setminus T$ and $\theta$ be the angle between $\phi_1$ and $\phi_j$ ($0 \leq \theta \leq \pi$). Then, by [22, Lemma 2.1], we have

$$
|\cos \theta| \leq \delta_2
$$

and hence

$$
|\langle \mathbf{y}, \phi_j \rangle | = |\langle c \phi_1, \phi_j \rangle | = |c| \cos \theta \leq |c| \delta_2. \tag{31}
$$

Using this together with (26), we obtain the desired result (30).

(ii) $K \in \{2, 3\}$

Let $j \in \Omega \setminus T$ and

$$
q = \sqrt{K - k} \| x_{T \setminus T^k} \|_2 / \| \mathbf{P}^k_{T^k} \phi_j \|_2.
$$

Then, our task is to show that

$$
q < \| \mathbf{r}^k \|_2. \tag{32}
$$
Let $\Psi = P_{T^k}^{\perp} \Phi_{T \setminus T^k} \phi_j$ and

$$w = \left[ -\frac{\text{sgn}(\phi_j^T r^k) \sqrt{K - k}}{\|x_{T \setminus T^k}\|_2} \right].$$

(33)

Then, by noting that $r^k = P_{T^k}^{\perp} \Phi_{T \setminus T^k} x_{T \setminus T^k}$ and $P_{T^k}^{\perp} = (P_{T^k}^{\perp})^T = (P_{T^k})^2$, we have

$$\|\Psi w\|_2^2 = \left| r^k - \frac{\text{sgn}(\phi_j^T r^k) \sqrt{K - k}}{2} \|x_{T \setminus T^k}\|_2 \right|^2,$$

(34)

Also, from Lemma 2, we have

$$\|\Psi w\|_2^2 \geq (1 - \delta_{K+1}) \left( 1 + \frac{K - k}{4 \|P_{T^k}^{\perp} \phi_j\|_2^2} \right) \|x_{T \setminus T^k}\|_2^2,$$

(35)

where (a) is because $\|P_{T^k}^{\perp} \phi_j\|_2 \leq \|\phi_j\|_2 = 1$. By combining (34) and (35), we obtain

$$\|r^k\|_2^2 - q \geq \left( 1 - \left( 1 + \frac{K - k}{4} \right) \delta_{K+1} \right) \|x_{T \setminus T^k}\|_2^2,$$

(36)

$$\geq \left( 1 - \left( 1 + \frac{K - k}{4} \right) C_K \right) \|x_{T \setminus T^k}\|_2^2,$$

(37)

where (a) follows from (26) and (b) is because

$$\left( 1 + \frac{K}{4} \right) C_K = 1$$

for each of $K \in \{2, 3\}$ (see (26)). Thus, we have (32), which is the desired result.

(iii) $K \geq 4$

In this case, the RIP constant $\delta_{K+1}$ of $\Phi$ satisfies

$$\delta_{K+1} \leq \frac{1}{\sqrt{K}} \leq \frac{1}{2},$$

where (a) is from (26). Then, by applying Lemma 4, we obtain

$$\max_{j \in \Omega \setminus T} \|P_{T^k}^{\perp} \phi_j\|_2 \leq \frac{\delta_{K+1} \|r^k\|_2^2}{\|x_{T \setminus T^k}\|_2^2},$$

(38)

where (a) is from (26). This completes the proof. □

One can observe from (31) that (30) holds if $\phi_1$ and $\phi_j$ are linearly independent.\(^\#\) This implies that when $K = 1$, OLS ensures the perfect recovery of $x$ if any two columns of $\Phi$ are linearly independent (i.e., $\text{rank}(\Phi) \geq 2$). We note that this condition is the fundamental limit (i.e., the minimum requirement on $\Phi$) to guarantee the exact sparse recovery [15, Theorem 2].

There have been previous efforts to analyze a recovery guarantee of the OLS algorithm [3], [6]–[8]. So far, the best result states that OLS guarantees the exact reconstruction under [6, Theorem 1]

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

(39)

Clearly, the proposed guarantee (26) is less restrictive than this condition in all sparsity region. One might wonder about the key difference between our analysis and previous ones. One major concern in the analysis of OLS lies in dealing with the denominator term $\|P_{T^k}^{\perp} \phi_j\|_2$ in (7), and various efforts have been made to handle this term. In [3, eq. (E.7)], the inequality

$$\|P_{T^k}^{\perp} \phi_j\|_2 \geq \sqrt{1 - \frac{\delta_{T^k+1}^2}{1 - \delta_{T^k}}}$$

was developed and employed to establish (3). This inequality was recently improved to [6, 16]

$$\|P_{T^k}^{\perp} \phi_j\|_2 \geq \sqrt{1 - \frac{\delta_{T^k+1}^2}{1 - \delta_{T^k}}},$$

which leads to the improved performance guarantee (38). While previous studies have focused on constructing a tight lower bound of $\|P_{T^k}^{\perp} \phi_j\|_2$ and using the bound as an estimate of $\|P_{T^k}^{\perp} \phi_j\|_2$, we do not approximate this term to prevent the loss, if any, caused by its relaxation. In fact, by properly defining some vectors incorporated with $\|P_{T^k}^{\perp} \phi_j\|_2$ (e.g., $u$ and $v$ in (12)), the inequality (10) could be obtained without any relaxation on $\|P_{T^k}^{\perp} \phi_j\|_2$, which eventually leads us to obtain the improved guarantee (26).

We now demonstrate that the proposed guarantee (26) is optimal. This argument is established by showing that if $\delta_{K+1} \geq C_K$, then there exists a counterexample for which OLS fails the recovery.

**Theorem 2.** For any positive integer $K$ and constant $\delta^* \in [C_K, 1],$

$$\delta^* \leq \frac{1}{\sqrt{K}} \leq \frac{1}{2},$$

there always exist an $f_2$-normalized matrix $\Phi$ with $\delta_{K+1} = \delta^*$ and a $K$-sparse vector $x$ such that the OLS algorithm fails to recover $x$ from $y = \Phi x$ in $K$ iterations.

**Proof.** It is enough to consider the case where $K > 1$ since there is no $\delta^*$ satisfying $\delta^* \in [C_K, 1]$ when $K = 1$. In our proof, we consider three cases: (i) $K = 2$, (ii) $K = 3$, and (iii) $K \geq 4$.

(i) $K = 2$

\(^\#\)If $\phi_1$ and $\phi_j$ are linearly independent, then $\theta$ cannot be zero or $\pi$. Thus, $0 < \theta < \pi$ and then $|\cos \theta| < 1$.\(^\#\)
We consider $x = [0 1 1]'$ and an $\ell_2$-normalized matrix $\Phi$ satisfying (see Appendix A for details)

$$\Phi'\Phi = \begin{bmatrix} 1 & \frac{\delta^*}{2} & \frac{\delta^*}{4} \\ \frac{\delta^*}{2} & 1 & -\frac{\delta^*}{4} \\ \frac{\delta^*}{4} & -\frac{\delta^*}{4} & 1 \end{bmatrix}. \quad (40)$$

First, we compute the RIP constant of $\Phi$. Note that the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of $\Phi'\Phi$ are

$$\lambda_1 = \lambda_2 = 1 + \frac{\delta^*}{2}, \quad \lambda_3 = 1 - \delta^*.$$  

Then, by exploiting the connection between the eigenvalues of $\Phi'\Phi$ and the RIP constant of $\Phi$ [13, Remark 1], we have

$$\delta_3 = \max_{i \in \{1,2,3\}} |\lambda_i - 1| = \delta^*.$$  

In short, $\Phi$ is an $\ell_2$-normalized matrix satisfying the RIP with $\delta_{K+1} = \delta^*$. We now take a look at the first iteration of the OLS algorithm. Note that

$$|\langle y, \phi_j \rangle| = |(\Phi'\Phi x)_j| = \begin{cases} \delta^*, & j = 1, \\ 1 - \frac{\delta^*}{2}, & j = 2, 3, \end{cases}$$

and

$$\delta^* \geq C_2 = \frac{1}{\sqrt{K + \frac{1}{4}}}_{|K=2} = \frac{2}{3}.$$  

by (39). Thus, the index $t^1$ chosen in the first iteration would be

$$t^1 = \arg \max_{j \in \{1,2,3\}} |\langle r^0, \phi_j \rangle| = \arg \max_{j \in \{1,2,3\}} |\langle y, \phi_j \rangle| = 1,$$  

and hence OLS cannot recover $x$ in $K(=2)$ iterations.  

(ii) $K = 3$

In this case, we consider $x = [0 1 1 1]'$ and $\Phi$ satisfying

$$\Phi'\Phi = \begin{bmatrix} 1 & \frac{\delta^*}{2} & \frac{\delta^*}{4} & \frac{\delta^*}{4} \\ \frac{\delta^*}{2} & 1 & -\frac{\delta^*}{4} & -\frac{\delta^*}{4} \\ \frac{\delta^*}{4} & -\frac{\delta^*}{4} & 1 & -\frac{\delta^*}{8} \\ \frac{\delta^*}{8} & \frac{\delta^*}{8} & -\frac{\delta^*}{8} & 1 \end{bmatrix}. \quad (41)$$

Since the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ of $\Phi'\Phi$ are

$$\lambda_1 = 1 + \frac{3\delta^*}{4}, \quad \lambda_2 = \lambda_3 = 1 + \frac{\delta^*}{8}, \quad \lambda_4 = 1 - \delta^*,$$

the RIP constant $\delta_4$ of $\Phi$ is given by

$$\delta_4 = \max_{i \in \{1,2,3,4\}} |\lambda_i - 1| = \delta^*.$$  

Note that

$$|\langle y, \phi_j \rangle| = |(\Phi'\Phi x)_j| = \begin{cases} \frac{3\delta^*}{4}, & j = 1, \\ 1 - \frac{\delta^*}{4}, & j = 2, 3, 4, \end{cases}$$

and

$$\delta^* \geq C_3 = \frac{1}{\sqrt{K + \frac{1}{16}}}_{|K=3} = \frac{4}{7}.$$  

by (39). Then, by (7), the OLS algorithm would pick the first index in the first iteration (i.e., $t^1 = 1$) and hence cannot recover $x$ in $K(=3)$ iterations.  

(iii) $K \geq 4$

In this case, let

$$x = \begin{bmatrix} 0 \\ 1_{K \times 1} \end{bmatrix}, \Phi'\Phi = \begin{bmatrix} 1 & \frac{\delta^*}{\sqrt{K}} & 1_{K \times 1} & \frac{\delta^*}{\sqrt{K}} \end{bmatrix}. \quad (42)$$

Since the eigenvalues $\lambda_1, \ldots, \lambda_{K+1}$ of $\Phi'\Phi$ are

$$\lambda_1 = 1 + \delta^*, \quad \lambda_2 = \ldots = \lambda_K = 1, \quad \lambda_{K+1} = 1 - \delta^*,$$

the RIP constant $\delta_{K+1}$ of $\Phi$ is

$$\delta_{K+1} = \max_{i \in \{1,\ldots,K+1\}} |\lambda_i - 1| = \delta^*.$$  

Also, by noting that

$$|\langle y, \phi_j \rangle| = |(\Phi'\Phi x)_j| = \begin{cases} \sqrt{K} \delta^*, & j = 1, \\ 1, & j = 2, \ldots, K \end{cases} \quad (43)$$

and

$$\delta^* \geq C_K = \frac{1}{\sqrt{K}},$$

we know that the first index would be chosen in the first iteration of OLS. Therefore, OLS cannot recover $x$ in $K$ iterations.  

Remark 3. A performance limit of OLS over which the exact reconstruction is not uniformly ensured has also been studied in [6]. The main difference of our work over this is that our result holds for all $K$ and each of $\delta_{K+1} \in [C_K, 1)$. This is in contrast to the result in [6], which is limited to the case where $K = 2$ and $\delta_{K+1} = C_K$.  

Theorem 2 implies that a performance guarantee of the OLS algorithm cannot be less restrictive than

$$\delta_{K+1} < C_K.$$  

Combining this with the result in Theorem 1, we conclude that the proposed guarantee (26) is optimal.  

IV. DISCUSSIONS

So far, we have shown that if the columns of a sampling matrix are $\ell_2$-normalized, then (26) is an optimal recovery guarantee of OLS. In this section, we discuss some issues that arise from our analysis.

A. Extension for General Sampling Matrices

We note that our result is based on the assumption that the columns of a sampling matrix are $\ell_2$-normalized. This assumption, however, might not hold in some applications (e.g., $\Phi$ is Gaussian random [9]). For this case, we build a condition of general (not necessarily $\ell_2$-normalized) sampling matrices for the success of the OLS algorithm. Our result is formally described in the following theorem.

3When $\delta^* = \frac{3}{4}$, $t^1 = 1$ by the tie-breaking rule of OLS (see Table I).
Theorem 3. Let $\Phi \in \mathbb{R}^{m \times n}$ be a sampling matrix and $M = \max_{j \in \Omega} \|\phi_j\|_2$. If $\Phi$ satisfies the RIP of order $K + 1$ with
\[
\delta_{K+1} < \min \left\{ \frac{1}{1 + M^2}, \frac{2}{M(1 + M^2)\sqrt{K}} \right\},
\]
then the OLS algorithm exactly reconstructs any $K$-sparse vector $x \in \mathbb{R}^n$ from its samples $y = \Phi x$ in $K$ iterations.

Proof. Similar to the proof of Theorem 1, we show that if $T^k \subset T$ for some $k$ ($0 \leq k < K$), then the OLS algorithm chooses a support index in the $(k + 1)$-th iteration. Recall that $T^{k+1} \in T$ if and only if (8) holds. In our proof, we build a lower bound of
\[
p_1 = \max_{j \in T \setminus T^k} \frac{|\langle r^k, \phi_j \rangle|}{\|P_{T^k}^\perp \phi_j\|_2}
\]
and an upper bound of
\[
q_1 = \max_{j \in \Omega \setminus T} \frac{|\langle r^k, \phi_j \rangle|}{\|P_{T^k} \phi_j\|_2},
\]
and then show that the former is larger than the latter under (44).

- **Lower bound of $p_1$:**
  Since $\|P_{T^k}^\perp \phi_j\|_2 \leq \|\phi_j\|_2 \leq M$ for each of $j \in T \setminus T^k$, we have
  \[
p_1 \geq \frac{1}{M} \max_{j \in T \setminus T^k} |\langle r^k, \phi_j \rangle|.
  \]
  Also, since $T^k \subset T$, we have [24, p.1375]
  \[
  \max_{j \in T \setminus T^k} |\langle r^k, \phi_j \rangle| \geq \frac{\|r^k\|_2^2}{\sqrt{K - k}\|x_{T \setminus T^k}\|_2}.
  \]
  Using this together with (47), we obtain
  \[
p_1 \geq \frac{\|r^k\|_2^2}{M \sqrt{K - k}\|x_{T \setminus T^k}\|_2}.
  \]

- **Upper bound of $q_1$:**
  Let $j \in \Omega \setminus T$. Since $\|P_{T^k}^\perp \phi_j\|_2 \leq \|\phi_j\|_2 \leq M^2$, we have
  \[
  1 - (1 + \|P_{T^k}^\perp \phi_j\|_2^2) \delta_{K+1} \geq 1 - (1 + M^2)\delta_{K+1} > 0,
  \]
  where (a) follows from (44). Then, from (17) and (18), we obtain
  \[
  \frac{2\|x_{T \setminus T^k}\|_2 |\langle r^k, \phi_j \rangle|}{\|P_{T^k}^\perp \phi_j\|_2} \leq (1 + \|P_{T^k}^\perp \phi_j\|_2^2)\delta_{K+1}\|r^k\|_2^2 \\
  \leq (1 + M^2)\delta_{K+1}\|r^k\|_2^2.
  \]
  Since (49) holds for arbitrary $j \in \Omega \setminus T$, we have
  \[
  q_1 < \frac{(1 + M^2)\delta_{K+1}\|r^k\|_2^2}{2\|x_{T \setminus T^k}\|_2}.
  \]

- **When $p_1 > q_1$:**
  From (48) and (50), we have
  \[
p_1 - q_1 \geq \left( \frac{1}{M \sqrt{K - k}} - \frac{(1 + M^2)\delta_{K+1}}{2} \right) \frac{\|r^k\|_2^2}{\|x_{T \setminus T^k}\|_2}.
  \]
  One can easily check that the right-hand side of (51) is strictly larger than zero under (44). Therefore, $p_1 > q_1$, and hence OLS picks a support index in the $(k+1)$-th iteration. \hfill \Box

Note that if $\Phi$ has unit $\ell_2$-norm columns, then $M = 1$ so that (44) is satisfied under
\[
\delta_{K+1} < \min \left\{ \frac{1}{2}, \frac{1}{\sqrt{K}} \right\}.
\]
This implies that if $\Phi$ has unit $\ell_2$-norm columns and $K \geq 4$, then OLS guarantees the exact recovery under $\delta_{K+1} < \frac{1}{\sqrt{K}}$, which coincides with the result in Theorem 1. Also, since $M = \max_{j \in \Omega} \|\phi_j\|_2 = \sqrt{1 + \delta_{K+1}}$ by the RIP, (44) is satisfied under
\[
\delta_{K+1} < \min \left\{ \frac{1}{2 + \delta_{K+1}}, \frac{2}{\sqrt{1 + \delta_{K+1}(2 + \delta_{K+1})\sqrt{K}}} \right\},
\]
which is equivalent to
\[
\delta_{K+1} < \left\{ \frac{\sqrt{2} - 1}{\sqrt{1 + \delta_{K+1}(2 + \delta_{K+1})\sqrt{K}}} \right\}, \quad 1 \leq K \leq 2,
\]
\[
K \geq 3.
\]
We compare our result with existing ones. In [6, Corollary 2], it has been shown that OLS perfectly recovers any $K$-sparse vector in $K$ iterations under
\[
\delta_{K+1} < \frac{1}{2\sqrt{K + 1}}.
\]
In Fig. 1, we plot the upper bounds in (52) and (53) as a function of $K$. From the figure, one can observe that our bound is roughly two times larger than the conventional bound. Note that an $m \times n$ matrix whose entries are drawn i.i.d. from a Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$ obeys the RIP with $\delta_{K+1} \leq \epsilon \in (0, 1)$ with overwhelming probability $\chi$ under
\[
m \geq C_{\chi} K \log \frac{n}{\epsilon^2},
\]
where $C_{\chi}$ is the constant depending on $\chi$ [9]. Using this result, one can deduce from (52) that OLS requires a four times smaller number of random (Gaussian) measurements for accurate recovery than that expected by the previous result in (53).

B. Exact Support Recovery With OLS From Noisy Measurements

Thus far, we have focused on the performance guarantee of OLS in the noiseless case. In this subsection, we analyze the performance in the more realistic scenario where the measurement vector $y$ is contaminated by the noise $v \in \mathbb{R}^m$.

\[
y = \Phi x + v.
\]

The noise vector $v$ is often assumed to be bounded (i.e., $\|v\|_2 \leq \epsilon$ for some $\epsilon > 0$) or Gaussian (i.e., $v_i \sim \mathcal{N}(0, \sigma^2)$). In this paper, we focus on the bounded scenario, but our analysis can be readily extended to the Gaussian scenario by exploiting the fact that a Gaussian noise is essentially...
bounded with high probability [17, Lemma 3]. Note that the perfect recovery of $x$ is not possible in the noisy case. Thus, we analyze the condition under which OLS identifies the support exactly. Our result is formally described in the following theorem.

**Theorem 4.** Let $\Phi \in \mathbb{R}^{m \times n}$ be a sampling matrix having unit $\ell_2$-norm columns. Then the OLS algorithm accurately reconstructs the support of any $K$-sparse vector $x$ from its noisy measurements $y = \Phi x + v$ in $K$ iterations, provided that $\Phi$ obeys the RIP with (26) and $\min_{i \in T} |x_i|$ satisfies

$$\min_{i \in T} |x_i| > \left\{ \begin{array}{ll} \frac{2 \|v\|_2}{1 - \delta_{K+1}/(2\|v\|_2)}, & 1 \leq K \leq 3, \\ \frac{1}{(1 - \delta_K)(1 - \sqrt{K\delta_{K+1}})}, & K \geq 4. \end{array} \right. \quad \text{(55)}$$

**Proof.** We focus on the case where $K \geq 4$. For the proofs of Theorem 1 when $K = 1$ and $K \in \{2, 3\}$, see Appendices B and C, respectively.

Similar to the proof of Theorem 1, we show that if $T^j \subset T$ for some $K \geq 4$, then OLS chooses a support index in the $(k+1)$-th iteration. To this end, we construct a lower bound of $p_1$ defined in (45) and an upper bound of $q_1$ defined in (46), and then show that the former is larger than the latter.

- **Lower bound of $p_1$:**
  For each of $j \in T \setminus T^k$, we have
  $$|\langle r^k, \phi_j \rangle| = |\langle P_{T^k}^+ (\Phi x + v), \phi_j \rangle|$$
  $$(a) \geq |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| - |\langle P_{T^k}^+ v, \phi_j \rangle|$$
  $$(b) \geq |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| - \|v\|_2\|P_{T^k}^+ \phi_j\|_2, \quad \text{(56)}$$
  where (a) follows from the triangle inequality and (b) is because
  $$|\langle P_{T^k}^+ v, \phi_j \rangle| = |\langle v, P_{T^k}^+ \phi_j \rangle| \leq \|v\|_2\|P_{T^k}^+ \phi_j\|_2$$
  by Cauchy-Schwarz inequality. As a result, we have
  $$p_1 = \max_{j \in T \setminus T^k} |\langle r^k, \phi_j \rangle|$$
  $$(a) \geq \max_{j \in T \setminus T^k} |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| - \|v\|_2$$
  $$(a) \geq \frac{\|P_{T^k}^+ \Phi x\|_2^2}{\sqrt{K - k}\|x_{T^k}^\perp\|_2^2} - \|v\|_2, \quad \text{(57)}$$
  where (a) follows from [6, Proposition 2].

- **Upper bound of $q_1$:**
  For each of $j \in \Omega \setminus T$, we have
  $$|\langle r^k, \phi_j \rangle| = |\langle P_{T^k}^+ (\Phi x + v), \phi_j \rangle|$$
  $$(a) \leq |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| + |\langle P_{T^k}^+ v, \phi_j \rangle|$$
  $$(b) \leq |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| + \|v\|_2\|P_{T^k}^+ \phi_j\|_2,$$
  where (a) and (b) follow from the triangle inequality and (56), respectively. Thus, $q_1$ satisfies
  $$q_1 = \max_{j \in \Omega \setminus T} |\langle r^k, \phi_j \rangle| \leq \frac{\max_{j \in \Omega \setminus T} |\langle P_{T^k}^+ \Phi x, \phi_j \rangle|}{\sqrt{K - k}\|x_{T^k}^\perp\|_2^2} + 2\|v\|_2. \quad \text{(58)}$$
  Note that when $K \geq 4$, the RIP constant $\delta_{K+1}$ of $\Phi$ satisfies
  $$\delta_{K+1} < C_K = \frac{1}{\sqrt{K}} \leq \frac{1}{2}.$$
  Then, from Lemma 2, we obtain
  $$\max_{j \in \Omega \setminus T} |\langle P_{T^k}^+ \Phi x, \phi_j \rangle| \leq \delta_{K+1} \|P_{T^k}^+ \Phi x\|_2^2 / \|x_{T^k}^\perp\|_2^2.$$
  Combining this together with (58), we have
  $$q_1 \leq \frac{\delta_{K+1} \|P_{T^k}^+ \Phi x\|_2^2}{\|x_{T^k}^\perp\|_2^2} + \|v\|_2. \quad \text{(59)}$$

- **When $p_1 > q_1$:**
  From (57) and (59), we have
  $$p_1 - q_1 \geq \frac{\|P_{T^k}^+ \Phi x\|_2^2}{\sqrt{K - k}\|x_{T^k}^\perp\|_2^2} (1 - \sqrt{K - k}\delta_{K+1}) - 2\|v\|_2$$
  $$\geq \frac{\|P_{T^k}^+ \Phi x\|_2^2}{\sqrt{K - k}\|x_{T^k}^\perp\|_2^2} (1 - \sqrt{K}\delta_{K+1}) - 2\|v\|_2$$
  $$(a) \geq \frac{\|x_{T^k}^\perp\|_2^2}{\sqrt{K - k}} (1 - \delta_K)(1 - \sqrt{K}\delta_{K+1}) - 2\|v\|_2$$
  $$\geq \min_{i \in T} |x_i|(1 - \delta_K)(1 - \sqrt{K}\delta_{K+1}) - 2\|v\|_2; p_1 - q_1$$
  $$\geq \frac{\|x_{T^k}^\perp\|_2^2}{\sqrt{K - k}} (1 - \delta_K)(1 - \sqrt{K}\delta_{K+1}) - 2\|v\|_2; p_1 - q_1$$
  $$(60)$$
  where (a) is because $1 - \sqrt{K}\delta_{K+1} > 0$ by (26) and
  $$\|P_{T^k}^+ \Phi x\|_2^2 \geq (1 - \delta_K)\|x_{T^k}^\perp\|_2^2$$
  by Lemma 2. Clearly, the right-hand side of (60) is strictly larger than zero under $\delta_{K+1} < \frac{1}{\sqrt{K}}$ and (55). Therefore, $p_1 > q_1$ and hence OLS chooses a support index in the $(k+1)$-th iteration. \[\square\]
Note that if the support of $x$ is determined accurately, then the estimated signal $x^K$ satisfies [25, eq. (26)]
\[
\|x - x^K\|_2 \leq \frac{\|v\|_2}{\sqrt{1 - \delta_K}}.
\]
Thus, Theorem 4 implies that under some proper condition on $\min_{i \in T} |x_i|$, the proposed RIP condition ($\delta_{K+1} < C_K$) ensures that the $\ell_2$-norm of reconstruction error of OLS is upper bounded by a constant multiple of the noise power.

In [8, Theorem 3], it has been shown that OLS picks all the support indices in $K$ iterations if
\[
\delta_{K+1} < \frac{1}{\sqrt{K + 1}},
\]  
(61a)
\[
\min_{i \in T} |x_i| > \frac{2\|v\|_2}{1 - \sqrt{K + 1} \delta_{K+1}}.
\]  
(61b)

Clearly, the proposed RIP condition ($\delta_{K+1} < C_K$) is less restrictive than the conventional one (61a) in all sparsity region. Also, when $1 \leq K \leq 3$, the lower bound in (55) is uniformly smaller than that in (61b) for all of $\delta_{K+1}$, which implies that our minimum magnitude condition (55) is less restrictive. To compare (55) with (61b) when $K \geq 4$, we define
\[
b_{\text{prop}} = \frac{2\|v\|_2}{(1 - \delta_K)(1 - \sqrt{K} \delta_{K+1})},
\]
\[
b_{\text{conv}} = \frac{2\|v\|_2}{1 - \sqrt{K + 1} \delta_{K+1}},
\]
and plot $b_{\text{prop}}/b_{\text{conv}}$ as a function of $\delta_{K+1}$ in Fig. 2. From the figure, one can observe that the proposed bound $b_{\text{prop}}$ is slightly larger than the conventional bound $b_{\text{conv}}$ when $\delta_{K+1}$ is small. However, when $\delta_{K+1}$ approaches $1/\sqrt{K + 1}$, $b_{\text{prop}}/b_{\text{conv}}$ goes to zero, which indicates that the proposed condition (55) is far more relaxed than the conventional condition (61b).

V. Recovery of Sparse Signals Using OMP

In the sparse recovery literature, OLS is similar in spirit to the OMP algorithm [10], [11]. While two algorithms are similar in many aspects (e.g., algorithm inputs, initialization, augmentation, estimation, and residual update), they differ primarily in their greedy rules of updating the estimated support in each iteration. Specifically, in contrast to OLS that identifies the column of $\Phi$ minimizing the residual power, OMP finds the column maximally correlated with the residual [10]. In [5], it has been empirically observed that OLS requires slightly more computation but is more reliable than OMP (see [5], [12] for more details).

Over the years, various RIP-based recovery guarantees have been proposed for the OMP algorithm [18]–[23]. The best result demonstrates that in the noiseless case, OMP exactly recovers any $K$-sparse vector in $K$ iterations under
\[
\delta_{K+1} < \frac{1}{\sqrt{K + 1}}.
\]  
(63)

This condition has also been shown to be optimal through counterexamples for which OMP fails the reconstruction under [21], [23]
\[
\delta_{K+1} \in \left[\frac{1}{\sqrt{K + 1}}, 1\right).
\]

In this section, by exploiting results developed for OLS in previous sections, we show that the optimal recovery guarantee (63) of OMP can be improved further if an $\ell_2$-normalized sampling matrix is employed. Our result is formally described in the following theorem.

**Theorem 5.** Let $\Phi \in \mathbb{R}^{m \times n}$ be a sampling matrix having unit $\ell_2$-norm columns. If $\Phi$ obeys the RIP with (26), then the OMP algorithm perfectly recovers any $K$-sparse vector $x \in \mathbb{R}^n$ from its measurements $y = \Phi x$ in $K$ iterations.

**Proof.** Let $T_{\text{OMP}}^k$ be the estimated support generated in the $k$-th iteration of the OMP algorithm. In our proof, we show that if $T_{\text{OMP}}^k \subset T$ for some $k$ ($0 \leq k < K$), then OMP chooses a support index in the $(k+1)$-th iteration. In doing so, we can show that $T_{\text{OMP}}^K = T$,

\[
T_{\text{OMP}}^K = \bigcap_{j \in \Omega} \left\{ j \in T_{\text{OMP}}^j \right\}.
\]

where $r_{\text{OMP}}^j = \frac{1}{T_{\text{OMP}}^j} y$. Therefore, to show $T_{\text{OMP}}^{k+1} \subset T$, it suffices to show that
\[
\max_{j \in T \setminus T_{\text{OMP}}^k} |\langle r_{\text{OMP}}^j, \phi_j \rangle| > \max_{j \in \Omega \setminus T} |\langle r_{\text{OMP}}^j, \phi_j \rangle|.
\]  
(65)

The argument for the base case ($k = 0$) clearly holds true since $T_{\text{OMP}}^0 = \emptyset \subset T$. 

![Figure 2. Comparison of the minimum magnitude constraints in (55) and (61b) when $K \geq 4$.](image-url)
Note that the left-hand side of (65) satisfies [24, p.1375]
\[
\max_{j \in T \setminus T^k_{\text{OMP}}} |\langle r^k_{\text{OMP}}, \phi_j \rangle| \geq \frac{\|r^k_{\text{OMP}}\|_2^2}{\sqrt{K - k}\|x_{T \setminus T^k_{\text{OMP}}}\|_2^2}.
\] (66)

Also, the right-hand side of (65) satisfies
\[
\max_{j \in \Omega \setminus T} |\langle r^k_{\text{OMP}}, \phi_j \rangle| \leq \max_{j \in \Omega \setminus T} |\langle r^k_{\text{OMP}}, \phi_j \rangle| \leq \frac{\|r^k_{\text{OMP}}\|_2^2}{\sqrt{K - k}\|x_{T \setminus T^k_{\text{OMP}}}\|_2^2},
\] (67)

where (a) is because \(\|P^\perp_{T^k_{\text{OMP}}} \phi_j\|_2 \leq \|\phi_j\|_2 = 1\) for each of \(j \in \Omega \setminus T\) and (b) follows from (28).\(^5\) By combining (66) and (67), we obtain the desired result (65). \(\Box\)

Note that if \(k = 0\) and \(\|\phi_j\|_2 = 1\) for each of \(j \in \Omega\), then the selection rule of OLS in (7) reduces to (64). This in turn implies that if a sampling matrix has unit \(\ell_2\)-norm columns, then the outputs of OMP and OLS are the same in the first iteration. Thus, counterexamples in (40), (41), and (42) for which OLS fails to choose a support index in the first iteration also apply to OMP, which leads us to the following theorem.

**Theorem 6.** For any positive integer \(K\) and constant \(\delta^* \in [C_K, 1]\), there always exist an \(\ell_2\)-normalized matrix \(\Phi\) with \(\delta_{K+1} = \delta^*\) and a \(K\)-sparse vector \(x\) such that the OMP algorithm fails to recover \(x\) from \(y = \Phi x + v\) in \(K\) iterations.

Combining the results in Theorems 5 and 6, we conclude that if the columns of a sampling matrix are \(\ell_2\)-normalized, then (26) is also an optimal performance guarantee for the OMP algorithm.

By taking similar steps to the proofs of Theorems 4 and 5, one can obtain the following theorem, which gives an sufficient condition for OMP to ensure the exact support recovery in the noisy scenario.

**Theorem 7.** Let \(x \in \mathbb{R}^n\) be an arbitrary \(K\)-sparse vector and \(\Phi \in \mathbb{R}^{m \times n}\) be a sampling matrix having unit \(\ell_2\)-norm columns. If \(\min_{i \in T} |x_i|\) satisfies (55) and \(\Phi\) obeys the RIP with (26), then the OMP algorithm perfectly identifies the support of \(x\) from its noisy measurements \(y = \Phi x + v\) in \(K\) iterations.

We define the signal-to-noise ratio (SNR) and the minimum-to-average ratio (MAR) as
\[
\text{SNR} = \frac{\|\Phi x\|_2^2}{\|v\|_2^2} \quad \text{and} \quad \text{MAR} = \min_{i \in T} |x_i|/\|x\|_2/\sqrt{K}.
\]
Then, using [8, eq. (62)]
\[
\sqrt{\text{SNR}} \leq \sqrt{1 + \delta_{K+1}} \text{MAR} \cdot |\min_{i \in T} |x_i|/\|v\|_2| \sqrt{K},
\]
\(\Box\)

\(5\) We note that the proof of (28) holds whenever \(\delta_{K+1} < C_K\) and \(T^k \subset T\).

Therefore, Theorem 7 implies that OMP ensures the exact support recovery under \(\delta_{K+1} < C_K\) and (68). This outperforms the existing result in [25, Theorem 3.1], which states that the condition
\[
\sqrt{\text{SNR}} > \frac{2(1 + \delta_{K+1})}{1 - (\sqrt{K} + 1)\delta_{K+1}} \sqrt{K}
\]
is sufficient for OMP to identify the support accurately.

**VI. Conclusion**

In this paper, we analyzed the performance guarantee of the OLS algorithm. Firstly, we showed that if the columns of a sampling matrix are \(\ell_2\)-normalized, then OLS ensures the exact recovery of any \(K\)-sparse vector in \(K\) iterations under
\[
\delta_{K+1} < C_K = \begin{cases} \sqrt{K}, & K = 1, \\ 1/\sqrt{K}, & K = 2, \\ 1/(\sqrt{K+1}), & K = 3, \\ \sqrt{K}/\sqrt{K}, & K \geq 4. \end{cases}
\]

Secondly, we showed that (69) is an optimal RIP condition of \(\ell_2\)-normalized sampling matrices for the success of OLS. Thirdly, we extended our results to the more general and involved cases. Specifically, we built a condition of general (not necessarily \(\ell_2\)-normalized) sampling matrices and also studied the performance of OLS in the noisy scenario. Our result demonstrates that under a proper condition on the minimum magnitude of nonzero entries in an input signal, the proposed RIP condition (69) ensures OLS to identify the support perfectly. Finally, we extended our results for the conventional OMP algorithm. Specifically, we showed that if the columns of a sampling matrix are \(\ell_2\)-normalized, then (69) is also an optimal recovery guarantee for OMP.

**APPENDIX A**

**Main Steps for Constructing the Counterexample in (40)**

In this section, we demonstrate how we construct the counterexample in (40).

First of all, we construct a general 2-sparse vector \(x = [0 \ p \ q]^T\) and an \(\ell_2\)-normalized sampling matrix \(\Phi\) such that
\[
\Phi \Phi' = \begin{bmatrix} 1 & c_1 & c_2 \\ c_1 & 1 & c_3 \\ c_2 & c_3 & 1 \end{bmatrix}.
\]

Our main goal is to determine the values of \(p, q, c_1, c_2, c_3\) such that \(\Phi\) obeys the RIP with \(\delta_{K+1} = \delta^* \in [C_K, 1]\) and OLS fails to pick a support index in the first iteration, i.e.,
\[
|\langle y, \phi_1 \rangle| \geq \max_{j \in \{2, 3\}} |\langle y, \phi_j \rangle|.
\]

one can easily show that (55) is satisfied under
\[
\sqrt{\text{SNR}} > \begin{cases} \frac{2\sqrt{1 + \delta_{K+1}}}{(1 - \delta_{K+1})/\delta_{K+1}} \sqrt{K}, & 1 \leq K \leq 3, \\ \frac{2\sqrt{1 + \delta_{K+1}}}{(1 - \delta_{K+1})/\delta_{K+1}} \sqrt{K}, & K \geq 4. \end{cases}
\]

(68)
By [6, Proposition 2], the right-hand side of (70) satisfies
\[
\max_{j \in \{2,3\}} |\langle y, \phi_j \rangle| \geq \frac{||y||_2^2}{\sqrt{2}||x_T||_2}. \tag{71}
\]
Also, if \( \Phi \) obeys the RIP with \( \delta_{K+1} = \delta^* \), then we obtain from (36) that
\[
|\langle y, \phi_1 \rangle| \leq \frac{||y||_2^2}{\sqrt{2}||x_T||_2} - \frac{||x_T||_2}{\sqrt{2}} \left(1 - \frac{3}{2}\delta^*\right). \tag{72}
\]
Note that if the equalities of (71) and (72) hold, i.e.,
\[
\max_{j \in \{2,3\}} |\langle y, \phi_j \rangle| = \frac{||y||_2^2}{\sqrt{2}||x_T||_2}, \tag{73a}
\]
\[
|\langle y, \phi_1 \rangle| = \frac{||y||_2^2}{\sqrt{2}||x_T||_2} - \frac{||x_T||_2}{\sqrt{2}} \left(1 - \frac{3}{2}\delta^*\right), \tag{73b}
\]
then (70) clearly holds true since \( \delta^* \in [C_K,1) \) = (0, 1). Motivated by this fact, we aim to find \( p, q, c_1, c_2, c_3 \) such that (73a) and (73b) hold. To this end, we investigate the condition under which (73a) and (73b) hold.

We first examine the condition under which (73a) holds. Note that
\[
\max_{j \in \{2,3\}} |\langle y, \phi_j \rangle| \geq \frac{1}{\sqrt{2}} ||\Phi_T^* y||_2 \geq \frac{1}{\sqrt{2}} \frac{||y||_2^2}{||x_T||_2},
\]
where the last inequality follows from the Cauchy-Schwarz inequality.\(^6\) Thus, to find out \( x \) and \( \Phi \) satisfying (73a), we need to investigate the condition under which
\[
\max_{j \in \{2,3\}} |\langle y, \phi_j \rangle| = \frac{1}{\sqrt{2}} ||\Phi_T^* y||_2, \tag{74a}
\]
\[
||\Phi_T^* y||_2 = \frac{||y||_2^2}{||x_T||_2}. \tag{74b}
\]
Note that (74a) holds if and only if \( |\langle y, \phi_2 \rangle| = |\langle y, \phi_3 \rangle| \), i.e.,
\[
|p + c_3 q| = |c_2 p + q|. \tag{75}
\]
Also, (74b) holds true if and only if \( \Phi_T^* y = \lambda x_T \) for some \( \lambda > 0, \)\(^7\) that is,
\[
\begin{bmatrix} p + c_3 q \\ c_2 p + q \end{bmatrix} = \lambda \begin{bmatrix} p \\ q \end{bmatrix}. \tag{76}
\]
One can see that to satisfy (75) and (76), the condition
\[
|p| = |q|
\]
should be satisfied. In our construction, we set \( p = q = 1 \), i.e., \( x = [0 \ 1 \ 1]^T \).

We next analyze the condition under which (73b) holds. Note that
\[
\left|\frac{\langle y, \phi_1 \rangle}{\frac{\langle y, \phi_2 \rangle}{\sqrt{2}||x_T||_2}}\right|^2 \geq \frac{3||x_T||_2^2}{2} \tag{77}
\]
Also, if \( \Phi \) obeys the RIP with \( \delta_{K+1} = \delta^* \), then
\[
\left|\frac{\langle y, \phi_1 \rangle}{\frac{\langle y, \phi_2 \rangle}{\sqrt{2}||x_T||_2}}\right|^2 \geq \frac{3||x_T||_2^2}{2}, \tag{78}
\]
where (a) is because
\[
\sigma_{\min}^2(\Phi_T \Phi_1) \geq 1 - \delta_{K+1}
\]
by [13, Remark 1]. Therefore, to satisfy (77b), the equalities of (77) and (78) should hold. One can see that the equalities hold if \( [x_T^T - \frac{\langle y, \phi_1 \rangle}{\frac{\langle y, \phi_2 \rangle}{\sqrt{2}||x_T||_2}}] \) is the eigenvector of \( \Phi_T \Phi_1 \) corresponding to the smallest eigenvalue \( \sigma_{\min}^2(\Phi_T \Phi_1) \) and \( \sigma_{\min}^2(\Phi_T \Phi_1) = 1 - \delta^* \), i.e.,
\[
\begin{bmatrix} x_T & \frac{\langle y, \phi_1 \rangle}{\frac{\langle y, \phi_2 \rangle}{\sqrt{2}||x_T||_2}} \end{bmatrix} = (1 - \delta^*) \begin{bmatrix} x_T \\ \frac{\langle y, \phi_1 \rangle}{\frac{\langle y, \phi_2 \rangle}{\sqrt{2}||x_T||_2}} \end{bmatrix}.
\]
This means that (73b) is satisfied if
\[
(c_1, c_2, c_3) = (\frac{\delta^*}{2}, \delta^*, \frac{\delta^*}{2}) \text{ or } (-\frac{\delta^*}{2}, \frac{\delta^*}{2}, \frac{\delta^*}{2}).
\]
In our construction, we choose \( (c_1, c_2, c_3) = (\frac{\delta^*}{2}, \frac{\delta^*}{2}, -\frac{\delta^*}{2}) \), i.e.,
\[
\Phi^* \Phi = \begin{bmatrix} 1 & 1 \\ \frac{\delta^*}{2} & \frac{\delta^*}{2} & -\frac{\delta^*}{2} \\ \frac{\delta^*}{2} & -\frac{\delta^*}{2} & 1 \end{bmatrix}.
\]
Finally, what remains is to examine if the constructed matrix \( \Phi \) indeed satisfies the RIP with \( \delta_{K+1} = \delta^* \). This has already been done in the proof of Theorem 2.

\section*{Appendix B}
\textbf{Proof of Theorem 4 When} \( K = 1 \)

\textit{Proof.} It suffices to show that OLS chooses the support index in the first iteration. Without loss of generality, we assume that \( x = [c \ 0 \ldots \ 0]^T \in \mathbb{R}^n \) for some \( c \in \mathbb{R} \setminus \{0\}. \) Then we have
\[
|\langle y, \phi_1 \rangle| = |\langle c \phi_1 + v, \phi_1 \rangle| \geq |c| - |\langle v, \phi_1 \rangle| \geq |c| - ||v||_2, \tag{79}
\]
where (a) is from the triangle inequality and (b) is because
\[
|\langle v, \phi_1 \rangle| \leq ||v||_2 ||\phi_1||_2 = ||v||_2. \tag{80}
\]
for each of $i \in \Omega$ by Cauchy-Schwarz inequality. Also, we have
\[
\max_{j \in \Omega \setminus T} |\langle y, \phi_j \rangle| = \max_{j \in \Omega \setminus T} |\langle c\phi_1 + v, \phi_j \rangle| \\
\leq \max_{j \in \Omega \setminus T} \left( |\langle c\phi_1, \phi_j \rangle| + |\langle v, \phi_j \rangle| \right) \\
\leq \max_{j \in \Omega \setminus T} |\langle c\phi_1, \phi_j \rangle| + \|v\|_2 \\
\leq |c|\delta_2 + \|v\|_2, 
\]
where (a), (b), and (c) follow from the triangle inequality, (80), and (31), respectively. From (79) and (81), we obtain
\[
|\langle y, \phi_1 \rangle| - \max_{j \in \Omega \setminus T} |\langle y, \phi_j \rangle| \geq |c|(1 - \delta_2) - 2\|v\|_2. 
\]
Note that the right-hand side of (82) is strictly larger than zero since
\[
|c| = \min_{i \in T} |x_i| > \frac{2\|v\|_2}{1 - \delta_2 + 1/C_K} = \frac{2\|v\|_2}{1 - \delta_2} 
\]
by (55). As a result, OLS identifies the support index in the first iteration, which completes the proof. \(\square\)

**APPENDIX C**

**Proof of Theorem 4 When K \in \{2,3\}**

**Proof.** Similar to the proof of Theorem 1, we show that if $T^k \subset T$ for some $k (0 \leq k < K)$, then the OLS algorithm chooses a support index in the $(k+1)$-th iteration. Recall that $T^{k+1} \subset T$ if and only if (8) holds. Since the left-hand side of (8) satisfies
\[
\max_{j \in T \setminus T^k} \frac{|\langle r^k, \phi_j \rangle|}{\|P_{T^k}\Phi x\|_2^2} > \frac{\|P_{T^k}\Phi x\|_2^2}{\sqrt{K-k}\|x_{T^k}\|_2} - \|v\|_2, 
\]
by (57), it suffices to show that
\[
\max_{j \in \Omega \setminus T} \frac{|\langle r^k, \phi_j \rangle|}{\|P_{T^k}\Phi x\|_2^2} < \frac{\|P_{T^k}\Phi x\|_2^2}{\sqrt{K-k}\|x_{T^k}\|_2} - \|v\|_2. 
\]
Note that
\[
\max_{j \in \Omega \setminus T} \frac{|\langle r^k, \phi_j \rangle|}{\|P_{T^k}\Phi x\|_2^2} \leq \max_{j \in \Omega \setminus T} \frac{|\langle P_{T^k}\Phi x, \phi_j \rangle|}{\|P_{T^k}\Phi x\|_2} + \|v\|_2 
\]
by (58). Also, for each of $j \in \Omega \setminus T$, we have
\[
\frac{|\langle P_{T^k}\Phi x, \phi_j \rangle|}{\|P_{T^k}\Phi x\|_2} \\
\leq \frac{|\langle P_{T^k}\Phi x, \phi_j \rangle|}{\sqrt{K-k}\|x_{T^k}\|_2^2} - \left( 1 - \frac{K}{4} \frac{\delta_{K+1}}{C_K} \right) \frac{\|x_{T^k}\|_2}{\sqrt{K-k}} \\
\leq \frac{\|P_{T^k}\Phi x\|_2^2}{\sqrt{K-k}\|x_{T^k}\|_2^2} - \left( 1 - \frac{\delta_{K+1}}{C_K} \right) \frac{\|x_{T^k}\|_2}{\sqrt{K-k}} \\
\leq \frac{\|P_{T^k}\Phi x\|_2^2}{\sqrt{K-k}\|x_{T^k}\|_2^2} - \left( 1 - \frac{\delta_{K+1}}{C_K} \right) \min_{i \in T} |x_i| \\
\leq \frac{\|P_{T^k}\Phi x\|_2^2}{\sqrt{K-k}\|x_{T^k}\|_2^2} - 2\|v\|_2, 
\]
where (a) follows from (36), (b) is because $(1 + \frac{K}{4})C_K = 1$ for $K \in \{2,3\}$ (see (26)), (c) is because $\delta_{K+1} < C_K$ and
\[
\|x_{T^k}\|_2 \geq \sqrt{K-k} \min_{i \in T} |x_i| \geq \sqrt{K-k} \min_{i \in T} |x_i|, 
\]
and (d) is because $\delta_{K+1} < C_K$ and $\min_{i \in T} |x_i|$ obeys (55). By combining (84) and (85), we obtain (83), which is the desired result. \(\square\)

**References**

[1] 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.

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

[3] J. Wang and P. Li, “Recovery of sparse signals using multiple orthogonal least squares,” IEEE Trans. Signal Process., vol. 65, no. 8, pp. 2049–2062, Aug. 2017.

[4] 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. 7599–7624, Nov. 2013.

[5] 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.

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

[7] J. Kim and B. Shim, “Multiple orthogonal least squares for joint sparse recovery,” in Proc. IEEE Int. Symp. Inf. Theory, Vail, CO, USA, Jun. 2018, pp. 61–65.

[8] J. Kim and B. Shim, “A near-optimal restricted isometry condition of multiple orthogonal least squares,” IEEE Access, vol. 7, no. 1, pp. 46822–46830, Mar. 2019.

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

[10] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. IEEE 27th Annu. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, vol. 1, Nov. 1993, pp. 40–44.

[11] A. J. 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.

[12] T. Blumensath and M. E. Davies, “On the difference between orthogonal matching pursuit and orthogonal least squares,” Tech. Rep., 2007. [Online]. Available: https://eprints.soton.ac.uk/142469/1/BIDOMPsOLS97.pdf

[13] 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.

[14] 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, Mar. 2015.

[15] M. E. Davies and Y. C. Eldar, “Rank awareness in joint sparse recovery,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[16] J. Kim, J. Wang and B. Shim, “Nearly optimal restricted isometry condition for rank aware order recursive matching pursuit,” IEEE Trans. Signal Process., vol. 67, no. 17, pp. 4449–4463, Sep. 2019.

[17] 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.

[18] 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.

[19] Q. Mei 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.
[20] 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.

[21] 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, Nov. 2013.

[22] 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.

[23] Q. Mo, “A sharp restricted isometry constant bound of orthogonal matching pursuit,” *arXiv:1501.01708*, 2015.

[24] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, “Sharp conditions for exact support recovery with noise via OMP,” *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1370–1382, Mar. 2017.

[25] 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.