Canad. Math. Bull. Vol. 62 (4), 2019 pp. 780–797
http://dx.doi.org/10.4153/S0008439519000134

Perturbation Analysis of Orthogonal Least Squares

Pengbo Geng, Wengu Chen, and Huanmin Ge

Abstract. The Orthogonal Least Squares (OLS) algorithm is an efficient sparse recovery algorithm that has received much attention in recent years. On one hand, this paper considers that the OLS algorithm recovers the supports of sparse signals in the noisy case. We show that the OLS algorithm exactly recovers the support of K-sparse signal $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$ in K iterations, provided that the sensing matrix $\boldsymbol{\Phi}$ satisfies the restricted isometry property (RIP) with restricted isometry constant (RIC) $\delta_{K+1} < 1/\sqrt{K+1}$, and the minimum magnitude of the nonzero elements of $\boldsymbol{x}$ satisfies some constraint. On the other hand, this paper demonstrates that the OLS algorithm exactly recovers the support of the best K-term approximation of an almost sparse signal $\boldsymbol{x}$ in the general perturbations case, which means both $\boldsymbol{y}$ and $\boldsymbol{\Phi}$ are perturbed. We show that the support of the best K-term approximation of $\boldsymbol{x}$ can be recovered under reasonable conditions based on the restricted isometry property (RIP).

1 Introduction

As a new sampling theory, compressive sensing (CS) has a significant impact in signal processing and biomedical imagery [3, 4, 9]. A central aim of CS is to reconstruct a high-dimensional signal $\boldsymbol{x} \in \mathbb{R}^n$ from low-dimensional measurements

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}, \tag{1.1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n} (m \ll n)$ is called a *measurement matrix*, and $\boldsymbol{e} \in \mathbb{R}^m$ is a *vector of measurement errors*. In the noiseless case, $\boldsymbol{e} = \boldsymbol{0}$. Without any prior assumption on $\boldsymbol{x}$ in (1.1), this could not work as expected, but one can obtain the feasibility of such a program when the signal $\boldsymbol{x}$ is sparse. To recover the K-sparse $\boldsymbol{x}$ (*i.e.,* it has at most K nonzero entries), a natural method is to solve the l_0-minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi x} - \boldsymbol{y} \in \mathfrak{B}, \tag{1.2}$$

where $\|\boldsymbol{x}\|_0 = |\{i : x_i \neq 0\}|$ (the number of non-zero entries of $\boldsymbol{x}$), and the set $\mathfrak{B}$ is determined by the noise structure. Particularly, in the noiseless case, $\mathfrak{B} = \{\boldsymbol{0}\}$. However, as a combinatorial optimization problem, the above l_0-minimization problem is NP-hard in general.

Fortunately, many alternative approaches have been proposed for (1.2). For example, the nonconvex l_0-norm in (1.2) can be replaced by its convex relaxation l_1-norm.

Received by the editors September 5, 2019.
Published online on Cambridge Core July 10, 2019.
W. Chen is the corresponding author. This work was supported by the NSF of China (No. 11871109) and NSAF (Grant No. U1830107).
AMS subject classification: **65D15**, 65J22, 68W40.
Keywords: orthogonal least square (OLS), sparse signal, restricted isometry property (RIP), general perturbation.

That is,

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi x} - \boldsymbol{y} \in \mathfrak{B},$$

which is Basis Pursuit (BP) [6].

Compared to BP, greedy algorithms have the advantage of low computational cost and simple geometric interpretation, including Orthogonal Least Square (OLS) [5, 11, 25], Orthogonal Matching Pursuit (OMP) [8, 21, 25, 26, 28], generalized Orthogonal Matching Pursuit (gOMP) [7, 22–24, 27].

In practical applications, not only is the measurement vector $\boldsymbol{y}$ in (1.1) often contaminated by the noise vector $\boldsymbol{e}$, but the measurement matrix $\boldsymbol{\Phi}$ in (1.1) is also often perturbed. Such perturbations are called *general perturbations*. It is important to consider these perturbations for $\boldsymbol{y}$ and $\boldsymbol{\Phi}$, since they can account for precision errors when applications call for physically implementing the matrix $\boldsymbol{\Phi}$ in a sensor [14].

According to different perspective of views, the general perturbations can be classified into two situations [8]. From a user's point of view, when $\boldsymbol{\Phi}$ represents a system model, the whole sensing process is

$$\widehat{\boldsymbol{y}} = \boldsymbol{\Phi x} + \boldsymbol{e}, \quad \widehat{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \boldsymbol{E}, \tag{1.3}$$

where $\boldsymbol{E}$ is involved due to mismodeling of the actual system $\boldsymbol{\Phi}$, which can be found in source separation [1], remote sensing [10], radar [13], and countless other problems. Both the contaminated sensing matrix $\widehat{\boldsymbol{\Phi}}$ and contaminated measurement vector $\widehat{\boldsymbol{y}}$ are available for recovery.

The other scenario is from the designer's perspective; the system perturbation $\boldsymbol{E}$ is introduced when the system is physically implemented. Thus, the sensing process is

$$\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{\Phi}}\boldsymbol{x} + \boldsymbol{e}, \quad \widehat{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \boldsymbol{E}. \tag{1.4}$$

Only the nominal sensing matrix $\boldsymbol{\Phi}$ and contaminated measurement vector $\widehat{\boldsymbol{y}}$ are available for recovery.

And we are aware of quite a few researchers who studied the recovery of $\boldsymbol{x}$ under general perturbations for the past few years. Herman et al. used BP [14] and CoSaMP [12], respectively, to study the recovery of the signals. [8] and [16], respectively, studied the recovery of $\boldsymbol{x}$ from (1.3) and (1.4) via the OMP algorithm.

In this paper, we first consider that the OLS algorithm recovers the support of $\boldsymbol{x}$ in the noisy case, *i.e.*, $\boldsymbol{e} \neq \boldsymbol{0}$ in (1.1). Second, we consider the recovery of $\boldsymbol{x}$ from (1.3) and (1.4) via the OLS algorithm in Table 1, where $\boldsymbol{P}_{\bullet}^{\perp}$ is an orthogonal projection (see Section 2). In fact, the OLS algorithm is one of the most effective algorithm that is computationally similar to the OMP algorithm. The main difference between OMP and OLS lies in the greedy rule of updating the support at each iteration. In the selection procedure, while OMP seeks a column that is most strongly correlated with the residual, OLS chooses a candidate that leads to the minimum residual error. Compared with OMP, OLS has better convergence properties but is computationally a little more complicated [20]. For (1.1) with $\boldsymbol{e} = \boldsymbol{0}$, [5, 11, 15, 25] considered the exact recovery of the support for the K-sparse signal $\boldsymbol{x}$ via OLS in K iterations. Particularly, Wen et al. [29] have shown nearly optimal sufficient condition based on RIC (see Definition 1.1) for the exact recovery of the support of any K-sparse signal $\boldsymbol{x}$ via OLS

Table 1: The OLS Algorithm

Input: $\boldsymbol{\Phi}$, $\boldsymbol{y}$.
Initialize: $k = 0$, $\boldsymbol{r}^0 = \boldsymbol{y}$, $T^0 = \varnothing$.
Repeat until the stopping criterion is met
1: $k = k + 1$;
2: $t^k = \arg\min_{i \in \{1,\cdots,n\}} \lVert \boldsymbol{P}^{\perp}_{T^{k-1} \cup \{i\}} \boldsymbol{y} \rVert_2^2$;
3: $T^k = T^{k-1} \cup \{t^k\}$;
4: $\boldsymbol{x}^k = \arg\min_{\boldsymbol{u}:\mathrm{supp}(\boldsymbol{u}) = T^k} \lVert \boldsymbol{y} - \boldsymbol{\Phi u} \rVert_2$;
5: $\boldsymbol{r}^k = \boldsymbol{y} - \boldsymbol{\Phi x}^k$.
Output: T^k and $\boldsymbol{x}^k$.

in K iterations. That is, if $\boldsymbol{\Phi}$ with unit l_2-norm columns satisfies the RIP with $\delta_{K+1} < 1/\sqrt{K+1}$, which is nearly optimal, then OLS exactly recovers the support of any K-sparse signal $\boldsymbol{x}$ from (1.1) with $\boldsymbol{e} = \boldsymbol{0}$, in K iterations.

Definition 1.1 ([2]) For the given $m \times n$ matrix $\boldsymbol{\Phi}$ and an integer K, the restricted isometry constant (RIC) of order K can be defined as the smallest constant $\delta_K \in [0, 1)$ such that

$$(1 - \delta_K) \lVert \boldsymbol{x} \rVert_2^2 \le \lVert \boldsymbol{\Phi x} \rVert_2^2 \le (1 + \delta_K) \lVert \boldsymbol{x} \rVert_2^2 \tag{1.5}$$

holds for all K-sparse signals $\boldsymbol{x}$.

Denote $\boldsymbol{x}_{\max(K)} \in \mathbb{R}^n$ as the K-sparse signal that contains the K largest magnitude entries of $\boldsymbol{x}$, *i.e.*, the best K-term approximation of $\boldsymbol{x}$ and $T = \mathrm{supp}(\boldsymbol{x}_{\max(K)})$. In order to delineate the compressibility of signals $\boldsymbol{x}$ in (1.3) and (1.4), let

$$\beta = \frac{\lVert \boldsymbol{x}_{T^c} \rVert_2}{\lVert \boldsymbol{x}_T \rVert_2}, \quad \gamma = \frac{\lVert \boldsymbol{x}_{T^c} \rVert_1}{\sqrt{K} \lVert \boldsymbol{x}_T \rVert_2}. \tag{1.6}$$

A vector $\boldsymbol{x}$ is almost sparse if β and γ are far less than 1. When $\boldsymbol{x}_{T^c} = \boldsymbol{0}$, *i.e.*, $\boldsymbol{x}$ is K-sparse, then $\beta = \gamma = 0$.

The symbol $\lVert \boldsymbol{\Phi} \rVert_2$ denotes the spectral norm of a matrix $\boldsymbol{\Phi}$, and $\lVert \boldsymbol{\Phi} \rVert_2^{(K)}$ denotes the largest spectral norm taken over all K-column submatrices of $\boldsymbol{\Phi}$. The measurement noise $\boldsymbol{e}$ and system perturbation $\boldsymbol{E}$ can be quantified as

$$\frac{\lVert \boldsymbol{e} \rVert_2}{\lVert \boldsymbol{\Phi x} \rVert_2} \le \epsilon_e, \quad \frac{\lVert \boldsymbol{e} \rVert_2}{\lVert \widehat{\boldsymbol{\Phi}} \boldsymbol{x} \rVert_2} \le \widehat{\epsilon}_e, \quad \frac{\lVert \boldsymbol{E} \rVert_2^{(K)}}{\lVert \boldsymbol{\Phi} \rVert_2^{(K)}} \le \epsilon, \tag{1.7}$$

where $\lVert \boldsymbol{\Phi x} \rVert_2, \lVert \widehat{\boldsymbol{\Phi}} \boldsymbol{x} \rVert_2, \lVert \boldsymbol{\Phi} \rVert_2^{(K)} \ne 0$. In this paper, ϵ, $\widehat{\epsilon}_e$, and ϵ_e are assumed to be far less than 1.

In this paper, our contributions can be stated as follows.

(i) For (1.1), let $\boldsymbol{x}$ be K-sparse and $T = \mathrm{supp}(\boldsymbol{x})$ with $|T| \le K$. Suppose $\|\boldsymbol{e}\|_2 \le \eta$ and $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ with unit l_2-norm columns satisfies the RIP with

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

Then OLS with the stopping rule $\|\boldsymbol{r}^k\|_2 \le \eta$ exactly recovers the support T from (1.1) in K iterations, provided that

$$\min_{i\in T} |x_i| > \frac{2\eta}{\sqrt{1-\delta_{K+1}}(1-\sqrt{K+1}\delta_{K+1})}.$$

(ii) Let $t_0 = \min_{i\in T} |x_i|$ and

$$\epsilon_0 = \frac{1.31}{1-\epsilon}\big(\epsilon + \epsilon_e + (1+\epsilon_e)(\beta+\gamma)\big)\,\|\boldsymbol{x}_T\|_2,$$

where $\beta, \gamma, \epsilon, \epsilon_e$ are defined in (1.6) and (1.7), and they are far less than 1. If $\widehat{\boldsymbol{\Phi}}$ in (1.3) with unit l_2-norm columns satisfies the RIP of order $K+1$ with

$$\widehat{\delta}_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}},$$

then OLS exactly recovers the support of $\boldsymbol{x}_{\max(K)}$ from (1.3) in K iterations.

(iii) Let $t_0 = \min_{i\in T} |x_i|$ and

$$\epsilon_0 = 1.31\big(\epsilon + \widehat{\epsilon}_e + \epsilon\widehat{\epsilon}_e + (1+\widehat{\epsilon}_e)(1+\epsilon)(\beta+\gamma)\big)\,\|\boldsymbol{x}_T\|_2.$$

If $\boldsymbol{\Phi}$ in (1.4) with unit l_2-norm columns satisfies the RIP of order $K+1$ with

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}},$$

then OLS exactly recovers the support of $\boldsymbol{x}_{\max(K)}$ from (1.4) in K iterations for any almost sparse signal $\boldsymbol{x}$.

2 Notation and Preliminaries

Before moving on to the main results of this paper, we need some preliminaries and notations.

Throughout this paper, let $\Omega = \{1,\dots,n\}$. For $S \subseteq \Omega$, $|S|$ is the cardinality of S, $\boldsymbol{\Phi}_S$ denotes the submatrix of S that contains only the columns indexed by S, and $\boldsymbol{x}_S$ denotes the subvector of $\boldsymbol{x}$ that contains only the entries indexed by S.

Suppose $\boldsymbol{\Phi}$ is normalized to have unit columns (*i.e.*, $\|\boldsymbol{\Phi}_i\|_2 = 1$ for all $i \in \Omega$), and $\boldsymbol{\Phi}_S^\dagger$ represents the pseudo-inverse of $\boldsymbol{\Phi}_S$. When $\boldsymbol{\Phi}_S$ is full column rank, $\boldsymbol{\Phi}_S^\dagger = (\boldsymbol{\Phi}_S'\boldsymbol{\Phi}_S)^{-1}\boldsymbol{\Phi}_S'$; $\mathrm{span}(\boldsymbol{\Phi}_S)$ is the span of columns in $\boldsymbol{\Phi}_S$. $\boldsymbol{P}_S = \boldsymbol{\Phi}_S\boldsymbol{\Phi}_S^\dagger$ is the projection onto $\mathrm{span}(\boldsymbol{\Phi}_S)$. Then $\boldsymbol{P}_S^\perp = I - \boldsymbol{P}_S$ is the projection onto the orthogonal complement of $\mathrm{span}(\boldsymbol{\Phi}_S)$, where I is the identity matrix.

We first recall the following lemmas, which will be used in the proof of the main results.

Lemma 2.1 ([2]) *If a matrix satisfies the RIP of both orders K_1 and K_2 where $K_1 \le K_2$, then $\delta_{K_1} \le \delta_{K_2}$.*

Lemma 2.2 ([18, Proposition 3.1]) *Let $S \subset \Omega$. If $\delta_{|S|} < 1$, then for any vector $\boldsymbol{y} \in \mathbb{R}^m$,*

$$\|\boldsymbol{\Phi}_S' \boldsymbol{y}\|_2 \le \sqrt{1+\delta_{|S|}}\|\boldsymbol{y}\|_2.$$

Lemma 2.3 ([19, Lemma 1]) *For any $\boldsymbol{x} \in \mathbb{R}^{|S_1 \setminus S_2|}$, the following inequality holds:*

$$(1-\delta_{|S_1 \cup S_2|})\|\boldsymbol{x}\|_2^2 \le \|\boldsymbol{P}_{S_2}^{\perp}\boldsymbol{\Phi}_{S_1 \setminus S_2}\boldsymbol{x}\|_2^2 \le (1+\delta_{|S_1 \cup S_2|})\|\boldsymbol{x}\|_2^2.$$

Lemma 2.4 ([29, Lemma 3]) *Suppose that $S \subset \Omega$ and let $\boldsymbol{\Phi}$ have unit l_2-norm columns and satisfy the RIP of order $|S|+1$. Then for any $i \in \Omega \setminus S$,*

$$\|\boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_i\|_2 \ge \sqrt{1-\delta_{|S|+1}^2}.$$

Lemma 2.5 ([18, Proposition 3.5]) *Suppose that $\boldsymbol{\Phi}$ satisfies the RIP of K with δ_K. Then for every signal $\boldsymbol{x}$,*

$$\|\boldsymbol{\Phi x}\|_2 \le \sqrt{1+\delta_K}\Big(\|\boldsymbol{x}\|_2 + \frac{1}{\sqrt{K}}\|\boldsymbol{x}\|_1\Big).$$

Lemma 2.6 *Consider the system model (1.1) in the noisy case and the OLS algorithm. Let $\boldsymbol{r}^k$ be the residual produced in the k-th ($0 \le k < K$) iteration of OLS. Then OLS selects in the $(k+1)$-th iteration the index*

$$t^{k+1} = \arg\max_{i \in \Omega \setminus T^k} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}_{T^k}^{\perp}\boldsymbol{\Phi}_i\|_2}. \tag{2.1}$$

Proof This lemma is an extension of [29, Proposition 1]. As shown in Table 1, at the $(k+1)$-th iteration ($0 \le k < K$) OLS selects an index that results in the maximum reduction of the residual power, *i.e.*,

$$t^{k+1} = \arg\min_{i \in \Omega}\|\boldsymbol{P}_{T^k \cup \{i\}}^{\perp}\boldsymbol{y}\|_2^2. \tag{2.2}$$

Notice that $\boldsymbol{P}_{T^k \cup \{i\}}\boldsymbol{y}$ and $\boldsymbol{P}_{T^k \cup \{i\}}^{\perp}\boldsymbol{y}$ are orthogonal to each other, so we have

$$\|\boldsymbol{P}_{T^k \cup \{i\}}^{\perp}\boldsymbol{y}\|_2^2 = \|\boldsymbol{y}\|_2^2 - \|\boldsymbol{P}_{T^k \cup \{i\}}\boldsymbol{y}\|_2^2,$$

and it follows that (2.2) is equivalent to

$$t^{k+1} = \arg\max_{i \in \Omega}\|\boldsymbol{P}_{T^k \cup \{i\}}\boldsymbol{y}\|_2^2. \tag{2.3}$$

Observe the fact that $\|\boldsymbol{P}_{T^k \cup \{i\}}\boldsymbol{y}\|_2^2$ can be decomposed as follows (see [25]):

$$\|\boldsymbol{P}_{T^k \cup \{i\}}\boldsymbol{y}\|_2^2 = \|\boldsymbol{P}_{T^k}\boldsymbol{y}\|_2^2 + \Big(\frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}_{T^k}^{\perp}\boldsymbol{\Phi}_i\|_2}\Big)^2. \tag{2.4}$$

Combining (2.3) and (2.4), we obtain

$$t^{k+1} = \arg\max_{i \in \{1,\cdots,n\}} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}_{T^k}^{\perp}\boldsymbol{\Phi}_i\|_2}. \tag{2.5}$$

https://doi.org/10.4153/S0008439519000134 Published online by Cambridge University Press

Observe that

$$\begin{aligned}
\boldsymbol{r}^k &= \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^k = \boldsymbol{y} - \boldsymbol{\Phi}_{T^k}\boldsymbol{\Phi}^{\dagger}_{T^k}\boldsymbol{y} \\
&= \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{y} = \boldsymbol{P}^{\perp}_{T^k}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}) = \boldsymbol{P}^{\perp}_{T^k}(\boldsymbol{\Phi}_T\boldsymbol{x}_T + \boldsymbol{e}) \\
&= \boldsymbol{P}^{\perp}_{T^k}(\boldsymbol{\Phi}_{T^k}\boldsymbol{x}_{T^k} + \boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{e}) \\
&\overset{(a)}{=} \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e},
\end{aligned} \tag{2.6}$$

where (a) follows from

$$\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T^k} = 0. \tag{2.7}$$

Hence, by (2.6) and (2.7), for $i \in T^k$, we have

$$\langle \boldsymbol{r}^k, \boldsymbol{\Phi}_i\rangle = \boldsymbol{\Phi}'_i\boldsymbol{r}^k = 0.$$

Therefore, (2.5) is equivalent to (2.1). ∎

Next, we show the main lemma that will play a key role during our analysis.

Lemma 2.7 *Suppose that S is any subset of T ($S \subseteq T$) and $\boldsymbol{\Phi}$ satisfies the RIP of order $|T| + 1$. Then for all $\mu > 0$,*

$$\sqrt{|T| - |S|}\|\boldsymbol{\Phi}'_{T\setminus S}P^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty - \mu\|\boldsymbol{\Phi}'_{T^c}P^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty \geq \Big(1 - \sqrt{\mu^2+1}\delta_{|T|+1}\Big)\|\boldsymbol{x}_{T\setminus S}\|_2.$$

Proof The idea of the proof comes from [29, Lemma 4].
Obviously, it suffices to show for each $j \in T^c$,

$$\sqrt{|T| - |S|}\|\boldsymbol{\Phi}'_{T\setminus S}\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty - \mu|\boldsymbol{\Phi}'_j\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}| \geq \Big(1 - \sqrt{\mu^2+1}\delta_{|T|+1}\Big)\|\boldsymbol{x}_{T\setminus S}\|_2. \tag{2.8}$$

We can apply the fact that $\boldsymbol{P}^{\perp}_S = (\boldsymbol{P}^{\perp}_S)' = (\boldsymbol{P}^{\perp}_S)^2$ to yield that

$$\begin{aligned}
&\sqrt{|T| - |S|}\|\boldsymbol{x}_{T\setminus S}\|_2\|\boldsymbol{\Phi}'_{T\setminus S}\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty \\
&\geq \|\boldsymbol{x}_{T\setminus S}\|_1\|\boldsymbol{\Phi}'_{T\setminus S}\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty \\
&= \Big(\sum_{l\in T\setminus S}|\boldsymbol{x}_l|\Big)\|\boldsymbol{\Phi}'_{T\setminus S}\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_\infty \\
&\geq \sum_{l\in T\setminus S}(\boldsymbol{x}_l\boldsymbol{\Phi}'_l\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}) \\
&= \Big(\sum_{l\in T\setminus S}\boldsymbol{x}_l\boldsymbol{\Phi}_l\Big)'\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \\
&= (\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S})'\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \\
&= (\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S})'\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \\
&= \|\boldsymbol{P}^{\perp}_S\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}\|_2^2.
\end{aligned} \tag{2.9}$$

https://doi.org/10.4153/S0008439519000134 Published online by Cambridge University Press

Let

$$\alpha = \frac{1 - \sqrt{\mu^2 + 1}}{\mu}.$$

It is easily verified that $\alpha^2 < 1$,

$$\frac{2\alpha}{1-\alpha^2} = -\mu, \tag{2.10}$$

$$\frac{1+\alpha^2}{1-\alpha^2} = \sqrt{\mu^2+1}. \tag{2.11}$$

For simplicity of notation, given $j \in T^c$, denote

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{P}_S^{\perp}[\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{\Phi}_j], \\ \boldsymbol{u} &= [\boldsymbol{x}'_{T\setminus S} 0]' \in \mathbb{R}^{|T\setminus S|+1}, \\ \boldsymbol{v} &= [\boldsymbol{0}' \alpha t \|\boldsymbol{x}_{T\setminus S}\|_2]' \in \mathbb{R}^{|T\setminus S|+1}, \end{aligned}$$

where

$$t = \begin{cases} 1 & \text{if } \boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp} \boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \geq 0, \\ -1 & \text{if } \boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp} \boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} < 0. \end{cases}$$

Then

$$\boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} = \boldsymbol{A}\boldsymbol{u}, \quad \boldsymbol{P}_S^{\perp}\alpha t\|\boldsymbol{x}_{T\setminus S}\|_2 = \boldsymbol{A}\boldsymbol{v}, \tag{2.12}$$

and

$$\|\boldsymbol{u}+\boldsymbol{v}\|_2^2 = (1+\alpha^2)\|\boldsymbol{x}_{T\setminus S}\|_2^2, \tag{2.13}$$

$$\|\alpha^2\boldsymbol{u}-\boldsymbol{v}\|_2^2 = \alpha^2(1+\alpha^2)\|\boldsymbol{x}_{T\setminus S}\|_2^2. \tag{2.14}$$

By the definition of t, we have

$$\begin{aligned} \boldsymbol{v}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{u} &= \alpha t\|\boldsymbol{x}_{T\setminus S}\|_2 \boldsymbol{\Phi}'_j (\boldsymbol{P}_S^{\perp})' \boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \\ &= \alpha t\|\boldsymbol{x}_{T\setminus S}\|_2 \boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S} \\ &= \alpha\|\boldsymbol{x}_{T\setminus S}\|_2 |\boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}|. \end{aligned}$$

Therefore, for $j \in T^c$, we have

$$\begin{aligned} \|\boldsymbol{A}(\boldsymbol{u}+\boldsymbol{v})\|_2^2 &= \|\boldsymbol{A}\boldsymbol{u}\|_2^2 + \|\boldsymbol{A}\boldsymbol{v}\|_2^2 + 2\boldsymbol{v}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{u} \\ &= \|\boldsymbol{A}\boldsymbol{u}\|_2^2 + \|\boldsymbol{A}\boldsymbol{v}\|_2^2 + 2\alpha\|\boldsymbol{x}_{T\setminus S}\|_2 |\boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}| \end{aligned}$$

and

$$\|\boldsymbol{A}(\alpha^2\boldsymbol{u}-\boldsymbol{v})\|_2^2 = \alpha^4\|\boldsymbol{A}\boldsymbol{u}\|_2^2 + \|\boldsymbol{A}\boldsymbol{v}\|_2^2 - 2\alpha^3\|\boldsymbol{x}_{T\setminus S}\|_2 |\boldsymbol{\Phi}'_j \boldsymbol{P}_S^{\perp}\boldsymbol{\Phi}_{T\setminus S}\boldsymbol{x}_{T\setminus S}|.$$

https://doi.org/10.4153/S0008439519000134 Published online by Cambridge University Press

Moreover, by direct computation, we have

$$\begin{aligned}(2.15)\qquad &\|A(u+v)\|_2^2 - \|A(\alpha^2 u - v)\|_2^2\\ &= (1-\alpha^4)\|Au\|_2^2 + 2\alpha(1+\alpha^2)\|x_{T\setminus S}\|_2 |\Phi_j' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}|\\ &= (1-\alpha^4)\Big(\|Au\|_2^2 + \frac{2\alpha}{1-\alpha^2}\|x_{T\setminus S}\|_2 |\Phi_j' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}|\Big)\\ &= (1-\alpha^4)\Big(\|P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}\|_2^2 - \mu\|x_{T\setminus S}\|_2 |\Phi_j' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}|\Big),\end{aligned}$$

where the first part of last equality follows from (2.12), and the second part follows from (2.10).

From Lemma 2.3, (2.13), and (2.14), we can get

$$\begin{aligned}(2.16)\qquad &\|A(u+v)\|_2^2 - \|A(\alpha^2 u - v)\|_2^2\\ &\geq (1-\delta_{|T|+1})\|u+v\|_2^2 - (1+\delta_{|T|+1})\|\alpha^2 u - v\|_2^2\\ &= (1-\delta_{|T|+1})(1+\alpha^2)\|x_{T\setminus S}\|_2^2\\ &\quad - (1+\delta_{|T|+1})\alpha^2(1+\alpha^2)\|x_{T\setminus S}\|_2^2\\ &= (1+\alpha^2)\|x_{T\setminus S}\|_2^2\big((1-\delta_{|T|+1}) - (1+\delta_{|T|+1})\alpha^2\big)\\ &= (1-\alpha^4)\|x_{T\setminus S}\|_2^2\Big(1 - \frac{1+\alpha^2}{1-\alpha^2}\delta_{|T|+1}\Big)\\ &= (1-\alpha^4)\|x_{T\setminus S}\|_2^2\big(1 - \sqrt{\mu^2+1}\,\delta_{|T|+1}\big),\end{aligned}$$

where the last equality is due to (2.11).

By combining (2.15), (2.16), and the fact that $1-\alpha^4 > 0$, we obtain

$$(2.17)\qquad \|P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}\|_2^2 - \mu\|x_{T\setminus S}\|_2 |\Phi_j' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}| \geq \|x_{T\setminus S}\|_2^2\big(1 - \sqrt{\mu^2+1}\,\delta_{|T|+1}\big).$$

Combining (2.17) with (2.9), we have

$$\begin{aligned}\sqrt{|T|-|S|}\|x_{T\setminus S}\|_2 \|\Phi_{T\setminus S}' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}\|_\infty - \mu\|x_{T\setminus S}\|_2 |\Phi_j' P_S^{\perp}\Phi_{T\setminus S}x_{T\setminus S}|\\ \geq \|x_{T\setminus S}\|_2^2\big(1 - \sqrt{\mu^2+1}\,\delta_{|T|+1}\big).\end{aligned}$$

Therefore, (2.8) holds, which completes the proof of the lemma. ■

Remark 2.8 In [17], Mo established a key lemma (Lemma *II*.1). In Lemma 2.7, α is defined in the same way as in [17], in order to ensure that (2.15) holds. When $\mu = \sqrt{|T|-|S|}$, Lemma 2.7 reduces to [29, Lemma 1].

3 Main Results

In this section, we study the recovery performance of OLS under the RIP condition. We first consider (1.1) in the noisy case, and the result is the following theorem.

Theorem 3.1 *Let $\boldsymbol{x}$ be K-sparse with $T = \operatorname{supp}(\boldsymbol{x})$, $|T| \le K$, and $\|\boldsymbol{e}\|_2 \le \eta$. Suppose that the measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ with unit l_2-norm columns satisfies the RIP with*

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}}. \tag{3.1}$$

Then OLS with the stopping rule $\|\boldsymbol{r}^k\|_2 \le \eta$ exactly recovers the support T from (1.1) in K iterations, provided that

$$\min_{i\in T} |x_i| > \frac{2\eta}{\sqrt{1-\delta_{K+1}}(1-\sqrt{K+1}\delta_{K+1})}. \tag{3.2}$$

Proof We prove the result by induction. Suppose that OLS makes the correct selection in each of the previous k iterations with $0 \le k < K$, *i.e.*, $T^k \subset T$. Then we need to show that the OLS algorithm also selects a correct index at the $(k+1)$-th iteration, *i.e.*, $t^{k+1} \in T \setminus T^k$. Thus, the proof of the first selection corresponds to the case where $k = 0$. Clearly the induction hypothesis $T^k \subset T$ holds for this case since $T^0 = \varnothing$.

In order to show that $t^{k+1} \in T \setminus T^k$, by Lemma 2.6, we need to prove that

$$\max_{i\in T\setminus T^k} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2} > \max_{i\in T^c} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2}. \tag{3.3}$$

We now analyze (3.3). According to the fact that $\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2 \le \|\boldsymbol{\Phi}_i\|_2 = 1$, we have

$$\begin{aligned}
\max_{i\in T\setminus T^k} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2} &\ge \max_{i\in T\setminus T^k} |\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle| \\
&= \|\boldsymbol{\Phi}'_{T\setminus T^k}(\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e})\|_\infty \\
&\ge \|\boldsymbol{\Phi}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty - \|\boldsymbol{\Phi}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty,
\end{aligned} \tag{3.4}$$

where the equality is from (2.6).

Let $j_0 := \arg\max_{i\in T^c} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2}$; then

$$\begin{aligned}
\max_{i\in T^c} \frac{|\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_i\|_2} &= \frac{|\langle \boldsymbol{\Phi}_{j_0}, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \le \frac{\max_{i\in T^c} |\langle \boldsymbol{\Phi}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&= \frac{\|\boldsymbol{\Phi}'_{T^c}(\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e})\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&\le \frac{\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty + \|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2}.
\end{aligned} \tag{3.5}$$

Therefore, combining (3.4) with (3.5), (3.3) is guaranteed if

$$\begin{aligned}
&\|\boldsymbol{\Phi}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty - \frac{\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&\qquad > \|\boldsymbol{\Phi}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty + \frac{\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2}.
\end{aligned} \tag{3.6}$$

https://doi.org/10.4153/S0008439519000134 Published online by Cambridge University Press

Next, we build a lower bound on the left-hand side of (3.6). It follows from Lemma 2.7 with

$$\mu = \frac{\sqrt{|T| - |T^k|}}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \quad \text{and} \quad S = T^k,$$

we have

$$\begin{aligned}
&\sqrt{|T| - |T^k|}\|\boldsymbol{\Phi}'_{T\smallsetminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\smallsetminus T^k}\boldsymbol{x}_{T\smallsetminus T^k}\|_\infty \\
&\qquad - \frac{\sqrt{|T| - |T^k|}}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2}\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\smallsetminus T^k}\boldsymbol{x}_{T\smallsetminus T^k}\|_\infty \\
&\qquad \geq \Big(1 - \sqrt{\frac{|T| - |T^k|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2^2} + 1}\,\delta_{|T|+1}\Big)\|\boldsymbol{x}_{T\smallsetminus T^k}\|_2 \\
&\qquad \overset{(a)}{\geq} \Big(1 - \sqrt{\frac{|T| - |T^k|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2^2} + 1}\,\delta_{|T|+1}\Big)\sqrt{|T| - |T^k|}\min_{i\in T}|x_i| \\
&\qquad \overset{(b)}{\geq} (1 - \sqrt{K+1}\,\delta_{K+1})\sqrt{|T| - |T^k|}\min_{i\in T}|x_i|;
\end{aligned}$$

i.e.,

$$\|\boldsymbol{\Phi}'_{T\smallsetminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\smallsetminus T^k}\boldsymbol{x}_{T\smallsetminus T^k}\|_\infty - \frac{\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\smallsetminus T^k}\boldsymbol{x}_{T\smallsetminus T^k}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \geq (1 - \sqrt{K+1}\,\delta_{K+1})\min_{i\in T}|x_i|, \tag{3.7}$$

where (a) is due to

$$\|\boldsymbol{x}_{T\smallsetminus T^k}\|_2 \geq \sqrt{|T| - |T^k|}\min_{i\in T\smallsetminus T^k}|x_i| \geq \sqrt{|T| - |T^k|}\min_{i\in T}|x_i|,$$

and (b) follows from $(|T| - |T^k|)/(\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2^2) \leq K$ (*i.e.*, [29, (C.4)]) under $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ and Lemma 2.1 with $|T| \leq K$.

We next give an upper bound on the right-hand side of (3.6). Obviously, there exist $i_0 \in T \smallsetminus T^k$ and $k_0 \in T^c$ such that

$$\|\boldsymbol{\Phi}'_{T\smallsetminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty = |\boldsymbol{\Phi}'_{i_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}|, \quad \|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty = |\boldsymbol{\Phi}'_{k_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}|.$$

Thus, by $\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2 \leq \|\boldsymbol{\Phi}_{j_0}\|_2 = 1$, we have

$$\begin{aligned}
&\|\boldsymbol{\Phi}'_{T\smallsetminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty + \frac{\|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&\quad \leq \frac{\|\boldsymbol{\Phi}'_{T\smallsetminus T^k}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty + \|\boldsymbol{\Phi}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&\quad = \frac{|\boldsymbol{\Phi}'_{i_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}| + |\boldsymbol{\Phi}'_{k_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}|}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} = \frac{\|\boldsymbol{\Phi}'_{i_0\cup k_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_1}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \\
&\quad \leq \frac{\sqrt{2}\|\boldsymbol{\Phi}'_{i_0\cup k_0}\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_2}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2} \leq \frac{\sqrt{2(1+\delta_{K+1})}}{\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{j_0}\|_2}\|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_2 \leq \frac{2\eta}{\sqrt{1-\delta_{K+1}}},
\end{aligned} \tag{3.8}$$

where the last inequality is from Lemma 2.4 and the assumed l_2-bound of the perturbation $\boldsymbol{e}$.

By combining (3.7) and (3.8), (3.3) is guaranteed by

$$(1-\sqrt{K+1}\delta_{K+1})\min_{i\in T}|x_i| > \frac{2\eta}{\sqrt{1-\delta_{K+1}}},$$

i.e.,

$$\min_{i\in T}|x_i| > \frac{2\eta}{\sqrt{1-\delta_{K+1}}(1-\sqrt{K+1}\delta_{K+1})}.$$

Thus, if (3.2) holds, then the OLS algorithm selects a correct index in each iteration.

What remains to show is that the OLS algorithm performs exact $|T|$ iterations, which is equivalent to showing that $\|\boldsymbol{r}^k\|_2 > \eta$ for $1 \le k < |T|$ and $\|\boldsymbol{r}^{|T|}\|_2 \le \eta$.

Since OLS selects a correct index in each iteration under (3.2), for $1 \le k < |T|$, according to Lemma 2.3 and (3.2), we have

$$\begin{aligned}\|\boldsymbol{r}^k\|_2 &= \|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_2\\ &\ge \|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_2 - \|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{e}\|_2\\ &\ge \|\boldsymbol{P}^{\perp}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_2 - \eta\\ &\ge \sqrt{1-\delta_{|T|}}\|\boldsymbol{x}_{T\setminus T^k}\|_2 - \eta\\ &\ge \sqrt{1-\delta_{K+1}}\sqrt{|T|-k}\min_{i\in T}|x_i| - \eta\\ &\ge \sqrt{1-\delta_{K+1}}\min_{i\in T}|x_i| - \eta > \eta.\end{aligned}$$

Therefore, if (3.2) holds, $\|\boldsymbol{r}^k\|_2 > \eta$ for $1 \le k < |T|$, *i.e.*, the OLS algorithm does not terminate before the $|T|$-th iteration.

Owing to $T^{|T|} = T$, we can get

$$\|\boldsymbol{r}^{|T|}\|_2 = \|\boldsymbol{P}^{\perp}_{T^{|T|}}\boldsymbol{\Phi}_{T\setminus T^{|T|}}\boldsymbol{x}_{T\setminus T^{|T|}} + \boldsymbol{P}^{\perp}_{T^{|T|}}\boldsymbol{e}\|_2 = \|\boldsymbol{P}^{\perp}_{T^{|T|}}\boldsymbol{e}\|_2 \le \eta.$$

Consequently, by the stopping rule, the OLS algorithm terminates after the $|T|$-th iteration. Hence, the OLS algorithm performs $|T|$ iterations. The proof is now completed. ∎

Remark 3.2 In Theorem 3.1, we provide a sufficient condition for the support recovery of sparse signals via OLS. In the noiseless case (*i.e.*, $\boldsymbol{e} = \boldsymbol{0}$), the condition (3.1) is nearly optimal for the recovery of K-sparse signal $\boldsymbol{x}$ (see [29]). Recently, Kim, Wang, and Shim, in an unpublished manuscript, provided a sharp RIP bound for OLS:

$$\delta_{K+1} < \frac{1}{\sqrt{K+\frac{1}{4}}}.$$

It is an interesting question to consider the support recovery of sparse signals in the noisy case on this RIP bound. We will deal with this problem in our future work.

Taking a completely perturbed model (1.3) into account, we show that the support of the best K-term approximation $\boldsymbol{x}_{\max(K)}$ of $\boldsymbol{x}$ can be exactly recovered under the RIP-based condition, if $\boldsymbol{x}$ is assumed to be almost sparse. See the following theorem.

Theorem 3.3 *Let* $t_0 = \min_{i\in T}|x_i|$ *and*

$$\epsilon_0 = \frac{1.31}{1-\epsilon}\big(\epsilon + \epsilon_{\boldsymbol{e}} + (1+\epsilon_{\boldsymbol{e}})(\beta+\gamma)\big)\,\|\boldsymbol{x}_T\|_2,$$

where $\beta, \gamma, \epsilon, \epsilon_e$ *are defined in* (1.6) *and* (1.7)*, and they are far less than* 1*. If* $\widehat{\boldsymbol{\Phi}}$ *in* (1.3) *has unit* l_2*-norm columns and satisfies the RIP of order* $K+1$ *with*

$$\widehat{\delta}_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}}, \tag{3.9}$$

then OLS will exactly recover the support $T = \operatorname{supp}(\boldsymbol{x}_{\max(K)})$ *of* $\boldsymbol{x}_T$ *from* (1.3) *in* K *iterations.*

Proof From (1.3) and $\boldsymbol{x} = \boldsymbol{x}_T + \boldsymbol{x}_{T^c}$, it follows that

$$\begin{aligned}\widehat{\boldsymbol{y}} &= \boldsymbol{\Phi x} + \boldsymbol{e} = (\widehat{\boldsymbol{\Phi}} - \boldsymbol{E})(\boldsymbol{x}_T + \boldsymbol{x}_{T^c}) + \boldsymbol{e}\\ &= \widehat{\boldsymbol{\Phi}}\boldsymbol{x}_T + (\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c} - \boldsymbol{Ex} + \boldsymbol{e}) = \widehat{\boldsymbol{\Phi}}\boldsymbol{x}_T + \widehat{\boldsymbol{e}},\end{aligned} \tag{3.10}$$

where $\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c} - \boldsymbol{Ex} + \boldsymbol{e}$.

We will give an upper bound of $\|\widehat{\boldsymbol{e}}\|_2$. According to (1.7), we have

$$\|\boldsymbol{E}\|_2^{(K)} \le \epsilon\|\boldsymbol{\Phi}\|_2^{(K)} = \epsilon\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{E}\|_2^{(K)} \le \epsilon\|\widehat{\boldsymbol{\Phi}}\|_2^{(K)} + \epsilon\|\boldsymbol{E}\|_2^{(K)}.$$

Since ϵ is assumed to be far less than 1,

$$\|\boldsymbol{E}\|_2^{(K)} \le \frac{\epsilon}{1-\epsilon}\|\widehat{\boldsymbol{\Phi}}\|_2^{(K)}. \tag{3.11}$$

Using Lemma 2.5, it follows that

$$\begin{aligned}\|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c}\|_2 &\le \sqrt{1+\widehat{\delta}_K}\Big(\|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &\overset{(a)}{=} \sqrt{1+\widehat{\delta}_K}(\beta+\gamma)\|\boldsymbol{x}_T\|_2\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}\|\boldsymbol{Ex}\|_2 &\le \|\boldsymbol{Ex}_T\|_2 + \|\boldsymbol{Ex}_{T^c}\|_2\\ &\le \|\boldsymbol{E}\|_2^{(K)}\|\boldsymbol{x}_T\|_2 + \|\boldsymbol{E}\|_2^{(K)}\Big(\|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &= \|\boldsymbol{E}\|_2^{(K)}\Big(\|\boldsymbol{x}_T\|_2 + \|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &\overset{(b)}{\le} \frac{\epsilon}{1-\epsilon}\sqrt{1+\widehat{\delta}_K}(1+\beta+\gamma)\|\boldsymbol{x}_T\|_2,\end{aligned} \tag{3.13}$$

where (a) is from (1.6), and (b) is from (1.6), (3.11), and $\|\widehat{\boldsymbol{\Phi}}\|_2^{(K)} \le \sqrt{1+\widehat{\delta}_K}$.

Furthermore, by the triangle inequality and $\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c} - \boldsymbol{E}\boldsymbol{x} + \boldsymbol{e}$, we obtain

$$\begin{aligned}
\|\widehat{\boldsymbol{e}}\|_2 &\leq \|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2 + \|\boldsymbol{e}\|_2 \\
&\overset{(a)}{\leq} \|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2 + \epsilon_e\|\boldsymbol{\Phi}\boldsymbol{x}\|_2 \\
&\overset{(b)}{\leq} \|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2 + \epsilon_e\|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}\|_2 + \epsilon_e\|\boldsymbol{E}\boldsymbol{x}\|_2 \\
&\overset{(c)}{\leq} (1+\epsilon_e)(\|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_{T^c}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2) + \epsilon_e\|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}_T\|_2 \\
&\overset{(d)}{\leq} \frac{\sqrt{1+\widehat{\delta}_{K+1}}}{1-\epsilon}\big(\epsilon + \epsilon_e + (1+\epsilon_e)(\beta+\gamma)\big)\|\boldsymbol{x}_T\|_2,
\end{aligned}$$

where (a) and (b) follow from (1.7) and $\widehat{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \boldsymbol{E}$ in (1.3), respectively; (c) is due to $\boldsymbol{x} = \boldsymbol{x}_T + \boldsymbol{x}_{T^c}$ and the triangle inequality; (d) is from (1.5), (3.12), (3.13) and $|T| \leq K$. Noticing that $\widehat{\delta}_{K+1} < \frac{1}{\sqrt{K+1}} \leq \frac{1}{\sqrt{2}}$, one has $\|\widehat{\boldsymbol{e}}\|_2 < \epsilon_0$.

Similar to Theorem 3.1, to guarantee that OLS chooses a correct index in every iteration, i.e., $t^{k+1} \in T$ with $0 \leq k < K$, we only need to show that

$$\max_{i\in T\setminus T^k} \frac{|\langle\widehat{\boldsymbol{\Phi}}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_i\|_2} > \max_{i\in T^c} \frac{|\langle\widehat{\boldsymbol{\Phi}}_i, \boldsymbol{r}^k\rangle|}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_i\|_2}, \tag{3.14}$$

where $\boldsymbol{r}^k = \boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{e}}$.

We find (3.14) is guaranteed by

$$\begin{aligned}
&\|\widehat{\boldsymbol{\Phi}}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty - \frac{\|\widehat{\boldsymbol{\Phi}}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{j_0}\|_2} \\
&\qquad > \|\widehat{\boldsymbol{\Phi}}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{e}}\|_\infty + \frac{\|\widehat{\boldsymbol{\Phi}}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{e}}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{j_0}\|_2}.
\end{aligned} \tag{3.15}$$

The left-hand side of (3.15) satisfies

$$\begin{aligned}
&\|\widehat{\boldsymbol{\Phi}}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty - \frac{\|\widehat{\boldsymbol{\Phi}}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{j_0}\|_2} \geq \\
&\qquad (1-\sqrt{K+1}\widehat{\delta}_{K+1})\min_{i\in T}|x_i|,
\end{aligned} \tag{3.16}$$

and the right-hand side of (3.15) satisfies

$$\|\widehat{\boldsymbol{\Phi}}'_{T\setminus T^k}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{e}}\|_\infty + \frac{\|\widehat{\boldsymbol{\Phi}}'_{T^c}\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{e}}\|_\infty}{\|\boldsymbol{P}^{\perp}_{T^k}\widehat{\boldsymbol{\Phi}}_{j_0}\|_2} \leq \frac{\sqrt{2}\|\widehat{\boldsymbol{e}}\|_2}{\sqrt{1-\widehat{\delta}_{K+1}}} \leq \frac{\sqrt{2}\|\widehat{\boldsymbol{e}}\|_2}{1-\widehat{\delta}_{K+1}}. \tag{3.17}$$

By (3.16) and (3.17), we can show that (3.15) holds true if

$$\widehat{\delta}_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\|\widehat{\boldsymbol{e}}\|_2}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}}. \tag{3.18}$$

Consequently, under (3.18), a correct index is chosen at the every iteration of OLS.

Combining (3.18) and the fact that $\|\widehat{\boldsymbol{e}}\|_2 < \epsilon_0$, (3.9) ensures the selecting of all support indices with OLS in K iterations. ∎

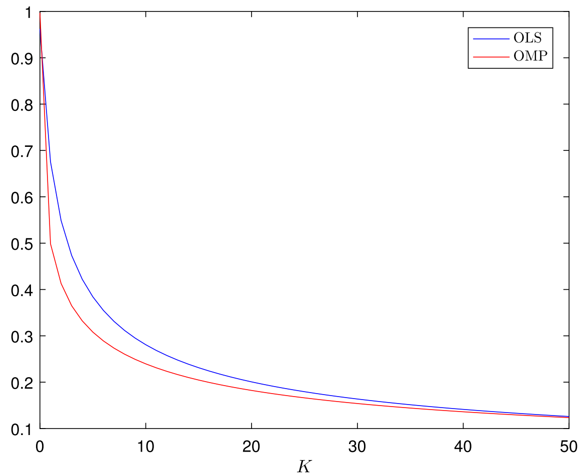


Figure 1: The cures of the RIC bounds in (3.9) and (3.19) as functions of sparsity.

Remark 3.4 In [8], the OMP algorithm was studied under the general perturbation setting, in which the authors obtained the condition

$$\widehat{\delta}_{K+1} < \frac{1}{\sqrt{K}+1} - \frac{3\epsilon_0}{(\sqrt{K}+1)t_0}. \tag{3.19}$$

Since OMP and OLS share many aspects in common and differ only in the identification principle, we provide some comparison between condition (3.9) and (3.19). When $\epsilon_0 = 0$, (3.9) is less restrictive than the condition (3.19). For the same $\epsilon_0 > 0$, t_0 and K, when $\frac{\epsilon_0}{t_0} = 0.001$, we plot the cures of (3.9) and (3.19) as a function of sparsity in Figure 1.

For the model (1.3), if $\boldsymbol{E} = \boldsymbol{0}$, then (1.3) turns to (1.1). Based on Theorem 3.3, we present the recovery condition of OLS for (1.1) from the other perspective.

Corollary 3.5 *For* (1.3)*, let* $\boldsymbol{E} = \boldsymbol{0}$ *and* $\boldsymbol{x}$ *be* K*-sparse with* $\operatorname{supp}(\boldsymbol{x}) = T$ *and* $|T| \leq K$*. Let* $\epsilon_0 = 1.31\epsilon_e \|\boldsymbol{x}_T\|_2$*. If* $\boldsymbol{\Phi}$ *with unit* l_2*-norm columns satisfies the RIP of order* $K+1$ *with*

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}},$$

then OLS will exactly recover the support T *from* (1.3) *in* K *iterations.*

When neither the measurement vector nor the sampling matrix is perturbed (*i.e.*, $\boldsymbol{e} = \boldsymbol{0}$, $\boldsymbol{E} = \boldsymbol{0}$ in (1.3), we have the following corollary by Theorem 3.3, which shows that the support of the best K-term approximation $\boldsymbol{x}_{\max(K)}$ for the almost sparse signal $\boldsymbol{x}$ can be exactly recovered.

Corollary 3.6 *For the model* (1.3)*, let* $\boldsymbol{e} = \mathbf{0}, \boldsymbol{E} = \mathbf{0}$*, and let* $\boldsymbol{x}$ *be almost sparse with* $\operatorname{supp}(\boldsymbol{x}_{\max(K)}) = T$*. Let*

$$\epsilon_0 = 1.31(\beta + \gamma)\|\boldsymbol{x}_T\|_2.$$

If $\boldsymbol{\Phi}$ *with unit* l_2*-norm columns satisfies the RIP of order* $K+1$ *with*

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}},$$

then OLS will exactly recover the support T *from* (1.3) *in* K *iterations.*

Similarly, we consider the perturbed model (1.4).

Theorem 3.7 *Let* $t_0 = \min_{i\in T}|x_i|$ *and*

$$\epsilon_0 = 1.31\big(\epsilon + \widehat{\epsilon}_e + \epsilon\widehat{\epsilon}_e + (1+\widehat{\epsilon}_e)(1+\epsilon)(\beta+\gamma)\big)\|\boldsymbol{x}_T\|_2,$$

where β *and* γ *are defined in* (1.6)*,* ϵ *and* $\widehat{\epsilon}_{\boldsymbol{e}}$ *are defined in* (1.7)*, and they are far less than* 1*. If* $\boldsymbol{\Phi}$ *in* (1.4) *with unit* l_2*-norm columns satisfies the RIP of order* $K+1$ *with*

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}} - \Big(\frac{\sqrt{2}\epsilon_0}{\sqrt{K+1}t_0}\Big)^{\frac{1}{2}},$$

then OLS exactly recovers the support of $\boldsymbol{x}_{\max(K)}$ *from* (1.4) *in* K *iterations.*

Proof The idea of the proof is similar to that of Theorem 3.3, so we only present the outline of the proof. By (1.4), one has that

$$\begin{aligned}\widehat{\boldsymbol{y}} &= \widehat{\boldsymbol{\Phi}}\boldsymbol{x} + \boldsymbol{e} = (\boldsymbol{\Phi} + \boldsymbol{E})\boldsymbol{x} + \boldsymbol{e}\\ &= \boldsymbol{\Phi}\boldsymbol{x}_T + \boldsymbol{\Phi}\boldsymbol{x}_{T^c} + \boldsymbol{E}\boldsymbol{x} + \boldsymbol{e} = \boldsymbol{\Phi}\boldsymbol{x}_T + \bar{\boldsymbol{e}}.\end{aligned}$$

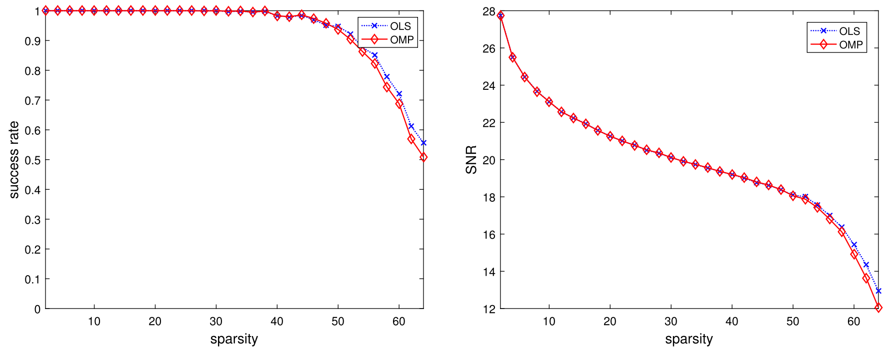
where $\bar{\boldsymbol{e}} = \boldsymbol{\Phi}\boldsymbol{x}_{T^c} + \boldsymbol{E}\boldsymbol{x} + \boldsymbol{e}$. Clearly, $\boldsymbol{\Phi}$ and $\bar{\boldsymbol{e}}$ in the above equality correspond to $\widehat{\boldsymbol{\Phi}}$ and $\widehat{\boldsymbol{e}}$ in (3.10), respectively. Thus, we only need to give an upper bound of $\|\bar{\boldsymbol{e}}\|_2$.

According to Lemma 2.5, we have

$$\begin{aligned}\|\boldsymbol{\Phi}\boldsymbol{x}_{T^c}\|_2 &\le \sqrt{1+\delta_K}\Big(\|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &= \sqrt{1+\delta_K}(\beta+\gamma)\|\boldsymbol{x}_T\|_2,\end{aligned} \tag{3.20}$$

and

$$\begin{aligned}\|\boldsymbol{E}\boldsymbol{x}\|_2 &\le \|\boldsymbol{E}\boldsymbol{x}_T\|_2 + \|\boldsymbol{E}\boldsymbol{x}_{T^c}\|_2\\ &\le \|\boldsymbol{E}\|_2^{(K)}\|\boldsymbol{x}_T\|_2 + \|\boldsymbol{E}\|_2^{(K)}\Big(\|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &= \|\boldsymbol{E}\|_2^{(K)}\Big(\|\boldsymbol{x}_T\|_2 + \|\boldsymbol{x}_{T^c}\|_2 + \frac{\|\boldsymbol{x}_{T^c}\|_1}{\sqrt{K}}\Big)\\ &\le \epsilon\sqrt{1+\delta_K}(1+\beta+\gamma)\|\boldsymbol{x}_T\|_2.\end{aligned} \tag{3.21}$$



(a) (b)

Figure 2: Comparison of success rates of OMP and OLS. (a) Noise-free case. (b) Noisy case.

Therefore, by the triangle inequality and $\bar{\boldsymbol{e}} = \boldsymbol{\Phi x}_{T^c} + \boldsymbol{Ex} + \boldsymbol{e}$, we get

$$\begin{aligned}\|\bar{\boldsymbol{e}}\|_2 &\le \|\boldsymbol{\Phi x}_{T^c}\|_2 + \|\boldsymbol{Ex}\|_2 + \|\boldsymbol{e}\|_2\\ &\overset{(a)}{\le} \|\boldsymbol{\Phi x}_{T^c}\|_2 + \|\boldsymbol{Ex}\|_2 + \widehat{\epsilon}_e\|\widehat{\boldsymbol{\Phi}}\boldsymbol{x}\|_2\\ &\overset{(b)}{\le} \|\boldsymbol{\Phi x}_{T^c}\|_2 + \|\boldsymbol{Ex}\|_2 + \widehat{\epsilon}_e\|\boldsymbol{\Phi x}\|_2 + \widehat{\epsilon}_e\|\boldsymbol{Ex}\|_2\\ &\overset{(c)}{\le} (1+\widehat{\epsilon}_e)(\|\boldsymbol{\Phi x}_{T^c}\|_2 + \|\boldsymbol{Ex}\|_2) + \widehat{\epsilon}_e\|\boldsymbol{\Phi x}_T\|_2\\ &\overset{(d)}{\le} \sqrt{1+\delta_{K+1}}\big(\epsilon + \widehat{\epsilon}_e + \epsilon\widehat{\epsilon}_e + (1+\widehat{\epsilon}_e)(1+\epsilon)(\beta+\gamma)\big)\|\boldsymbol{x}_T\|_2\\ &< \epsilon_0,\end{aligned}$$

where (a) and (b) follow from (1.7) and $\widehat{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \boldsymbol{E}$ in (1.4), respectively; (c) is due to $\boldsymbol{x} = \boldsymbol{x}_T + \boldsymbol{x}_{T^c}$ and the triangle inequality; (d) is from (1.5), (3.20), (3.21) and $|T| \le K$. ■

Remark 3.8 While this paper focuses exclusively on recovering real sparse signals, we can extend the obtained results to the complex setting by using methods similar to those in [8].

4 Numerical Experiments

This section presents some numerical experiments to demonstrate our theorems. The experiments compare the performance of OMP and OLS under the noise-free and noisy cases. In each trial, we construct Gauss matrix $\boldsymbol{\Phi}$ of size $m \times n$ (where $m = 128$ and $n = 256$). The perturbation matrix $\boldsymbol{E}$ is a random Gaussian matrix. For each value of K varying from 2 to 64 with step size 2, we generate a K-sparse signal of size $n \times 1$, whose support is chosen uniformly at random and drawn independently from a standard Gaussian distribution, and the simulation is repeated for 1000 trails. As shown in Figure 2 (a), we plot the success rate as a function of the sparsity. It can be

https://doi.org/10.4153/S0008439519000134 Published online by Cambridge University Press

seen that sparse signals can be exactly recovered via OMP and OLS under the noise-free case; Figure 2 (b) presents the SNR versus sparsity K under Gaussian noise, from which one can see that sparse signals can be still stably recovered. The SNR of the recovered signals is given by

$$\mathrm{SNR} = 10\log_{10}\Big(\frac{\|\boldsymbol{x}^0\|_2}{\|\widetilde{\boldsymbol{x}} - \boldsymbol{x}^0\|_2}\Big),$$

where $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}^0$ denote the recovered signal and the true signal, respectively.

Acknowledgments The authors thank the referees for their valuable suggestion and comments that greatly improved the readability of this paper.

References

[1] T. Blumensath and M. Davies, *Compressed sensing and source separation*. In: *Int. Conf. on Independent Component Analysis and Signal Separation*, Springer-Verlag Berlin, Heidelberg, 2007, pp. 341–348.
[2] E. J. Candès and T. Tao, *Decoding by linear programming*. IEEE Trans. Inform. Theory 51(2005), no. 12, 4203–4215. https://doi.org/10.1109/TIT.2005.858979
[3] E. J. Candès and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies*. IEEE Trans. Inform. Theory 52(2006), no. 12, 5406–5425. https://doi.org/10.1109/TIT.2006.885507
[4] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*. IEEE Trans. Inform. Theory 52(2006), no. 2, 489–509. https://doi.org/10.1109/TIT.2005.862083
[5] S. Chen, S. A. Billings, and W. Luo, *Orthogonal least squares methods and their application to non-linear system identification*. Internat. J. Control. 50(1989), no. 5, 1873–1896. https://doi.org/10.1080/00207178908953472
[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*. SIAM J. Sci. Comput. 20(1998), 33–61. https://doi.org/10.1137/S1064827596304010
[7] W. Chen and H. Ge, *A sharp bound on RIC in generalized orthogonal maching pursuit*. Canad. Math. Bull. 61(2017), no. 1, 40–54. https://doi.org/10.4153/CMB-2017-009-6
[8] J. Ding, L. Chen, and Y. Gu, *Perturbation analysis of orthogonal matching pursuit*. IEEE Trans. Signal Process. 61(2013), no. 2, 398–410. https://doi.org/10.1109/TSP.2012.2222377
[9] D. L. Donoho, *Compressed sensing*. IEEE Trans. Inform. Theory 56(2006), no. 4, 1289–1306. https://doi.org/10.1109/TIT.2006.871582
[10] A. C. Fannjiang, T. Strohmer, and P. Yan, *Compressed remote sensing of sparse objects*. SIAM J. Imaging Sci. 3(2010), no. 3, 595–618. https://doi.org/10.1137/090757034
[11] S. Foucart, *Stability and robustness of weak orthogonal matching pursuits*. In: *Recent advances in harmonic analysis and applications*, Springer Proc. Math. Stat., 25, Springer, New York, 2013. https://doi.org/10.1007/978-1-4614-4565-4_30
[12] M. A. Herman and D. Needell, *Mixed operators in compressed sensing*. In: *Proceedings IEEE 44th Ann. Conf. Inf. Sci. Syst.*, Princeton, NJ, 2010, pp. 1–6.
[13] M. A. Herman and T. Strohmer, *High-resolution radar via compressed sensing*. IEEE Trans. Signal Process. 57(2009), no. 6, 2275–2284. https://doi.org/10.1109/TSP.2009.2014277
[14] M. A. Herman and T. Strohmer, *General deviants: An analysis of perturbations in compressed sensing*. IEEE J. Sel. Topics Signal Process. 4(2010), no. 2, 342–349.
[15] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, *Exact recovery conditions for sparse representations with partial support information*. IEEE Trans. Inform. Theory 59(2013), no. 11, 7509–7524.
[16] H. Li and G. Liu, *An improved analysis for support recovery with orthogonal matching pursuit under general perturbations*. IEEE Access. 99(2018), no. 6, 18856–18867.
[17] Q. Mo, *A sharp restricted isometry constant bound of orthogonal matching pursuit*. 2015. arxiv:1501.01708.
[18] D. Needell and J. A. Troop, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*. Appl. Comput. Harmon. Anal. 26(2009), no. 3, 301–321. https://doi.org/10.1016/j.acha.2008.07.002

[19] Y. Shen, B. Li, W. Pan, and J. Li, *Analysis of generalized orthogonal matching pursuit using restricted constant*. Electron. Lett. 50(2014), no. 14, 1020–1022.

[20] 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 59(2013), no. 5, 3158–3174. https://doi.org/10.1109/TIT.2013.2238606

[21] J. A. Tropp and A. C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*. IEEE Trans. Inform. Theory 53(2007), no. 12, 4655–4666. https://doi.org/10.1109/TIT.2007.909108

[22] J. Wang, *Support recovery With orthogonal matching pursuit in the presence of noise*. IEEE Trans. Signal Process. 63(2015), no. 21, 5868–5877. https://doi.org/10.1109/TSP.2015.2468676

[23] J. Wang, S. Kwon, P Li, and B. Shim, *Recovery of sparse signals via generalized orthogonal matching pursuit: a new analysis*. IEEE Trans. Signal Process. 64(2015), no. 4, 1076–1089. https://doi.org/10.1109/TSP.2015.2498132

[24] J. Wang, S. Kwon, and B. Shim, *Generalized orthogonal matching pursuit*. IEEE Trans. Singnal Process. 60(2012), no. 12, 6202–6216.

[25] J. Wang and P. Li, *Recovery of sparse signals using multiple orthogonal least squares*. IEEE Trans. Signal Process. 65(2017), no. 8, 2049–2062. https://doi.org/10.1109/TSP.2016.2639467

[26] J. Wang and B. Shim, *On the recovery limit of sparse signals using orthogonal matching pursuit*. IEEE Trans. Signal Process. 60(2012), no. 9, 4973–4976. https://doi.org/10.1109/TSP.2012.2203124

[27] J. Wen, Z. Zhou, D. Li, and X. Tang, *A novel sufficient condition for generalized orthogonal matching pursuit*. IEEE Comm. Lett. 21(2017), no. 4, 805–808.

[28] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, *A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit*. IEEE Trans. Signal Process. 65(2017), 1370–1382. https://doi.org/10.1109/TSP.2016.2634550

[29] J. Wen, J. Wang, and Q. Zhang, *Nearly optimal bounds for orthogonal least squares*. IEEE Trans. Signal Proces. 65(2017), no. 20, 5347–5356. https://doi.org/10.1109/TSP.2017.2728502

Graduate School, China Academy of Engineering Physics, Beijing 100088, China
e-mail: gpb1990@126.com

Institute of Applied Physics and Computational Mathematics, Beijing, 100088, China
e-mail: chenwg@iapcm.ac.cn

Sports Engineering College, Beijing Sport University, Beijing, 100088, China
e-mail: gehuanmin@163.com