Journal of Computational Mathematics
Vol.41, No.1, 2023, 1–17.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2104-m2020-0093

REQUIRED NUMBER OF ITERATIONS FOR SPARSE SIGNAL RECOVERY VIA ORTHOGONAL LEAST SQUARES*

Haifeng Li and Jing Zhang
Henan Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China
Email: lihaifengxx@126.com, a1293651766@126.com
Jinming Wen[1)]
College of Information Science and Technology, Jinan University, Guangzhou 510632, China
Email: jinming.wen@mail.mcgill.ca
Dongfang Li
School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China;
Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan 430074, China
Email: dfli@mail.hust.edu.cn

Abstract

In countless applications, we need to reconstruct a K-sparse signal $\mathbf{x} \in \mathbb{R}^n$ from noisy measurements $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$, where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ is a sensing matrix and $\mathbf{v} \in \mathbb{R}^m$ is a noise vector. Orthogonal least squares (OLS), which selects at each step the column that results in the most significant decrease in the residual power, is one of the most popular sparse recovery algorithms. In this paper, we investigate the number of iterations required for recovering $\mathbf{x}$ with the OLS algorithm. We show that OLS provides a stable reconstruction of all K-sparse signals $\mathbf{x}$ in $\lceil 2.8K \rceil$ iterations provided that $\mathbf{\Phi}$ satisfies the restricted isometry property (RIP). Our result provides a better recovery bound and fewer number of required iterations than those proposed by Foucart in 2013.

Mathematics subject classification: 94A12, 65F22, 65J22.
Key words: Sparse signal recovery, Orthogonal least squares (OLS), Restricted isometry property (RIP).

1. Introduction

Compressed sensing (CS) has been attracted considerable attention in numerous fields [1–5]. The main task of CS is to recover a signal $\mathbf{x} \in \mathbb{R}^n$ from

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}, \tag{1.1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}(m \ll n)$ is a sensing matrix with ℓ_2-normalized columns, $\mathbf{x}$ is a K-sparse (i.e., $\|\mathbf{x}\|_0 \leq K$, where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of $\mathbf{x}$) signal, and $\mathbf{v} \in \mathbb{R}^m$ is a noise vector.

There are many algorithms ([6–12]) for recovering $\mathbf{x}$ from (1.1). One of the popular one is the orthogonal least squares (OLS) [13–16] algorithm. It has been shown in [15] that OLS

* Received April 8, 2020 / Revised version received June 24, 2020 / Accepted April 23, 2021 / Published online March 29, 2022 /
1) Corresponding author

is computationally more expensive yet is more reliable than the orthogonal matching pursuit (OMP) algorithm [17, 18], hence it has been attracted much attention in recent years. OLS identifies the support of $\mathbf{x}$ by adding one index to the list at each iteration, and estimates the coefficients of the sparse vector over the enlarged support. Specifically, it adds to the estimated support an index which leads to the maximum reduction of the residual power in each iteration. The vestige of the active list is then eliminated from $\mathbf{y}$, yields a residual update for the next iteration.

Denote $\Omega = \{1, \cdots, n\}$ and $T = \text{supp}(\mathbf{x}) = \{i | x_i \neq 0, i \in \Omega\}$ as the support of K-sparse signal $\mathbf{x}$. Let Λ be a subset of Ω, $|\Lambda|$ be the cardinality of Λ, and $T \backslash \Lambda = \{i | i \in T, i \notin \Lambda\}$. Let $\mathbf{x}_\Lambda \in \mathbb{R}^n$ be the vector equal to $\mathbf{x}$ on the index set Λ and zero elsewhere. Throughout the paper, we assume that $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ is column normalized (i.e., $\|\mathbf{\Phi}_i\|_2 = 1$ for $i = 1, 2, \cdots, n$)[1)] . Let $\mathbf{\Phi}_\Lambda \in \mathbb{R}^{m \times |\Lambda|}$ be the submatrix of $\mathbf{\Phi}$ with index of its columns in set Λ. For any matrix $\mathbf{\Phi}_\Lambda$ of full column-rank, let $\mathbf{\Phi}_\Lambda^\dagger = (\mathbf{\Phi}'_\Lambda \mathbf{\Phi}_\Lambda)^{-1} \mathbf{\Phi}'_\Lambda$ be the pseudo-inverse of $\mathbf{\Phi}_\Lambda$, where $\mathbf{\Phi}'_\Lambda$ denotes the transpose of $\mathbf{\Phi}_\Lambda$. $\mathbf{P}_\Lambda = \mathbf{\Phi}_\Lambda \mathbf{\Phi}_\Lambda^\dagger$ and $\mathbf{P}_\Lambda^\perp = \mathbf{I} - \mathbf{P}_\Lambda$ denote the orthogonal projection onto $\text{span}(\mathbf{\Phi}_\Lambda)$ (i.e., the column space of $\mathbf{\Phi}_\Lambda$) and its orthogonal complement, respectively. OLS is mathematically described in Algorithm 1.1.

Algorithm 1.1. The OLS algorithm [19]

Input: $\mathbf{\Phi}$, $\mathbf{y}$, maximum iteration number $k_{\max}$.
Initialization: For $\mathbf{r}^0 = \mathbf{y}$, $k = 0$, and $S^0 = \emptyset$.
1: **while** $k < k_{\max}$ **do**
2: $k = k + 1$. 3: Choose the index s^k that satisfies
$s^k = \arg\min_{i \in \Omega} \|\mathbf{P}^\perp_{S^{k-1} \cup \{i\}} \mathbf{y}\|_2^2$.
4: Let $S^k = S^{k-1} \bigcup \{s^k\}$, and calculate
$\mathbf{x}^k = \arg\min_{\text{supp}(\mathbf{u}) = S^k} \|\mathbf{y} - \mathbf{\Phi u}\|_2$.
5: $\mathbf{r}^k = \mathbf{y} - \mathbf{\Phi x}^k = \mathbf{P}^\perp_{S^k} \mathbf{y}$.
6: **end while**
Output: $\mathbf{x}^k$ and S^k.

The performance analysis of OLS has been extensively studied. For example, Soussen *et al.* showed that OLS is guaranteed to exactly recover the support of $\mathbf{x}$ in at most K iterations when the exact recovery condition (ERC) is met [14]. Based on mutual coherence, Herzet *et al.* addressed the exact recovery of $\mathbf{x}$ in the noiseless setting when some partial information of its support is available [15]. Herzet *et al.* developed extended coherence-based sufficient conditions for exact sparse support recovery with OLS [16]. Wen *et al.* [19] and Geng *et al.* [20] utilized the restricted isometry property (RIP), which is defined as follows, to study the sufficient condition of exact recovery of $\mathbf{x}$ with OLS. Using the RIP, the authors in [21–24] discussed the performance of multiple OLS which is an extension of OLS.

Definition 1.1 ([25]). *A measurement matrix* $\mathbf{\Phi}$ *is said to satisfy the RIP of order* K *if there exists a constant* $\delta \in [0, 1)$ *such that,*

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \tag{1.2}$$

1) The behavior of OLS is unchanged whether columns of $\mathbf{\Phi}$ are normalized or not ([31]).

holds for all K-sparse vector x. The minimum δ satisfying (1.2) *is defined as the restricted isometry constant* (*RIC*) δ_K.

It has been shown in [21] that OLS recovers any K-sparse signal in exact K iterations provided that

$$\delta_{K+1} < \frac{1}{\sqrt{K}+2}. \tag{1.3}$$

The sufficient condition (1.3) has recently been improved to [19]

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}}. \tag{1.4}$$

One can interpret from (1.3) and (1.4) that exact recovery with OLS can be ensured when δ_{K+1} is inversely proportional to $\sqrt{K}$. Thus, these upper bounds will vanish when the sparsity K is large.

While the above works focused on the scenario where the number of iterations is limited to K [14], there are some works which investigate the behavior of OMP and OLS with more than K iterations [11, 27–30]. For example, it has been shown that if OLS runs $6K$ iterations [29], the stable reconstruction is guaranteed under

$$\delta_{10K} \leq \frac{1}{6}. \tag{1.5}$$

Note that running fewer iterations offers computational benefits, thus we aim to improve (1.5) in this paper. Specifically, our result is shown in Theorem 2.1 that OLS stably recovers any K-sparse vectors $\mathbf{x}$ from (1.1) in $\lceil 2.8K \rceil$ iterations.

The rest of this paper is organized as follows. Section 2 presents our main results. In Section 3, we provide some technical lemmas that are useful for our analysis and prove Theorem 2.1. Finally, this paper is summarized in Section 4.

2. Sparse Recovery with OLS

Before presenting our main results, we obtain an important observations on the OLS algorithm. As shown in Algorithm 1.1, in the $(k+1)$-th iteration ($k \geq 0$), OLS adds an index s^{k+1} to S^k that results in the maximum reduction of the residual power, i.e.,

$$s^{k+1} = \arg\min_{i \in \Omega} \|\mathbf{P}^{\perp}_{S^k \cup \{i\}} \mathbf{y}\|_2^2. \tag{2.1}$$

From (2.1), we can observe that to find s^{k+1}, we need to construct $n-k$ different orthogonal projections (i.e., $\mathbf{P}^{\perp}_{S^k \cup \{i\}}$). But this implementation is computationally expensive. In order to solve this problem, inspired by [31], Wang and Li [21] presented a cost-effective expression alternative to (2.1) for the identification step of Algorithm 1.1. The result is listed in the following Lemma 2.1.

Lemma 2.1 ([19, 21]). *At the* $(k+1)$*-th iteration, the OLS algorithm selects the index*

$$s^{k+1} = \arg\max_{i \in \Omega} \frac{|\langle \mathbf{\Phi}_i, \mathbf{r}^k \rangle|}{\|\mathbf{P}^{\perp}_{S^k} \mathbf{\Phi}_i\|_2}. \tag{2.2}$$

We can see from (2.2) that to find s^{k+1}, we only need to compute one projection operator (i.e., $\mathbf{P}_{S^k}^{\perp}$), hence it is much cheaper to find s^{k+1} based on (2.2) than that based on (2.1). Numerical experiments indicate that the simplification indeed offers massive reduction in the computational cost. Hence, Lemma 2.1 plays an important role in analyzing the number of iterations of OLS.

Moreover, note that the identification rule of OLS is akin to the OMP rule. Specifically, in the $(k+1)$-th iteration, OMP picks an index corresponding to the column which is most strongly correlated with the signal residual, i.e.,

$$s^{k+1} = \arg\max_{i\in\Omega} |\langle \mathbf{\Phi}_i, \mathbf{r}^k\rangle|.$$

Clearly, the rule of OLS differs from that of OMP only in that it has an extra normalization factor (i.e., $\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_i\|_2$). Thus, the greedy selection rule in OLS can also be viewed as an extension of the OMP rule. This arguments has been verified (see [14, 21, 31]). However, this property shows that the rule of OLS coincides with that of OMP only in the first iteration (because $S^0 = \emptyset$ leads to $\|\mathbf{P}_{S^0}^{\perp}\mathbf{\Phi}_i\|_2 = \|\mathbf{\Phi}_i\|_2$). For the subsequent iterations, it does make a difference since $\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_i\|_2 \le \|\mathbf{\Phi}_i\|_2, \forall\, k \ge 1$. In fact, as will be seen later, this factor makes the analysis of OLS different and more challenging than that of OMP.

Let

$$\lambda > -\frac{4}{(1-\delta^2_{k+K+\lfloor\lambda\theta^k\rfloor})(1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor})}\log\left(\frac{1}{2} - \sqrt{\frac{\delta_{k+K+\lfloor\lambda\theta^k\rfloor}}{2(1+\delta_{k+K+\lfloor\lambda\theta^k\rfloor})}}\right), \tag{2.3}$$

$$\eta = \exp\Big(-\frac{\lambda(1-\delta^2_{k+K+\lfloor\lambda\theta^k\rfloor})}{4}\big(1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}\big)\Big), \tag{2.4}$$

$$\xi_k = 2\left(1 - 2\Big(\frac{(1+\delta_{k+K+\lfloor\lambda\theta^k\rfloor})(1-\eta)\eta}{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}}\Big)^{\frac{1}{2}}\right)^{-1} - 1. \tag{2.5}$$

In the following, we introduce our main results.

Theorem 2.1. *Let $\theta^k = |T\backslash S^k|$ be the number of remaining support set after running k $(k \ge 0)$ iterations of OLS, if $\mathbf{\Phi}$ obeys the RIP of the order $k + K + \lfloor\lambda\theta^k\rfloor$, then the residual of OLS satisfies*

$$\|\mathbf{r}^{k+\lceil\lambda\theta^k\rceil}\|_2 \le \xi_k\|\mathbf{v}\|_2, \tag{2.6}$$

where λ and $\xi_k \ge 1$ are constants which are defined in (2.3) *and* (2.5)*, respectively. They depend only on $\delta_{k+K+\lfloor\lambda\theta^k\rfloor}$.*

Proof. See Section III. □

From Theorem 2.1, we observe that after running $k(k \ge 0)$ iterations, OLS requires at most $\lceil\lambda\theta^k\rceil$ additional iterations to ensure that the condition (2.6) is fulfilled, and the ℓ_2-norm of residual is upper bounded by the product of a constant and $\|\mathbf{v}\|_2$.

In particular, when $k = 0$, $\theta^0 = |T\backslash S^0| = K$, hence we can obtain the following corollary which shows that the ℓ_2-norm of residual falls below $\xi_0\|\mathbf{v}\|_2$:

Corollary 2.1. *Let $\mathbf{\Phi}$ satisfies the RIP of order $\lfloor(\lambda+1)K\rfloor$, then the residual of OLS satisfies*

$$\|\mathbf{r}^{\lceil\lambda K\rceil}\|_2 \le \xi_0\|\mathbf{v}\|_2, \tag{2.7}$$

where ξ_0 is defined as (2.5) *when $k = 0$.*

In the following, we give some remarks.

Remark 2.1. It has been shown that $\|\mathbf{r}^{6K}\|_2$ is upper bounded by the product of a constant and $\|\mathbf{v}\|_2$ provided that $\delta_{10K} \le \frac{1}{6}$ [29]. By the relation λ and $\delta_{\lfloor(\lambda+1)K\rfloor}$ in (2.3), and setting $k = 0$, we get

$$\lambda > -\frac{4}{(1-\delta^2_{\lfloor(\lambda+1)K\rfloor})(1-\delta_{\lfloor(\lambda+1)K\rfloor})}\log(\frac{1}{2}-\sqrt{\frac{\delta_{\lfloor(\lambda+1)K\rfloor}}{2(1+\delta_{\lfloor(\lambda+1)K\rfloor})}}). \tag{2.8}$$

Let $\lambda = 3$ and solve the above inequality, we obtain

$$\delta_{4K} \le 0.001. \tag{2.9}$$

Corollary 2.1 shows that $\|\mathbf{r}^{3K}\|_2$ is upper bounded by the product of a constant and $\|\mathbf{v}\|_2$ when (2.9) holds. Hence, compared to the results in [29], our result shows that OLS needs fewer number of iterations for stable sparse signal recovery.

Remark 2.2. The authors in [19] showed that OLS can exactly recover all K-sparse $\mathbf{x}$ from the samples $\mathbf{y} = \mathbf{\Phi x}$ in K iterations if $\mathbf{\Phi}$ satisfies (1.4). It is easy to see that (1.4) is inversely proportional to K. The upper bound will vanish when K is large. Whereas, the upper bound in (2.9) is an absolute constant and is independent of the sparsity K.

Let $\mathbf{\Phi}$ be a random measurement matrix whose entries independent and identically follow the Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$. Then by the proof of [26, Theorem 5.2], $\mathbf{\Phi}$ satisfies the RIP with $\delta_K \le \epsilon$ with overwhelming probability if $m = \mathcal{O}(\frac{K \log \frac{n}{K}}{\epsilon^2})$. Hence, to satisfy (1.4), the number of required measurements is $m = \mathcal{O}(K^2 \log \frac{n}{K})$. Whereas, the proposed condition (2.9) requires $m = \mathcal{O}(K \log \frac{n}{K})$ which is significantly smaller than the previous result.

Remark 2.3. In [11, (12)], the authors proved that, in the noiseless case, OMP can accurately recover all K-sparse signal within $\lceil cK \rceil$ iterations where

$$c \ge -\frac{4(1+\delta_{\lfloor(c+1)K\rfloor})}{1-\delta_{\lfloor(c+1)K\rfloor}}\log\left(\frac{1}{2}-\sqrt{\frac{\delta_{\lfloor(c+1)K\rfloor}}{2(1+\delta_{\lfloor(c+1)K\rfloor})}}\right). \tag{2.10}$$

By setting $\delta_{\lfloor(c+1)K\rfloor} = 0$ in (2.10), the authors in [11] claimed that OMP can uniformly recover all K-sparse signals using at least $2.8K$ iterations.

From Corollary 2.1, set $\delta_{\lfloor(\lambda+1)K\rfloor} = 0$, then we have

$$\lambda > 4\log 2 \approx 2.77,$$

or,

$$\lambda \ge 2.8.$$

So, our result is the same as [11] when $\delta_{\lfloor(\lambda+1)K\rfloor} = 0$. However, our bound is smaller than Wang's result when $\delta_{\lfloor(\lambda+1)K\rfloor} > 0$. Fig. 2.1 shows that the comparison between our bound and Wang's result [11].

Remark 2.4. In [29, Remark], the authors proved that

$$\|\mathbf{r}^{6K}\|_2 \le C\|\mathbf{v}\|_2,$$

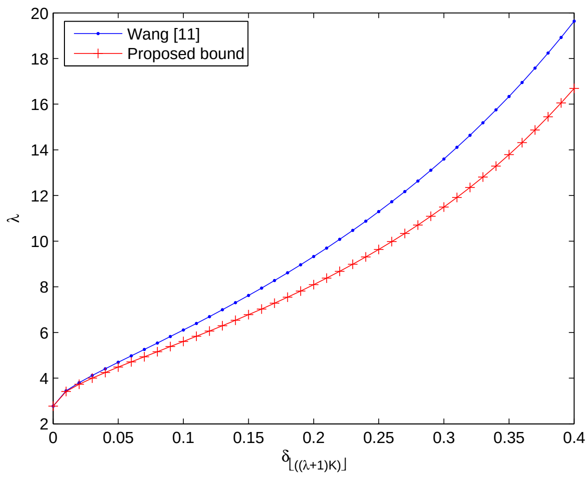

Fig. 2.1. Comparison between the proposed bound and Wang [11].

where

$$C = \frac{\sqrt{\frac{8(1+\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K})}{(1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K})exp(\lceil\frac{3}{\rho^2}\rceil\rho^2(1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K}))}} + \sqrt{\frac{2}{1-\frac{1}{exp(\lceil\frac{3}{\rho^2}\rceil\rho^2(1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K}))}}}}{1-\sqrt{\frac{8(1+\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K})}{(1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K})exp(\lceil\frac{3}{\rho^2}\rceil\rho^2(1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K}))}}},$$

and

$$\rho^2 = 1 - \frac{\delta^2_{(1+3\lceil\frac{3}{\rho^2}\rceil)K}}{1-\delta_{(1+3\lceil\frac{3}{\rho^2}\rceil)K}}.$$

By (2.5), it is easy to see that $\xi_k > C$ when $\lambda = 3$. Note that the residual power of OLS is non-increasing, this result is not surprising because it performs only $3K$ iterations while it performs $6K$ iterations in [29].

The following theorem shows that the ℓ_2-norm of the recovery error can also be upper bounded by the product of a constant and $\|\mathbf{v}\|_2$.

Theorem 2.2. *If* $\mathbf{\Phi}$ *obeys the RIP of the order* $\lceil(\lambda+1)K\rceil$*, then*

$$\|\mathbf{x}^{\lceil\lambda K\rceil} - \mathbf{x}\|_2 \le (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}(\xi_0+1)\|\mathbf{v}\|_2,$$

where λ *and* ξ_0 *are defined in* (2.3) *and* (2.5)*, respectively.*

Proof. See Appendix A. □

Remark 2.5. In [29], the authors proved that the norm of the recovery error is upper bounded by the product of a constant and $\|\mathbf{v}\|_2$ after at most $12K$ iterations, provided that $\delta_{20K} \le \frac{1}{6}$.

By Theorem 2.2, using the relation λ and $\delta_{\lfloor(\lambda+1)K\rfloor}$, our result shows that the norm of the recovery error is upper bounded by the product of a constant and $\|\mathbf{v}\|_2$ after at most $3K$ iterations (set $\lambda = 3$), provided that $\delta_{4K} \le 0.001$.

3. Proof of Theorem 2.1

Proof. The main idea of our proof is inspired by [30]. Here, we first denote $F^k = T \backslash S^k$ and $\theta^k = |F^k|$. For notational convenience, assume that x_i is arranged in descending order of their magnitudes, i.e., $|x_1| \geq |x_2| \geq \cdots \geq |x_{\theta^k}|$. Now, we define the subset F_j^k of F^k as

$$F_j^k = \begin{cases} \emptyset, & j = 0, \\ \{1, 2, \cdots, 2^j - 1\}, & j = 1, \cdots, \lfloor \log_2 \theta^k \rfloor, \\ F^k, & j = \lfloor \log_2 \theta^k \rfloor + 1. \end{cases} \tag{3.1}$$

For constant $\tau > 1$, let $L \in \{1, 2, \cdots, \lfloor \log_2 \theta^k \rfloor + 1\}$ be the minimum positive integer satisfying

$$\|\mathbf{x}_{F^k \backslash F_0^k}\|_2^2 < \tau \|\mathbf{x}_{F^k \backslash F_1^k}\|_2^2, \tag{3.2}$$

$$\|\mathbf{x}_{F^k \backslash F_1^k}\|_2^2 < \tau \|\mathbf{x}_{F^k \backslash F_2^k}\|_2^2, \tag{3.3}$$

$$\cdots,$$

$$\|\mathbf{x}_{F^k \backslash F_{L-2}^k}\|_2^2 < \tau \|\mathbf{x}_{F^k \backslash F_{L-1}^k}\|_2^2, \tag{3.4}$$

$$\|\mathbf{x}_{F^k \backslash F_{L-1}^k}\|_2^2 \geq \tau \|\mathbf{x}_{F^k \backslash F_L^k}\|_2^2. \tag{3.5}$$

Then we have

$$\|\mathbf{x}_{F^k \backslash F_j^k}\|_2^2 < \tau^{L-1-j} \|\mathbf{x}_{F^k \backslash F_{L-1}^k}\|_2^2, \quad \text{for } j = 0, 1, \cdots, L-2. \tag{3.6}$$

Note that the last set $F^k_{\lfloor \log_2 \theta^k \rfloor + 1} (= F^k)$ may have no more than $2^{\lfloor \log_2 \theta^k \rfloor + 1} - 1$ elements. It is also noted that if (3.5) holds true for all $L \geq 1$, then we ignore (3.2)–(3.4) and simply take $L = 1$. Besides, note that L always exists because

$$\|\mathbf{x}_{F^k \backslash F^k_{\lfloor \log_2 \theta^k \rfloor + 1}}\|_2^2 = 0$$

so that (3.6) holds true at least for $L = \lfloor \log_2 \theta^k \rfloor + 1$.

In consideration of the selection rule of OLS viewed as an extension of the OMP rule, we will prove Theorem 2.1 by using mathematical induction in θ^k. In fact, the mathematical induction has been proposed in [30]. Being here, θ^k stands for the number of remaining indices after k iterations of OLS. We first select $\theta^k = 0$, and then no more iteration is needed, i.e., $T \subseteq S^k$, then we have

$$\|\mathbf{r}^k\|_2 = \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k\|_2 = \min_{\text{supp}(\mathbf{u}) = S^k} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2 \leq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2 = \|\mathbf{v}\|_2 \leq \xi_k \|\mathbf{v}\|_2.$$

Now we suppose that the conclusion holds up to $\theta^k - 1$, where $\theta^k \geq 1$ is a positive integer. Then, we need to prove that (2.6) holds true, that is,

$$\|\mathbf{r}^{k + \lceil \lambda \theta^k \rceil}\|_2 \leq \xi_k \|\mathbf{v}\|_2. \tag{3.7}$$

In order to prove (3.7), we will choose a decent amount of support indices in F^k, which must be selected within a specified number of additional iterations. Then the number of remaining support indices is upper bounded.

Now we define that

$$k_i = \sum_{j=1}^{i} \lceil \frac{\lambda |F_j^k|}{4} \rceil, \qquad i = 1, \cdots, L. \tag{3.8}$$

According to the definition of F_j^k, we have $|F_j^k| \le 2^j - 1$ for $j = 1, \cdots, L$, and then

$$k_i \le k_L = \sum_{j=1}^{L} \lceil \frac{\lambda |F_j^k|}{4} \rceil \le \sum_{j=1}^{L} \lceil \frac{\lambda (2^j - 1)}{4} \rceil \overset{(a)}{\le} \lceil \lambda 2^{L-1} \rceil - 1, \tag{3.9}$$

where (a) is according to [11, (A.18)]. Let

$$k' = \lceil \lambda 2^{L-1} \rceil - 1 \tag{3.10}$$

be a specified additional iterations after running k iterations of OLS.

Now if we suppose that the number of remaining support indices satisfies

$$\theta^{k+k'} = |F^{k+k'}| \le \theta^k - 2^{L-1} \tag{3.11}$$

after running $k + k'$ iterations. Then the inequality (3.7) holds when we requires at most $\lceil \lambda \theta^{k+k'} \rceil$ additional iterations. Thus, our proof is completed. In the following, we will explain it. In fact, being here, the total number of iterations of OLS is

$$k + k' + \lceil \lambda \theta^{k+k'} \rceil \le k + \lceil \lambda 2^{L-1} \rceil - 1 + \lceil \lambda (\theta^k - 2^{L-1}) \rceil \overset{(a)}{\le} k + \lceil \lambda \theta^k \rceil, \tag{3.12}$$

where (a) follows from $\lceil a \rceil + \lceil b \rceil - 1 \le \lceil a + b \rceil$.

Since the residual power of OLS is non-increasing (i.e., $\|\mathbf{r}^i\|_2 \le \|\mathbf{r}^j\|_2$ for $i \ge j$, see Lemma C.1), we obtain

$$\|\mathbf{r}^{k+\lceil \lambda \theta^k \rceil}\|_2 \le \|\mathbf{r}^{k+k'+\lceil \lambda \theta^{k+k'} \rceil}\|_2 \le \|\mathbf{r}^{k+k'}\|_2. \tag{3.13}$$

It follows from (3.11) that the index number of remaining support is no more than $\theta^k - 1$, i.e.,

$$\theta^{k+k'} = |F^{k+k'}| \le \theta^k - 2^{L-1} \le \theta^k - 1.$$

From the induction hypothesis, we have

$$\|\mathbf{r}^{k+k'+\lceil \lambda \theta^{k+k'} \rceil}\|_2 \le \xi_k \|\mathbf{v}\|_2. \tag{3.14}$$

Then we combine (3.13) with (3.14) and have

$$\|\mathbf{r}^{k+\lceil \lambda \theta^k \rceil}\|_2 \le \|\mathbf{r}^{k+k'+\lceil \lambda \theta^{k+k'} \rceil}\|_2 \le \xi_k \|\mathbf{v}\|_2.$$

In summary, it remains to prove that (3.11) holds true. By the definition of F_j^k in (3.1), we have

$$F_{L-1}^k = \{1, 2, \cdots, 2^{L-1} - 1\}$$

and

$$|F^k \setminus F_{L-1}^k| = |\{2^{L-1}, 2^{L-1} + 1, \cdots, \theta^k\}| = \theta^k - 2^{L-1} + 1.$$

Then (3.11) can be rewritten as

$$\theta^{k+k'} = |F^{k+k'}| < |F^k \backslash F^k_{L-1}|. \tag{3.15}$$

Since $\mathbf{x}_{F^k \backslash F^k_{L-1}}$ consists of $|F^k \backslash F^k_{L-1}|$ smallest non-zero elements (in magnitude) of $\mathbf{x}_{F^k}$, instead of proving it directly, we show that a sufficient condition of (3.15) is true. That is,

$$\|\mathbf{x}_{F^{k+k'}}\|_2 < \|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2. \tag{3.16}$$

Hence, we need to prove that the inequality (3.16) holds true.

By the result in Proposition D.1 (see Appendix D), i.e.,

$$\|\mathbf{x}_{F^{k+k'}}\|_2 < \alpha \|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2 + \beta \|\mathbf{v}\|_2, \tag{3.17}$$

where α and β are defined in (D.1) and (D.2), respectively.

It follows from (2.3) that $\alpha < 1$. Then we discuss two cases in the following.

If $\beta\|\mathbf{v}\|_2 < (1-\alpha)\|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2$, it is easy to see that (3.16) holds true.

If $\beta\|\mathbf{v}\|_2 \geq (1-\alpha)\|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2$, it follows from (3.13) that

$$\begin{aligned}
\|\mathbf{r}^{k+\lceil\lambda\theta^k\rceil}\|_2 &\leq \|\mathbf{r}^{k+k'+\lceil\lambda\theta^{k+k'}\rceil}\|_2 \leq \|\mathbf{r}^{k+k'}\|_2 \\
&\overset{(a)}{<} \sqrt{4\eta(1-\eta)(1+\delta_{k+K+\lfloor\lambda\theta^k\rfloor})}\|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2 + \|\mathbf{v}\|_2 \\
&\overset{(b)}{=} \alpha\sqrt{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}}\|\mathbf{x}_{F^k \backslash F^k_{L-1}}\|_2 + \|\mathbf{v}\|_2 \\
&\leq \alpha\sqrt{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}} \times \frac{\beta}{1-\alpha}\|\mathbf{v}\|_2 + \|\mathbf{v}\|_2 \\
&\overset{(c)}{=} (\frac{2}{1-\alpha} - 1)\|\mathbf{v}\|_2 = \xi_k\|\mathbf{v}\|_2,
\end{aligned}$$

where (a) follows from (D.12), (b) is from (D.1), (c) follows from (D.2). Here, ξ_k has been defined in (2.5). Then we can prove that (3.7) directly holds true. □

4. Conclusion

In this paper, we evaluate the performance of OLS when the number of iterations exceeds the sparsity K of the signal $\mathbf{x}$. Compared to the state-of-the-art results, our results reduce the required number of iterations for stable sparse signal recovery. This advantage provides computational benefits as well as relaxations in the measurement size and the sparsity range of the sparse signals need to be recovered.

Appendix A

Proof of Theorem 2.2.

Proof. Since $\mathbf{r}^{\lceil\lambda K\rceil} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{\lceil\lambda K\rceil} = \mathbf{\Phi}(\mathbf{x} - \mathbf{x}^{\lceil\lambda K\rceil}) + \mathbf{v}$, we have

$$\begin{aligned}
\|\mathbf{x}^{\lceil\lambda K\rceil} - \mathbf{x}\|_2 = \|\mathbf{x} - \mathbf{x}^{\lceil\lambda K\rceil}\|_2 &\overset{(a)}{\leq} (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}^{\lceil\lambda K\rceil})\|_2 \\
= (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}\|\mathbf{r}^{\lceil\lambda K\rceil} - \mathbf{v}\|_2 &\overset{(b)}{\leq} (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}(\|\mathbf{r}^{\lceil\lambda K\rceil}\|_2 + \|\mathbf{v}\|_2)
\end{aligned}$$

$$\overset{(c)}{\leq} (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}(\xi_0\|\mathbf{v}\|_2+\|\mathbf{v}\|_2) = (1-\delta_{\lceil(\lambda+1)K\rceil})^{-\frac{1}{2}}(\xi_0+1)\|\mathbf{v}\|_2,$$

where (a) is based on the RIP and $|T\cup S^{\lceil\lambda K\rceil}|\leq K+\lceil\lambda K\rceil=\lceil(\lambda+1)K\rceil$, (b) uses the norm inequality, (c) is from Corollary 2.1. □

Appendix B

Lemma B.1 ([19]). *Suppose that* $\Lambda\subseteq\Omega$ *and* $\mathbf{\Phi}\in\mathbb{R}^{m\times n}$ *satisfies the RIP of order* $|\Lambda|+1$. *Then, for any* $i\in\Omega\backslash\Lambda$,

$$\|\mathbf{P}_{\Lambda}^{\perp}\mathbf{\Phi}_i\|_2\geq\sqrt{1-\delta_{|\Lambda|+1}^2}.$$

Lemma B.2. *According to* (2.2) *and the third step in Algorithm* 1.1*, we have*

$$|\langle\mathbf{\Phi}_{s^{k+1}},\mathbf{r}^k\rangle|\geq\varepsilon\max_{1\leq s\leq n}|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|. \tag{B.1}$$

where $\varepsilon=\sqrt{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}^2}$.

Proof. The proof of Lemma B.2 is similar to that of [29, Theorem 2]. But Lemma B.2 improves [29, Theorem 2] which shows that the parameter is

$$\varepsilon=\sqrt{1-\frac{\delta_{k+K+\lfloor\lambda\theta^k\rfloor}^2}{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}}}.$$

According to the proof of [29, Theorem 2], for any $s\notin S^k$, we have

$$\frac{|\langle\mathbf{\Phi}_{s^{k+1}},\mathbf{r}^k\rangle|^2}{\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_{s^{k+1}}\|_2^2}\overset{(a)}{=}\max_{i\notin S^k}\frac{|\langle\mathbf{\Phi}_i,\mathbf{r}^k\rangle|^2}{\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_i\|_2^2}\geq\frac{|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2}{\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_s\|_2^2}\overset{(b)}{\geq}|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2,$$

where (a) follows from Lemma 2.1, (b) is based on $\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_s\|_2\leq 1$. Thus we get

$$\begin{aligned}&|\langle\mathbf{\Phi}_{s^{k+1}},\mathbf{r}^k\rangle|^2\geq\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_{s^{k+1}}\|_2^2|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2\\ \overset{(a)}{\geq}\ &(1-\delta_{|S^k|+1}^2)|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2\overset{(b)}{\geq}(1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}^2)|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2,\end{aligned} \tag{B.2}$$

where (a) follows from Lemma B.1, (b) follows from $|S^k|\leq k$ and the monotonicity of the RIP.

For $s\in S^k$, since $|\langle\mathbf{\Phi}_s,\mathbf{r}^k\rangle|^2=0$, (B.2) also holds. Hence (B.1) is established. □

Appendix C

Lemma C.1 ([29]). Note that the third step in Algorithm 1.1. The residual decreases at each iteration.

$$\|\mathbf{r}^{k+1}\|_2^2=\|\mathbf{r}^k\|_2^2-\frac{|\langle\mathbf{\Phi}_{s^{k+1}},\mathbf{r}^k\rangle|^2}{\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}_{s^{k+1}}\|_2^2}.$$

Lemma C.2. Let $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ be a matrix with ℓ_2-normalized columns. For the $(\ell+1)$-th ($\ell \geq k$) iteration of OLS, let ν be the vector such that,

$$\nu = \begin{cases} \mathbf{x}_U, & U = T \cap S^k \cup F_j^k, \\ 0, & U = \Omega \backslash (T \cap S^k \cup F_j^k), \end{cases} \tag{C.1}$$

where $j \in \{1, \cdots, \lfloor \log_2 \theta^k \rfloor + 1\}$. When $\varepsilon = \sqrt{1 - \delta^2_{k+K+\lfloor \lambda\theta^k \rfloor}}$, we have

$$\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^{\ell+1}\|_2^2 \geq \frac{\varepsilon^2}{|F_j^k|}(1 - \delta_{|F_j^k \cup S^\ell|}) \times (\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k \backslash F_j^k} + \mathbf{v}\|_2^2). \tag{C.2}$$

Proof. From Lemma C.1, we have

$$\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^{\ell+1}\|_2^2 = \frac{|\langle \mathbf{\Phi}_{s^{\ell+1}}, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}^\perp_{S^\ell}\mathbf{\Phi}_{s^{\ell+1}}\|_2^2} \overset{(a)}{\geq} |\langle \mathbf{\Phi}_{s^{\ell+1}}, \mathbf{r}^\ell\rangle|^2 \overset{(b)}{\geq} \varepsilon^2 \max_{1\leq s\leq n} |\langle \mathbf{\Phi}_s, \mathbf{r}^\ell\rangle|^2 = \varepsilon^2 \|\mathbf{\Phi}'\mathbf{r}^\ell\|_\infty^2,$$

where (a) follows from $\|\mathbf{P}^\perp_{S^\ell}\mathbf{\Phi}_{s^{\ell+1}}\|_2 \leq 1$, (b) is from Lemma B.2. Now, we only show that

$$\|\mathbf{\Phi}'\mathbf{r}^\ell\|_\infty^2 \geq \frac{1 - \delta_{|F_j^k \cup S^\ell|}}{|F_j^k|}(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2). \tag{C.3}$$

Note that $\mathrm{supp}(\mathbf{\Phi}'\mathbf{r}^\ell) = \Omega\backslash S^\ell$. Then we have

$$\begin{aligned} \|\mathbf{\Phi}'\mathbf{r}^\ell\|_\infty &= \|(\mathbf{\Phi}'\mathbf{r}^\ell)_{\Omega\backslash S^\ell}\|_\infty \overset{(a)}{\geq} \frac{\langle (\mathbf{\Phi}'\mathbf{r}^\ell)_{\Omega\backslash S^\ell}, \nu_{\Omega\backslash S^\ell}\rangle}{\|\nu_{\Omega\backslash S^\ell}\|_1} = \frac{\langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu\rangle}{\|\nu_{\Omega\backslash S^\ell}\|_1} \overset{(b)}{\geq} \frac{\langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu\rangle}{\|\nu_{F_j^k}\|_1} \\ &\overset{(c)}{\geq} \frac{\langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu\rangle}{\sqrt{|F_j^k|}\|\nu_{\Omega\backslash S^\ell}\|_2} \overset{(d)}{=} \frac{\langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu - \mathbf{x}^\ell\rangle}{\sqrt{|F_j^k|}\|\nu_{\Omega\backslash S^\ell}\|_2}, \end{aligned} \tag{C.4}$$

where (a) is from Hölder's inequality, (b) is true since (C.1) and $\|\nu_{\Omega\backslash S^\ell}\|_1 = \|\nu_{T\backslash S^\ell}\|_1 \leq \|\nu_{F_j^k}\|_1$, (c) follows from the norm inequality ($\|\omega\|_1 \leq \sqrt{\|\omega\|_0}\|\omega\|_2$), and (d) is true since $\mathrm{supp}(\mathbf{\Phi}'\mathbf{r}^\ell) \cap \mathrm{supp}(\mathbf{x}^\ell) = \emptyset$.

We observe further that

$$\begin{aligned} \langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu - \mathbf{x}^\ell\rangle &= \langle \mathbf{r}^\ell, \mathbf{\Phi}(\nu - \mathbf{x}^\ell)\rangle = \frac{1}{2}(\|\mathbf{\Phi}(\nu - \mathbf{x}^\ell)\|_2^2 + \|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^\ell - \mathbf{\Phi}(\nu - \mathbf{x}^\ell)\|_2^2) \\ &\overset{(a)}{=} \frac{1}{2}(\|\mathbf{\Phi}(\nu - \mathbf{x}^\ell)\|_2^2 + \|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2) \\ &\overset{(b)}{\geq} \|\mathbf{\Phi}(\nu - \mathbf{x}^\ell)\|_2 \sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2} \\ &\overset{(c)}{\geq} \sqrt{1 - \delta_{|F_j^k \cup S^\ell|}}\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2}\|\nu - \mathbf{x}^\ell\|_2 \\ &\geq \sqrt{(1 - \delta_{|F_j^k \cup S^\ell|})(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2)}\|(\nu - \mathbf{x}^\ell)_{\Omega\backslash S^\ell}\|_2 \\ &\overset{(d)}{=} \sqrt{(1 - \delta_{|F_j^k \cup S^\ell|})(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{v}\|_2^2)}\|\nu_{\Omega\backslash S^\ell}\|_2, \end{aligned} \tag{C.5}$$

where (a) is according to

$$\mathbf{r}^\ell = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^\ell = \mathbf{\Phi}(\mathbf{x} - \mathbf{x}^\ell) + \mathbf{v}$$

$$= \mathbf{\Phi}(\mathbf{x} - \nu + \nu - \mathbf{x}^\ell) + \mathbf{v} = \mathbf{\Phi}(\nu - \mathbf{x}^\ell + \mathbf{x}_{F^k \setminus F_j^k}) + \mathbf{v} = \mathbf{\Phi}(\nu - \mathbf{x}^\ell) + \mathbf{\Phi}\mathbf{x}_{F^k \setminus F_j^k} + \mathbf{v},$$

(b) is true since we only consider $\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 \geq 0$ and use $a^2 + b^2 \geq 2ab$ (when $\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 < 0$, (C.3) holds trivially, i.e., $\|\mathbf{\Phi}'\mathbf{r}^\ell\|_\infty^2 \geq 0$), (c) uses the condition of RIP and $\mathrm{supp}(\nu - \mathbf{x}^\ell) = (T \cap S^k \cup F_j^k) \cup S^\ell \subseteq F_j^k \cup S^\ell$, (d) is due to $(\mathbf{x}^\ell)_{\Omega\setminus S^\ell} = \mathbf{0}$.

Finally, plugging (C.5) into (C.4) and have

$$\begin{aligned}\|\mathbf{\Phi}'\mathbf{r}^\ell\|_\infty &\geq \frac{\langle \mathbf{\Phi}'\mathbf{r}^\ell, \nu - \mathbf{x}^\ell\rangle}{\sqrt{|F_j^k|}\|\nu_{\Omega\setminus S^\ell}\|_2} \geq \frac{\sqrt{1-\delta_{|F_j^k\cup S^\ell|}}\|\nu_{\Omega\setminus S^\ell}\|_2}{\sqrt{|F_j^k|}\|\nu_{\Omega\setminus S^\ell}\|_2}\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2} \\ &= \sqrt{\frac{1-\delta_{|F_j^k\cup S^\ell|}}{|F_j^k|}}\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2}.\end{aligned}$$

This completes the proof. □

Appendix D

Proposition D.1 Let $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ be a matrix with ℓ_2-normalized columns. Let $\theta^k = |F^k| = |T\setminus S^k|$ be the index number of remaining support set after running k ($k \geq 0$) iterations of OLS. Let $\mathbf{x}_{F^{k+k'}}$ and $\mathbf{x}_{F^k\setminus F_{L-1}^k}$ be two truncated vectors of $\mathbf{x}$, where k' is defined in (3.10) and $L \in \{1, 2, \cdots, \lfloor \log_2 \theta^k \rfloor + 1\}$. Then we have

$$\|\mathbf{x}_{F^{k+k'}}\|_2 < \alpha\|\mathbf{x}_{F^k\setminus F_{L-1}^k}\|_2 + \beta\|\mathbf{v}\|_2,$$

where

$$\alpha = \sqrt{\frac{4\eta(1-\eta)(1+\delta_{k+K+\lfloor\lambda\theta^k\rfloor})}{(1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor})}}, \tag{D.1}$$

$$\beta = \frac{2}{\sqrt{1-\delta_{k+K+\lfloor\lambda\theta^k\rfloor}}}\ , \tag{D.2}$$

and η is defined in (2.4).

Proof. According to Lemma C.2 (see Appendix C), let

$$\beta_\ell = \frac{\varepsilon^2}{|F_j^k|}(1 - \delta_{|F_j^k\cup S^\ell|}). \tag{D.3}$$

(C.2) can be rewritten as

$$\|\mathbf{r}^{\ell+1}\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 \leq (1-\beta_\ell)(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2).$$

Using $1 - \beta_\ell \leq e^{-\beta_\ell}$, we have

$$\|\mathbf{r}^{\ell+1}\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 \leq \exp(-\beta_\ell)(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2). \tag{D.4}$$

For $\ell' > \ell \geq k$, we also have

$$\|\mathbf{r}^{\ell'}\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 \leq \exp(-\beta_{\ell'-1})(\|\mathbf{r}^{\ell'-1}\|_2^2 - \|\mathbf{\Phi}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2), \tag{D.5}$$

$$\cdots,$$
$$\|\mathbf{r}^{\ell+2}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 \le \exp(-\beta_{\ell+1})(\|\mathbf{r}^{\ell+1}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2). \quad \text{(D.6)}$$

Thus, from (D.4)-(D.6), we have further

$$\begin{aligned}\|\mathbf{r}^{\ell'}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2 &\le \prod_{\eta=\ell}^{\ell'-1} \exp(-\beta_\eta)(\|\mathbf{r}^{\ell}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2)\\ &\overset{(a)}{\le} \exp(-(\ell'-\ell)\beta_{\ell'-1})(\|\mathbf{r}^{\ell}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2), \quad \text{(D.7)}\end{aligned}$$

where (a) follows since β_η is non-increasing.

Let $\ell' = k + k_i$ and $\ell = k + k_{i-1}, i = 1, \cdots, L$. By (D.7), we have

$$\begin{aligned}&\|\mathbf{r}^{k+k_i}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2\\ \overset{(a)}{\le}& \exp(-\frac{\varepsilon^2(k_i-k_{i-1})}{|F_j^k|}(1-\delta_{|F_j^k\cup S^{k+k_{i-1}}|}))(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2)\\ \overset{(b)}{\le}& \exp(-\frac{\varepsilon^2(k_i-k_{i-1})}{|F_i^k|}(1-\delta_{|F_i^k\cup S^{k+k_{i-1}}|}))(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2)\\ \overset{(c)}{\le}& \exp(-\frac{\lambda\varepsilon^2}{4}(1-\delta_{|F_i^k\cup S^{k+k_{i-1}}|}))(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2)\\ \overset{(d)}{\le}& \exp(-\frac{\lambda\varepsilon^2}{4}(1-\delta_{K+k+\lfloor\lambda\theta^k\rfloor}))(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2), \quad \text{(D.8)}\end{aligned}$$

where (a) is due to (D.3), (b) is from $j \le i$, (c) is true since (3.8), (d) is because

$$\begin{aligned}|F_i^k \cup S^{k+k_i-1}| &\overset{(3.9)}{\le} |T \cup S^{k+\lceil\lambda 2^{L-1}\rceil-1-1}| \le |T \cup S^{k+\lfloor\lambda 2^{L-1}\rfloor}|\\ &\overset{(3.12)}{\le} |T \cup S^{k+\lfloor\lambda\theta^k\rfloor}| \le K + k + \lfloor\lambda\theta^k\rfloor, \quad \text{(D.9)}\end{aligned}$$

and the monotonicity of the RIP.

From (2.4), (D.8) can be rewritten as

$$\|\mathbf{r}^{k+k_i}\|_2^2 \le \eta\|\mathbf{r}^{k+k_{i-1}}\|_2^2 + (1-\eta)\|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2,$$

where $i = 1, \cdots, L$. Note that $k_0 = 0$. Then we have

$$\|\mathbf{r}^{k+k_L}\|_2^2 \le \eta\|\mathbf{r}^{k+k_{L-1}}\|_2^2 + (1-\eta)\|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2, \quad \text{(D.10)}$$
$$\cdots,$$
$$\|\mathbf{r}^{k+k_1}\|_2^2 \le \eta\|\mathbf{r}^{k}\|_2^2 + (1-\eta)\|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2. \quad \text{(D.11)}$$

From (D.10)–(D.11), we get

$$\begin{aligned}&\|\mathbf{r}^{k+k_L}\|_2^2 \le \eta^L\|\mathbf{r}^{k}\|_2^2 + (1-\eta)\sum_{j=1}^{L}\eta^{L-j}\|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2\\ \overset{(a)}{\le}& \eta^L\|\mathbf{\Phi x}_{F^k} + \mathbf{v}\|_2^2 + (1-\eta)\sum_{j=1}^{L}\eta^{L-j}\|\mathbf{\Phi x}_{F^k\setminus F_j^k} + \mathbf{v}\|_2^2\end{aligned}$$

$$
\begin{aligned}
&\leq \eta^L(\|\mathbf{\Phi x}_{F^k}\|_2+\|\mathbf{v}\|_2)^2+(1-\eta)\sum_{j=1}^{L}\eta^{L-j}(\|\mathbf{\Phi x}_{F^k\backslash F_j^k}\|_2+\|\mathbf{v}\|_2)^2\\
&\overset{(b)}{\leq}\eta^L(\sqrt{1+\delta_{|F^k|}}\|\mathbf{x}_{F^k\backslash F_0^k}\|_2+\|\mathbf{v}\|_2)^2+(1-\eta)\sum_{j=1}^{L}\eta^{L-j}(\sqrt{1+\delta_{|F^k\backslash F_j^k|}}\|\mathbf{x}_{F^k\backslash F_j^k}\|_2+\|\mathbf{v}\|_2)^2\\
&\overset{(c)}{\leq}\eta^L(\sqrt{1+\delta_{\theta^k}}\|\mathbf{x}_{F^k\backslash F_0^k}\|_2+\|\mathbf{v}\|_2)^2+(1-\eta)\sum_{j=1}^{L}\eta^{L-j}(\sqrt{1+\delta_{\theta^k}}\|\mathbf{x}_{F^k\backslash F_j^k}\|_2+\|\mathbf{v}\|_2)^2\\
&\overset{(d)}{<}\eta^L(\sqrt{(1+\delta_{\theta^k})\tau^{L-1}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2+\|\mathbf{v}\|_2)^2\\
&\qquad+(1-\eta)\sum_{j=1}^{L}\eta^{L-j}(\sqrt{(1+\delta_{\theta^k})\tau^{L-j-1}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2+\|\mathbf{v}\|_2)^2\\
&=((\tau\eta)^L+(1-\eta)\sum_{j=1}^{L}(\tau\eta)^{L-j})\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2^2+(\eta^L+(1-\eta)\sum_{j=1}^{L}\eta^{L-j})\|\mathbf{v}\|_2^2+2((\sqrt{\tau}\eta)^L\\
&\qquad+(1-\eta)\sum_{j=1}^{L}(\sqrt{\tau}\eta)^{L-j})\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2\|\mathbf{v}\|_2,
\end{aligned}
$$

where (a) is due to $\|\mathbf{r}^k\|_2^2 \leq \|\mathbf{\Phi x}_{F^k}+\mathbf{v}\|_2^2$ which is from Proposition 1 in [30], (b) holds due to the RIP, (c) is according to $|F^k\backslash F_j^k| < |F^k| = \theta^k$ for $j = 1, \cdots, L$, (d) is from (3.6) (i.e., $\|\mathbf{x}_{F^k\backslash F_j^k}\|_2 < \sqrt{\tau^{L-1-j}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2$). Since

$$
\begin{aligned}
&(\tau\eta)^L < \frac{1-\eta}{1-\tau\eta}(\tau\eta)^L = (1-\eta)\sum_{j=L}^{\infty}(\tau\eta)^j, \quad \eta^L = (1-\eta)\sum_{j=L}^{\infty}\eta^j,\\
&(\sqrt{\tau}\eta)^L < \frac{1-\eta}{1-\sqrt{\tau}\eta}(\sqrt{\tau}\eta)^L = (1-\eta)\sum_{j=L}^{\infty}(\sqrt{\tau}\eta)^j
\end{aligned}
$$

when $\tau > 1, \tau\eta < 1$, and $\eta < 1$, then we have

$$
\begin{aligned}
\|\mathbf{r}^{k+k_L}\|_2^2 &< ((1-\eta)\sum_{j=L}^{\infty}(\tau\eta)^j+(1-\eta)\sum_{j=0}^{L-1}(\tau\eta)^j)\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2^2+((1-\eta)\sum_{j=L}^{\infty}\eta^j\\
&\qquad+(1-\eta)\sum_{j=0}^{L-1}\eta^j)\|\mathbf{v}\|_2^2+2((1-\eta)\sum_{j=L}^{\infty}(\sqrt{\tau}\eta)^j\\
&\qquad+(1-\eta)\sum_{j=1}^{L}(\sqrt{\tau}\eta)^{L-j})\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2\|\mathbf{v}\|_2\\
&=(1-\eta)\sum_{j=0}^{\infty}(\tau\eta)^j\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2^2+(1-\eta)\sum_{j=0}^{\infty}\eta^j\|\mathbf{v}\|_2^2\\
&\qquad+2(1-\eta)\sum_{j=0}^{\infty}(\sqrt{\tau}\eta)^j\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2\|\mathbf{v}\|_2\\
&=\frac{1-\eta}{1-\tau\eta}\times\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2^2+\|\mathbf{v}\|_2^2+\frac{2(1-\eta)}{1-\sqrt{\tau}\eta}\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2\|\mathbf{v}\|_2\\
&\overset{(a)}{<}\frac{1-\eta}{1-\tau\eta}\times\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2^2+\|\mathbf{v}\|_2^2+2\sqrt{\frac{1-\eta}{1-\tau\eta}}\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2\|\mathbf{v}\|_2
\end{aligned}
$$

$$
\begin{aligned}
&=\Big(\sqrt{\frac{(1-\eta)(1+\delta_{\theta^k})}{\tau(1-\tau\eta)}}\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\|\mathbf{v}\|_2\Big)^2\\
&\overset{(b)}{=}\Big(\sqrt{4\eta(1-\eta)(1+\delta_{\theta^k})}\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\|\mathbf{v}\|_2\Big)^2\\
&\overset{(c)}{\le}\Big(\sqrt{4\eta(1-\eta)(1+\delta_{K+k+\lfloor\lambda\theta^k\rfloor})}\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\|\mathbf{v}\|_2\Big)^2,
\end{aligned}
$$

where (a) is from

$$
\begin{aligned}
\Big(\frac{1-\eta}{1-\sqrt{\tau}\eta}\Big)^2-\Big(\sqrt{\frac{1-\eta}{1-\tau\eta}}\Big)^2&=\frac{(1-\eta)^2(1-\tau\eta)-(1-\eta)(1-\sqrt{\tau}\eta)^2}{(1-\sqrt{\tau}\eta)^2(1-\tau\eta)}\\
&=\frac{-\eta(1-\eta)(\sqrt{\tau}-1)^2}{(1-\sqrt{\tau}\eta)^2(1-\tau\eta)}<0,
\end{aligned}
$$

(b) chooses $\tau=\frac{1}{2\eta}$, (c) follows from $\theta^k=|T\backslash S^k|\le|T\cup S^{k+\lfloor\lambda\theta^k\rfloor}|\le K+k+\lfloor\lambda\theta^k\rfloor$ and the monotonicity of the RIP. Note that

$$
k+k_L\overset{(3.9)}{\le}k+\lceil\lambda 2^{L-1}\rceil-1\overset{(3.10)}{=}k+k',
$$

and $\|\mathbf{r}^k\|_2$ is always non-increasing for $k\ge 0$, then we have

$$
\|\mathbf{r}^{k+k'}\|_2\le\|\mathbf{r}^{k+k_L}\|_2<\sqrt{4\eta(1-\eta)(1+\delta_{K+k+\lfloor\lambda\theta^k\rfloor})}\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\|\mathbf{v}\|_2. \tag{D.12}
$$

On the other hand, we have

$$
\begin{aligned}
\|\mathbf{r}^{k+k'}\|_2&=\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}^{k+k'}\|_2=\|\mathbf{\Phi}(\mathbf{x}-\mathbf{x}^{k+k'})+\mathbf{v}\|_2\ge\|\mathbf{\Phi}(\mathbf{x}-\mathbf{x}^{k+k'})\|_2-\|\mathbf{v}\|_2\\
&\overset{(a)}{\ge}\sqrt{1-\delta_{|T\cup S^{k+k'}|}}\|\mathbf{x}-\mathbf{x}^{k+k'}\|_2-\|\mathbf{v}\|_2\ge\sqrt{1-\delta_{|T\cup S^{k+k'}|}}\|\mathbf{x}_{F^{k+k'}}\|_2-\|\mathbf{v}\|_2\\
&\overset{(b)}{\ge}\sqrt{1-\delta_{K+k+\lfloor\lambda\theta^k\rfloor}}\|\mathbf{x}_{F^{k+k'}}\|_2-\|\mathbf{v}\|_2,
\end{aligned} \tag{D.13}
$$

where (a) is due to the RIP, (b) follows from (D.9) and the monotonicity of the RIP. By relating (D.12) and (D.13), we obtain

$$
\begin{aligned}
\|\mathbf{x}_{F^{k+k'}}\|_2&\le\frac{1}{\sqrt{1-\delta_{K+k+\lfloor\lambda\theta^k\rfloor}}}(\|\mathbf{r}^{k+k'}\|_2+\|\mathbf{v}\|_2)\\
&<\sqrt{\frac{4\eta(1-\eta)(1+\delta_{K+k+\lfloor\lambda\theta^k\rfloor})}{(1-\delta_{K+k+\lfloor\lambda\theta^k\rfloor})}}\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\frac{2}{\sqrt{1-\delta_{K+k+\lfloor\lambda\theta^k\rfloor}}}\|\mathbf{v}\|_2\\
&=\alpha\|\mathbf{x}_{F^k\backslash F^k_{L-1}}\|_2+\beta\|\mathbf{v}\|_2,
\end{aligned} \tag{D.14}
$$

where α and β has been defined in (D.1) and (D.2), respectively. □

Acknowledgments. This study was partially supported by the National Natural Science Foundation of China (grant nos. 61907014, 11871248, 11701410, 61901160), the Natural Science Foundation of Guangdong province (No. 2021A1515010857), Youth Science Foundation of Henan Normal University (grant no. 2019QK03), China Postdoctoral Science Foundation (grant no. 2019M660557), and Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019).

References

[1] E. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, **52** (2006), 489–509.

[2] D. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, **52** (2006), 1289–1306.

[3] E. Candès and J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Problems*, **23** (2007), 969–985.

[4] Y. Zhao, C. Xu, 1-Bit compressive sensing: Reformulation and RRSP-based sign recovery theory, *Science China Mathematics*, **59** (2016) 2049–2074.

[5] B. Han, Z. Xu, Robustness properties of dimensionality reduction with Gaussian random matrices, *Science China Mathematics*, **60** (2017) 1753–1778.

[6] S. Chen, D. Donoho, Saunders M, Atomic decomposition by basis pursuit, *SIAM Review*, **43** (2001), 129–159.

[7] W. Chen, H. Ge, A sharp recovery condition for block sparse signals by block orthogonal multi-matching pursuit, *Science China Mathematics*, **60** (2017), 1325–1340.

[8] Z. Zhao, F. Xu, X. Li, Adaptive projected gradient thresholding methods for constrained ℓ_0 problems, *Science China Mathematics*, **58** (2015), 2205–2224.

[9] W. Dan, R. Wang, Robustness of orthogonal matching pursuit under restricted isometry property, *Science China Mathematics*, **57** (2014), 627–634.

[10] H. Li, J. Wen, Generalized covariance-assisted matching pursuit, *Signal Processing,* **163** (2019), 232–237.

[11] J. Wang, B. Shim, Exact recovery of sparse signals using orthogonal matching pursuit: How many iterations do we need? *IEEE Transactions on Signal Processing*, **64** (2016), 4194–4202.

[12] H. Li, L. Wang, X. Zhan, et al. On the fundamental limit of orthogonal matching pursuit for multiple measurement vector, *IEEE Access*, **7** (2019), 48860–48866.

[13] S. Chen, S. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, *International Journal of Control*, **50** (1989), 1873–1896.

[14] C. Soussen, R. Gribonval, J. Idier, et al, Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares, *IEEE Transactions on Information Theory*, **59** (2013), 3158–3174.

[15] C. Herzet, C. Soussen, J. Idier, et al, Exact recovery conditions for sparse representations with partial support information, *IEEE Transactions on Information Theory*, **59** (2013), 7509–7524.

[16] C. Herzet, A. Drémeau, C. Soussen, Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay, *IEEE Transactions on Information Theory*, **62** (2015), 459–470.

[17] Y. Pati, R. Rezaiifar, P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in Proc. *27th Annu. Asilomar Conf. Signals, Systems, and Computers. IEEE, Pacific Grove, CA*, **1** (1993), 40–44.

[18] H. Li, G. Liu, An improved analysis for support recovery with orthogonal matching pursuit under general perturbations, *IEEE Access*, **6** (2018), 18856–18867.

[19] J. Wen, J. Wang, Q. Zhang, Nearly optimal bounds for orthogonal least squares, *IEEE Transactions on Signal Processing*, **65** (2017), 5347–5356.

[20] P. Geng, W. Chen, H. Ge, Perturbation analysis of orthogonal least squares, *Canadian Mathematical Bulletin*, **62** (2019), 780–797.

[21] J. Wang, P. Li, Recovery of sparse signals using multiple orthogonal least squares, *IEEE Transactions on Signal Processing*, **65** (2017), 2049–2062.

[22] H. Li, J. Zhang, J. Zou, Improving the bound on the restricted isometry property constant in multiple orthogonal least squares, *IET Signal Processing*, **12** (2018), 666–671.

[23] J. Kim, B. Shim, A near-optimal restricted isometry condition of multiple orthogonal least squares, *IEEE Access*, **7** (2018) 46822–46830.

[24] S. Mukhopadhyay, S. Satpathi, M. Chakraborty, A modified multiple OLS (m^2OLS) algorithm for signal recovery in compressive sensing, *Signal Processing*, **168** (2020), 1–10.

[25] E. Candes, T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, **51** (2005), 4203–4215.

[26] R. Baraniuk, M. Davenport, R. Devore, et al. A Simple Proof of the Restricted Isometry Property for Random Matrices, *Constructive Approximation*, **28** (2008), 253–263.

[27] T. Zhang, Sparse recovery with orthogonal matching pursuit under RIP, *IEEE Transactions on Information Theory*, **57** (2011), 6215–6221.

[28] A. Cohen, W. Dahmen, R. DeVore, Orthogonal matching pursuit under the restricted isometry property, *Constructive Approximation*, **45** (2017), 113–127.

[29] S. Foucart, Stability and robustness of weak orthogonal matching pursuits, *Recent Advances in Harmonic Analysis and Applications*, Springer, 2013.

[30] J. Wang, S. Kwon, P. Li, et al, Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis, *IEEE Transactions on Signal Processing*, **64** (2016), 1076–1089.

[31] T. Blumensath, M. Davies, On the difference between orthogonal matching pursuit and orthogonal least squares, *Technical Report, University of Southampton, Southampton*, UK, 2007.