Acta Mathematica Scientia, 2022, **42B**(3): 941–956
https://doi.org/10.1007/s10473-022-0308-4

Mathematica Acta Scientia 数学物理学报
http://actams.apm.ac.cn

A NEW SUFFICIENT CONDITION FOR SPARSE RECOVERY WITH MULTIPLE ORTHOGONAL LEAST SQUARES*

Haifeng LI (李海锋)† *Jing ZHANG* (张静)
Henan Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China
E-mail: *lihaifengxx@126.com; a1293651766@126.com*

Abstract A greedy algorithm used for the recovery of sparse signals, multiple orthogonal least squares (MOLS) have recently attracted quite a big of attention. In this paper, we consider the number of iterations required for the MOLS algorithm for recovery of a K-sparse signal $\mathbf{x} \in \mathbb{R}^n$. We show that MOLS provides stable reconstruction of all K-sparse signals $\mathbf{x}$ from $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ in $\lceil \frac{6K}{M} \rceil$ iterations when the matrix $\mathbf{A}$ satisfies the restricted isometry property (RIP) with isometry constant $\delta_{7K} \le 0.094$. Compared with the existing results, our sufficient condition is not related to the sparsity level K.

Key words Sparse signal recovery; multiple orthogonal least squares (MOLS); sufficient condition; restricted isometry property (RIP)

2010 MR Subject Classification 94A12; 65F22; 65J22

1 Introduction

The orthogonal least squares (OLS) algorithm [1–11] is a classical greedy algorithm for recovering K-sparse signal $\mathbf{x} \in \mathbb{R}^n$ from

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1.1}$$

where $\mathbf{x}$ has at most K nonzero entries (i.e., $\|\mathbf{x}\|_0 \le K$), $\mathbf{A} \in \mathbb{R}^{m\times n} (m \ll n)$ and $\mathbf{w}$ is a noise vector. OLS identifies the support of the underlying sparse signal by adding one index to the list at a time, and estimates the sparse coefficients 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 $\mathbf{y}$, yielding a residual update for the next iteration. See [9] for a mathematical description of OLS.

In [6] and [9], we observe that the OLS algorithm has a better convergence property, while it is computationally more expensive than OMP (orthogonal matching pursuit) (see [12–17]). The OLS algorithm has attracted much attention over the course of last several years.

*Received September 2, 2020; revised May 28, 2021. This work was partially supported by the National Natural Science Foundation of China (61907014, 11871248, 11701410, 61901160), Youth Science Foundation of Henan Normal University (2019QK03).

†Corresponding author

Recently, many efforts have also been made to study an extension of OLS, referred to as multiple orthogonal least squares (MOLS) (see [6, 7]). Compared to OLS, MOLS selects M indices at a time (see Table 1), which reduces the computational complexity and greatly improves the computational speed. An important challenge is to characterize the exact recovery conditions of MOLS using the properties of measurement matrices such as the restricted isometry property (RIP). The definition of RIP is as follows:

Definition 1.1 ([18]) A measurement matrix $\mathbf{A}$ satisfies the RIP of order K if there exists a constant $\delta \in (0,1)$ such that

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

holds for all K-sparse vectors $\mathbf{x}$. The minimum δ satisfying (1.2) is defined as the restricted isometry constant (RIC) constant δ_K.

Table 1 The MOLS algorithm

Input: $\mathbf{A}$, $\mathbf{y}$, sparsity level K, residual tolerant ϵ, and selection parameter $M \le K$

Initialization: $\mathbf{r}^0 = \mathbf{y}$, $k = 0$, and $T^0 = \emptyset$

while $k < \frac{m}{M}$ and $\|\mathbf{r}^k\|_2 > \epsilon$ **do**

$k = k + 1$

Identify $S^k = \underset{S:|S|=M}{\arg\min} \sum_{i\in S} \|\mathbf{P}^{\perp}_{T^{k-1}\cup\{i\}}\mathbf{y}\|_2^2$

Enlarge $T^k = T^{k-1} \cup \{S^k\}$

Estimate $\mathbf{x}^k = \underset{\text{supp}(\mathbf{u})\subseteq T^k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2$

Update $\mathbf{r}^k = \mathbf{y} - \mathbf{A}\mathbf{x}^k$

end while

Output: the estimated support $\hat{T} = \underset{\Lambda:|\Lambda|=K}{\arg\min} \|\mathbf{x}^k - \mathbf{x}^k_\Lambda\|_2$ and signal $\hat{\mathbf{x}}$ satisfying

$\hat{\mathbf{x}} = \underset{\text{supp}(\mathbf{u})\subseteq \hat{T}}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2$

At present, most studies indicate that MOLS recovers K-sparse signals in at most K iterations. For example, in [6], it was shown that MOLS recovers K-sparse signals in at most K iterations under

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

In [7], the condition was improved to

$$\delta_{MK+1} < \frac{1}{\sqrt{\frac{K}{M} + 2}}. \tag{1.4}$$

Recently, Kim and Shim [8] presented a near-optimal restricted isometry condition (RIC) of MOLS as follows:

$$\delta_{LK-L+2} < \frac{1}{\sqrt{\frac{K}{M} + 2 - \frac{1}{M}}}. \tag{1.5}$$

One can conclude from (1.3)–(1.5) that exact recovery with MOLS can be ensured when RIC is inversely proportional to $\sqrt{K}$.

In this paper, we go further, and study the performance analysis of the MOLS algorithm. More specifically, as is shown in Corollary 2.3, MOLS achieves stable recovery of the K-sparse signal $\mathbf{x}$ from $\mathbf{y}=\mathbf{A}\mathbf{x}+\mathbf{w}$ in $\lceil \frac{6K}{M} \rceil$ iterations with

$$\delta_{7K} \leq 0.094. \tag{1.6}$$

It is important to stress the novel aspects of our contribution. The works [18, 19] claimed that a random matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with entries drawn i.i.d. from Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$ obeys the RIP with $\delta_K \leq \epsilon$ with overwhelming probability if

$$m = \mathcal{O}\bigg(\frac{K \log \frac{n}{K}}{\epsilon^2}\bigg). \tag{1.7}$$

The number of required measurements is $m=\mathcal{O}(K^2 \log \frac{n}{K})$. On the other hand, our condition (1.6) requires that $m=\mathcal{O}(K \log \frac{n}{K})$; this is obviously smaller than the previous results, in particular for large K.

The rest of this paper is organized as follows: in Section 2, we present some observations and our main results. In Section 3, we provide some technical lemmas that are useful for our analysis and prove Theorem 2.2. Finally, we summarize our results in Section 4.

Notation: Denote $\Omega=\{1,\dots,n\}$. Let $T=\mathrm{supp}(\mathbf{x})=\{i|x_i \neq 0, i\in\Omega\}$ be the support of a K-sparse vector $\mathbf{x}$ (i.e., the set of the positions of its K nonzero elements). Let Λ be a subset of Ω and let $|\Lambda|$ be the cardinality of Λ. $T\backslash\Lambda=\{i|i\in T, i\notin\Lambda\}$. $\mathbf{x}_\Lambda \in \mathbb{R}^n$ denotes the vector equal to $\mathbf{x}$ on an index set Λ and zero elsewhere. Throughout the paper, we assume that $\mathbf{A}\in\mathbb{R}^{m\times n}$ is normalized to have a unit column norm (i.e., $\|\mathbf{A}_i\|_2=1$ for $i=1,2,\dots,n$).[1] Let $\mathbf{A}_\Lambda \in \mathbb{R}^{m\times|\Lambda|}$ be a sub-matrix of $\mathbf{A}$ with an index of its columns in set Λ. For any matrix $\mathbf{A}_\Lambda$ of full column-rank, let $\mathbf{A}_\Lambda^\dagger=(\mathbf{A}'_\Lambda\mathbf{A}_\Lambda)^{-1}\mathbf{A}'_\Lambda$ be the pseudo-inverse of $\mathbf{A}_\Lambda$, where $\mathbf{A}'_\Lambda$ denotes the transpose of $\mathbf{A}_\Lambda$. $\mathbf{P}_\Lambda=\mathbf{A}_\Lambda\mathbf{A}_\Lambda^\dagger$ and $\mathbf{P}_\Lambda^\perp=\mathbf{I}-\mathbf{P}_\Lambda$ stand for the projector and orthogonal complement projector , respectively, onto $\mathrm{span}(\mathbf{A}_\Lambda)$ (i.e., the column space of $\mathbf{A}_\Lambda$).

2 Sparse Recovery With MOLS

2.1 Observations

Before giving the details of Theorem 2.2, we obtain an important observation on the MOLS algorithm. As shown in Table 1, in the $(k+1)$-th iteration ($k\geq 0$), MOLS adds to T^k a set of M indices that results in the maximum reduction of the residual power, i.e.,

$$S^{k+1}=\underset{S:|S|=M}{\arg\min}\sum_{i\in S}\|\mathbf{P}^\perp_{T^k\cup\{i\}}\mathbf{y}\|_2^2. \tag{2.1}$$

We intuitively observe that it requires the construction of $n-Mk$ different orthogonal projections (i.e., $\mathbf{P}^\perp_{T^k\cup\{i\}}$) to identify S^{k+1}. This implementation of MOLS is, however, computationally expensive. In order to solve this problem, inspired by the technical report [20], Wang and

[1] [20] has shown that the behavior of OLS is unchanged whether columns of $\mathbf{A}$ are normalized or not. As MOLS is a direct extension of OLS, one can verify that the normalization does not matter either for the behavior of MOLS.

Li [6] presented a cost-effective alternative to (2.1) for the identification step of MOLS. The result is given in the following lemma:

Lemma 2.1 ([6], Proposition 1) Consider the MOLS algorithm. At the $(k+1)$-th iteration, the MOLS algorithm selects the index

$$S^{k+1}=\underset{S:|S|=M}{\arg\max}\sum_{i\in S}\frac{|\langle \mathbf{A}_i,\mathbf{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\mathbf{A}_i\|_2}. \tag{2.2}$$

We can see from (2.2) that it suffices to find the M largest values in $\frac{|\langle \mathbf{A}_i,\mathbf{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\mathbf{A}_i\|_2}$, which is much simpler than (2.1), as it involves only one projection operator (i.e., $\mathbf{P}^{\perp}_{T^k}$). By numerical experiments, we have indeed confirmed that the simplification offers a massive reduction in the computational cost. Hence, Lemma 2.1 plays an important role in analyzing the recovery condition of MOLS.

Moreover, note that the identification rule of MOLS is akin to that of a generalized orthogonal matching pursuit (gOMP). Specifically, in the $(k+1)$-th iteration, gOMP picks a set of M indices corresponding to the column which is most strongly correlated with the signal residual, i.e.,

$$S^{k+1}=\underset{S:|S|=M}{\arg\max}\sum_{i\in S}|\langle \mathbf{A}_i,\mathbf{r}^k\rangle|.$$

Clearly, the rule of MOLS differs from that of gOMP only in that it has an extra normalization factor (i.e., $\|\mathbf{P}^{\perp}_{T^k}\mathbf{A}_i\|_2$). Thus, the greedy selection rule in MOLS can also be viewed as an extension of the gOMP rule. This argument has been verified (see [6, 20, 21]). However, this property shows that the rule of MOLS coincides with that of gOMP only in the first iteration because $S^0=\emptyset$ leads to $\|\mathbf{P}^{\perp}_{S^0}\mathbf{A}_i\|_2=\|\mathbf{A}_i\|_2$. For the subsequent iterations, MOLS does make a difference, since $\|\mathbf{P}^{\perp}_{S^k}\mathbf{A}_i\|_2\le\|\mathbf{A}_i\|_2, \forall k\ge 1$. In fact, as will be seen later, this factor makes the analysis of MOLS different and more challenging than that of gOMP.

2.2 Main Results

Theorem 2.2 Let $\mathbf{y}=\mathbf{A}\mathbf{x}+\mathbf{w}$ be the noisy measurement, where $\mathbf{x}\in\mathbb{R}^n$ is any K-sparse signal supported on T, $\mathbf{A}\in\mathbb{R}^{m\times n}$ is a measurement matrix with ℓ_2-normalized columns, and $\mathbf{w}$ is a noisy vector. Let $\theta^k=|T\backslash S^k|$ be the index number of a remaining support set after performing k $(k\ge 0)$ iterations of MOLS. If $\mathbf{A}$ obeys the RIP with the isometry constant

$$\delta_{Mk+7\theta^k}\le 0.094, \tag{2.3}$$

the residual of MOLS satisfies

$$\|\mathbf{r}^{k+\lceil\frac{6\theta^k}{M}\rceil}\|_2\le\xi_k\|\mathbf{w}\|_2, \tag{2.4}$$

where

$$\xi_k=2\left(1-\left(\frac{4\mu(1-\mu)(1+\delta_{Mk+7\theta^k})}{(1-\delta_{Mk+7\theta^k})}\right)^{\frac{1}{2}}\right)^{-1}-1\ge 1 \tag{2.5}$$

is a constant depending only on $\delta_{Mk+7\theta^k}$, and

$$\mu=e^{-\frac{3}{2}(1-\delta_{Mk+7\theta^k})^2}. \tag{2.6}$$

Proof See Section III. □

We observe from Theorem 2.2 that MOLS requires at most $\lceil \frac{6\theta^k}{M} \rceil$ additional iterations after running $k (k \geq 0)$ iterations to ensure that the condition in (2.4) is fulfilled. In other words, the ℓ_2-norm of the residual is upper bounded by the product of a constant and $\|\mathbf{w}\|_2$.

In particular, when $k = 0$ and $\theta^0 = |T \backslash S^0| = K$ in (2.4), we obtain that the ℓ_2-norm of the residual falls below $\xi_0 \|\mathbf{w}\|_2$. The result is as follows:

Corollary 2.3 Let $\mathbf{x}$ be any K-sparse signal and let $\mathbf{A}$ be a matrix with ℓ_2-normalized columns. If $\mathbf{A}$ satisfies the RIP with

$$\delta_{7K} \leq 0.094, \tag{2.7}$$

the residual of MOLS satisfies

$$\|\mathbf{r}^{\lceil \frac{6K}{M} \rceil}\|_2 \leq \xi_0 \|\mathbf{w}\|_2, \tag{2.8}$$

where ξ_0 has been defined in (2.5) with $k = 0$.

Next, we show that the ℓ_2-norm of the recovery error is also upper bounded by the product of a constant and $\|\mathbf{w}\|_2$.

Theorem 2.4 Let $\mathbf{x} \in \mathbb{R}^n$ be any K-sparse signal, let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix with ℓ_2-normalized columns, and let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ be the noisy sampling model. If $\mathbf{A}$ obeys RIP with (2.7), MOLS satisfies

$$\|\mathbf{x}^{\lceil \frac{6K}{M} \rceil} - \mathbf{x}\|_2 \leq (\xi_0 + 1) \|\mathbf{w}\|_2, \tag{2.9}$$

and

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq 2(1 - \delta_{8K})^{-\frac{1}{2}} ((\xi_0 + 1)(1 + \delta_{8K})^{\frac{1}{2}} + 1) \|\mathbf{w}\|_2, \tag{2.10}$$

where ξ_0 has been defined in (2.5) with $k = 0$.

Proof See Appendix A. □

Remark 2.5 (Comparison with [8]) In [8], the authors showed that the MOLS algorithm ensures the accurate recovery of any K-sparse signal, provided that $\mathbf{A}$ satisfies the restricted isometry property (RIP) with (1.5), which is the best existing result for MOLS. However, the sufficient condition (1.5) is inversely proportional to $\sqrt{K}$; it will vanish as K increases. Our sufficient condition (2.7) is upper bounded by a constant.

Remark 2.6 (Comparison with [22]) The authors in [22] claimed that gOMP can exactly recover the supports of all K-sparse vectors from the samples $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ in $\max\{K, \lfloor \frac{8K}{M} \rfloor\}$ iterations if $\mathbf{A}$ satisfies the RIP with

$$\delta_{\max\{9K, (M+1)K\}} \leq \frac{1}{8} = 0.125. \tag{2.11}$$

It is easy to see that the number of iterative steps in our results are less than those of [22].

Remark 2.7 (Comparison with [23]) In [23], the authors claimed that MOLS ensures the exact recovery of any K-sparse vector in $\max\{K, \lfloor \frac{8K}{M} \rfloor\}$ iterations under $\delta_{\lfloor 7.8K \rfloor} \leq 0.155$. According to (2.8), our results show that MOLS can recover any K-sparse signals within $\lceil \frac{6K}{M} \rceil$. Thus, our result is better than that in [23].

Remark 2.8 (Comparison with [9, 24]) MOLS reduces to OLS when $M = 1$. From Corollary 2.3, our result indicates that OLS can exactly recover the K-sparse signals $\mathbf{x}$ within

$6K$ iteration with (2.7). The work [9] stated that OLS exactly recovers the support of any K-sparse vector $\mathbf{x}$ with

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}}. \tag{2.12}$$

The reference [24] provides a sufficient condition for OLS:

$$\delta_{K+1} = \begin{cases} \dfrac{1}{\sqrt{K}}, & K = 1, \\ \dfrac{1}{\sqrt{K+\frac{1}{4}}}, & K = 2, \\ \dfrac{1}{\sqrt{K+\frac{1}{16}}}, & K = 3, \\ \dfrac{1}{\sqrt{K}}, & K \geq 4. \end{cases} \tag{2.13}$$

It is easy to see that the upper bounds of (2.12) and (2.13) are inversely proportional to $\sqrt{K}$, which requires that m should scale with $K^2 \log \frac{n}{K}$. On the other hand, the upper bound of the proposed guarantee (2.7) is independent of K.

3 Proof of Theorem 2.2

For the proof of the analysis of Theorem 2.2, our idea is related to [22]. Here, we first denote $F^k = T \backslash T^k$ 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

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

Then we observe that the last set

$$F^k_{\max\{0, \lfloor \log_2 \frac{\theta^k+1}{M} \rfloor\}+1} = F^k$$

may have less than $2^{\max\{0, \lfloor \log_2 \frac{\theta^k+1}{M} \rfloor\}+1} M - 1$ elements. On the other hand, we can obtain

$$2|F^k| > 2^L M - 1 \tag{3.2}$$

for $L \in \{1, \ldots, \max\{0, \lfloor \log_2 \frac{\theta^k+1}{M} \rfloor\} + 1\}$.

Next, for constant $\tau > 1$, let $L \in \{1, 2, \ldots, \max\{0, \lfloor \log_2 \frac{\theta^k+1}{M} \rfloor\} + 1\}$ be the minimum positive integer satisfying

$$\|\mathbf{x}_{F^k \backslash F_0^k}\|_2^2 < \tau \|\mathbf{x}_{F^k \backslash F_1^k}\|_2^2, \tag{3.3}$$

$$\|\mathbf{x}_{F^k \backslash F_1^k}\|_2^2 < \tau \|\mathbf{x}_{F^k \backslash F_2^k}\|_2^2, \tag{3.4}$$

$$\cdots,$$

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

$$\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2^2 \geq \tau\|\mathbf{x}_{F^k\setminus F^k_L}\|_2^2. \tag{3.6}$$

Then we have

$$\|\mathbf{x}_{F^k\setminus F^k_j}\|_2^2 < \tau^{L-1-j}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2^2 \tag{3.7}$$

for $j = 0, 1, \ldots, L$. Here we note that if (3.6) holds true for all $L \geq 1$, we ignore (3.3)–(3.5) and simply take $L = 1$. In addition, L always exists because $\Big\|\mathbf{x}_{F^k\setminus F^k_{\max\{0,\lfloor\log_2\frac{\theta^k+1}{M}\rfloor\}+1}}\Big\|_2^2 = 0$, so that (3.6) holds true at least for $L = \max\{0, \lfloor\log_2\frac{\theta^k+1}{M}\rfloor\} + 1$.

In consideration of the selection rule of MOLS viewed as an extension of the gOMP rule, we will prove Theorem 2.2 by using mathematical induction in θ^k. In fact, the mathematical induction was proposed in [22]. Here, θ^k stands for the number of remaining indices after k iterations of MOLS. We first select $\theta^k = 0$, and then no more iteration is needed, i.e., $T \subseteq T^k$, so we have

$$\|\mathbf{r}^k\|_2 = \|\mathbf{y}-\mathbf{A}\mathbf{x}^k\|_2 = \min_{\mathrm{supp}(\mathbf{u})=T^k}\|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{w}\|_2 \leq \xi_k\|\mathbf{w}\|_2.$$

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 (2.4) holds true, i.e.,

$$\|\mathbf{r}^{k+\lceil\frac{6\theta^k}{M}\rceil}\|_2 \leq \xi_k\|\mathbf{w}\|_2 \tag{3.8}$$

holds true.

In order to prove (3.8), we will choose a decent number 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

$$k_0 = 0 \tag{3.9}$$

and

$$k_i = \frac{3}{2}\sum_{j=1}^{i}\lceil\frac{|F^k_j|}{M}\rceil, \quad i = 1, \ldots, L. \tag{3.10}$$

Because of the definition of F^k_j, we have $|F^k_j| \leq 2^jM - 1$ for $j = 1, \ldots, L$, and

$$\begin{aligned} k_i \leq k_L &= \frac{3}{2}\sum_{j=1}^{L}\lceil\frac{|F^k_j|}{M}\rceil \leq \frac{3}{2}\sum_{j=1}^{L}\lceil\frac{2^jM-1}{M}\rceil \\ &= \frac{3}{2}\sum_{j=1}^{L}\lceil 2^j - \frac{1}{M}\rceil \overset{(a)}{\leq} \frac{3}{2}\sum_{j=1}^{L}2^j = 3\times 2^L - 3, \end{aligned} \tag{3.11}$$

where (a) follows from $0 < \frac{1}{M} \leq 1$. Let

$$k' = 3\times 2^L - 3 \tag{3.12}$$

indicate specified additional iterations after running k iterations of OLS.

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

$$\theta^{k+k'} = |F^{k+k'}| \leq \theta^k - 2^{L-1}M \tag{3.13}$$

after running $k+k'$ iterations, then inequality (3.8) holds when we require at most $\lceil \frac{6\theta^{k+k'}}{M} \rceil$ additional iterations. Our proof is completed. In fact, the total number of iterations of MOLS is

$$\begin{aligned} k+k'+\lceil \frac{6\theta^{k+k'}}{M} \rceil &\le k+3\times 2^L-3+\lceil \frac{6}{M}(\theta^k-2^{L-1}M)\rceil \\ &\le k+3\times 2^L-3+\lceil \frac{6\theta^k}{M}\rceil-3\times 2^L+1<k+\lceil \frac{6\theta^k}{M}\rceil. \end{aligned} \tag{3.14}$$

Then we obtain

$$\|\mathbf{r}^{k+\lceil \frac{6\theta^k}{M}\rceil}\|_2 \le \|\mathbf{r}^{k+k'+\lceil \frac{6\theta^k}{M}\rceil}\|_2. \tag{3.15}$$

Since the index number of remaining support is no more than θ^k-1, i.e.,

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

by the induction hypothesis we have that

$$\|\mathbf{r}^{k+k'}\|_2 \le \xi_k\|\mathbf{w}\|_2. \tag{3.16}$$

Then we combine (3.15) with (3.16) and get

$$\|\mathbf{r}^{k+\lceil \frac{6\theta^k}{M}\rceil}\|_2 \le \|\mathbf{r}^{k+k'}\|_2 \le \xi_k\|\mathbf{w}\|_2.$$

Thus, we require it to be ensured that (3.13) holds true. By the definition of F_j^k in (3.1), we have

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

and

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

Then (3.13) can be rewritten as

$$\theta^{k+k'}=|F^{k+k'}|<|F^k\backslash F_{L-1}^k|. \tag{3.17}$$

Since $\mathbf{x}_{F^k\backslash F_{L-1}^k}$ consists of $|F^k\backslash F_{L-1}^k|$ smallest non-zero elements (in magnitude) of $\mathbf{x}_{F^k}$, instead of proving things directly, we show that a sufficient condition of (3.17) is true; that is,

$$\|\mathbf{x}_{F^{k+k'}}\|_2<\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2. \tag{3.18}$$

Hence, we need to prove that inequality (3.18) holds true.

We have, by the result in Proposition E.6 (see Appendix E),

$$\|\mathbf{x}_{F^{k+k'}}\|_2<\alpha\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2+\beta\|\mathbf{w}\|_2, \tag{3.19}$$

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

It follows from (2.7) that $\alpha<1$. Then we discuss two cases.

If $\beta\|\mathbf{w}\|_2<(1-\alpha)\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2$, it is easy to see that (3.18) holds true.

If $\beta\|\mathbf{w}\|_2\ge(1-\alpha)\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2$, (3.8) holds true directly, due to

$$\begin{aligned} \|\mathbf{r}^{k+\lceil \frac{6\theta^k}{M}\rceil}\|_2 &\overset{(3.15)}{\le} \|\mathbf{r}^{k+k'+\lceil \frac{6\theta^k}{M}\rceil}\|_2 \overset{(a)}{\le} \|\mathbf{r}^{k+k'}\|_2 \\ &\overset{(E.14)}{<} \sqrt{4\mu(1-\mu)(1+\delta_{Mk+7\theta^k})}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2+\|\mathbf{w}\|_2 \\ &\overset{(E.1)}{=} \alpha\sqrt{1-\delta_{Mk+7\theta^k}}\|\mathbf{x}_{F^k\backslash F_{L-1}^k}\|_2+\|\mathbf{w}\|_2 \end{aligned}$$

$$
\begin{aligned}
&\le \alpha\sqrt{1-\delta_{Mk+7\theta^k}} \times \frac{\beta}{1-\alpha}\|\mathbf{w}\|_2 + \|\mathbf{w}\|_2 \\
&\overset{(E.2)}{=} \Big(\frac{2}{1-\alpha} - 1\Big)\|\mathbf{w}\|_2 = \xi_k\|\mathbf{w}\|_2,
\end{aligned}
\tag{3.20}
$$

where (a) follows from the fact that the residual power of MOLS is non-increasing, and where ξ_k has been defined in (2.5).

4 Conclusion

As an extension of OLS, MOLS is effective in reconstructing sparse signals and enhancing recovery performance. In this paper, we have presented an improved recovery guarantee of MOLS, which can stability recover K-sparse signals from the noisy measurements in $\lceil\frac{6K}{M}\rceil$ iterations under $\delta_{7K} \le 0.094$.

Appendix A

Proof of Theorem 2.4 Since $\mathbf{r}^{\lceil\frac{6K}{M}\rceil} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\lceil\frac{6K}{M}\rceil} = \mathbf{A}(\mathbf{x} - \mathbf{x}^{\lceil\frac{6K}{M}\rceil}) + \mathbf{w}$, we have

$$
\begin{aligned}
\|\mathbf{x}^{\lceil\frac{6K}{M}\rceil} - \mathbf{x}\|_2 &= \|\mathbf{x} - \mathbf{x}^{\lceil\frac{6K}{M}\rceil}\|_2 \overset{(a)}{\le} (1-\delta_{|T\cup T^{\lceil\frac{6K}{M}\rceil}|})^{-\frac{1}{2}}\|\mathbf{A}(\mathbf{x} - \mathbf{x}^{\lceil\frac{6K}{M}\rceil})\|_2 \\
&= (1-\delta_{|T\cup T^{\lceil\frac{6K}{M}\rceil}|})^{-\frac{1}{2}}\|\mathbf{r}^{\lceil\frac{6K}{M}\rceil} - \mathbf{w}\|_2 \overset{(b)}{\le} (1-\delta_{8K})^{-\frac{1}{2}}(\|\mathbf{r}^{\lceil\frac{6K}{M}\rceil}\|_2 + \|\mathbf{w}\|_2) \\
&\overset{(c)}{\le} (1-\delta_{8K})^{-\frac{1}{2}}(\xi_0\|\mathbf{w}\|_2 + \|\mathbf{w}\|_2) = (1-\delta_{8K})^{-\frac{1}{2}}(\xi_0+1)\|\mathbf{w}\|_2,
\end{aligned}
\tag{A.1}
$$

where (a) is based on the RIP, (b) uses the norm inequality and $|T\cup T^{\lceil\frac{6K}{M}\rceil}| \le K + \lceil\frac{6K}{M}\rceil M \le 7K + M \le 8K$, and (c) is according to Corollary 2.3.

Now, we need to prove that (2.10) is true. Since the best K-term approximation $(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K$ of $\mathbf{x}^{\lceil\frac{6K}{M}\rceil}$ is supported on $\hat{T}$ and

$$
\hat{\mathbf{x}} = \underset{\mathrm{supp}(\mathbf{u})=\hat{T}}{\arg\min}\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2,
\tag{A.2}
$$

we have

$$
\begin{aligned}
\|\hat{\mathbf{x}} - \mathbf{x}\|_2 &\overset{(a)}{\le} (1-\delta_{2K})^{-\frac{1}{2}}\|\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x})\|_2 = (1-\delta_{2K})^{-\frac{1}{2}}\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y} + \mathbf{w}\|_2 \\
&\le (1-\delta_{2K})^{-\frac{1}{2}}(\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2 + \|\mathbf{w}\|_2) \overset{(b)}{\le} (1-\delta_{2K})^{-\frac{1}{2}}(\|\mathbf{A}(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{y}\|_2 + \|\mathbf{w}\|_2) \\
&= (1-\delta_{2K})^{-\frac{1}{2}}(\|\mathbf{A}(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{A}\mathbf{x} - \mathbf{w}\|_2 + \|\mathbf{w}\|_2) \\
&\le (1-\delta_{2K})^{-\frac{1}{2}}(\|\mathbf{A}(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{A}\mathbf{x}\|_2 + 2\|\mathbf{w}\|_2),
\end{aligned}
\tag{A.3}
$$

where (a) is based on RIP, and (b) is from (A.2).

Using the RIP and the property of norm, (A.3) can be changed into

$$
\begin{aligned}
\|\hat{\mathbf{x}} - \mathbf{x}\|_2 &\le (1-\delta_{2K})^{-\frac{1}{2}}((1+\delta_{2K})^{\frac{1}{2}}\|(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{x}\|_2 + 2\|\mathbf{w}\|_2) \\
&= (1-\delta_{2K})^{-\frac{1}{2}}((1+\delta_{2K})^{\frac{1}{2}}\|(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{x}^{\lceil\frac{6K}{M}\rceil} + \mathbf{x}^{\lceil\frac{6K}{M}\rceil} - \mathbf{x}\|_2 + 2\|\mathbf{w}\|_2) \\
&\le (1-\delta_{2K})^{-\frac{1}{2}}((1+\delta_{2K})^{\frac{1}{2}}(\|(\mathbf{x}^{\lceil\frac{6K}{M}\rceil})_K - \mathbf{x}^{\lceil\frac{6K}{M}\rceil}\|_2 + \|\mathbf{x}^{\lceil\frac{6K}{M}\rceil} - \mathbf{x}\|_2) + 2\|\mathbf{w}\|_2) \\
&\overset{(a)}{\le} (1-\delta_{2K})^{-\frac{1}{2}}(2(1+\delta_{2K})^{\frac{1}{2}}\|\mathbf{x}^{\lceil\frac{6K}{M}\rceil} - \mathbf{x}\|_2 + 2\|\mathbf{w}\|_2)
\end{aligned}
$$

$$\begin{aligned}
&\overset{(b)}{\le} (1-\delta_{2K})^{-\frac{1}{2}}(2(1+\delta_{2K})^{\frac{1}{2}}(\xi_0+1)\|\mathbf{w}\|_2+2\|\mathbf{w}\|_2)\\
&\le (1-\delta_{8K})^{-\frac{1}{2}}(2(1+\delta_{8K})^{\frac{1}{2}}(\xi_0+1)\|\mathbf{w}\|_2+2\|\mathbf{w}\|_2)\\
&= 2(1-\delta_{8K})^{-\frac{1}{2}}((\xi_0+1)(1+\delta_{8K})^{\frac{1}{2}}+1)\|\mathbf{w}\|_2, \qquad \text{(A.4)}
\end{aligned}$$

where (a) is because $(\mathbf{x}^{\lceil \frac{6K}{M}\rceil})_K$ is the best K-term approximation of $\mathbf{x}^{\lceil \frac{6K}{M}\rceil}$, and (b) follows from (2.9). □

Appendix B

Lemma B.1 ([9], Lemma 3) Suppose that $\Lambda \subseteq \Omega$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ satisfy the RIP of order $|\Lambda|+1$. Then, for any $i \in \Omega\backslash\Lambda$,

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

Lemma B.2 ([22], Lemma 1) Let $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$ be two distinct vectors and let $W = \mathrm{supp}(\mathbf{u}) \cap \mathrm{supp}(\mathbf{v})$. Let U be the set of M indices corresponding to the M most significant elements in $\mathbf{u}$. For any integer $M \ge 1$, we have

$$\langle \mathbf{u},\mathbf{v}\rangle \le \Big(\lceil \frac{|W|}{M}\rceil\Big)^{\frac{1}{2}}\|\mathbf{u}_U\|_2\|\mathbf{v}_W\|_2.$$

Lemma B.3 Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a matrix with ℓ_2-normalized columns. According to Lemma 2.1, for the $(\ell+1)$-th ($\ell \ge k$) iteration of MOLS, we have

$$\sum_{i\in S^{\ell+1}} |\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2 \ge (1-\delta_{|T^\ell|+1}^2)\|\mathbf{A}_{S^*}'\mathbf{r}^\ell\|_2^2,$$

where

$$S^* = \underset{S:|S|=M}{\arg\max}\, \|\mathbf{A}_S'\mathbf{r}^\ell\|_2^2. \qquad \text{(B.1)}$$

Proof Note that this proof technique is similar in spirit to the work of [23, Lemma 8]. By Lemma 2.1, we have

$$\begin{aligned}
\sum_{i\in S^{\ell+1}} |\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2 &\overset{(a)}{\ge} \min_{i\in S^{\ell+1}}\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2 \sum_{i\in S^{\ell+1}} \frac{|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2}\\
&\overset{(b)}{\ge} (1-\delta_{|T^\ell|+1}^2)\sum_{i\in S^{\ell+1}} \frac{|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2} \overset{(c)}{=} (1-\delta_{|T^\ell|+1}^2)\max_{S:|S|=M}\sum_{i\in S} \frac{|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2}\\
&\ge (1-\delta_{|T^\ell|+1}^2)\sum_{i\in S^*} \frac{|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2} \overset{(d)}{\ge} (1-\delta_{|T^\ell|+1}^2)\sum_{i\in S^*}|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2\\
&= (1-\delta_{|T^\ell|+1}^2)\max_{S:|S|=M}\|\mathbf{A}_S'\mathbf{r}^\ell\|_2^2 \overset{(e)}{=} (1-\delta_{|T^\ell|+1}^2)\|\mathbf{A}_{S^*}'\mathbf{r}^\ell\|_2^2,
\end{aligned}$$

where (a) is because

$$\sum_{i\in S^{\ell+1}} \frac{|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2} \le \frac{\sum\limits_{i\in S^{\ell+1}} |\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{\min\limits_{i\in S^{\ell+1}}\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2},$$

(b) follows from Lemma B.1, (c) is from Lemma 2.1, (d) is according to $\|\mathbf{P}_{T^\ell}^\perp \mathbf{A}_i\|_2^2 \le \|\mathbf{A}_i\|_2^2 = 1$, and (e) is based on (B.1). □

Appendix C

Lemma C.4 For the $(\ell+1)$-th $(\ell \geq k)$ iteration of the MOLS Algorithm, the residual of MOLS has $\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^{\ell+1}\|_2^2 \geq \frac{1-\delta^2_{|T^\ell|+1}}{1+\delta_{|S^{\ell+1}|}}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2^2$.

Proof Note that this proof technique is similar in spirit to the work of [23, Lemma 8]. For any integer $\ell \geq k$, we have

$$\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^{\ell+1}\|_2^2 \overset{(a)}{\geq} \frac{\|\mathbf{A}'_{S^{\ell+1}}\mathbf{r}^\ell\|_2^2}{1+\delta_{|S^{\ell+1}|}} = \frac{\sum\limits_{i\in S^{\ell+1}}|\langle \mathbf{A}_i, \mathbf{r}^\ell\rangle|^2}{1+\delta_{|S^{\ell+1}|}} \overset{(b)}{\geq} \frac{(1-\delta^2_{|T^\ell|+1})}{1+\delta_{|S^{\ell+1}|}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2^2}, \tag{C.1}$$

where (a) follows from [6, (E.4)], and (b) follows from Lemma B.3. □

Appendix D

Lemma D.5 Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a matrix with ℓ_2-normalized columns. For the $(\ell+1)$-th $(\ell \geq k)$ iteration of MOLS, we have

$$\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^{\ell+1}\|_2^2 \geq \frac{(1-\delta^2_{|T^\ell|+1})(1-\delta_{|F_j^k\cup T^\ell|})}{(1+\delta_{|S^{\ell+1}|})\lceil\frac{|F_j^k|}{M}\rceil}\left(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2\right). \tag{D.1}$$

Proof Note that this proof technique is similar in spirit to the work of [23, Lemma 8]. According to Lemma C.4, we only give that

$$\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2^2 \geq \frac{1-\delta_{|F_j^k\cup T^\ell|}}{\lceil\frac{|F_j^k|}{M}\rceil}\left(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2\right). \tag{D.2}$$

Let $\mathbf{u} = \mathbf{A}'\mathbf{r}^\ell$ and $\mathbf{v} \in \mathbb{R}^n$ be the vectors satisfying $\mathbf{v}_{T\cap T^k\cup F_j^k} = \mathbf{x}_{T\cap T^k\cup F_j^k}$ and $\mathbf{v}_{\Omega\backslash(T\cap T^k\cup F_j^k)} = 0$. Note that $\mathrm{supp}(\mathbf{u}) = \Omega\backslash T^\ell$, $\mathrm{supp}(\mathbf{v}) = T\cap T^k \cup F_j^k$ and $T^k \subseteq T^\ell$. Then we have $W = \mathrm{supp}(\mathbf{u}) \cap \mathrm{supp}(\mathbf{v}) = F_j^k\backslash T^\ell$. Recall that S^* contains the indices corresponding to the M most significant elements in $\mathbf{u} = \mathbf{A}'\mathbf{r}^\ell$. According to Lemma B.2, we have

$$\begin{aligned}\langle \mathbf{A}'\mathbf{r}^\ell, \mathbf{v}\rangle &\leq (\lceil\frac{|F_j^k\backslash T^\ell|}{M}\rceil)^{\frac{1}{2}}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2\|\mathbf{v}_{F_j^k\backslash T^\ell}\|_2 \leq (\lceil\frac{|F_j^k\backslash T^\ell|}{M}\rceil)^{\frac{1}{2}}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2\|\mathbf{v}_{\Omega\backslash T^\ell}\|_2\\ &\leq (\lceil\frac{|F_j^k|}{M}\rceil)^{\frac{1}{2}}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2\|\mathbf{v}_{\Omega\backslash T^\ell}\|_2.\end{aligned} \tag{D.3}$$

On the other hand, We observe further that

$$\begin{aligned}\langle \mathbf{A}'\mathbf{r}^\ell, \mathbf{v}\rangle &\overset{(a)}{=} \langle \mathbf{A}'\mathbf{r}^\ell, \mathbf{v}-\mathbf{x}^\ell\rangle = \langle \mathbf{r}^\ell, \mathbf{A}(\mathbf{v}-\mathbf{x}^\ell)\rangle\\ &= \frac{1}{2}(\|\mathbf{A}(\mathbf{v}-\mathbf{x}^\ell)\|_2^2 + \|\mathbf{r}^\ell\|_2^2 - \|\mathbf{r}^\ell - \mathbf{A}(\mathbf{v}-\mathbf{x}^\ell)\|_2^2)\\ &\overset{(b)}{=} \frac{1}{2}(\|\mathbf{A}(\mathbf{v}-\mathbf{x}^\ell)\|_2^2 + \|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2)\\ &\overset{(c)}{\geq} \|\mathbf{A}(\mathbf{v}-\mathbf{x}^\ell)\|_2\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2}\\ &\overset{(d)}{\geq} \sqrt{1-\delta_{|F_j^k\cup T^\ell|}}\|\mathbf{v}-\mathbf{x}^\ell\|_2\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2}\\ &\geq \sqrt{(1-\delta_{|F_j^k\cup T^\ell|})(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2)} \times \|(\mathbf{v}-\mathbf{x}^\ell)_{\Omega\backslash T^\ell}\|_2\\ &\overset{(e)}{=} \|\mathbf{v}_{\Omega\backslash T^\ell}\|_2\sqrt{(1-\delta_{|F_j^k\cup T^\ell|})(\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\backslash F_j^k} + \mathbf{w}\|_2^2)},\end{aligned} \tag{D.4}$$

where (a) is because $\mathrm{supp}(\mathbf{A}'\mathbf{r}^\ell) \cap \mathrm{supp}(\mathbf{x}^\ell) = \emptyset$ so that $\langle \mathbf{A}'\mathbf{r}^\ell, \mathbf{x}^\ell \rangle = 0$, (b) is according to

$$\begin{aligned}\mathbf{r}^\ell &= \mathbf{y} - \mathbf{A}\mathbf{x}^\ell = \mathbf{A}(\mathbf{x} - \mathbf{x}^\ell) + \mathbf{w} = \mathbf{A}(\mathbf{x} - \nu + \nu - \mathbf{x}^\ell) + \mathbf{w} \\ &= \mathbf{A}(\nu - \mathbf{x}^\ell + \mathbf{x}_{F^k \setminus F_j^k}) + \mathbf{w} = \mathbf{A}(\nu - \mathbf{x}^\ell) + \mathbf{A}\mathbf{x}_{F^k \setminus F_j^k} + \mathbf{w},\end{aligned}$$

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

Finally, by combining (D.3) with (D.4), we have

$$\begin{aligned}\|\mathbf{A}'_{S^*}\mathbf{r}^\ell\|_2 &\geq \frac{\langle \mathbf{A}'\mathbf{r}^\ell, \mathbf{v}\rangle}{\sqrt{\lceil \frac{|F_j^k|}{M} \rceil}\|\mathbf{v}_{\Omega \setminus T^\ell}\|_2} \geq \frac{\sqrt{1 - \delta_{|F_j^k \cup T^\ell|}}\|\mathbf{v}_{\Omega \setminus T^\ell}\|_2}{\sqrt{\lceil \frac{|F_j^k|}{M} \rceil}\|\mathbf{v}_{\Omega \setminus T^\ell}\|_2}\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k \setminus F_j^k} + \mathbf{w}\|_2^2} \\ &= \frac{\sqrt{1 - \delta_{|F_j^k \cup T^\ell|}}}{\sqrt{\lceil \frac{|F_j^k|}{M} \rceil}}\sqrt{\|\mathbf{r}^\ell\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k \setminus F_j^k} + \mathbf{w}\|_2^2}.\end{aligned}$$

This completes the proof. □

Appendix E

Proposition E.6 Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix. Let $\theta^k = |F^k| = |T \setminus T^k|$ be the index number of a remaining support set after running k ($k \geq 0$) iterations of MOLS. 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' indicates specified additional iterations after running k iterations and $L \in \{1, 2, \ldots, \max\{0, \lfloor \log_2 \frac{\theta^k + 1}{M} \rfloor\} + 1\}$. Then we have

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

where

$$\alpha = \sqrt{\frac{4\mu(1 - \mu)(1 + \delta_{Mk + 7\theta^k})}{(1 - \delta_{Mk + 7\theta^k})}}, \tag{E.1}$$

$$\beta = \frac{2}{\sqrt{1 - \delta_{Mk + 7\theta^k}}}, \tag{E.2}$$

and μ has been defined in (2.6).

Proof According to (D.1), let

$$\beta_\ell = \frac{(1 - \delta_{|T^\ell| + 1}^2)(1 - \delta_{|F_j^k \cup T^\ell|})}{(1 + \delta_{|S^{\ell+1}|})\lceil \frac{|F_j^k|}{M} \rceil}. \tag{E.3}$$

Then, (D.1) can be rewritten as

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

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

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

and for $\ell' > \ell \geq k$, we also have

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

$$\ldots,$$
$$\|\mathbf{r}^{\ell+2}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2 \le \exp(-\beta_{\ell+1})(\|\mathbf{r}^{\ell+1}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2). \tag{E.6}$$

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

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

where (a) holds, since β_η is non-increasing.

Let $\ell' = k + k_i$ and $\ell = k + k_{i-1}, i = 1, \ldots, L$. Then we have

$$\begin{aligned}&\|\mathbf{r}^{k+k_i}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\\ &\overset{(a)}{\le} \exp\bigg(-\frac{(k_i-k_{i-1})(1-\delta^2_{|T^{k+k_i-1}|+1})(1-\delta_{|F_j^k\cup T^{k+k_i-1}|})}{(1+\delta_{|S^{k+k_i}|})\lceil\frac{|F_j^k|}{M}\rceil}\bigg)\\ &\quad\times\Big(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\Big)\\ &\overset{(b)}{\le} \exp\bigg(-\frac{(k_i-k_{i-1})(1-\delta^2_{|T^{k+k_i-1}|+1})(1-\delta_{|F_i^k\cup T^{k+k_i-1}|})}{(1+\delta_{|S^{k+k_i}|})\lceil\frac{|F_i^k|}{M}\rceil}\bigg)\\ &\quad\times\Big(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\Big)\\ &\overset{(c)}{\le} \exp\bigg(-\frac{\frac{3}{2}(1-\delta^2_{|T^{k+k_i-1}|+1})(1-\delta_{|F_i^k\cup T^{k+k_i-1}|})}{(1+\delta_{|S^{k+k_i}|})}\bigg)\times\Big(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\Big)\\ &\overset{(d)}{\le} \exp\bigg(-\frac{\frac{3}{2}(1-\delta^2_{Mk+7\theta^k})(1-\delta_{Mk+7\theta^k})}{(1+\delta_{Mk+7\theta^k})}\bigg)\times\Big(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\Big)\\ &= e^{-\frac{3}{2}(1-\delta_{Mk+7\theta^k})^2}\Big(\|\mathbf{r}^{k+k_{i-1}}\|_2^2 - \|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\Big),\end{aligned} \tag{E.7}$$

where (a) is due to (E.3), (b) is from $j \le i$, (c) is true since (3.10), (d) is because

$$\begin{aligned}|F_i^k \cup T^{k+k_i-1}| &\overset{(3.11)}{\le} |T \cup T^{k+k'}| = |T^{k+k'}| + |F^{k+k'}| \le M(k+k') + |F^k|\\ &\overset{(3.12)}{=} Mk + (3\times 2^L - 3)M + \theta^k = Mk + 3\times 2^L M + \theta^k - 3M\\ &\overset{(3.2)}{<} Mk + 3(2\theta^k+1) + \theta^k - 3M = Mk + 7\theta^k + 3 - 3M \le Mk + 7\theta^k,\\ |T^{k+k_i-1}| + 1 \le |T^{k+k'}| + 1 &\le M(k+k') + 1 \overset{(3.12)}{=} Mk + (3\times 2^L - 3)M + 1\\ &= Mk + 3\times 2^L M + 1 - 3M \overset{(3.2)}{<} Mk + 3(2\theta^k+1) + 1 - 3M\\ &= Mk + 6\theta^k + 4 - 3M \le Mk + 6\theta^k + 1 \le Mk + 7\theta^k,\end{aligned} \tag{E.8}$$

and $|S^{k+k_i}| = M < Mk + 7\theta^k$.

According to (2.6), (E.7) can be rewritten as

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

where $i = 1, \ldots, L$.

Note that $k_0 = 0$. Then we have

$$\|\mathbf{r}^{k+k_L}\|_2^2 \le \mu\|\mathbf{r}^{k+k_{L-1}}\|_2^2 + (1-\mu)\|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2, \tag{E.9}$$

$$\begin{aligned}&\dots,\\ &\|\mathbf{r}^{k+k_1}\|_2^2 \le \mu\|\mathbf{r}^k\|_2^2 + (1-\mu)\|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2.\end{aligned} \tag{E.10}$$

From (E.9)–(E.10), we get that

$$\begin{aligned}\|\mathbf{r}^{k+k_L}\|_2^2 &\le \mu^L\|\mathbf{r}^k\|_2^2 + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\\ &\overset{(a)}{\le} \mu^L\|\mathbf{A}\mathbf{x}_{F^k} + \mathbf{w}\|_2^2 + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k} + \mathbf{w}\|_2^2\\ &\le \mu^L(\|\mathbf{A}\mathbf{x}_{F^k}\|_2 + \|\mathbf{w}\|_2)^2 + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\times(\|\mathbf{A}\mathbf{x}_{F^k\setminus F_j^k}\|_2 + \|\mathbf{w}\|_2)^2,\end{aligned} \tag{E.11}$$

where (a) is due to $\|\mathbf{r}^k\|_2^2 \le \|\mathbf{A}\mathbf{x}_{F^k} + \mathbf{w}\|_2^2$, which is from Proposition 1 in [22].

According to the RIP and $F_0^k = \emptyset$, (E.11) can be changed into

$$\begin{aligned}\|\mathbf{r}^{k+k_L}\|_2^2 &\le \mu^L\Big(\sqrt{1+\delta_{|F^k|}}\|\mathbf{x}_{F^k\setminus F_0^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &\quad + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\Big(\sqrt{1+\delta_{|F^k\setminus F_j^k|}}\|\mathbf{x}_{F^k\setminus F_j^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &\overset{(a)}{\le} \mu^L\Big(\sqrt{1+\delta_{\theta^k}}\|\mathbf{x}_{F^k\setminus F_0^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &\quad + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\Big(\sqrt{1+\delta_{\theta^k}}\|\mathbf{x}_{F^k\setminus F_j^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &\overset{(b)}{<} \mu^L\Big(\sqrt{(1+\delta_{\theta^k})\tau^{L-1}}\|\mathbf{x}_{F^k\setminus F_{L-1}^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &\quad + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\Big(\sqrt{(1+\delta_{\theta^k})\tau^{L-j-1}}\|\mathbf{x}_{F^k\setminus F_{L-1}^k}\|_2 + \|\mathbf{w}\|_2\Big)^2\\ &= \Big((\tau\mu)^L + (1-\mu)\sum_{j=1}^{L}(\tau\mu)^{L-j}\Big)\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\setminus F_{L-1}^k}\|_2^2\\ &\quad + \Big(\mu^L + (1-\mu)\sum_{j=1}^{L}\mu^{L-j}\Big)\|\mathbf{w}\|_2^2 + 2\Big((\sqrt{\tau}\mu)^L + (1-\mu)\sum_{j=1}^{L}(\sqrt{\tau}\mu)^{L-j}\Big)\\ &\quad \times\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\setminus F_{L-1}^k}\|_2\|\mathbf{w}\|_2,\end{aligned} \tag{E.12}$$

where (a) is according to $|F^k\setminus F_j^k| < |F^k| = \theta^k$ for $j = 1,\dots,L$, and (b) is from (3.7).

Note that

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

when $\tau > 1, \tau\mu < 1$, and $\mu < 1$. Thus, (E.12) can be changed into

$$\begin{aligned}
\|\mathbf{r}^{k+k_L}\|_2^2 <& \Big((1-\mu)\sum_{j=L}^{\infty}(\tau\mu)^j + (1-\mu)\sum_{j=0}^{L-1}(\tau\mu)^j\Big)\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2^2 \\
&+ \Big((1-\mu)\sum_{j=L}^{\infty}\mu^j + (1-\mu)\sum_{j=0}^{L-1}\mu^j\Big)\|\mathbf{w}\|_2^2 \\
&+ 2\Big((1-\mu)\sum_{j=L}^{\infty}(\sqrt{\tau}\mu)^{L-j} + (1-\mu)\sum_{j=1}^{L}(\sqrt{\tau}\mu)^{L-j}\Big)\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2\|\mathbf{w}\|_2 \\
=& \frac{1-\mu}{1-\tau\mu}\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2^2 + \|\mathbf{w}\|_2^2 + 2\frac{1-\mu}{1-\sqrt{\tau}\mu}\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2\|\mathbf{w}\|_2^2 \\
\overset{(a)}{<}& \frac{1-\mu}{1-\tau\mu}\frac{1+\delta_{\theta^k}}{\tau}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2^2 + \|\mathbf{w}\|_2^2 + 2\sqrt{\frac{1-\mu}{1-\tau\mu}}\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2\|\mathbf{w}\|_2^2 \\
=& \Big(\sqrt{\frac{1-\mu}{1-\tau\mu}}\sqrt{\frac{1+\delta_{\theta^k}}{\tau}}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2 + \|\mathbf{w}\|_2\Big)^2 \\
\overset{(b)}{\le}& \Big(\sqrt{4\mu(1-\mu)(1+\delta_{Mk+7\theta^k})}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2 + \|\mathbf{w}\|_2\Big)^2,
\end{aligned} \tag{E.13}$$

where (a) is from

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

(b) chooses $\tau = \frac{1}{2\mu}$, and $\theta^k = |T\setminus T^k| \le |T\cup T^{k+k'}| < Mk + 7\theta^k$.

Since $k + k_L \overset{(3.11)}{\le} k + k'$, we have that

$$\|\mathbf{r}^{k+k'}\|_2 \le \|\mathbf{r}^{k+k_L}\|_2 < \sqrt{4\mu(1-\mu)(1+\delta_{Mk+7\theta^k})}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2 + \|\mathbf{w}\|_2. \tag{E.14}$$

On the other hand, we have that

$$\begin{aligned}
\|\mathbf{r}^{k+k'}\|_2 &= \|\mathbf{y} - \mathbf{A}\mathbf{x}^{k+k'}\|_2 = \|\mathbf{A}(\mathbf{x}-\mathbf{x}^{k+k'}) + \mathbf{w}\|_2 \\
&\ge \|\mathbf{A}(\mathbf{x}-\mathbf{x}^{k+k'})\|_2 - \|\mathbf{w}\|_2 \overset{(a)}{\ge} \sqrt{1-\delta_{|T\cup T^{k+k'}|}}\|\mathbf{x}-\mathbf{x}^{k+k'}\|_2 - \|\mathbf{w}\|_2 \\
&\ge \sqrt{1-\delta_{|T\cup T^{k+k'}|}}\|\mathbf{x}_{F^{k+k'}}\|_2 - \|\mathbf{w}\|_2 \overset{(b)}{\ge} \sqrt{1-\delta_{Mk+7\theta^k}}\|\mathbf{x}_{F^{k+k'}}\|_2 - \|\mathbf{w}\|_2,
\end{aligned} \tag{E.15}$$

where (a) is due to the RIP, and (b) follows from (E.8).

By relating (E.14) and (E.15), we obtain that

$$\begin{aligned}
\|\mathbf{x}_{F^{k+k'}}\|_2 &\le \frac{1}{\sqrt{1-\delta_{Mk+7\theta^k}}}(\|\mathbf{r}^{k+k'}\|_2 + \|\mathbf{w}\|_2) \\
&< \sqrt{\frac{4\mu(1-\mu)(1+\delta_{Mk+7\theta^k})}{(1-\delta_{Mk+7\theta^k})}}\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2 + \frac{2}{\sqrt{1-\delta_{Mk+7\theta^k}}}\|\mathbf{w}\|_2 \\
&= \alpha\|\mathbf{x}_{F^k\setminus F^k_{L-1}}\|_2 + \beta\|\mathbf{w}\|_2,
\end{aligned} \tag{E.16}$$

where α and β are as defined in (E.1) and (E.2), respectively. $\square$

References

[1] Chen S, Billings S, Luo W. Orthogonal least squares methods and their application to non-linear system identification. Int J Control, 1989, **50**(5): 1873–1896

[2] Foucart S. Stability and robustness of weak orthogonal matching pursuits//Recent Adv Harmonic Anal and App. Springer, 2013: 395–405

[3] Geng P, Chen W, Ge H. Perturbation analysis of orthogonal least squares. Canadian Mathematical Bulletin, 2019, **62**(4): 780–797

[4] Herzet C, Soussen C, Idier J, Gribonval R. Exact recovery conditions for sparse representations with partial support information. IEEE Trans Inform Theory, 2013, **59**(11): 7509–7524

[5] Herzet C, Drémeau A, Soussen C. Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay. IEEE Transactions on Information Theory, 2015, **62**(1): 459–470

[6] Wang J, Li P. Recovery of sparse signals using multiple orthogonal least squares. IEEE Trans. Signal Process, 2017, **65**(8): 2049–2062

[7] Li H, Zhang J, Zou J. Improving the bound on the restricted isometry property constant in multiple orthogonal least squares. IET Signal Processing, 2018, **12**(5): 666–671

[8] Kim J, Shim B. A Near-Optimal Restricted Isometry Condition of Multiple Orthogonal Least Squares. IEEE Access, 2019, **7**: 46822–46830

[9] Wen J, Wang J, Zhang Q. Nearly optimal bounds for orthogonal least squares. IEEE Trans Signal Process, 2017, **65**(20): 5347–5356

[10] Chen W, Li Y. Stable recovery of signals with the high order D-RIP condition. Acta Matematica Scientia, 2016, **36**(6): 1721–1730

[11] Abdillahi-Ali D, Azzaout N, Guillin A, Le Mailloux G, Matsui T. Penalized least square in sparse setting with convex penalty and non Gaussian errors. Acta Matematica Scientia, 2021, **41**(6): 2198–2216

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

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

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

[15] Dan W, Wang R. Robustness of orthogonal matching pursuit under restricted isometry property. Science China Mathematics, 2014, **57**(3): 627–634

[16] Zhang T. Sparse recovery with orthogonal matching pursuit under RIP. IEEE Transactions on Information Theory, 2011, **57**(9): 6215–6221

[17] Wang J, Li P, Shim B. Exact recovery of sparse signals using orthogonal matching pursuit: How many iterations do we need?. IEEE Transactions on Signal Processing, 2016, **64**(16): 4194–4202

[18] Candes E, Tao T. Decoding by linear programming. IEEE Trans Inf Theory, 2005, **51**(12): 4203–4215

[19] Baraniuk R, Davenport M, DeVore R, Wakin M. A simple proof of the restricted isometry property for random matrices. Construct Approx, 2008, **28**(3): 253–263

[20] Blumensath T, Davies M. On the difference between orthogonal matching pursuit and orthogonal least squares//Technical Report. Southampton: University of Southampton, 2007. https://eprints.soton.ac.uk/142469/1/BDOMPvsOLS07.pdf.

[21] Soussen C, Gribonval R, Idier J, Herzet C. Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares. IEEE Trans Inf Theory, 2013, **59**(5): 3158–3174

[22] Wang J, Kwon S, Li P, Shim B. Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis. IEEE Transactions on Signal Process, 2016, **64**(4): 1076–1089

[23] Kim J, Wang J, Nguyen L, Shim B. Joint sparse recovery using signal space matching pursuit. IEEE Transactions on Information Theory, 2020, **66**(8): 5072–5096

[24] Kim J, Wang J, Shim B. Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares. IEEE Transaction on Signal Processing, 2021, **69**(1): 1521–1536