THE PERFORMANCE OF ORTHOGONAL MULTI-MATCHING PURSUIT UNDER RIP

ZHIQIANG XU

Abstract. The orthogonal multi-matching pursuit (OMMP) is a natural extension of the orthogonal matching pursuit (OMP). We denote the OMMP with the parameter \( M \) as OMMP(\( M \)) where \( M \geq 1 \) is an integer. The main difference between OMP and OMMP(\( M \)) is that OMMP(\( M \)) selects \( M \) atoms per iteration, while OMP only adds one atom to the optimal atom set. In this paper, we study the performance of orthogonal multi-matching pursuit under RIP. In particular, we show that, when the measurement matrix \( A \) satisfies \((9s, 1/10)-RIP\), there exists an absolute constant \( M_0 \leq 8 \) so that OMMP(\( M_0 \)) can recover \( s \)-sparse signal within \( s \) iterations. We furthermore prove that OMMP(\( M \)) can recover \( s \)-sparse signal within \( O(\frac{s}{M}) \) iterations for a large class of \( M \) provided the signal is slowly-decaying. In particular, for \( M = s^a \) with \( a \in [0, 1/2] \), OMMP(\( M \)) can recover slowly-decaying \( s \)-sparse signals within \( O(s^{1-a}) \) iterations. The result implies that OMMP can reduce the computational complexity heavily.

1. Introduction

1.1. Orthogonal Matching Pursuit. Orthogonal matching pursuit (OMP) is a popular algorithm for the recovery of sparse signals and it is also commonly used in compressed sensing. Let \( A \) be a matrix of size \( m \times N \) and \( y \) be a vector of size \( m \). The aim of OMP is to find the approximate solution to the following \( \ell_0 \)-minimization problem:

\[
\min_{x \in \mathbb{C}^N} \|x\|_0 \quad \text{s.t.} \quad Ax = y,
\]

where \( \|x\|_0 \) denotes the number of non-zero entries in \( x \). In compressed sensing and the sparse representation of signals, we often have \( m \ll N \). Throughout this paper, we suppose that the sampling matrix \( A \in \mathbb{C}^{m \times N} \) whose columns \( a_1, \ldots, a_N \) are \( \ell_2 \)-normalized.

To introduce the performance of OMP, we first recall the definition of the restricted isometry property (RIP) [6] which is frequently used in the analysis of the recovering algorithm in compressed sensing. Following Candès and Tao, for \( 1 \leq s \leq N \) and \( \delta \in [0, 1) \), we say that the matrix \( A \) satisfies \((s, \delta)-\text{RIP}\) if

\[
(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2
\]

holds for all \( s \)-sparse signals \( x \). We say that the signal \( x \) is \emph{\( s \)-sparse} if \( \|x\|_0 \leq s \) and use \( \Sigma_s \) to denote the set of \( s \)-sparse signals, i.e.,

\[
\Sigma_s = \{ x \in \mathbb{C}^N : \|x\|_0 \leq s \}.
\]

We next state the definition of the spark (see also [1]).

Supported by the National Natural Science Foundation of China (11171336).
Definition 1. The spark of a matrix $A$ is the size of the smallest linearly dependent subset of columns, i.e.,

$$\text{Spark}(A) := \min \{ \|x\|_0 : Ax = 0, x \neq 0 \}.$$ 

Theoretical analysis of OMP has concentrated primarily on two directions. The first one is to study the condition for the matrix $A$ under which OMP can recover $s$-sparse signals in exactly $s$ iterations. In this direction, one uses the coherence and RIP to analyze the performance of OMP. In particular, Davenport and Wakin showed that, when the matrix $A$ satisfies $(s + 1, \sqrt{s})$-RIP, OMP can recover $s$-sparse signal in exactly $s$ iterations [8]. The sufficient condition is improved to $(s + 1, 1/\sqrt{s})$-RIP in [12, 13] (see also [10, 11]). However, it was observed in [16], when the matrix $A$ satisfies $(c_0 s, \delta_0)$-RIP for some fixed constants $c_0 > 1$ and $0 < \delta_0 < 1$, that $s$ iterations of OMP is not enough to uniformly recover $s$-sparse signals, which implies that OMP has to run for more than $s$ iterations to uniformly recover the $s$-sparse signals. Hence, one investigates the performance of OMP along the second line with allowing to OMP run more than $s$ iterations. For this case, it is possible that OMP add wrong atoms to the optimal atom set, but one can identify the correct atoms by the least square. A main result in this direction is presented by Zhang [20] with proving that when $A$ satisfies $(31s, 1/3)$-RIP OMP can recover the $s$-sparse signal in at most $30s$ iterations.

The other type of greedy algorithms, which are based on OMP, have been proposed including the regularized orthogonal matching pursuit (ROMP) [14], subspace pursuit (SP) [7], CoSaMP [15], and many other variants. For each of these algorithms, it has been shown that, under a natural RIP setting, they can recover the $s$-sparse signals in $s$ iterations.

1.2. Orthogonal Multi-matching Pursuit and Main Results. A more natural extension of OMP is the orthogonal multi-matching pursuit (OMMP) [11]. We denote the OMMP with the parameter $M$ as OMMP($M$) where $M \geq 1$ is an integer. The main difference between OMP and OMMP($M$) is that OMMP($M$) selects $M$ atoms per iteration, while OMP only adds one atom to the optimal atom set. The Algorithm 1 outlines the procedure of OMMP($M$) with initial feature set $\Lambda^0$.

In comparison with OMP, OMMP has fewer iterations and computational complexity [10]. We note that, when $M = 1$, OMMP($M$) is identical to OMP. OMMP is also studied in [10, 12, 18] under the names of KOMP, MOMP and gOMP, respectively. These results show that, when RIP constant $\delta = O(\sqrt{M}/s)$, OMMP($M$) can recover the $s$-sparse signal in at most $30s$ iterations.

The aim of this paper is to study the performance of OMMP($M$) under a more natural setting of RIP (the RIP constant is an absolute constant). Particularly, we also would like to understand the relation between the number of iterations and the parameter $M$. So, we are interested in the following questions:

Question 1 Does there exist an absolute constant $M_0$ so that OMMP($M_0$) can recover all the $s$-sparse signals within $s$ iterations?

Question 2 For $1 \leq M \leq s$, can OMMP($M$) recover the $s$-sparse signals within $O(s/M)$ iterations?

We next state one of our main results which gives an affirmative answer to Question 1.
Algorithm 1 OMMP($M$)

**Input:** sampling matrix $A$, samples $y = Ax$, candidate number $M$ for each step, stopping iteration index $H$, initial feature set $\Lambda^0 \subset \{1, \ldots, N\}$

**Output:** the $x^*$.

**Initialize:** $\ell = 0$.

$x^0 = \arg\min_{z: \text{supp}(z) \subset \Lambda^0} \| y - Az \|_2$, $r^0 = y - Ax^0$

while $\ell < H$

  match: $h^\ell = A^T r^\ell$

  calculate: $T^\ell = M$ indices corresponding to the largest magnitude entries in the vector $h^\ell$

  identity: $\Lambda^{\ell+1} = \Lambda^\ell \cup T^\ell$

  update: $x^{\ell+1} = \arg\min_{z: \text{supp}(z) \subset \Lambda^{\ell+1}} \| y - Az \|_2$

  $r^{\ell+1} = y - Ax^{\ell+1}$

  $\ell = \ell + 1$

end while

$x^* = x^H$

**Theorem 1.** Let $x \in \Sigma_s$ and $S = \text{supp}(x)$. Suppose that the sampling matrix $A \in \mathbb{C}^{m \times N}$ satisfies $(9s, 1/10)$-RIP and $\text{Spark}(A) > \max\{Ms', 8s'\} + \#\Lambda^0$ where $\Lambda^0$ is the initial feature set in OMMP algorithm. Then OMMP($M$) can recover the signal $x$ within, at most, $\max\{s', (4M + 2)s'\}$ iterations, where $s' := \#(S \setminus \Lambda^0)$.

The above theorem shows that, when $M \geq 8$, OMMP($M$) with the initial feature set $\Lambda^0 = \emptyset$ can recover all the $s$-sparse signal within, at most, $s$ iterations. It implies that there exists an absolute constant $M_0 \leq 8$ so that OMMP($M_0$) can recover all the $s$-sparse signals within $s$ iterations. We believe that the constant $M_0 = 8$ is not optimal. The numerical experiments make us conjecture that the optimal number is 2, i.e., under RIP, OMMP(2) can recover the $s$-sparse signal within $s$ iterations.

We next turn to Question 2. The following theorem shows that, when $1 \leq M \leq \sqrt{s}$, OMMP($M$) can recover slowly-decaying signal within $O(s/M)$ iterations.

**Theorem 2.** Let $x \in \Sigma_s$, $S = \text{supp}(x)$ and $s' = \#(S \setminus \Lambda^0)$. Consider the OMMP($M$) algorithm with $1 \leq M \leq \sqrt{s}$ and the initial feature set $\Lambda^0$. If the sampling matrix $A \in \mathbb{C}^{m \times N}$ satisfies $(9s, 1/10)$-RIP and

$$\text{Spark}(A) > 8(C_0^2 + 2)s' + \#\Lambda^0,$$

then OMMP($M$) recovers the $x$ within $\lfloor 8(C_0^2 + 2)s'/M \rfloor$ iterations where $C_0 = \max_{j \in S} |x_j| / \min_{j \in S} |x_j|.$

The theorem above shows that, for $1 \leq M \leq \sqrt{s}$, OMMP($M$) can recover $s$-sparse signals within $C_1 s/M$ iterations. Here, the constant $C_1$ depends on the signal $x$. In particular, if we take $M = \lfloor s/\alpha \rfloor$ in Theorem 2 we have

**Corollary 1.** Under the condition of Theorem 2 if $M = \lfloor s/\alpha \rfloor$ with $\alpha \in [0, 1/2]$, then OMMP($M$) with the initial feature set $\Lambda^0 = \emptyset$ recovers the $s$-sparse signal within $\lfloor 8(C_0^2 + 2)s^{1-\alpha} \rfloor$ iterations.

We next consider the case with $M = \alpha \cdot s$. In particular, for ‘small’ $\alpha$, we give an affirmative answer to Question 2 up to a log factor.
Theorem 3. Let \( x \in \Sigma_s \) and \( S = \text{supp}(x) \). Suppose that the sampling matrix \( A \in \mathbb{C}^{m \times N} \) satisfies \((14s, 1/10)\)-RIP and
\[
\text{Spark}(A) > 8s \log_2(2(s + 1)).
\]
Consider the OMMP\((M)\) algorithm with the initial feature set \( \Lambda^0 = \emptyset \). If \( M = \alpha \cdot s \), then OMMP\((M)\) recover the \( s \)-sparse signal \( x \) from \( y = Ax \) within \( \lceil \frac{8}{\alpha} \log_2(2(s + 1)) \rceil \) iterations, where \( 0 < \alpha \leq \frac{2}{(C_0^2 + 2)} \) and \( C_0 = \max_{j \in S} |x_j| / \min_{j \in S} |x_j| \).

Remark 1. We prove the main results using some of the techniques developed by Zhang in his study of OMP [20] (see also [9]). To make the paper more readable, we state our results for the strictly sparse signal. In fact, using a similar method, one also can extend the results in this paper to the case where the measurement vector \( y \) is subjected to an additive noise and \( x \) is not strictly sparse.

Remark 2. In [11], Liu and Tymlyakov proved that, when \( A \) satisfies \((M_0, \delta)\)-RIP with \( \delta = \sqrt{M_0 / ((2 + \sqrt{2}) \sqrt{s})} \), OMMP\((M_0)\) can recover \( s \)-sparse signal within, at most, \( s \) iterations. The result requires the RIP constant \( \delta \) depends on \( s = \|x\|_0 \). In Theorem 1 we require that the measurement matrix \( A \) satisfies \((9s, \delta)\)-RIP with \( \delta \) being an absolute constant 1/10. Hence, Theorem 1 gives an affirmative answer to Question 1 under the more natural setting for the measurement matrix \( A \).

Remark 3. It is of interest to know which matrices \( A \) obey the \((s, \delta)\)-RIP and the \( \text{Spark}(A) > K \) where \( K \) is a fixed constant. Much is known about finding matrices that satisfy the \((s, \delta)\)-RIP (see [2, 4, 5, 17, 19]). If we draw a random \( m \times N \) matrix \( A \) whose entries are i.i.d. Gaussian random variables, then \( \text{Spark}(A) = m \) with probability 1 (see [1, 3]). Moreover, the random matrix \( A \) also satisfies \((s, \delta)\)-RIP with high probability provided
\[
m = O \left( \frac{s \log(N/s)}{\delta^2} \right).
\]
So, to make the random matrices \( A \) obey the \((s, \delta)\)-RIP and the \( \text{Spark}(A) > K \), one can take
\[
m = \max \left\{ O \left( \frac{s \log(N/s)}{\delta^2} \right), K + 1 \right\}.
\]

2. Numerical experiments

The purpose of the experiment is the comparison for the reconstruction performances of and the iteration number of OMMP\((M)\) with different parameter \( M \). Given the parameters \( m = 300 \) and \( N = 1,500 \), we randomly generate a \( m \times N \) sampling matrix \( A \) from the standard i.i.d Gaussian ensemble. The support set \( S \) of the sparse signal \( x \) is drawn from the uniform distribution over the set of all subsets of \([1, N] \cap \mathbb{Z}\) of size \( s \). We then generate the sparse signal \( x \) according to the probability model: the entries \( x_j, j \in S \), are independent random variable having the Gaussian distribution with mean 5 and standard deviation 1.

We apply the OMMP\((M)\) to recover the sparse signal \( x \) from \( y = Ax \) for different parameters \( M \in \{1, \lceil \sqrt{s} \rceil, \lceil s \rceil \} \). Note that when \( M = 1 \), OMMP\((M)\) is identical with OMP. We repeat the experiment 200 times for each number \( s \in \{1, 2, \ldots, 80\} \) and calculate the success rate. When OMMP succeeds, we record the number of the iteration steps. The left graph in Fig. 1 depicts the success rate of the
reconstructing algorithm OMMP($M$) with $M \in \{1, \lceil \sqrt{s} \rceil, \lfloor \frac{s}{2} \rfloor \}$. The number of the average iteration steps of OMMP($M$) with $M \in \{1, \lceil \sqrt{s} \rceil, \lfloor \frac{s}{2} \rfloor \}$ are illustrated in the right graph in Fig. 1. The numerical results show that the performance of OMMP($M$), $M \in \{1, \lceil \sqrt{s} \rceil, \lfloor \frac{s}{2} \rfloor \}$, is similar with that of OMP, while the number of iteration steps of OMMP($M$), $M \in \{1, \lceil \sqrt{s} \rceil, \lfloor \frac{s}{2} \rfloor \}$, is far less than that of OMP, which agrees with the theoretical results presented in this paper.

3. Extension

According to Theorem 2 and Theorem 3, OMMP has a good performance for the slowly-decaying sparse signal $x$. Naturally, one may want to know whether OMMP($M$) can recover all the $s$-sparse signal within less than $s$ iterations for some $M \in [1, s] \cap \mathbb{Z}$. Numerical experiments show that, for some fast-decaying $s$-sparse signal $x$, OMMP($M$) has to run at least $s$ steps to recover $x$ for any $M \in [1, s] \cap \mathbb{Z}$. However, as shown in [8], when the $s$-sparse signal $x$ is fast-decaying, OMP has a good performance. To state the result in [8], we firstly introduce the definition of $\alpha$-decaying signals. For any $s$-sparse signal $x \in \mathbb{C}^N$, we denote by $S$ the support of $x$. Without loss of generality, we suppose that $S = \{j_1, \ldots, j_s\}$ and

$$|x_{j_1}| \geq |x_{j_2}| \geq \cdots \geq |x_{j_s}| > 0.$$ 

For $\alpha > 1$, we call the $x$ $\alpha$-decaying if $|x_{j_t}|/|x_{j_{t+1}}| \geq \alpha$ for all $t \in \{1, 2, \ldots, s-1\}$.

**Theorem 4.** ([8]) Suppose that $A$ satisfies $(s+1, \delta_{s+1})$-RIP with $\delta_{s+1} < \frac{1}{3}$. Suppose that $x$ with $\|x\|_0 \leq s$ is $\alpha$-decaying signal. If

$$\alpha > \frac{1 + 2 \cdot \delta_{s+1}}{1 - \delta_{s+1}} \sqrt{s - 1},$$

then OMP will recover $x$ exactly from $y = Ax$ in $s$ iterations.

In this paper, motivated by the proof of Theorem 1, we can improve Theorem 4 as follows:
Theorem 5. Suppose that \( A \) satisfies \((s, \delta_s)\)-RIP with \( \delta_s < \sqrt{2} - 1 \). Suppose that \( x \in \mathbb{C}^N \) with \( \|x\|_0 \leq s \) is \( \alpha \)-decaying. If

\[
\alpha > \sqrt{\frac{1 + \delta_s}{2 - (1 + \delta_s)^2}},
\]

then OMP can recover \( x \) exactly from \( y = Ax \) in \( s \) iterations.

Remark 4. In Theorem 4, the right side of (2) depends on RIP constant and \( s = \|x\|_0 \), while in Theorem 5, the right side of (3) only depends on the RIP constant. So, Theorem 5 is an improvement over Theorem 4.

Appendix A. Lemmas

In this section, we introduce many lemmas, which extend some results in \([9]\). To state conveniently, for any set \( T \subset \{1, \ldots, N\} \) of column indices, we denote by \( A_T \) the \( m \times \#T \) matrix composed of these columns. Similarly, for a vector \( x \in \mathbb{C}^N \), we use \( x_T \) to denote the vector formed by the entries of \( x \) with indices from \( T \). For \( u \in \mathbb{C}^N \) and \( t \in \mathbb{Z}_+ \), we extend the \( \ell_1 \)-norm to a generalized \( \ell_1 \)-norm defined as

\[
\|u\|_{t,1} := \sum_{j=0}^{\lfloor N/t \rfloor - 1} \sqrt{u_{jt+1}^2 + \cdots + u_{(j+1)t}^2} + \sqrt{u_{n_0 t+1}^2 + \cdots + u_N^2}.
\]

Similarly, we also can extend the \( \ell_{\infty} \)-norm as follows

\[
\|u\|_{t,\infty} := \max \left\{ \max_{0 \leq j \leq \lfloor N/t \rfloor - 1} \sqrt{u_{jt+1}^2 + \cdots + u_{(j+1)t}^2} \right\}.
\]

Then the following lemma presents some inequalities for the extension norm:

Lemma 1. Suppose that \( u \in \mathbb{C}^N \), \( v \in \mathbb{C}^N \) and \( t \in \mathbb{Z}_+ \). Then

(i) \( \Re(\langle u, v \rangle) \leq \|u\|_{t,\infty} \cdot \|v\|_{t,1} \),

where \( \Re(\cdot) \) denotes the real part;

(ii) \( \|u\|_{t,1}^2 \leq \left\lfloor \frac{N}{t} \right\rfloor \cdot \|u\|_2^2 \).

Proof. To state conveniently, we set \( T_j := \{j \cdot t, \ldots, j \cdot t + t\}, \) \( j = 0, \ldots, n_0 - 1 \) and \( T_{n_0} := \{n_0 t + 1, \ldots, N\} \), where \( n_0 = \left\lfloor \frac{N}{t} \right\rfloor \). Then

\[
\Re(\langle u, v \rangle) = \sum_{j=0}^{n_0} \Re(\langle u_{T_j}, v_{T_j} \rangle)
\]

\[
\leq \sum_{j=0}^{n_0} \|u_{T_j}\|_2 \cdot \|v_{T_j}\|_2 \leq \|u\|_{t,\infty} \sum_{j=0}^{n_0} \|v_{T_j}\|_2 
\]

\[
\leq \|u\|_{t,\infty} \cdot \|v\|_{t,1}.
\]

We now consider (ii). Note that

\[
\|u\|_2^2 = \sum_{j=0}^{n_0} \|u_{T_j}\|_2^2.
\]
To this end, we consider \( (A^6(7)(A^7)) \). According to \( (6) \), we obtain that \( A^6(7)(A^7) \). Furthermore, \( (4) \) implies that \( \supp(z) \subseteq \Lambda^6+1 \). Let

\[
\begin{align*}
x^n & := \arg\min_{\supp(z) \subseteq \Lambda^n} \|y - Az\|_2, \\
x^{n+1} & := \arg\min_{\supp(z) \subseteq \Lambda^{n+1}} \|y - Az\|_2,
\end{align*}
\]

and

\[
V^n := A^H_{T^n}(y - Ax^n),
\]

where \( A^H_{T^n} := (A_{T^n})^H \). Then

\[
\|y - Ax^{n+1}\|_2^2 \leq \|y - Ax^n\|_2^2 - \frac{1}{1+\delta_t} \|V^n\|_2^2.
\]

**Proof.** The definition of \( x^{n+1} \) implies that the residuality \( y - Ax^{n+1} \) is orthogonal to the space \( \text{span}(A_{\Lambda^{n+1}}) \). Noting \( A(x^{n+1} - x^n) \in \text{span}(A_{\Lambda^{n+1}}) \), we obtain that

\[
\langle y - Ax^{n+1}, A(x^{n+1} - x^n) \rangle = 0,
\]

which implies that

\[
\|y - Ax^n\|_2^2 = \|y - Ax^{n+1} + A(x^{n+1} - x^n)\|_2^2
\]

\[
= \|y - Ax^{n+1}\|_2^2 + \|A(x^{n+1} - x^n)\|_2^2.
\]

Furthermore, \( A^H_{\Lambda^{n+1}}(y - Ax^n) = 0 \) implies that

\[
(A^H_{\Lambda^{n+1}}y)_{\Lambda^{n+1}} = (A^HAx^{n+1})_{\Lambda^{n+1}}.
\]

Similarly, we have

\[
(A^H_{\Lambda^{n}}y)_{\Lambda^{n}} = (A^H_{\Lambda^{n}}Ax^n)_{\Lambda^{n}}.
\]

According to \( (6) \), we obtain that

\[
(A^H_{\Lambda^{n+1}}A(x^{n+1} - x^n))_{\Lambda^{n+1}} = (A^H_{\Lambda^{n+1}}(y - Ax^n))_{\Lambda^{n+1}},
\]

since

\[
(A^H_{\Lambda^{n+1}}A(x^{n+1} - x^n))_{\Lambda^{n+1}} = (A^H_{\Lambda^{n+1}}y)_{\Lambda^{n+1}} - (A^H_{\Lambda^{n+1}}Ax^n)_{\Lambda^{n+1}} = (A^H_{\Lambda^{n+1}}(y - Ax^n))_{\Lambda^{n+1}}.
\]

To this end, we consider

\[
\|A(x^{n+1} - x^n)\|_2^2 = \langle x^{n+1} - x^n, A^H_{\Lambda^{n+1}}A(x^{n+1} - x^n) \rangle
\]

\[
= \langle (x^{n+1} - x^n)_{\Lambda^{n+1}}, (A^H_{\Lambda^{n+1}}A(x^{n+1} - x^n))_{\Lambda^{n+1}} \rangle
\]

\[
= \langle (x^{n+1} - x^n)_{\Lambda^{n+1}}, (A^H_{\Lambda^{n+1}}(y - Ax^n))_{\Lambda^{n+1}} \rangle
\]

\[
= \langle (x^{n+1} - x^n)_{T^n}, (A^H_{\Lambda^{n}}y)_{T^n} \rangle
\]

\[
= \langle (x^{n+1})_{T^n}, (A^H_{\Lambda^{n+1}}(y - Ax^n))_{T^n} \rangle
\]

\[
= \langle (A^H_{\Lambda^{n+1}}(y - Ax^n))_{T^n}, (x^{n+1})_{T^n} \rangle
\]

\[
= \langle V^n, (x^{n+1})_{T^n} \rangle,
\]
where the third and the fourth equality follow from (8) and (7), respectively. According to (4),
\[ x^{n+1} = A_{n+1}^T y, \]
where \( A_{n+1}^+ = (A_{n+1}^H A_{n+1})^{-1} A_{n+1}^H \) is the Moore-Penrose pseudoinverse of \( A_{n+1} \).
And hence
\[ x^{n+1} = (A_{n+1}^H A_{n+1})^{-1} A_{n+1}^H y. \]
We can write \( A_{n+1} \) as \( [A_{n}, A_T] \). Then
\[ A_{n+1}^H A_{n+1} = \begin{bmatrix} A_{n}^H A_{n} & A_{n}^H A_T \\ A_T^H A_{n} & A_T^H A_T \end{bmatrix}. \]
We next consider
\[ (A_{n+1}^H A_{n+1})^{-1} = \begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix}, \]
where
\[ M_4 = (A_T^H A_T - A_{n+1}^H A_{n} A_{n+1}^H A_T) \]
\[ M_3 = -M_4 (A_{n+1}^H A_{n})^{-1}. \]
Noting (10) and that
\[ A_{n+1}^H y = \begin{bmatrix} A_{n}^H y \\ A_T^H y \end{bmatrix}, \]
we obtain that
\[ (x^{n+1})_T = ((A_{n+1}^H A_{n+1})^{-1} A_{n+1}^H y)_T = M_4 A_{n}^H y + M_4 A_T^H y = -M_4 (A_{n+1}^H A_{n})^{-1} A_{n+1}^H y + M_4 A_T^H y = M_4 A_T^H (-A_{n} A_{n}^T y + y) = M_4 A_T^H (y - A x^n) = M_4 \langle v^n \rangle. \]
Combining (4) and (11) we have
\[ (V^n)^H M_4 V^n = \langle V^n, (x^{n+1})_T \rangle = \| A(x^{n+1} - x^n) \|_2^2. \]
To this end, we consider \( u^H M_4^{-1} u \) for any \( u \in \mathbb{C}^l \). Note that
\[ u^H M_4^{-1} u = u^H A_{n}^T A_{n} u - u^H A_{n}^T A_{n+1}^{-1} A_{n}^H A_T u = u^H A_{n}^T A_{n} u - \langle A_{n} (A_{n} A_{n}^H), (u, P_{A_{n}} (A_{n}^H u)) \rangle \]
\[ \leq \| A_{n} u \|_2^2 - \| P_{A_{n}} (A_{n}^H u) \|_2^2 \]
\[ \leq (1 + \delta_t) \| u \|_2^2, \]
where \( P_{A_{n}} (A_{n}^H u) \) denotes the orthogonal projection of \( A_{n} u \) in the subspace \( \text{span}(A_{n}) \). The last inequality follows from the RIP property of \( A \). Since \( \text{Spark}(A) > \# \{ n+1 \} \), we have \( A_{n} u \notin \text{span}(A_{n}) \) which implies that \( \| P_{A_{n}} (A_{n}^H u) \|_2^2 < \| A_{n} u \|_2^2 \) provided \( u \neq 0 \). And hence, according to (13),
\[ u^H M_4^{-1} u = \| A_{n} u \|_2^2 - \| P_{A_{n}} (A_{n}^H u) \|_2^2, \]
which implies that $M_4$ is a positive-definite matrix since $u^H M_4^{-1} u > 0$ provided $u \neq 0$. Combining (12) and (14), we obtain that

$$\|A(x^{n+1} - x^n)\|_2^2 = (V^n)^H M_4 V^n \geq \frac{1}{1 + \delta_t} \| V^n \|_2^2.$$ 

Then the (5) implies that

$$\|y - Ax^{n+1}\|_2^2 = \|y - Ax^n\|_2^2 - \|A(x^{n+1} - x^n)\|_2^2 \leq \|y - Ax^n\|_2^2 - \frac{1}{1 + \delta_t} \| V^n \|_2^2.$$

\[\square\]

**Lemma 3.** Consider OMMP($M$) and $\Lambda^n \subset \Lambda^{n+1} \subset \{1, \ldots, N\}$. Set $T^n := \Lambda^{n+1} \setminus \Lambda^n$ and $t := \#T^n$. Suppose that the sampling matrix $A \in \mathbb{C}^{n \times N}$ whose columns $a_1, \ldots, a_N$ are $\ell_2$-normalized. Then for any $u \in \mathbb{C}^N$ whose support $U := \text{supp}(u)$ not included in $\Lambda^n$, we have

$$\| V^n \|_2^2 \geq \frac{\|A(u - x^n)\|_2^2 (\|y - Ax^n\|_2^2 - \|y - Au\|_2^2)}{\|u_{\text{Xn}}\|_{t,1}^2},$$

where $V^n := A_{T^n}^H (y - Ax^n)$.

**Proof.** To this end, we only need prove that

$$\| V^n \|_2^2 \cdot \| u_{\text{Xn}} \|_{t,1}^2 \geq \|A(u - x^n)\|_2^2 \cdot (\|y - Ax^n\|_2^2 - \|y - Au\|_2^2).$$

When

$$\|y - Ax^n\|_2^2 - \|y - Au\|_2^2 < 0,$$

the conclusion holds. So, we only consider the case where

$$\|y - Ax^n\|_2^2 - \|y - Au\|_2^2 \geq 0.$$

Recall that $T^n$ is the $t$ indices corresponding to the largest magnitude entries in the vector $(A^H (y - Ax^n))^\Lambda^n$. Then

$$\| V^n \|_2 \geq \|(A^H (y - Ax^n))_{\text{Xn}}\|_{t,\infty}.$$ 

Noting that $(x^n)_{\text{Xn}} = 0$ and $(A^H (y - Ax^n))_{\Lambda^n} = 0$, we have

$$\| V^n \|_2 \cdot \| u_{\text{Xn}} \|_{t,1} \geq \|(A^H (y - Ax^n))_{\text{Xn}}\|_{t,\infty} \cdot \|(u - x^n)_{\text{Xn}}\|_{t,1} \geq \mathcal{R} \left( \langle (u - x^n)_{\text{Xn}}, (A^H (y - Ax^n))_{\text{Xn}} \rangle \right) \geq \mathcal{R} \left( \langle (u - x^n), A^H (y - Ax^n) \rangle \right) \geq \frac{1}{2} \left( \| A(u - x^n) \|_2^2 + \| y - Ax^n \|_2^2 - \| A(u - x^n) - (y - Ax^n) \|_2^2 \right) \geq \| A(u - x^n) \|_2 \cdot \sqrt{\| y - Ax^n \|_2^2 - \|y - Au\|_2^2},$$

which implies the result, where the second inequality follows from Lemma [1]  \[\square\]
Lemma 4. Under the conditions of Lemma 3, we have
\[
\|y - Ax^{n+1}\|_2^2 \leq \|y - Ax^n\|_2^2 - \frac{(1 - \delta)}{1 + \delta} \max \{0, \|y - Ax^n\|_2^2 - \|y - Au\|_2^2\},
\]
where \(\delta = \delta_{\#(U \setminus \Lambda^n)}\).

Proof. According to Lemma 2 and Lemma 3, we have
\[
\|y - Ax^{n+1}\|_2^2 \leq \|y - Ax^n\|_2^2 - \frac{1}{1 + \delta} \|V_n\|_2^2.
\]
(16)
From Lemma 1, we have
\[
\|u_{\Lambda^n}\|_2^2 \leq \left\lfloor \frac{\#(U \setminus \Lambda^n)}{t} \right\rfloor \cdot \|u_{\Lambda^n}\|_2^2.
\]
(17)
Also,
\[
\|A(u - x^n)\|_2^2 \geq (1 - \delta)\|u - x^n\|_2^2 \geq (1 - \delta)\|u - x^n\|_{\Lambda^t}\|_2^2 \geq (1 - \delta)\|u_{\Lambda^n}\|_2^2.
\]
(18)
Putting (16), (17) and (18) together, we arrive at the conclusion. \(\square\)

Remark 5. Lemma 4 extends some results in [9], where Foucart considered the case with \(t = \#(\Lambda^{n+1} \setminus \Lambda^n) = 1\), to the general case. In fact, if takes \(t = 1\) in Lemma 4, one can obtain Lemma 4 in [9].

Appendix B. PROOF OF THEOREM 4

Proof of Theorem 4. To state conveniently, we set \(x' := x_{\Lambda^n}\) and \(K := \max \{s', \frac{M}{s'} s'\}\). We claim that the conclusion follows provided \(S \subset \Lambda^K\). Indeed, since
\[
\#\Lambda^K \leq \max \{Ms', 8s'\} + \#\Lambda^0 < \text{Spark}(A),
\]
one can recover \(x\) by solving the least square, i.e.,
\[
x = \arg \min_{z: \text{supp}(z) \subset \Lambda^K} \|y - Az\|_2.
\]
Thus, to this end, we only need prove that \(S \subset \Lambda^K\), i.e. \(\#(S \setminus \Lambda^K) = 0\). The proof is by induction on \(s' = \#(S \setminus \Lambda^0)\). If \(s' = 0\), then the conclusion holds. For the induction step, we assume that the result holds up to an integer \(s' - 1\). We next show that it holds for \(s'\).

Without loss of generality, we suppose that
\[
|x'_1| \geq |x'_2| \geq \cdots \geq |x'_{s'}| > 0.
\]
For \(\ell = 1, \ldots, \max \{0, \left\lceil \log_2 \frac{M}{s'} \right\rceil\} + 1\), we set
\[
\tilde{x}'_j := \begin{cases} x'_j & \text{if } j \geq 2^{\ell-1} \cdot M + 1, \\ 0 & \text{else}, \end{cases}
\]
and $\tilde{x}^0 := x'$. Suppose that $L \in \mathbb{Z}$ such that

\begin{equation}
||\tilde{x}^0||^2 < \mu ||\tilde{x}^1||^2, \ldots, ||\tilde{x}^{L-2}||^2 < \mu ||\tilde{x}^{L-1}||^2
\end{equation}

and

\begin{equation}
||\tilde{x}^{L-1}||^2 \geq \mu ||\tilde{x}^L||^2.
\end{equation}

And hence, $L$ is the least integer such that $||\tilde{x}^{L-1}||^2 \geq \mu ||\tilde{x}^L||^2$ and we will choose $\mu > 2$ late. The existence of such a $L$ can follow from $||\tilde{x}^\ell||^2 = 0$ when $\ell = \max\{0, \log_2 \frac{S}{\lambda}\} + 1$. And hence, we have

\[ 1 \leq L \leq \max \left\{ 0, \left\lfloor \log_2 \frac{S}{\lambda} \right\rfloor \right\} + 1. \]

We first consider the case where $L = 1$. We take $u = u^1 := x - \tilde{x}^1$ and $t = M$ in (15). Then a simple observation is that

\[ \#(\text{supp}(u^1) \setminus A^0) = \min\{M, s'\}. \]

Noting that $\left\lfloor \#(\text{supp}(u^1) \setminus A^0) \right\rfloor = 1$ and

\[ ||y - Au^1||^2 = ||A\tilde{x} - Au^1||^2 = ||A\tilde{x}^1||^2. \]

By subtracting $||y - Au^1||^2 = ||A\tilde{x}^1||^2$ on both sides of (15), we can obtain that

\[ \max\{0, ||y - Ax^1||^2 - ||Ax^1||^2\} \leq \left(1 - \frac{1 - \delta_s}{1 + \delta_s}\right) \max\{0, ||y - Ax^0||^2 - ||Ax^1||^2\}, \]

which implies that

\begin{align*}
||y - Ax^1||^2 &\leq \left(1 - \frac{1 - \delta_s}{1 + \delta_s}\right) \max\{0, ||y - Ax^0||^2 - ||Ax^1||^2\} + ||Ax^1||^2 \\
&= \left(1 - \frac{1 - \delta_s}{1 + \delta_s}\right) \max\{0, ||Ax^0||^2 - ||Ax^1||^2\} + ||Ax^1||^2 \\
&\leq \left(1 - \frac{1 - \delta_s}{1 + \delta_s}\right) \parallel Ax^0\parallel^2 + ||Ax^1||^2 \\
&\leq (1 + \delta_s) \left(1 - \frac{1 - \delta_s}{1 + \delta_s}\right) \parallel Ax^0\parallel^2 + \parallel Ax^1\parallel^2 \\
&\leq 2\delta_s\parallel Ax^0\parallel^2 + \frac{1 + \delta_s}{\mu} \parallel Ax^0\parallel^2 = \left(2\delta_s + \frac{1 + \delta_s}{\mu}\right) \parallel Ax^0\parallel^2, \tag{21}
\end{align*}

where the last inequality uses the fact that $L = 1$ and hence $\parallel Ax^1\parallel^2 \leq \parallel Ax^0\parallel^2/\mu$. On the other hand, we note that

\begin{align*}
||y - Ax^1||^2 &= \parallel A(x - x^1)\parallel^2 \\
&\geq (1 - \delta_{2s}) \parallel x - x^1\parallel^2 \\
&\geq (1 - \delta_{2s}) \parallel x_{\Lambda_{\delta}}\parallel^2. \tag{22}
\end{align*}

Then, combining (21) and (22), we obtain that

\[ \parallel x_{\Lambda_{\delta}}\parallel^2 \leq \frac{1}{1 - 3\delta_{2s}} \left(2\delta_s + \frac{1 + \delta_s}{\mu}\right) \parallel Ax^0\parallel^2. \]

Noting $\delta_s \leq \delta_{2s} \leq \delta_s \leq \frac{1}{17}$, we have

\[ \frac{1 + \delta_s}{1 - 3\delta_{2s}} \leq 2 < \mu, \]
which implies that
\[
\frac{1}{1 - \delta_{2s}} \left(2\delta_s + \frac{1 + \delta_s}{\mu}\right) < 1.
\]
And hence,
\[
\|x_{\Lambda}^0\|_2^2 < \|\tilde{x}\|_2^2,
\]
i.e.
\[
\#(S \setminus \Lambda^1) \leq s' - 1.
\]
Now we continue the algorithm with the initial feature set \(\Lambda^1\). According to the induction assumption, we can recover the \(s\)-sparse signal \(x\) within \(\max\{s' - 1, \frac{s}{M}(s' - 1)\}\) iterations provided the initial feature set is \(\Lambda^1\). Thus, if one chooses the initial feature set as \(\Lambda^0\) then \(x\) can be recovered within \(1 + \max\{s' - 1, \frac{s}{M}(s' - 1)\}\) iterations.

Then, the conclusion follows since
\[
1 + \max\{s' - 1, \frac{s}{M}(s' - 1)\} \leq \max\{s, \frac{8}{M}s\}.
\]

We next consider the case where \(L \geq 2\). We take \(u = u^\ell := x - \tilde{x}_\ell\) and \(t = M\) in (15). Then a simple observation is that
\[
\#(\text{supp}(u^\ell) \setminus \Lambda^n) = \#(\text{supp}(u^\ell) \cap \Lambda^0) + \min\{2, \frac{s}{M}s\} - \#(\text{supp}(u^\ell) \cap \Lambda^n).
\]
Thus, for any \(n \geq 0\),
\[
\#(\text{supp}(u^\ell) \setminus \Lambda^n) = \#(\text{supp}(u^\ell) \cap \Lambda^0) + \min\{2, \frac{s}{M}s\} - \#(\text{supp}(u^\ell) \cap \Lambda^n) \leq \min\{2, \frac{s}{M}s\}.
\]
To state conveniently, we set
\[
\bar{U}^\ell := \left\lfloor \frac{\min\{2, \frac{s}{M}s\}}{M} \right\rfloor \in \mathbb{Z}.
\]
If \(\text{supp}(u^\ell) \not\subset \Lambda^n\) then we obtain that
\[
\max\{0, \|y - Ax^{n+1}\|_2^2 - \|Ax^\ell\|_2^2\} \leq \left(1 - \frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot \bar{U}^\ell}\right) \max\{0, \|y - Ax^n\|_2^2 - \|Ax^\ell\|_2^2\} \leq \exp\left(-\frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot \bar{U}^\ell}\right) \max\{0, \|y - Ax^n\|_2^2 - \|Ax^\ell\|_2^2\},
\]
which follows by subtracting
\[
\|y - Ax^\ell\|_2^2 = \|Ax - Ax^\ell\|_2^2 = \|Ax^\ell\|_2^2
\]
on both sides of (15) in Lemma 4. For the case \(\text{supp}(u^\ell) \subset \Lambda^n\), (23) still holds since both sides of (23) are equal to 0. Iterating (23) \(k\) times leads to
\[
\max\{0, \|y - Ax^{n+k}\|_2^2 - \|Ax^\ell\|_2^2\} \leq \exp\left(-k\frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot \bar{U}^\ell}\right) \max\{0, \|y - Ax^n\|_2^2 - \|Ax^\ell\|_2^2\}
\]
which implies that
\[
\|y - Ax^{n+k}\|_2^2 \\
\leq \exp\left(-k \frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot U^r}\right) \max\{0, \|y - Ax^n\|_2^2 - \|Ax^\ell\|_2^2\} + \|Ax^\ell\|_2^2.
\]
(25)
\[
\leq \exp\left(-k \frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot U^r}\right) \|y - Ax^n\|_2^2 + \|Ax^\ell\|_2^2.
\]

Here, if the left side of (24) is 0, then
\[
\|y - Ax^{n+k}\|_2^2 \leq \|Ax^\ell\|_2^2.
\]
Thus, (25) still holds since \(\|y - Ax^n\|_2^2 \geq 0\). To state conveniently, for \(\ell = 1, \ldots, L\), we set \(k_\ell := k \cdot U^\ell\), \(k_0 := 0\), \(K := k_1 + \cdots + k_L\) and \(\nu := \exp\left(-k \frac{1 - k_{\ell-1}M}{1 + \delta_M}\right)\), and we will choose \(k\) late. For \(\ell = 1, \ldots, L\), we take \(n := k_0 + \cdots + k_{\ell-1}\) and \(k := k_\ell\) in (26) and arrive at
\[
\|y - Ax^{k_1 + \cdots + k_\ell}\|_2^2 \leq \exp\left(-k \frac{1 - \delta_{s+(k_0 + \cdots + k_{\ell-1})M}}{(1 + \delta_M) \cdot U^r}\right) \|y - Ax^{k_1 + \cdots + k_{\ell-1}}\|_2^2 + \|Ax^\ell\|_2^2.
\]
(26)
\[
\leq \nu \|y - Ax^{k_1 + \cdots + k_{\ell-1}}\|_2^2 + \|Ax^\ell\|_2^2.
\]
Then, using the inequality (26) for \(L\) times, we can obtain that
\[
\|y - Ax^K\|_2^2 \leq \nu^L \|y - Ax^0\|_2^2 + \nu \|Ax^1\|_2^2 + \cdots + \nu \|Ax^{L-1}\|_2^2 + \|Ax^L\|_2^2
\]
\[
\leq \nu^L \|Ax^0\|_2^2 + \cdots + \nu \|Ax^{L-1}\|_2^2 + \|Ax^L\|_2^2.
\]
Here, for the second relation, we use the fact of
\[
\|y - Ax^0\|_2^2 = \min_{\text{supp(x) \subset} \Lambda^0} \|y - Ax\|_2^2 \leq \|y - A(x - \hat{x}^0)\|_2^2 = \|Ax^0\|_2^2
\]
with supp\((x - \hat{x}^0) \subset \Lambda^0\). Combining RIP property of \(A\), (19) and (20), we obtain that
\[
\|Ax^\ell\|_2^2 \leq (1 + \delta_s)\|\hat{x}^\ell\|_2^2 \leq (1 + \delta_s)\mu^{L-1-\ell}\|\hat{x}^{L-1}\|_2^2
\]
for \(\ell = 0, 1, \ldots, L\). Note that
\[
\|y - Ax^K\|_2^2 \leq \sum_{\ell=0}^L \nu^{L-\ell} \|Ax^\ell\|_2^2
\]
\[
\leq \frac{(1 + \delta_s)\|\hat{x}^{L-1}\|_2^2}{\mu} \sum_{\ell=0}^L (\mu \nu)^{L-\ell}
\]
\[
\leq \frac{(1 + \delta_s)\|\hat{x}^{L-1}\|_2^2}{\mu (1 - \mu \nu)},
\]
(27)
and
\[
\|y - Ax^K\|_2^2 \geq \|A(x - x^K)\|_2^2
\]
\[
\geq (1 - \delta_{s+K \cdot M})\|x - x^K\|_2^2
\]
\[
\geq (1 - \delta_{s+K \cdot M})\|x - x^K\|_2^2.
\]
(28)
Combining (27) and (28), we have
\[
\|x - x^K\|_2^2 \leq \frac{(1 + \delta_s)}{(1 - \delta_{s+K \cdot M})\mu(1 - \mu \nu)} \|\hat{x}^{L-1}\|_2^2.
\]
(29)
We can choose $\hat{k} = 2$, $\mu = \frac{1}{2^{\nu}}$, and $\delta_{s+K \cdot M} \leq \delta_{9s} \leq \frac{1}{10}$ with

$$K = k_1 + \cdots + k_L \leq 2^L \hat{k} \leq 8 \frac{s}{M}.$$ 

Noting that $\nu \leq \exp(-18/11)$ and $\mu = \frac{1}{2^{\nu}} > 2$, we have

$$\frac{(1 + \delta_s)}{(1 - \delta_{s+K \cdot M})\mu(1 - \mu\nu)} < 1.$$ 

Combining (29) and (30), we obtain that

$$\|x_{\Lambda^k}\|_2^2 < \|\tilde{x}_{L-1}\|_2^2.$$ 

As a result, after $K$ iterations, we have

$$\#(S \setminus \Lambda^k) = 0,$$

with

$$K = k_1 + \cdots + k_L \leq 2^L \hat{k}.$$ 

Now we continue the algorithm with the initial feature set $\Lambda^K$. According to the induction assumption, we can recover the $s$-sparse signal $x$ within $\bar{n}$ iterations provided the initial feature set is $\Lambda^K$, where $\bar{n} = \max\{s' - 2^{L-2} \cdot M - 1, \frac{8}{M}(s' - 2^{L-2} \cdot M - 1)\}$.

Thus, if one chooses the initial feature set as $\Lambda^0$ then $x$ can be recovered within $K + \bar{n}$ iterations. Then, the conclusion follows since $K + \bar{n} \leq \max\{s', \frac{8}{M}s'\}$.

**Appendix C. Proofs of Theorem 2 and Theorem 3**

To prove Theorem 2 and Theorem 3, we first introduce two lemmas.

**Lemma 5.** Consider the OMMP$(M)$ algorithm with $1 \leq M \leq s$. Suppose that the sampling matrix $A \in \mathbb{C}^{m \times N}$ satisfies $(9s, \frac{1}{10})$-RIP. Suppose that $x \in \Sigma_s$, $S = \text{supp}(x)$. Then

$$\#(S \setminus \Lambda^K) = 0,$$

where $K = \left[ \frac{8s'}{M} + 8(C_0^2 + 1)M \right]$, $s' = \#(S \setminus \Lambda^0)$ and $C_0 = \max_{j \in S} |x_j| / \min_{j \in S} |x_j|$.

**Proof.** To state conveniently, we set

$$x' := x_{\Lambda^0}$$

and

$$C_2 := \frac{C_0^2}{\mu - 1} + 1.$$ 

We will choose $\mu > 2$ late so that $C_2 < C_0^2 + 1$. To this end, we will prove that $\#(S \setminus \Lambda^{K_1}) = 0$ with $K_1 = \left[ \frac{8s'}{M} + 8C_2M \right]$, which implies the result. The proof is by induction on $s' = \#(S \setminus \Lambda^0)$. We first consider the case where $s' \leq C_2M$. According to Theorem 1, OMMP$(M)$ recover the $s$-sparse signal within $8C_2M < 8(C_0^2 + 1)M$ iterations. Thus, we arrive at the result provided $s' \leq C_2M$.

We next consider the case where $s' > C_2M$. Without loss of generality, we suppose that

$$|x'_1| \geq |x'_2| \geq \cdots \geq |x'_{s'}| > 0.$$
To state conveniently, for \( \ell = 1, \ldots, \lceil \log_2 \left( \frac{s'}{M} \right) \rceil + 1 \), we set
\[
\mathbf{x}_j^\ell := \begin{cases} 
\mathbf{x}_j' & \text{if } 2^{\ell-1}M + 1 \leq j, \\
0 & \text{else}.
\end{cases}
\]
and \( \mathbf{x}^0 := \mathbf{x}' \). Suppose that \( L \in \mathbb{Z} \) such that
\[
||\mathbf{x}^0||^2 < \mu||\mathbf{x}^1||^2, \ldots, ||\mathbf{x}^{L-2}||^2 < \mu||\mathbf{x}^{L-1}||^2
\]
and
\[
||\mathbf{x}^{L-1}||^2 \geq \mu||\mathbf{x}^L||^2.
\]
And hence, \( L \) is the least integer such that \( ||\mathbf{x}^{L-1}||^2 \geq \mu||\mathbf{x}^L||^2 \). The existence of such a \( L \) can follow from \( ||\mathbf{x}^\ell||^2 = 0 \) when \( \ell = \lceil \log_2 \left( \frac{s'}{M} \right) \rceil + 1 \). We next show that the assumption of \( s' > C_2M \) implies that \( ||\mathbf{x}^0||^2 < \mu||\mathbf{x}^1||^2 \) and hence \( L \geq 2 \). Indeed, \( ||\mathbf{x}^0||^2 < \mu||\mathbf{x}^1||^2 \) is equivalent to
\[
x_1^2 + \cdots + x_M^2 < (\mu - 1)\|\mathbf{x}^1\|_2^2.
\]
Hence, we only need argue (33). Note that
\[
x_1^2 + \cdots + x_M^2 \leq M \max_{j \in S} x_j^2 < (\mu - 1)(s' - M) \min_{j \in S} x_j^2 \leq (\mu - 1)||\mathbf{x}^1||^2,
\]
where the second relation uses the fact of
\[
s' > C_2M = \left( \frac{C_2}{\mu - 1} + 1 \right) M.
\]
And hence, we have \( 2 \leq L \leq \lceil \log_2 \left( \frac{s'}{M} \right) \rceil + 1 \). We take
\[
\mathbf{u} = \mathbf{u}^\ell := \mathbf{x} - \mathbf{x}^\ell
\]
and \( t = M \) in (15). Then a simple observation is that
\[
\#\text{supp}(\mathbf{u}^\ell) = \#\Lambda^0 + \min\{2^{\ell-1}M, s'\}.
\]
For any \( n \geq 0 \),
\[
\#(\text{supp}(\mathbf{u}^\ell) \setminus \Lambda^n) = \#(\text{supp}(\mathbf{u}^\ell) \cap \Lambda^n) + \min\{2^{\ell-1}M, s'\} - \#(\text{supp}(\mathbf{u}^\ell) \cap \Lambda^n) \\
\leq \min\{2^{\ell-1}M, s'\}.
\]
To state conveniently, we set
\[
U^\ell := \left\lfloor \frac{\min\{2^{\ell-1}M, s'\}}{M} \right\rfloor.
\]
Noting that
\[
||\mathbf{y} - A\mathbf{u}^\ell||_2^2 = ||A\mathbf{x} - A\mathbf{u}^\ell||_2^2 = ||A\mathbf{x}^\ell||_2^2,
\]
by (14), we obtain that
\[
\max\{0, ||\mathbf{y} - A\mathbf{x}^n||_2^2 - ||A\mathbf{x}^\ell||_2^2\} \\
\leq \left( 1 - \frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot U^\ell} \right) \max\{0, ||\mathbf{y} - A\mathbf{x}^n||_2^2 - ||A\mathbf{x}^\ell||_2^2\} \\
\leq \exp\left( \frac{1 - \delta_{s+nM}}{(1 + \delta_M) \cdot U^\ell} \right) \max\{0, ||\mathbf{y} - A\mathbf{x}^n||_2^2 - ||A\mathbf{x}^\ell||_2^2\}.
\]
Iterating (34) for \( k \) times leads to
\[
\max\{0, ||\mathbf{y} - A\mathbf{x}^{n+k}||_2^2 - ||A\mathbf{x}^\ell||_2^2\} \leq \exp\left( -k \frac{(1 - \delta_{s+KM})}{(1 + \delta_M) \cdot U^\ell} \right) \max\{0, ||\mathbf{y} - A\mathbf{x}^n||_2^2 - ||A\mathbf{x}^\ell||_2^2\}.
\]
which implies that
\begin{equation}
\|y - Ax^{n+k}\|_2^2 \leq \exp\left(-k \frac{1 - \delta_{s+KM}}{1 + \delta_M} \cdot \bar{U}^\ell\right) \|y - Ax^n\|_2^2 + \|Ax^\ell\|_2^2
\end{equation}
where \(k\) and \(K\) are integers satisfying \(K \geq n + k\).

To state conveniently, for \(\ell = 1, \ldots, L\), we set \(k_\ell := \bar{k} \cdot \bar{U}^\ell\), \(K := k_1 + \cdots + k_L\) and
\[v := \exp\left(-\bar{k} \frac{1 - \delta_{s+KM}}{1 + \delta_M}\right),\]
and we will choose \(\bar{k}\) late. We use (35) and a similar argument in the proof of Theorem 1 to obtain that
\begin{equation}
\|y - Ax^K\|_2^2 \leq \sum_{\ell=0}^L v^{L-\ell} \|Ax^\ell\|_2^2
\end{equation}
Note that
\begin{equation}
\|y - Ax^K\|_2^2 \geq \|A(x - x^K)\|_2^2
\end{equation}
\begin{align*}
&\geq (1 - \delta_{s+KM})\|x - x^K\|_2^2 \\
&\geq (1 - \delta_{s+KM})\|x_{\Lambda^K}\|_2^2.
\end{align*}
Combining (36) and (37), we arrive at
\begin{equation}
\|x_{\Lambda^K}\|_2^2 \leq \frac{(1 + \delta_s)}{(1 - \delta_{s+KM})\mu(1 - \mu v)}\|x^{L-1}\|_2^2.
\end{equation}
We can choose \(\bar{k} = 2\), \(\mu = \frac{1}{2v}\), and
\[\delta_{s+KM} \leq \delta_{9s} \leq \frac{1}{10},\]
and therefore \(v \leq \exp(-18/11)\) and
\[\mu = \frac{1}{2v} > 2.\]
Here, we use \(s + KM \leq s + 4\bar{k}s' \leq 9s\) since
\[K = k_1 + \cdots + k_L \leq \bar{k}(1 + \cdots + 2^{L-1}) \leq 2^L \bar{k} \leq 4 \cdot \bar{k} \cdot \frac{s'}{M}.\]
Then
\[\frac{(1 + \delta_s)}{(1 - \delta_{s+KM})\mu(1 - \mu v)} < 1,\]
which implies that
\[\|x_{\Lambda^K}\|_2^2 < \|x^{L-1}\|_2^2.\]
As a result, after \(K\) iterations, we have
\[\#(S \setminus \Lambda^K) < \#((S \setminus \Lambda^0) \setminus \text{supp}(u^{L-1})) = s' - 2^{L-2} M.\]
Now we continue the algorithm with the initial feature set $\Lambda^K$. According to the induction assumption, we can recover the $s$-sparse signal $x$ in $K + \bar{n}$ iterations where $$\bar{n} \leq \left\lfloor \frac{8s' - 2L - 2M}{M} + 8C_2M \right\rfloor.$$

Note that $L \geq 2$ and

$$K + \bar{n} \leq 2^{L+\bar{k}} + \left\lfloor \frac{8s' - 2L - 2M}{M} + 8C_2M \right\rfloor = 8 \cdot 2^{L+\bar{k}} - \left\lfloor \frac{8s' - 2L - 2M}{M} + 8C_2M \right\rfloor.$$

Then we arrive at

$$K + \bar{n} \leq \left\lfloor \frac{8s'}{M} + 8C_2M \right\rfloor,$$

which implies the result. \hfill \Box

**Lemma 6.** Suppose that $x$ is $s$-sparse, $S = \text{supp}(x)$ and $C_0 = \max_{j \in S} |x_j|/\min_{j \in S} |x_j|$. Consider the OMMP($M$) algorithm with $1 \leq M \leq \frac{2}{C_0^2 + 2} \cdot s$. Suppose that the sampling matrix $A \in \mathbb{C}^{m \times N}$ whose columns $a_1, \ldots, a_N$ are $\ell_2$-normalized, and that $A$ satisfies $(14s, \frac{1}{10})$-RIP. Set $s' := \#(S \setminus \Lambda^0)$ and

$$\bar{K} := \left\lfloor \frac{8s'}{M} + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2(s' + 1) \right\rfloor.$$

Then

$$\#(S \setminus \Lambda^{\bar{K}}) = 0.$$

**Proof.** To state conveniently, we set $x' := x_{\Lambda^0}$. The proof is by induction on $s' = \#(S \setminus \Lambda^0)$. When $s' = 0$, the conclusion holds trivially.

Without loss of generality, we suppose that

$$|x'_1| \geq |x'_2| \geq \cdots \geq |x'_{s'}| > 0.$$

For convenience, for $\ell = 1, \ldots, \left\lfloor \log_2\left(\frac{s'}{M}\right) \right\rfloor + 1$, we set

$$\hat{x}^\ell_j := \begin{cases} x'_j & \text{if } 2^{\ell-1} \frac{M}{s'} s' + 1 \leq j, \\ 0 & \text{else} \end{cases}$$

and $\hat{x}^0 := x'$. Similar with the proof of Lemma 5, suppose that $L$ is the least integer such that $\|\hat{x}^{L-1}\|_2^2 \geq \mu\|\hat{x}^{L}\|_2^2$. We will choose $\mu > 2$ late. The assumption of

$$M < \frac{2}{C_0^2 + 2} s$$

implies that

$$\|\hat{x}^0\|_2^2 < \mu\|\hat{x}^1\|_2^2.$$

And hence, we have $2 \leq L \leq \left\lfloor \log_2\left(\frac{s'}{M}\right) \right\rfloor + 1$. We take $u = u^\ell := x - \hat{x}^\ell$. 


and \( t = M \) in (15). Then a simple observation is that

\[
\#\text{supp}(u^t) = \#\text{supp}((u^t) \cap \Lambda^0) + \min \left\{ \left\lfloor 2^\ell - 1 \frac{M}{s} \right\rfloor, s' \right\}.
\]

For any \( n \geq 0 \),

\[
\#(\text{supp}(u^\ell) \setminus \Lambda^n) = \#\text{supp}((u^\ell) \cap \Lambda^0) + \min \left\{ \left\lfloor 2^\ell - 1 \frac{M}{s} \right\rfloor, s' \right\} - \#\text{supp}((u^\ell) \cap \Lambda^n).
\]

To state conveniently, we set \( \bar{U}^\ell := \left\lceil \min \left\{ \left\lfloor 2^\ell - 1 \frac{M}{s} \right\rfloor, s' \right\} \right\rceil \), \( k^\ell := \bar{k} \cdot \bar{U}^\ell \), \( K := k^1 + \cdots + k^L \) and \( v := \exp \left( -\bar{k} \frac{1 - \delta + KM}{1 + 3M} \right) \), and we will choose \( \bar{k} \) later. We use (35) and a similar argument in the proof of Theorem 1 to obtain that

\[
\|y - Ax^K\|_2^2 \leq \sum_{\ell=0}^L v^{L-\ell} \|Ax^\ell\|_2^2
\]

\[
\leq \frac{(1 + \delta_s)\|x^{L-1}\|_2^2}{\mu} \sum_{\ell=0}^L (\mu v)^{L-\ell}
\]

\[
(38)
\]

Note that

\[
\|y - Ax^K\|_2^2 \geq \|A(x - x^K)\|_2^2
\]

\[
\geq (1 - \delta_s + KM)\|x - x^K\|_2^2
\]

\[
\geq (1 - \delta_s + KM)\|x_{\Lambda^K}\|_2^2.
\]

(39)

Combining (38) and (39), we arrive at

\[
\|x_{\Lambda^K}\|_2^2 \leq \frac{(1 + \delta_s)}{(1 - \delta_s + KM)\mu(1 - \mu v)}\|x^{L-1}\|_2^2.
\]

We can choose \( \bar{k} = 2 \), \( \mu = 1/(2v) \), and \( \delta_s + KM \leq \delta_{14s} \leq 1/10 \). And hence \( v \leq \exp(-18/11) \) and \( \mu = 1/(2v) > 2 \). Here, we use \( s + KM \leq 13s \) with

\[
K = k^1 + \cdots + k^L
\]

\[
\leq \bar{k}(1 + \cdots + 2^L - 1) \frac{s'}{s} + \bar{k}L \leq 2^L \frac{\bar{k} s'}{s} + \bar{k}L
\]

\[
\leq 4\bar{k} \frac{s'}{M} + \bar{k}L \leq 8 \frac{s'}{M} + 4 + 2 \log_2 \frac{s}{M}.
\]

Then

\[
\frac{(1 + \delta_s)}{(1 - \delta_s + KM)\mu(1 - \mu v)} < 1,
\]

which implies that

\[
\|x_{\Lambda^K}\|_2^2 \leq \|x^{L-1}\|_2^2.
\]
As a result, after $K$ iterations, we have
\[
\#(S \setminus \Lambda^K) \leq \#((S \setminus \Lambda^0) \setminus \text{supp}(u^{L-1})) - 1 = s' - 2^{L-2}M \frac{s'}{s} - 1,
\]
with
\[
K = k_1 + \cdots + k_L \leq \tilde{k}(1 + \cdots + 2^{L-1}) \frac{s'}{s} + \tilde{k}L \leq 2^{L} \frac{s'}{s} + \tilde{k}L.
\]
Now we continue the algorithm from the iteration $K$. According to the induction assumption, we have
\[
\#(S \setminus \Lambda^{K+\bar{n}}) = 0
\]
with
\[
\bar{n} \leq \left\lfloor \frac{8s' - 2^{L-2}M s'}{M} + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 \left( s' - 2^{L-2}M \frac{s'}{s} \right) \right\rfloor.
\]
Note that $L \geq 2$ and that
\[
2^{L} \frac{s'}{s} + 8 \frac{s' - 2^{L-2}M s'}{M} \leq 8 \frac{s'}{M}.
\]
A simple calculation shows that
\[
\tilde{k}L + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 \left( s' - 2^{L-2}M \frac{s'}{s} \right) = 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 s' + \tilde{k}L + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 \left( 1 - 2^{L-2}M \right) \leq 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 (s' + 1).
\]
Then we arrive at
\[
K + \bar{n} \leq 2^{L} \frac{s'}{s} + \tilde{k}L + \bar{n} \leq \left\lfloor \frac{8s'}{M} + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2 (s' + 1) \right\rfloor,
\]
which implies the result. \hfill \Box

Proof of Theorem 2. According to Lemma 5, after OMMP($M$) running $\bar{K}$ steps, we have
\[
S \subset \Lambda^{\bar{K}}
\]
where
\[
\bar{K} = \left\lfloor \frac{8s'}{M} + 8(C_0^2 + 1)M \right\rfloor \leq \left\lfloor 8(C_0^2 + 2) \frac{s'}{M} \right\rfloor.
\]
Here we use the assumption of $M \leq \sqrt{s'}$. Since OMMP($M$) chooses $M$ atoms at each iteration, we have
\[
\#\Lambda^{\bar{K}} \leq \bar{K}M \leq 8(C_0^2 + 2)s' + \#\Lambda^0.
\]
Noting that $\text{Spark}(A) > 8(C_0^2 + 2)s' + \#\Lambda^0$, we obtain that
\[
\arg\min_{\mathbf{z} \in \mathbb{C}^N, \text{supp}(\mathbf{z}) \subset \Lambda^{\bar{K}}} \|A\mathbf{z} - \mathbf{y}\|_2 = \mathbf{x}
\]
which implies that OMMP(\(M\)) can recover the \(s\)-sparse signal \(x\) within \([8(C_0^2 + 2)^\frac{1}{\alpha^2}]\) iterations.

**Proof of Theorem 3.** By Lemma \(5\), we have
\[
S \subset \Lambda^\bar{K},
\]
since
\[
\bar{K} \leq \left[\frac{8s}{M} + 4 \cdot \ln 2 \cdot \frac{s}{M} \log_2(s + 1)\right] \leq \left[8\frac{s}{M} \log_2(2(s + 1))\right] = \left[\frac{8s}{M} \log_2(2(s + 1))\right].
\]
Here, we use the fact of \(\Lambda^0 = \emptyset\) and hence \(#(S \setminus \Lambda^0) = s\). Also, noting that
\[
#\Lambda^\bar{K} \leq \bar{K}M \leq 8s \log_2(2(s + 1))
\]
and
\[
\text{Spark}(A) > 8s \log_2(2(s + 1)),
\]
we obtain that
\[
\arg\min_{z \in \mathbb{C}^N, \supp(z) \subset \Lambda^\bar{K}} \|Az - y\|_2 = x,
\]
which implies the result. \(\square\)

**Appendix D. Proof of Theorem 5.**

**Proof.** The proof proceed by induction. We assume that \(\Lambda^\ell \subset \supp(x)\) holds for \(\ell = 0, \ldots, n - 1 \leq s - 1\). We next consider \(\Lambda^n\). Set
\[
\tilde{x}^{n-1} := x_{\Lambda^\ell \cup \{j^{n-1}\}}, \quad u := x_{\Lambda^n \setminus \{j^{n-1}\}},
\]
where \(j^{n-1}\) is the indices of the largest entries of \(x_{\Lambda^n \setminus \{j^{n-1}\}}\) in magnitude. Lemma \(3\) implies that
\[
\|y - Ax^n\|_2^2 \leq \|y - Ax^{n-1}\|_2^2 - \frac{(1 - \delta_s)}{\#(U \setminus \Lambda^{n-1})} \max\{0, \|y - Ax^{n-1}\|_2^2 - \|y - Au\|_2^2\},
\]
where \(U := \supp(u)\). Noting that \(#(U \setminus \Lambda^{n-1}) = 1\), we have
\[
\|y - Ax^n\|_2^2 \leq \|y - Ax^{n-1}\|_2^2 - (1 - \delta_n) \max\{0, \|y - Ax^{n-1}\|_2^2 - \|y - Au\|_2^2\} \leq \|y - Ax^{n-1}\|_2^2 - (1 - \delta_n) \max\{0, \|y - Ax^{n-1}\|_2^2 - \|A\tilde{x}^{n-1}\|_2^2\}.
\]
We claim that
\[
\|y - Ax^{n-1}\|_2^2 \geq \|A\tilde{x}^{n-1}\|_2^2.
\]
Then we have
\[
\|y - Ax^n\|_2^2 \leq \|y - Ax^{n-1}\|_2^2 - (1 - \delta_n) \max\{0, \|y - Ax^{n-1}\|_2^2 - \|A\tilde{x}^{n-1}\|_2^2\} \leq \delta_n \|y - Ax^{n-1}\|_2^2 + (1 - \delta_n) \|A\tilde{x}^{n-1}\|_2^2 \leq \delta_n \|A(x - x^{n-1})\|_2^2 + \|A\tilde{x}^{n-1}\|_2^2 \leq \delta_n (1 + \delta_s) \|x_{\Lambda^n \setminus \{j^{n-1}\}}\|_2^2 + (1 + \delta_s) \|\tilde{x}^{n-1}\|_2^2 \leq (1 + \delta_s) \left(\delta_n + \frac{1}{\alpha^2}\right) \|x_{\Lambda^n \setminus \{j^{n-1}\}}\|_2^2,
\]
Here, for the last inequality, we use the fact of \( \|\tilde{x}^{n-1}\|_2^2 \leq \|x^{\Lambda_{n-1}}\|_2^2/\alpha^2 \) since \( x \) is \( \alpha \)-decaying. On the other hand, we have

\[
\|y - Ax^n\|_2^2 = \|A(x - x^n)\|_2^2 \geq \|A(x - x^n)\|_2^2 \geq (1 - \delta_s)\|x^{\Lambda_{n-1}}\|_2^2.
\]

Combing the results above, we obtain that

\[
\|x^{\Lambda_{n}}\|_2^2 \leq \beta\|x^{\Lambda_{n-1}}\|_2^2
\]

where

\[
\beta = \frac{1 + \delta_s}{1 - \delta_s} \left( \delta_s + \frac{1}{\alpha^2} \right).
\]

Note that \( \delta_s < \sqrt{2} - 1 \) and hence \( 2 - (1 + \delta_s)^2 > 0 \). Then when

\[
\alpha > \sqrt{\frac{1 + \delta_s}{2 - (1 + \delta_s)^2}},
\]

we have

\[
\beta < 1.
\]

And hence,

\[
\|x^{\Lambda_{n}}\|_2^2 < \|x^{\Lambda_{n-1}}\|_2^2,
\]

which implies that \( \Lambda^n \subset \text{supp}(x) \).

We remain to argue that

\[
\|y - Ax^{n-1}\|_2^2 \geq \|Ax^{n-1}\|_2^2.
\]

We assume that

\[
\|y - Ax^{n-1}\|_2^2 < \|Ax^{n-1}\|_2^2,
\]

and we shall derive a contradiction. The RIP property of the matrix \( A \) implies that

\[
(1 - \delta_s)\|x^{\Lambda_{n-1}}\|_2^2 \leq \|y - Ax^{n-1}\|_2^2 < \|Ax^{n-1}\|_2^2 \leq (1 + \delta_s)\|\tilde{x}^{n-1}\|_2^2.
\]

And hence,

\[
\|x^{\Lambda_{n-1}}\|_2^2 \leq \frac{1 + \delta_s}{1 - \delta_s}\|\tilde{x}^{n-1}\|_2^2.
\]

Noting that \( \alpha^2\|\tilde{x}^{n-1}\|_2^2 \leq \|x^{\Lambda_{n-1}}\|_2^2 \), we have

\[
\alpha^2 \leq \frac{1 + \delta_s}{1 - \delta_s},
\]

which contradicts with \( \alpha^2 > \frac{1 + \delta_s}{2 - (1 + \delta_s)^2} \).

\[\square\]

**References**

[1] B. Alexeev, J. Cahill and D. G. Mixon, Full Spark Frames, J. Fourier Anal. Appl., DOI 10.1007/s00041-012-9235-4.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. vol.28, no.3, pp.253-263, 2008.

[3] Blumensath, T., Davies, M.E., Sampling Theorems for Signals From the Union of Finite-Dimensional Linear Subspaces, IEEE Trans. Inform. Theory, vol. 55, no.4, pp.1872-1882, 2009.

[4] T. Cai, L. Wang, and G. Xu, Shifting inequality and recovery of sparse signals. IEEE Trans. Signal Process, vol. 58, no. 3, pp. 1300-1308, 2010.

[5] T. Cai, L. Wang, and G. Xu, New Bounds for Restricted Isometry Constants, IEEE Transactions on Information Theory, vol. 56, no.9, pp. 4388 - 4394, 2010.
[6] E. J. Candès and T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory, vol.51, no. 12, pp.4203-4215, 2005.
[7] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory, vol. 55, no.5, pp.2230-49, 2009.
[8] M. A. Davenport and M. B. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, IEEE Transactions on Information Theory, vol. 56, no.9, pp. 4395-4401, 2010.
[9] S. Foucart, Stability and robustness of weak orthogonal matching pursuits, AMS Spring 2011 Southeastern Conference, Springer Proceedings in Mathematics, 2011.
[10] S. Huang and J. Zhu, Recovery of sparse signals using OMP and its variants: Convergence analysis based on RIP, Inverse problem, vol. 28, no.3, pp. 35003-35016, 2011.
[11] Entao Liu and V. N. Temlyakov, The orthogonal super greedy algorithm and applications in compressed sensing, IEEE Transactions on Information Theory, vol.58, no.4, pp.2040-2047, 2012.
[12] R. Malch, Improved RIP analysis of Orthogonal matching pursuit, arXiv:1102.4311.
[13] Qun Mo and Yi Shen, A remark on the Restricted Isometry Property in Orthogonal Matching Pursuit, IEEE Transaction on Information Theory, vol. 58, no. 6, pp.3654-3656, 2012.
[14] D. Needell and R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, Found. Comput. Math., vol.9, no.3, pp.317-334, 2009.
[15] D. Needell and J. Tropp, CoSamp: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal., vol.26, no.3, pp. 301-321, 2009.
[16] H. Rauhut, On the impossibility of uniform sparse reconstruction using greedy methods, Samppl. Theory Signal Image Process., vol.7, no.2, pp.197-215, 2008.
[17] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, arXiv:1011.3027.
[18] Jian Wang, Seokbeop Kwon and Byonghyo Shim, Generalized Orthogonal Matching Pursuit, IEEE Transaction on Signal Processing, vol. 60, no. 12, pp.6202-6216, 2012.
[19] Z. Xu, Deterministic sampling of sparse trigonometric polynomials, J. Complexity, vol.27, no.2, pp.133-140, 2011.
[20] Tong Zhang, Sparse recovery with orthogonal matching pursuit under RIP, IEEE Transactions on Information Theory, vol.57, No.9, pp.6215-6221, 2011.

Authors’ addresses:
Zhiquiang Xu, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science Chinese Academy of Sciences Beijing 100190, CHINA
Email: xuzq@lsec.cc.ac.cn