Almost Optimality of the Orthogonal Super Greedy Algorithm for $\mu$-Coherent Dictionaries†

Chunfang Shao 1,*, Jincai Chang 1, Peixin Ye 2, Wenhui Zhang 2 and Shuo Xing 2

1 College of Science, North China University of Science and Technology, Tangshan 063210, China; jincai@ncst.edu.cn
2 School of Mathematics and LPMC, Nankai University, Tianjin 300071, China; yepx@nankai.edu.cn (P.Y.); zhwnhui@163.com (W.Z.); xingshuo9853@163.com (S.X.)
* Correspondence: scfaku@163.com
† This work was supported by National Natural Science Foundation of China under Grant 11671213.

Abstract: We study the approximation capability of the orthogonal super greedy algorithm (OSGA) with respect to $\mu$-coherent dictionaries in Hilbert spaces. We establish the Lebesgue-type inequalities for OSGA, which show that the OSGA provides an almost optimal approximation on the first $[1/(18 \mu)]$ steps. Moreover, we improve the asymptotic constant in the Lebesgue-type inequality of OGA obtained by Livshitz E D.

Keywords: orthogonal super greedy algorithm; coherence; best $n$-term approximation; Lebesgue-type inequality

MSC: 41A65; 41A50; 41A46

1. Introduction

Approximation by the sparse linear combination of elements from a fixed redundant system continues to develop actively, which is driven not only by theoretical interest but also by frequent applications from areas such as signal processing and machine learning, cf. [1–7]. This type of approximation is called highly nonlinear approximation. Greedy-type algorithms have been used as a tool for generating such approximations. Among others, the orthogonal greedy algorithm (OGA) has been widely used in practice. In fact, the OGA is regarded as the most powerful algorithm to solve the problem of approximation with respect to redundant systems, cf. [8–10].

We recall some notations and definitions from the theory of greedy algorithms. Let $\mathcal{H}$ be a Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$. We say that a set $D$ of elements from $\mathcal{H}$ is a dictionary if $g \in D \Rightarrow \|g\| = 1$, and $\text{span}D = \mathcal{H}$.

We consider redundant dictionaries, which have been utilized frequently in the field of signal processing. Here, a redundant dictionary means that the elements of the dictionary may be linearly dependent.

We now recall the definition of the OGA from [1].

ORTHOGONAL GREEDY ALGORITHM (OGA)

Set $f_0 := f \in \mathcal{H}, G_0^{\text{OGA}}(f, D) := 0$. For each $m \geq 0$, we inductively find $g_{m+1} \in D$ such that

$$\langle f_m, g_{m+1} \rangle = \sup_{g \in D} |\langle f_m, g \rangle|$$

and define

$$G_m^{\text{OGA}}(f, D) := \text{proj}_{\text{span}\{g_1, g_2, \cdots, g_m\}}(f).$$
\[ f_{m+1} := f - G_{m+1}^{\text{OSGA}}(f, D), \]

where \( P_{\text{span}\{g_1, g_2, \ldots, g_m\}} \) is the operator of the orthogonal projection onto \( \text{span}\{g_1, g_2, \ldots, g_m\} \).

In [11], Liu and Temlyakov proposed the orthogonal super greedy algorithm (OSGA). The OSGA selects more than one element from a dictionary in each iteration step and hence reduces the computational burden of the conventional OGA. Therefore, the OSGA is more efficient than the OGA from the viewpoint of the computational complexity.

**ORTHOGONAL SUPER GREEDY ALGORITHM (OSGA)**

Set \( f_0 := f \in \mathcal{H}, G_0^{\text{OSGA}}(f, D) := 0. \) For a natural number \( s \geq 1 \) and each \( m \geq 0 \), we inductively define:

1. \( g_{s(m-1)+1}, g_{s(m-1)+2}, \ldots, g_{s m} \in \mathcal{D} \) are elements of the dictionary \( \mathcal{D} \), satisfying the following inequality. Denote \( I_m = [(m-1)s+1, ms] \) and assume that
   \[ \min_{i \in I_m} |\langle f_{m-1}, g_i \rangle| \geq \sup_{g \in \mathcal{D}, g \neq g, i \in I_m} |\langle f_{m-1}, g_i \rangle| \cdot \]

2. Let \( H_m := H_m(f) := \text{span}\{g_1, g_2, \ldots, g_{sm}\} \) and \( P_{H_m} \) denote the operator of the orthogonal projection onto \( H_m \). Define
   \[ G_m(f) := G_m(f, D) := G_n(f, D) := P_{H_m}(f). \]

3. Define the residual after the \( m \)-th iteration of the algorithm
   \[ f_m := f_m := f - G_m(f, D). \]

Note that, in the case \( s = 1 \), OSGA(s) coincides with OGA.

In this paper, we study the approximation capability of the OSGA with respect to \( \mu \)-coherent dictionaries in Hilbert spaces. We denote by

\[ \mu = \mu(\mathcal{D}) = \sup_{g \neq h, g, h \in \mathcal{D}} |\langle g, h \rangle| \]

the coherence of a dictionary. The coherence \( \mu \) is a blunt instrument to measure the redundancy of dictionaries. It is clear that if \( \mathcal{D} \) is an orthonormal basis, then \( \mu(\mathcal{D}) = 0 \). The smaller the \( \mu(\mathcal{D}) \), the more the \( \mathcal{D} \) resembles an orthonormal basis. We study dictionaries with small values of coherence \( \mu(\mathcal{D}) > 0 \), and call them \( \mu \)-coherent dictionaries.

In [11], the authors found that such computational burden reduction of OSGA does not degrade the approximation capability if \( f \) belongs to the closure of the convex hull of the symmetrized dictionary \( \mathcal{D}^\pm := \{ \pm g, g \in \mathcal{D} \} \), which is denoted by \( A_1(\mathcal{D}) \).

**Theorem 1.** Let \( \mathcal{D} \) be a dictionary with coherence parameter \( \mu := \mu(\mathcal{D}) \). Then, for \( s \leq (2\mu)^{-1} \), the algorithm OSGA(s) provides an approximation of \( f \in A_1(\mathcal{D}) \) with the following error bound:

\[ \|f_m^n\|^2 \leq 40.5(sm)^{-1}, \quad m = 1, 2, \ldots. \]

It seems that a dimensional independent convergence rate was deduced, but the condition that the target element belongs to \( A_1(\mathcal{D}) \) becomes more and more stringent as the number of the elements in \( \mathcal{D} \) grows, cf. [2].

Fang, Lin, and Xu [12] studied the behavior of OSGA for \( f \in \mathcal{H} \). They defined \( \mathcal{L}_1 := \{ f : f = \sum_{g \in \mathcal{D}} a_g g \} \) and \( \|f\|_{\mathcal{L}_1} := \inf \{ \sum_{g \in \mathcal{D}} |a_g| : f = \sum_{g \in \mathcal{D}} a_g g \} \) for \( f \in \mathcal{L}_1 \), and obtained the following theorem.

**Theorem 2.** Let \( \mathcal{D} \) be a dictionary with coherence \( \mu \). Then, for all \( f \in \mathcal{H}, h \in \mathcal{L}_1 \) and arbitrary \( s \leq (2\mu)^{-1} + 1 \), the OSGA(s) provides an approximation of \( f \) with the error bound:

\[ \|f_m^n\|^2 \leq \|f - h\|^2 + \frac{27}{2} \|h\|_{\mathcal{L}_1}^2 (sm)^{-1}, \quad k = 1, 2, \ldots. \]
The $\mu$-coherence of a dictionary is used in OSGA, which implies that computational burden reduction does not degenerate the approximation capability. Moreover, if $\mu > \frac{1}{2}$, then OSGA coincides with OGA.

Let $\Sigma_m$ denote the collection of elements in $\mathcal{H}$, which can be expressed as a linear combination of, at most, $m$ elements of the dictionary $\mathcal{D}$, namely

$$\Sigma_m := \Sigma_m(\mathcal{D}) = \{ g : g = \sum_{i \in \Lambda} c_i g_i, g_i \in \mathcal{D}, \Lambda \subset \mathbb{N}, \#(\Lambda) \leq m \}.$$ 

For an element $f \in \mathcal{H}$, we define its best $m$-term approximation error by

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{g \in \Sigma_m} \| f - g \|.$$ 

The inequality connecting the error of greedy approximation and the error of best $m$-term approximation is called the Lebesgue-type inequality, cf. [13–15]. In this paper, we will establish the Lebesgue-type inequalities for OSGA with respect to $\mu$-coherent dictionaries.

We first recall some results on the efficiency of OGA with respect to $\mu$-coherent dictionaries. These results relate the error of OGA’s $A(m)$-th approximation to the error of the best $m$-term approximation with an extra multiplier:

$$\| f_{A(m)}(f, \mathcal{D}) \| \leq B(m) \sigma_m(f, \mathcal{D}) \quad \text{for} \quad m \leq C(\mu),$$  

where $A(m) \in \mathbb{N}, B(m), C(\mu) \in \mathbb{R}$. Gillbert, Muthukrishnan, and Strauss [16] gave the first Lebesgue-type inequality for OGA. They proved

$$\| f_m \| \leq 8 m^{\frac{1}{2}} \sigma_m(f, \mathcal{D}) \quad \text{for} \quad 1 \leq m \leq \frac{1}{8\sqrt{2\mu}} - 1.$$ 

The constant in the above inequality was improved by Tropp in [17]:

$$\| f_m \| \leq (1 + 6m)^{\frac{1}{2}} \sigma_m(f) \quad \text{for} \quad 1 \leq m \leq \frac{1}{3\mu}.$$ 

Donoho, Elad, and Temlyakov [18] dramatically improved the factor in front of $\sigma_m$ and obtained that

$$\| f_m \| \leq 24 \sigma_m(f) \quad \text{for} \quad 1 \leq m \leq \frac{1}{20\mu^2},$$ 

where the constant 24 is not the best. Many researchers have sought to improve the factor $B(m)$. Temlyakov and Zheltov improved the above inequality in [4]. They obtained

$$\| f_{m\lfloor 2\sqrt{\log m} \rfloor}(f, \mathcal{D}) \| \leq 3 \sigma_m(f, \mathcal{D}) \quad \text{for} \quad m \lfloor 2\sqrt{\log m} \rfloor \leq \frac{1}{26\mu}.$$ 

Livshitz [19] took the parameters $A(m) := 2m, B(m) := 2.7, C(\mu) := \frac{1}{20}$ in (1) and obtained the following profound result.

**Theorem 3.** For every $\mu$-coherent dictionary $\mathcal{D}$ and any $f \in \mathcal{H}$, the OGA applied to $f$ provides

$$\| f_{2m} \| \leq 2.7 \sigma_m(f, \mathcal{D}) \quad \text{for} \quad 1 \leq m \leq \frac{1}{20\mu}.$$ 

By using the same method as in [19], Ye and Wei [20] improved slightly the constant 2.7. Based on the above works, we give the error bound of the form (1) for OSGA with respect to dictionaries with small but non-vanishing coherence.
Theorem 4. Let $D$ be a dictionary with coherence $\mu$. Then, for any $f \in \mathcal{H}$ and any $\epsilon > 0$, the OSGA(s) applied to $f$ provides

$$\|f_{Am}\| \leq 2.24(1 + \epsilon)\sigma_m(f)$$

for all $1 \leq m \leq \frac{1}{18\mu}$, $\frac{3}{100} \leq \mu \leq \frac{1}{18}$ and an absolute constant $A \geq 2$.

Remark 1.
1. We remark that the values of $\mu$ and $A$ for which (2) holds are coupled. For example, it is possible to obtain a smaller value of $\mu$ at the price of a larger value of $A$. Moreover, for sufficiently large $A$, $\mu$ can be arbitrarily close to zero.
2. Our results improve Theorem 3 only in the asymptotic constant and not in the rate. Under the condition of Theorem 4, for $s = 1$, taking $(A, \epsilon, \mu)$ as $(2, 0.1, 0.03)$, we can obtain $\|f_{2m}\| \leq 2.5\sigma_m(f, D)$. Comparing it with Theorem 3, the constant that we obtain is better.
3. The specific constant 2.24 in (2) is not the best. By adjusting parameters $A$ and $\mu$, we can obtain a more general estimation:

$$\|f_{Am}\| \leq C(A)(1 + \epsilon)\sigma_m(f)$$

for $\alpha_A \leq \mu \leq \frac{1}{18}$, where $C(A)$ and $\alpha_A$ are interdependent. Thus, Theorem 4 shows that OSGA(s) can achieve an almost optimal approximation on the first $\left[\frac{1}{18\mu s}\right]$ steps for dictionaries with small but non-vanishing coherence.

The paper is organized as follows. In Section 2, we establish several preliminary lemmas. In Section 3, for some closed subspace $L$ of $\mathcal{H}$, as defined below, we first give the estimations of $P_L^\perp(f_n)$ in different situations based on the lemmas in Section 2. Then, we estimate the $P_L(f_n)$. Finally, combining the above two estimations, we provide the detailed proof of Theorem 4. In Section 4, we test the performance of the OSGA in the case of finite dimensional Euclidean space. In Section 5, we make some concluding remarks on our work.

2. Preliminary Lemmas

In this section, we will introduce several quantities and discuss their properties, which are important to the proof of our main result. By the condition of Theorem 4, we have

$$\mu \leq m\mu s \leq Am\mu s \leq \frac{1}{18}.$$ 

We establish three preliminary lemmas.

Lemma 1. Let $n \leq Am$, $h \in \mathcal{H}$, $g_i \in D$, $1 \leq i \leq n$. Assume that

$$P_{\text{span}\{g_1, g_2, \ldots, g_n\}}^\perp h = \sum_{i=1}^{n} c_i g_i,$$

then, we have

$$\max_{1 \leq i \leq n} |c_i| \leq k_1 \max_{1 \leq i \leq n} |\langle h, g_i \rangle|,$$

where $k_1 = \frac{1}{4 - \mu Am} \leq \frac{18}{17}$. 

Proof. For any \( g_l \in \mathcal{D}, 1 \leq l \leq n \), we have
\[
|\langle h, g_l \rangle| = |\left\langle P_{\text{span}}(g_1, \ldots, g_n) h, g_l \right\rangle|
\]
\[
= \left| \sum_{i=1}^n c_i g_i, g_l \right|
\]
\[
\geq |c_l| - \sum_{i \neq l} |c_i| |\langle g_i, g_l \rangle|
\]
\[
\geq |c_l| - \mu \max_{1 \leq i \leq n} |c_i|(n - 1).
\]
This implies
\[
\max_{1 \leq i \leq n} |c_i| \leq \frac{1}{1 - \mu(n - 1)} \max_{1 \leq i \leq n} |\langle h, g_i \rangle| \leq \frac{1}{1 - \mu Ams} \max_{1 \leq i \leq n} |\langle h, g_i \rangle|,
\]
where \( k_1 = \frac{1}{1 - \mu Ams} \leq \frac{18}{17} \).

For \( \epsilon > 0 \), by the definition of \( \sigma_m(f) \), there exist \( b_j \in \mathbb{R}, \psi_j \in \mathcal{D}, 1 \leq j \leq m \) such that
\[
\| f - \sum_{j=1}^m b_j \psi_j \| \leq (1 + \epsilon) \sigma_m(f).
\]
(3)

For \( 1 \leq n \leq Am \), we set
\[
d_n := \sum_{j \in I_n} |\langle f_{n-1}, g_j \rangle|.
\]
(4)

Assume that \( x_{i,n} n \geq 1, 1 \leq i \leq ns \), satisfying the equation
\[
f_n = f_{n-1} - P_{H_n}(f_{n-1}) = f_{n-1} - \sum_{i=1}^{ns} x_{i,n} g_i.
\]
(5)

Next, we give the estimates of \( \{x_{i,n}\}_{i=1}^{ns} \) and \( \{d_n\} \) for \( n \geq 1 \) in turn. Applying Lemma 1, we have the following estimates for \( x_{i,n} \).

Lemma 2. For \( n \leq Am \), we have
\[
|x_{i,n}| \leq k_1 \mu d_n \quad \text{for} \quad 1 \leq i \leq (n - 1)s;
\]
\[
|x_{i,n} - \langle f_{n-1}, g_i \rangle| \leq k_1 \mu d_n \quad \text{for} \quad (n - 1)s + 1 \leq i \leq ns.
\]

Proof. Define
\[
h = f_{n-1} - \sum_{i \in I_n} \langle f_{n-1}, g_i \rangle g_i.
\]
Since
\[
f_n = f_{n-1} - P_{H_n}(f_{n-1})
\]
\[
= f_{n-1} - \sum_{i \in I_n} \langle f_{n-1}, g_i \rangle g_i - P_{H_n}(f_{n-1} - \sum_{i \in I_n} \langle f_{n-1}, g_i \rangle g_i)
\]
\[
= h - P_{H_n}(h)
\]
(6)

and, for any \( 1 \leq j \leq (n - 1)s \), \( \langle f_{n-1}, g_j \rangle = 0 \), we have
\[
|\langle h, g_j \rangle| = |\left\langle f_{n-1} - \sum_{i \in I_n} \langle f_{n-1}, g_i \rangle g_i, g_j \right\rangle| \leq \sum_{i \in I_n} |\langle f_{n-1}, g_i \rangle| \mu = \mu d_n.
\]
(7)
Proof. For $(n-1)s + 1 \leq j \leq ns$, we have

\[
\langle h, g_j \rangle = \left| \langle f_{n-1} - \sum_{i \in I_n} < f_{n-1}, g_i > g_i, h > \rangle \right|
\]

\[
= \langle f_{n-1}, g_j \rangle - \sum_{i \in I_n} \langle f_{n-1}, g_i \rangle \langle g_i, h \rangle
\]

\[
= \langle f_{n-1}, g_j \rangle - \langle f_{n-1}, g_i \rangle - \sum_{i,j \in I_n, i \neq j} \langle f_{n-1}, g_i \rangle \langle g_i, g_j \rangle
\]

\[
\leq \sum_{i,j \in I_n, i \neq j} \langle f_{n-1}, g_i \rangle \langle g_i, g_j \rangle
\]

\[
\leq \mu d_n.
\]

Let $P_{H_n}(h) = \sum_{i=1}^{ns} x_{i,n}^i g_i$. Combining (5) with (6), we have

\[
f_{n-1} = \sum_{i \in I_n} (f_{n-1}, g_i) g_i - \sum_{i=1}^{ns} x_{i,n}^i g_i = f_n - f_{n-1} - \sum_{i=1}^{ns} x_{i,n} g_i.
\]

Thus, for $1 \leq i \leq (n-1)s, x_{i,n} = x_{i,n}^i$; for $(n-1)s + 1 \leq i \leq ns, (f_{n-1}, g_i) + x_{i,n}^i = x_{i,n}$. By Lemma 1 and inequalities (7) and (8), we obtain

\[
\max_{1 \leq i \leq ns} |x_{i,n}^i| \leq k_1 \max_{1 \leq i \leq ns} |\langle h, g_i \rangle| \leq k_1 \mu d_n.
\]

Thus, for $1 \leq i \leq (n-1)s$,

\[
|x_{i,n}| = |x_{i,n}^i| \leq \max_{1 \leq i \leq ns} |x_{i,n}^i| \leq k_1 \mu d_n;
\]

for $(n-1)s + 1 \leq i \leq ns$,

\[
|x_{i,n} - (f_{n-1}, g_i)| = |x_{i,n}^i| \leq \max_{1 \leq i \leq ns} |x_{i,n}^i| \leq k_1 \mu d_n.
\]

We proceed to the estimate of $\{d_n\}$.

Lemma 3. For any $1 \leq l \leq n \leq Am + 1$, we have

\[
d_n \leq k_2 d_l,
\]

where $k_2 = \exp \left( Amsu \left( 1 + \frac{Amu}{1 - Amsu} \right) \right) \leq \exp \left( \frac{1}{17} \right)$.

Proof. For $1 \leq n \leq Am$, according to the definition of $d_n$, we have

\[
d_{n+1} = \sum_{i \in I_{n+1}} |\langle f_n, g_i \rangle| \leq s |\langle f_n, g_{ns+1} \rangle|
\]

\[
= s |(f_n - P_{H_n}(f_n-1), g_{ns+1})|
\]

\[
= s |\left( f_{n-1} - \sum_{i=1}^{ns} x_{i,n} g_i, g_{ns+1} \right)|
\]

\[
\leq s |\langle f_{n-1}, g_{ns+1} \rangle| + \sum_{i=1}^{ns} |x_{i,n}| |\langle g_i, g_{ns+1} \rangle|.
\]
We continue to estimate the two summands of the right-hand side of the above inequality. For the first summand, the greedy step implies
\[
|s\langle f_{n-1}, g_{ns+1} \rangle| \leq s|\langle f_{n-1}, g_i \rangle| \leq \sum_{i=1}^{l_n} |\langle f_{n-1}, g_i \rangle| = d_n. \tag{10}
\]

For the second summand, by Lemma 1, we have
\[
\sum_{i=1}^{n} |x_{i,n}||\langle g_i, g_{ns+1} \rangle| \leq s\mu \sum_{i=1}^{n} |x_{i,n}| \leq s\mu(\sum_{i=1}^{n} |x_{i,n}| + \sum_{i=(n-1)s+1}^{ns} |x_{i,n}|) \leq s\mu((n-1)s_k + \sum_{i \in I_n} (k_1\mu d_n + |\langle f_{n-1}, g_i \rangle|)) = s\mu((n-1)s_k d_n + s_1 = d_n \mu (1 + nsk_1\mu) \leq d_n \mu (1 + Ams_1 k_1). \tag{11}
\]

Combining inequalities (9)–(11) with Lemma 2, we conclude that
\[
d_{n+1} \leq d_n + \left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s = \left[1 + \left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s\right] d_n.
\]

Thus, for any \(n\) and \(1 \leq l \leq n \leq Am + 1\), we have
\[
d_n \leq \left[1 + \left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s\right] d_{n-1} \leq \cdots \leq \left[1 + \left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s\right]^{n-l} d_l \leq \left[1 + \left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s\right]^{Am} d_l \leq \exp\left(Am\left(1 + \frac{Ams_1}{1 - Ams_1}\right) d_n \mu s\right) d_l = k_2 d_l,
\]

where \(k_2 := \exp\left(Ams_1\left(1 + \frac{Ams_1}{1 - Ams_1}\right)\right) \leq \exp\left(\frac{1}{17}\right). \tag{12}\)

3. **Proof of Theorem 4**

Based on the above preliminary lemmas, we will prove Theorem 4 step by step. We first introduce some notations. Define
\[
L := \text{span}(\psi_1, \ldots, \psi_m), f_0 = f, \xi_n := P_L^\perp(f_n), 0 \leq n \leq Am.
\]
\[
T_1 := \{i \in \{1, \ldots, Am\} : g_i \in \{\psi_j\}_{j=1}^m\}, T_2 := \{1, \ldots, Am\} \setminus T_1.
\]

For \(d_n\), we define
\[
D := \sum_{1 \leq i \leq Am, 2 \cap i \neq 0} d_i^2.
\]

Let \(a_{j,n} \in \mathbb{R}, 1 \leq j \leq m, 0 \leq n \leq Am\) satisfy the following equations
\[
P_L(f_n) = \sum_{j=1}^{m} a_{j,n} \psi_j. \tag{12}
\]

Thus, for \(f_n, 0 \leq n \leq Am\), we have
\[
f_n := P_L(f_n) + P_L^\perp(f_n) = \sum_{j=1}^{m} a_{j,n} \psi_j + \xi_n.
\]
To obtain the upper bound of $\|f_n\|$, it suffices to estimate $\|\xi_n\|$ and $\|P_L(f_n)\|$. By the definitions of sets $T_1, T_2$ and $I_n$ in OSGA, we first give the estimate of $\|\xi_n\|$ according to whether the intersection of $T_2$ and $I_n$ is an empty set.

**Theorem 5.** Let $n$ satisfy $1 \leq n \leq Am$ and $I_n \cap T_2 = \emptyset$. Then,
\[
\|\xi_n\| \leq \|\xi_{n-1}\| + 0.22D\mu.
\]

**Proof.** Let
\[
\Lambda_n := \bigcup_{i=1}^n I_i, \quad T_2^n := T_2 \cap \Lambda_n,
\]
By Lemma 3, for $1 \leq l \leq n \leq Am$,
\[
d_n \leq k_2 \min_{l \cap T_2^l \neq \emptyset} d_l.
\]
Then, we have
\[
s^{-1}t_n(\min_{l \cap T_2^l \neq \emptyset} d_l)^2 \leq \sum_{l \cap T_2^l \neq \emptyset} (\sum_{i \in I_l} |\langle f_{l-1}, g_l \rangle|)^2 = \sum_{l \cap T_2^l \neq \emptyset} d_l^2 = D,
\]
so, we can obtain that
\[
s^{-1}t_n d_n^2 \leq k_2^2 D. \tag{13}
\]
Since $I_n \cap T_2 = \emptyset$, we obtain $t_n = t_{n-1}$. We define
\[
h := \sum_{i \in T_2^n} x_{i,n}g_i = \sum_{i \in T_2^{n-1}} x_{i,n}g_i. \tag{14}
\]
Note that
\[
f_n = f - P_{H_n}(f) = f_{n-1} + P_{H_{n-1}}(f) - P_{H_n}(f_{n-1} + P_{H_{n-1}}(f)) = f_{n-1} + P_{H_n}(f_{n-1}).
\]
By the definitions of $L, T_1, T_2, \Lambda_n$ and the expression of (14), we have $P_L^+(P_{H_n}(f_{n-1})) = P_L^+(h)$. Then, we obtain
\[
\|\xi_n\|^2 = \|P_L^+(f_n)\|^2 = \|P_L^+(f_{n-1} - P_{H_n}(f_{n-1}))\|^2 = \|\xi_{n-1} - P_L^+(h)\|^2 \leq \|\xi_{n-1}\|^2 + 2|\langle \xi_{n-1}, P_L^+(h) \rangle| + \|P_L^+(h)\|^2. \tag{15}
\]
To obtain the final result, it suffices to estimate the upper bounds of $|\langle \xi_{n-1}, P_L^+(h) \rangle|$ and $\|P_L^+(h)\|^2$.

For $|\langle \xi_{n-1}, P_L^+(h) \rangle|$, by (12) and (14), we have
\[
|\langle \xi_{n-1}, P_L^+(h) \rangle| = \left| \left\langle \sum_{j=1}^m a_{j,n-1} \psi_j, \sum_{i \in T_2^{n-1}} x_{i,n}g_i \right\rangle \right|
\leq \sum_{j=1}^m |a_{j,n-1}| \left| \sum_{i \in T_2^{n-1}} x_{i,n}(\psi_j, g_i) \right|,
\]
where we have used the fact $<f_{n-1}, h> = 0$. 
On the one hand, for any $1 \leq l \leq m$ and $n$ satisfying $T_2 \cap I_n = \emptyset$, we obtain
\begin{equation}
\left| \sum_{j=1}^{m} a_{j,n-1} \Psi_j \right| = \left| \sum_{j=1}^{m} a_{j,n-1} \Psi_j + \xi_{n-1} \Psi_l \right| = \left| \langle f_{n-1}, \Psi_l \rangle \right| \\
\leq \max_{I_n} \left| \langle f_{n-1}, \xi_i \rangle \right| \leq \sum_{I_n} \left| \langle f_{n-1}, \xi_i \rangle \right| = d_n.
\end{equation}

Thus, by Lemma 1 and inequality (17), we obtain
\begin{equation}
\sum_{j=1}^{m} |a_{j,n-1}| \leq m \cdot \max_{1 \leq j \leq m} |a_{j,n-1}| \leq mk_1 \max_{1 \leq j \leq m} \left| \langle f_{n-1}, \Psi_j \rangle \right| \\
\leq mk_1 \max_{1 \leq j \leq m} \left| \langle P_L (f_{n-1}) + \xi_{n-1}, \Psi_j \rangle \right| \\
= mk_1 \max_{1 \leq j \leq m} \left| \sum_{i=1}^{m} a_{i,n-1} \Psi_i \right| \leq mk_1 d_n.
\end{equation}

On the other hand, by Lemma 2, we have, for $1 \leq j \leq m$,
\begin{equation}
\sum_{T_2 \cap I_n} |x_{i,n} \langle \Psi_j, \xi_i \rangle | \leq \mu t_{n-1} \max_{i \in T_2 \cap I_n} |x_{i,n}| \leq \mu t_{n-1} k_1 \mu d_n.
\end{equation}

Thus, substituting (18) and (19) into (16), and then combining it with (13), we get the estimate
\begin{equation}
\left| \langle \xi_{n-1}, P_L^+ (h) \rangle \right| \leq \sum_{j=1}^{m} |a_{j,n-1}| \cdot \sum_{T_2 \cap I_n} |x_{i,n} \langle \Psi_j, \xi_i \rangle | \\
\leq m \mu^2 k_1^2 d_n^2 t_{n-1} \\
\leq m \mu^2 k_1^2 s^2 \mu^2 t_{n-1} \\
\leq m \mu^2 k_1^2 D \mu.
\end{equation}

Finally, we estimate $\|P_L^+ (h)\|^2$.

Note that
\begin{equation}
\|P_L^+ (h)\|^2 = \|h\|^2 = \| \sum_{i \in T_2 \cap I_n} x_{i,n} \xi_i \|^2 \\
= \sum_{i \in T_2 \cap I_n} x_{i,n}^2 + \sum_{i,j \in T_2 \cap I_n, i \neq j} |x_{i,n}| |x_{j,n}| |\langle \xi_i, \xi_j \rangle | \\
\leq \left( \max_{i \in T_2 \cap I_n} x_{i,n}^2 \right) (t_{n-1} + t_{n-1}^2) \\
\leq k_1^2 \mu^2 d_n^2 [t_n + t_n^2].
\end{equation}

By using (13), we have
\begin{equation}
\|P_L^+ (h)\|^2 \leq k_1^2 \mu^2 s k_2^2 D (1 + Ams \mu) = k_1^2 \mu s (1 + Ams \mu) k_2^2 D \mu.
\end{equation}

Combining (15) and (20) with (21), we get
\begin{equation}
\|\xi_n\|^2 \leq \|\xi_{n-1}\|^2 + 2 \langle \xi_{n-1}, P_L^+ (h) \rangle + \|P_L^+ (h)\|^2 \\
\leq \|\xi_{n-1}\|^2 + 2 m \mu s + \mu s (1 + Ams \mu) k_2^2 D \mu \\
\leq \|\xi_{n-1}\|^2 + 0.22 D \mu.
\end{equation}
Theorem 5 gives the estimation of \( \| \xi_n \| \) in the situation \( I_n \cap T_2 = \emptyset \). The following theorem deals with the situation \( I_n \cap T_2 \neq \emptyset \).

**Theorem 6.** Let \( n \) satisfy \( 1 \leq n \leq Am \) and \( I_n \cap T_2 = \emptyset \). Then,

\[
\| \xi_n \| \leq \| \xi_{n-1} \| ^2 - 0.8937d_n^2.
\]

**Proof.** Since

\[
\xi_n = P_L^\perp (f_n) = P_L^\perp (f_{n-1} - \sum_{i=1}^{ns} x_i \xi g_i)
\]

\[
= P_L^\perp (f_{n-1} - \sum_{i \in I_n} x_i \xi g_i - \sum_{i \in \Lambda_{n-1}} x_i \xi g_i)
\]

\[
= P_L^\perp (f_{n-1} - \sum_{i \in I_n} x_i \xi g_i) - P_L^\perp (\sum_{i \in \Lambda_{n-1}} x_i \xi g_i),
\]

we set \( \xi_n' = P_L^\perp (f_{n-1} - \sum_{i \in I_n} x_i \xi g_i) \), \( h = \sum_{i \in T_2^\perp} x_i \xi g_i \) and write \( \xi_n \) as

\[
\xi_n = \xi_n' - P_L^\perp (h).
\]

According to the following inequality,

\[
\| \xi_n \| ^2 = \| \xi_n' \| ^2 - 2 \langle \xi_n', P_L^\perp (h) \rangle + \| P_L^\perp (h) \| ^2 \leq \| \xi_n' \| ^2 - 2 \langle \xi_n', h \rangle + \| h \| ^2,
\]

we need to estimate \( \| \xi_n' \| ^2 \), \( \langle \xi_n', h \rangle \) and \( \| h \| ^2 \). We first estimate \( \| h \| ^2 \) by

\[
\| h \| ^2 = \| \sum_{i \in T_2^\perp} x_i \xi g_i \| ^2
\]

\[
= k_1^2 \mu^2 d_n^2 (t_n + t_n^2 \mu)
\]

\[
\leq k_1^2 \mu^2 d_n^2 (Ams + (Ams)^2 \mu)
\]

\[
= k_1^2 \mu Am's (1 + Am's) d_n^2
\]

\[
\leq \left( \frac{18}{17} \right) ^2 \frac{1}{18} \frac{19}{18} d_n^2 = \frac{19}{5202} d_n^2 \leq 0.0037 d_n^2.
\]

Next, we continue to estimate \( \| \xi_n' \| ^2 \). It is not difficult to see that

\[
\| \xi_n' \| ^2 = \| P_L^\perp (f_{n-1} - \sum_{i \in I_n} x_i \xi g_i) \| ^2
\]

\[
= \| \xi_{n-1} - P_L^\perp (\sum_{i \in I_n} x_i \xi g_i) \| ^2
\]

\[
= \| \xi_{n-1} \| ^2 - 2 \langle \xi_{n-1} - \sum_{i \in I_n} x_i \xi g_i, \sum_{i \in I_n} x_i \xi g_i \rangle + \| P_L^\perp (\sum_{i \in I_n} x_i \xi g_i) \| ^2
\]

\[
\leq \| \xi_{n-1} \| ^2 - 2 \sum_{i \in I_n} x_i \xi \langle \xi_{n-1}, \xi g_i \rangle + \| \sum_{i \in I_n} x_i \xi g_i \| ^2.
\]

Note that

\[
\sum_{i \in I_n} x_i \xi \langle \xi_{n-1}, \xi g_i \rangle
\]

\[
= \sum_{i \in I_n} (x_i \xi - \langle f_{n-1}, g_i \rangle) + \langle f_{n-1}, g_i \rangle \langle \langle \xi_{n-1}, g_i \rangle - \langle f_{n-1}, g_i \rangle, \langle f_{n-1}, g_i \rangle \rangle.
\]
By (18), for any $i \in T_2 \cap I_n$, we have

$$|\langle \xi_{n-1}, \xi_i \rangle - \langle f_{n-1}, \xi_i \rangle| = \left| f_{n-1} - \sum_{j=1}^{m} a_{j,n-1} \psi_j \xi_i \right| - |\langle f_{n-1}, \xi_i \rangle|$$

Thus, combining (27) with (28), for

$$\sum_{i \in I_n} x_{i,n} \langle \xi_{n-1}, \xi_i \rangle \geq \sum_{i \in I_n} (|\langle f_{n-1}, \xi_i \rangle| - k_1 \mu d_n) (|\langle f_{n-1}, \xi_i \rangle| - m k_1 \mu d_n)$$

$$= \sum_{i \in I_n} (|\langle f_{n-1}, \xi_i \rangle|)^2 - (k_1 \mu d_n + m k_1 \mu d_n) (|\langle f_{n-1}, \xi_i \rangle|) + m k_1^2 \mu^2 d_n^2$$

$$= \sum_{i \in I_n} (|\langle f_{n-1}, \xi_i \rangle|)^2 - \sum_{i \in I_n} (k_1 \mu d_n + m k_1 \mu d_n) (|\langle f_{n-1}, \xi_i \rangle|) + \sum_{i \in I_n} m k_1^2 \mu^2 d_n^2$$

$$\geq s^{-1} d_n^2 - (1 + m) k_1 \mu d_n^2$$

$$\geq (9A - k_1)(1 + m) \mu d_n^2 \geq 2 \mu (9A - k_1) d_n^2 \geq \frac{576}{17} \mu d_n^2$$

for $0 \leq s \leq \frac{18}{17A} \leq \frac{1}{9(1 + m)A}$, and $m \geq 1, A \geq 2$.

For the last summand of the right-hand side of the inequality in (24), we have

$$\left( \sum_{i \in I_n} |x_{i,n}| \right)^2 \leq \left( \sum_{i \in I_n} (|\langle f_{n-1}, \xi_i \rangle| + k_1 \mu d_n) \right)^2$$

$$\leq (1 + k_1 \mu s)^2 d_n^2$$

$$\leq \left( \frac{18}{17} \right)^2 d_n^2.$$
We next estimate $|\langle \xi_n', h \rangle|$. Since

$$ |\langle \xi_n', h \rangle| = |\left\langle P_L^{+} (f_{n-1} - \sum_{i \in I_n} x_{i,n} g_i), h \right\rangle| = |\left\langle \xi_{n-1} - \sum_{i \in I_n} x_{i,n} P_L^{+} (g_i), h \right\rangle| $$

$$ = |\left\langle f_{n-1} - \sum_{j=1}^{m} a_{j,n-1} \psi_j - \sum_{i \in I_n} x_{i,n} P_L^{+} (g_i), h \right\rangle| $$

$$ \leq \left| \sum_{j=1}^{m} a_{j,n-1} \langle \psi_j, h \rangle \right| + \left| \sum_{i \in I_n} x_{i,n} \langle P_L^{+} (g_i), h \rangle \right| $$

(30)

we need to give the upper bounds of $A$ and $B$. By (18) and (19), we have

$$ A = \sum_{j=1}^{m} |a_{j,n-1}||\langle \psi_j, h \rangle| \leq \sum_{j=1}^{m} |a_{j,n-1}| \sum_{i \in T_2^{-1}} |x_{i,n}||\langle \psi_j, g_i \rangle| $$

$$ \leq mk_1 d_n \mu - k_1 \mu d_n \leq mk_1 d_n \mu Amsk_1 \mu d_n \leq k_1^2 (Ams) (m \mu) d_n^2 $$

$$ \leq 0.0035 d_n^2. $$

As for $B$, since for $1 \leq j \leq (n-1)s < i \leq ns \leq Ams, i, j \in T_2, P_L (g_i) = \sum_{l=1}^{m} c_i \psi_l$, by Lemma 1, we know that

$$ \max_{1 \leq l \leq m} |c_l| \leq k_1 \max_{1 \leq l \leq m} |\langle g_i, \psi_l \rangle| \leq k_1 \mu $$

(32)

and

$$ |\langle P_L^{+} (g_i), g_j \rangle| = |\langle g_i - P_L (g_i), g_j \rangle| = |\left\langle g_i - \sum_{l=1}^{m} c_l \psi_l, g_j \right\rangle| $$

$$ \leq |\langle g_i, g_j \rangle| + \sum_{l=1}^{m} |c_l||\langle \psi_l, g_j \rangle|. $$

(33)

Combining (32) with (33), we have

$$ |\langle P_L^{+} (g_i), g_j \rangle| \leq \mu + m \max_{1 \leq l \leq m} |c_l| \mu \leq \mu + m k_1 \mu \leq \frac{18}{17} \mu. $$

(34)

Using Lemma 1 again, we obtain from (34) that

$$ B = \sum_{i \in I_n} |x_{i,n}||\langle P_L^{+} (g_i), h \rangle| = \sum_{i \in I_n} |x_{i,n}| \sum_{j \in T_2^{-1}} |x_{j,n}||\langle P_L^{+} (g_i), g_j \rangle| $$

$$ \leq \sum_{i \in I_n} (k_1 \mu d_n + |\langle f_{n-1}, g_i \rangle|) \sum_{j \in T_2^{-1}} |x_{j,n}| \frac{18}{17} \mu $$

$$ \leq (k_1 \mu d_n + d_n) k_1 \mu d_n (n-1) \frac{18}{17} \mu $$

$$ \leq (k_1 \mu s + 1) k_1 \mu (Ams) \frac{18}{17} \mu $$

$$ \leq \left( 1 + \frac{18}{17} \right) \frac{18}{17} \mu \frac{18}{17} \mu \frac{18}{17} \mu \frac{18}{17} \mu \leq 0.0035 d_n^2. $$

(35)
Thus, we get the upper bound of $|\langle \xi'_n, h \rangle|$ by (30), (31) and (35), i.e.,

$$|\langle \xi'_n, h \rangle| \leq A + B \leq 0.0072d^2_n. \quad (36)$$

Combining (22), (23) and (29) with (36), we have

$$\|\xi'_n\|^2 = \|\xi'_n\|^2 - 2\langle \xi'_n, h \rangle + \|h\|^2$$

$$\leq \|\xi'_n\|^2 + 2|\langle \xi'_n, h \rangle| + \|h\|^2$$

$$\leq \|\xi_{n-1}\|^2 - 0.9118d^2_n + 2(0.0072d^2_n) + 0.0037d^2_n$$

$$= \|\xi_{n-1}\|^2 - 0.8937d^2_n.$$

$\square$

It remains to estimate $\|P_L(f_n)\|$. We first recall a lemma proven by Fang, Lin and Xu in [12].

**Lemma 4.** Assume that a dictionary $D$ has coherence $\mu$. Then, we have, for any distinct $g_i \in D$, $a_i \in \mathbb{R}$, $i = 1, 2, \cdots, s$, the inequalities

$$(1 - \mu(s - 1))\sum_{i=1}^{s} a_i^2 \leq \|\sum_{i=1}^{s} a_i g_i\|^2 \leq (1 + \mu(s - 1))\sum_{i=1}^{s} a_i^2.$$

**Theorem 7.** For any $1 \leq n \leq Am$, we have

$$\|P_L(f_n)\|^2 = \|\sum_{j=1}^{m} a_{j,n} \psi_j\|^2 \leq 1.34D.$$

**Proof.** From Lemma 4, we know that

$$\|P_L(f_n)\|^2 = \|\sum_{j=1}^{m} a_{j,n} \psi_j\|^2 \leq \sum_{j=1}^{m} |a_{j,n}|^2 (1 + m\mu). \quad (37)$$

From Lemmas 1 and 2, we have, for any $1 \leq l \leq n + 1$,

$$\max_{1 \leq j \leq m} |a_{j,n}| \leq k_1 \max_{1 \leq j \leq m} |\langle f_n, \psi_j \rangle| \leq k_1 \sum_{i \in \ell_{n+1}} |\langle f_n, g_i \rangle| = k_1d_{n+1} \leq k_1k_2d_1.$$

Thus,

$$\sum_{j=1}^{m} |a_{j,n}|^2 \leq m \max_{1 \leq j \leq m} |a_{j,n}|^2 \leq (k_1k_2)^2 \sum_{i \in \ell_{n+1}} d_i^2 \leq (k_1k_2)^2 D. \quad (38)$$

Combining (37) with (38), we have

$$\|P_L(f_n)\|^2 \leq \sum_{j=1}^{m} |a_{j,n}|^2 (1 + m\mu) \leq (k_1k_2)^2 D \frac{19}{18} = (k_1k_2)^2 \frac{19}{18} D \leq 1.34D. \quad (39)$$

$\square$

Next, using Theorems 5 and 6, we give the estimation of $D$.

**Theorem 8.** For $A \geq 1$ and any positive integer $m$, the following inequalities hold.

$$D \geq 1.07(1 + \epsilon)\sigma_m(f), \forall \epsilon > 0,$$
\[ \| \xi Am \| \leq \| \xi 0 \|. \]

**Proof.** From (3), we have

\[ \| \xi 0 \|^2 = \| P_L(f) \|^2 = \| f - P_L(f) \|^2 \leq \| f - \sum_{j=1}^{m} b_j \psi_j \|^2 \leq (1 + \epsilon)^2 \sigma_m^2(f). \]

By using Theorems 5 and 6, we derive

\[ (1 + \epsilon)^2 \sigma_m^2(f) \geq \| \xi 0 \|^2 \geq \| \xi 0 \|^2 - \| \xi Am \|^2 = \sum_{n=1}^{Am} (\| \xi_{n-1} \|^2 - \| \xi_n \|^2) \]

\[ \geq \sum_{I_n \cap T_2 = \emptyset} \left( -0.22 D \mu \right) + \sum_{I_n \cap T_2 \neq \emptyset} 0.8937 d_n \]

\[ \geq -0.22 D Am \mu + 0.8937 D \]

\[ \geq -0.0123 D + 0.8937 D \]

\[ = 0.8814 D > 0, \]

which is equivalent to

\[ D^{\frac{1}{2}} \leq (0.8814)^{\frac{1}{2}} (1 + \epsilon) \sigma_m(f) \leq 1.07 (1 + \epsilon) \sigma_m(f). \]

Furthermore, we also have

\[ \| \xi Am \| \leq \| \xi 0 \|. \]

\[ \square \]

Now, we can give the proof of our main result.

**Proof of Theorem 4.** Note that

\[ \| f Am \| = \| P_L(f Am) + \xi Am \| \leq \| P_L(f Am) \| + \| \xi Am \|. \]

From Theorem 7 and Theorem 8, we obtain that

\[ \| f Am \| \leq \sqrt{1.34} D^{\frac{1}{2}} + \| \xi 0 \| \]

\[ \leq \sqrt{1.34} D^{\frac{1}{2}} + (1 + \epsilon) \sigma_m(f) \]

\[ \leq \sqrt{1.341.07} (1 + \epsilon) \sigma_m(f) + (1 + \epsilon) \sigma_m(f) \]

\[ \leq 2.24 (1 + \epsilon) \sigma_m(f). \]

Thus, we complete the proof of Theorem 4. \[ \square \]

4. Simulation Results

It is known from Theorem 4 that if \( f \in \Sigma_m \), then \( \sigma_m(f) = 0 \), and hence \( f = G_m(f) \). In this spirit, the OSGA can be used to recover sparse signals in compressed sensing, which is a new field of signal processing. We remark that in the field of signal processing, the orthogonal super greedy algorithm (OSGA) is also known as orthogonal multi-matching pursuit (OMMP). For the reader’s convenience, we will use the term OMMP instead of OSGA in what follows.
In this section, we test the performance of the orthogonal multi-matching pursuit with parameter $s$ (OMMP($s$)). We consider the following model. Suppose that $x \in \mathbb{R}^N$ is an unknown $N$-dimensional signal and we wish to recover it by the given data

$$ y = \Phi x, $$

(40)

where $\Phi \in \mathbb{R}^{M \times N}$ is a known measurement matrix with $M \ll N$. Furthermore, since $M \ll N$, the column vectors of $\Phi$ are linearly dependent and the collection of these columns can be viewed as a redundant dictionary.

By repeating the test 1000 times, we calculate the mean square error: $\text{MSE} = 1.1894$. The red line represents the original signal and the black squares represent the approximation.

A signal $x \in \mathbb{R}^N$ is said to be $K$-sparse if $\|x\|_0 := \#\text{supp}(x) = \#\{i : x_i \neq 0\} \leq K$. We will recover the support of a $K$-sparse signal via OMMP($s$) under the model (40). It is well known that OMMP takes the following form; see, for instance, [3].

ORTHOGONAL MULTI MATCHING PURSUIT (OMMP($s$))

Input: Measurement matrix $\Phi$, vector $y$, and $s$, the stopping criterion.

Step 1: Set the residual $r^0 = y$, an initial approximation $x^0 := 0$, the index set $\Lambda^0 := \emptyset$, and the iteration counter $l = 0$.

Step 2: Define $\Lambda^{l+1} := \Lambda^l \cup \{i_1, \cdots, i_s\}$ such that

$$ |\langle r^l, \varphi_{i_1} \rangle| \geq \cdots \geq |\langle r^l, \varphi_{i_s} \rangle| \geq \sup_{\varphi \in \Phi, \varphi \neq \varphi_{i_k}} |\langle r^l, \varphi \rangle|. $$

Then, $x^l := \arg \min_{z \in \text{supp}(z) \in \Lambda^{l+1}} \|y - \Phi z\|_2$ and update the residual $r^{l+1} := y - \Phi x^{l+1}$.

End if the stopping condition is achieved. Otherwise, we set $l := l + 1$ and turn to step 2.

Output: If the algorithm stops at the $k$th iteration, then output $\Lambda^k$ and $\tilde{x}_\Lambda^k = x^k$.

In the experiment, we set the measurement matrix $\Phi$ to be a Gaussian matrix where each entry is selected from the $\mathcal{N}(0, M^{-1})$ distribution and the density function of this distribution is $p(x) := \frac{1}{\sqrt{2\pi M}} e^{-x^2 M/2}$. We execute OMMP($s$) with the data vector $y = \Phi x$ and stop the algorithm when $\#\Lambda^l \geq K$. The mean square error(MSE) of $x$ is defined as follows:

$$ \text{MSE} = \frac{1}{N} \sum_{j=1}^{N} (x_j - \tilde{x}_{\Lambda^k,j})^2. $$

Figure 1 shows the performance of OMMP($s$) with $s = 5$ for an input signal in dimension $N = 512$ with sparsity level $K = 50$ and number of measurements $M = 200$, where the red line represents the original signal and the black squares represent the approximation. By repeating the test 1000 times, we calculate the mean square error: MSE = $1.1894 \times 10^{-8}$.
Figure 1. The recovery of an input signal in dimension $N = 512$ with sparsity level $K = 50$, number of measurements $M = 200$ and $s = 5$.

Figure 2 describes the case of the dimension $N = 256$. It displays which percentage (the average of 100 input signals) of the elements in support can be found correctly as a function of $M$ with $s = 3$. If the percentage equals 100%, it means that all the elements in support can be found, which implies that the input signal can be exactly recovered. As expected, Figure 2 shows that when the sparsity level $K$ increases, more measurements are necessary to guarantee signal recovery.

Figure 2. The average percentage of elements in support found correctly (100 input signals) as a function of the number of measurements $M$ for different sparsity levels $K$ in dimension $N = 256$ with $s = 3$.

5. Concluding Remarks

This paper investigates the error behavior of the orthogonal super greedy algorithm OSGA with respect to $\mu$-coherent dictionaries. The OSGA is simpler than the OGA from the viewpoint of the computational complexity. Under the assumption that the coherence parameter $\mu$ has a lower bound, we establish the ideal Lebesgue-type inequality for the OSGA, which shows that the OSGA provides an almost optimal approximation on the first $\left\lceil \frac{1}{18 \mu s} \right\rceil$ steps. Moreover, we improve the asymptotic constant in the Lebesgue-type inequality of the OGA obtained in [19]. We develop some new techniques to obtain our results. We found that there exists a strong dependency between the constant $A$ and the coherence parameter $\mu$ in (2). The specific constant 2.24 is not the best; we can change it by adjusting the values of $A$ and $\mu$, but the best one is still unknown. In fact, we do not even
know if such a constant exists. We will continue to study the improvement of the Lebesgue constant in our future work. As for the applications of the OSGA, our simulation results show that OSGA is very efficient for recovering sparse signals.

**Author Contributions:** Authors contribute evenly in this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to thank the referees, the editors, Zhang Haizhang and Xu Xu for their very useful suggestions, which significantly improved this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**