Estimates of the Restricted Isometry Constant in Super Greedy Algorithms

Wei Xiujie and Ye Peixin
School Of Mathematics and LPMC Nankai University Tianjin 300071, China yepx@nankai.edu.cn

Abstract

Orthogonal Super Greedy Algorithm (OSGA) is a super greedy-type algorithm for sparse approximation. We analyze the convergence of OSGA based on Restricted Isometry Property (RIP). Our main conclusion is that if a matrix $\Phi$ satisfies the Restricted Isometry Property of order $[sK]$ with isometry constant $\delta < \frac{\sqrt{s} + \sqrt{sK}}{\sqrt{sK}}$, then OSGA $(s)$ can exactly recover any $K$-sparse signal $x$ from $y = \Phi x$ in at most $K$ iterations.

Keywords: Compressed Sensing; $K$-Sparse Signal; Orthogonal Multi Matching Pursuit; Orthogonal Super Greedy Algorithm; Restricted Isometry Constant

1. Introduction

Very recently, considerable attention has focused on recovering sparse signals from a limited number of linear projection measurements at sub-Nyquist sampling rates. This topic is also known in the literature as Compressive Sensing or Compressed Sensing (CS). CS is a new paradigm in signal and image processing and it shows that it is high possibility to reconstruct sparse signals from their projection onto a small number of random vectors [2], [7-8]. Orthogonal Super Greedy Algorithm (OSGA) is the greedy algorithm for sparse approximation. Let $\Phi$ denote a measurement matrix of size $MN \times M$ (where typically $M < N$) and measurements $y$ denote a vector in $\mathbb{R}^M$, the goal of OSGA is to recover unknown signal $x \in \mathbb{R}^N$ from $y = \Phi x$. Here $M < N$ means we collect fewer data than unknowns. OSGA aims to achieve an approximation solution to $y = \Phi x$.

In the present paper, we investigate the efficiency of Orthogonal Super Greedy Algorithm (OSGA) in the compressed sensing problem. Let us begin with the demonstration of the use of super greedy algorithms in the compressed sensing problem. Now let us recall some necessary concepts of CS. Assume that $1 \leq K \leq M \leq N$ and $0 < \delta < 1$. Let $\|x\|_0$ denotes the number of nonzero coordinates of a signal $x = (x_j)_{j=1}^N \in \mathbb{R}^N$, we say that the signal $x$ is $K$-sparse if $\|x\|_0 \leq K$. Sparsity of $x$ allows for accurate reconstruction of the object from what would naively seem to be too few samples. A large number of algorithms within the broader field of CS have been studied using the Restricted Isometry Property (RIP) for the matrix $\Phi$. A matrix $\Phi$ is said to satisfy RIP ([5-6]) of order $K$ with isometry constant $\delta$ if, for all $K$-sparse vectors $x$, we have

$$(1 - \delta)\|x\|^2 \leq \|\Phi x\|^2 \leq (1 + \delta)\|x\|^2,$$

where $\|x\|^2 = \sum_{j=1}^N x_j^2$. (1)

In this article, we study signal recovery via Orthogonal Super Greedy Algorithm, Orthogonal Super Greedy Algorithm is also known in literature as Orthogonal Multi...
Matching Pursuit [11]. We give the definition of the Orthogonal Multi Matching Pursuit in terms of the theory of transmission of signals.

**Algorithm:** Orthogonal Matching Multi Pursuit (OMMP)

**Input:** \( \Phi, y \) and \( s \).

**Set:** \( r^0 := y, x^0 := 0, \Lambda^0 := \emptyset, l = 0. \)

**Iteration:** Define \( \Lambda^{t+1} := \Lambda^l \cup \{i_1, \ldots, i_t\} \) such that

\[
|\langle r^l, \phi_{i_1} \rangle| \geq \cdots \geq |\langle r^l, \phi_{i_t} \rangle| \geq \sup_{\phi \in \Phi, \phi \neq \phi_{i_1}, \ldots, \phi_{i_t}} |\langle r^l, \phi \rangle|.
\]

Then \( x^{t+1} := \arg\min_{z \in \text{supp}(\cdot) = \Lambda^{t+1}} \|y - \Phi z\| \) and \( r^{t+1} := y - \Phi x^{t+1} \).

If \( r^{t+1} = 0 \), stop. Otherwise, we set \( l := l + 1 \) and begin a new iteration.

**Output:** If the algorithm stops at the \( k \)-th iteration, output \( x^k \) such that \( \hat{x}_{\Lambda^k} = x^k \) and \( \hat{x}(\Lambda^k)^c = 0 \).

Now let us recall some results on recovery of sparse signals by the OSGA (s), when \( s = 1 \), OSGA is Orthogonal Matching Pursuit (OMP). Davenport and Wakin [8] have proved that \( \delta_{K+1} < \frac{1}{\sqrt{K}} \) is sufficient for OMP to recover any \( K \)-sparse signal in \( K \) iterations. Later, Liu and Temlyakov [10] have improved the condition to \( \delta_{K+1} < \frac{1}{(1+\sqrt{2})\sqrt{K}} \).

Recently, Mo and Shen [12] proved that \( \delta_{K+1} < \frac{1}{1+\sqrt{K}} \) is sufficient for OMP to recover any \( K \)-sparse signal in \( K \) iterations and showed that there exist a \( K \)-sparse signal \( x \in \mathbb{R}^N \) and a matrix \( \Phi \) with restricted isometry \( \delta_{K+1} = \frac{1}{\sqrt{K}} \) for which OMP fails in \( K \) iterations. For \( s \in \mathbb{N}, s \geq 1 \), Liu and Temlyakov [10] analyzed the convergence property of OSGA(s) based on Restricted Isometry Property. They proved that

**Theorem 1.** For \( s \in \mathbb{N} \) and \( s \leq K \), suppose \( \Phi \) satisfies the RIP of order \( sK \) with isometry constant \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}} \). Then OSGA (s) can exactly recover an arbitrary \( K \)-sparse signal \( x \in \mathbb{R}^N \) from \( y = \Phi x \) in at most \( K \) steps.

Wang, Kwon, and Shim [15] improved this bound to \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(3\sqrt{2}+\sqrt{K})} \). They showed that

**Theorem 2.** For \( s \in \mathbb{N} \) and \( s \leq K \), suppose \( \Phi \) satisfies the RIP of order \( sK \) with isometry constant \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(3\sqrt{2}+\sqrt{K})} \). Then OSGA (s) can exactly recover an arbitrary \( K \)-sparse signal \( x \in \mathbb{R}^N \) from \( y = \Phi x \) in at most \( K \) steps.

Theorem 2 shows that \( \delta := \delta_{sK} < \frac{s}{(3\sqrt{2}+\sqrt{K})} \) is a sufficient condition for OSGA (s) to recover every \( K \)-sparse signal in at most \( K \) iterations successfully. The aim of this article is to improve this sufficient condition to \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}} \). We proved here the following theorem

**Theorem 3.** For \( s \in \mathbb{N} \) and \( s \leq K \), suppose \( \Phi \) satisfies the RIP of order \( sK \) with isometry constant \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}} \). Then OSGA (s) can exactly recover an arbitrary \( K \)-sparse signal \( x \in \mathbb{R}^N \) from \( y = \Phi x \) in at most \( K \) steps.
Since $\sqrt{r/(3K+\sqrt{K})} > \sqrt{r/(3K+\sqrt{K})}$, Theorem 3 is an improvement of results of Wang, Kwon and Shim [15] Theorem 2. The rest part of the paper is organized as follows. In section 2, we will introduce some notations and investigate some restricted isometry constants. In section 3, we will give the proof of Theorem 3.

2. Preliminary Lemmas

Before going further, we introduce some notations.

For a $K$-sparse signal $x \in \mathbb{R}^N$, let $\Omega$ denote the index set of $x$ and $T$ the support of it, that is

$$\\Omega = \{1, \ldots, N\}$$

and

$$T = \text{supp}(x) = \{i : x_i \neq 0\},$$

where it holds $|T| \leq K$.

Suppose that $\phi_1, \phi_2, \ldots, \phi_N$ are the columns of a matrix $\Phi$, we assume that $\|\phi\|_2 = 1, 1 \leq i \leq N$. Suppose $\Lambda$ is a subset of $\Omega$, let $\Phi_\Lambda$ denotes the matrix consisted of the columns of $\Phi$ with indices $i \in \Lambda$. We use the same way to define $x_\Lambda$ for the vector $x \in \mathbb{R}^N$. Thus, we have

$$\Phi_\Lambda x = \Phi x_\Lambda = \Phi_\Lambda x_\Lambda.$$  \hspace{1cm} (2)

There are some useful properties of restricted isometry constant. Lemma 1 was obtained by D. Needell and J. A. Tropp in [13].

**Lemma 1.** Suppose $\Gamma$ and $\Lambda$ are two disjoint sets of indices. If the matrix $\Phi$ satisfies the RIP of order $|\Gamma \cup \Lambda|$ with isometry constant $\delta_{|\Gamma \cup \Lambda|}$. Then for any $x \in \mathbb{R}^{|\Lambda|}$ we have

$$\|\Phi^*_\Gamma \Phi_\Lambda x\| \leq \delta_{|\Gamma \cup \Lambda|} \|x\|$$  \hspace{1cm} (3)

and

$$\| (\Phi^*_\Lambda \Phi_\Lambda)^{-1} x\| \leq \frac{1}{1-\delta_{|\Lambda|}} \|x\|.$$  \hspace{1cm} (4)

For any two integers $K \leq K'$, it holds

$$\delta_K \leq \delta_{K'}.$$  \hspace{1cm} (5)

The following lemma was established by T. Cai and L. Wang in [3]. Suppose $\Phi^*$ represents the transpose of $\Phi$.

**Lemma 2.** Let $x \in \mathbb{R}^N$ be a $K$-sparse vector. Suppose the matrix $\Phi$ satisfies the RIP of order $K$ with isometry constant $\delta_K$. Then for $\Lambda \subset \Omega$, we have

$$(1-\delta_K) \|x_{\Gamma \setminus \Lambda}\| \leq \|\Phi^*_\Gamma \Phi_\Lambda (I-\Phi_\Lambda \Phi^*_\Lambda) \Phi_{\Gamma \setminus \Lambda} x_{\Gamma \setminus \Lambda}\| \leq (1+\delta_K) \|x_{\Gamma \setminus \Lambda}\|,$$  \hspace{1cm} (6)

where $\Phi^* = (\Phi^* \Phi)^{-1} \Phi^*$ is the pseudo inverse of matrix $\Phi$.

Let us look at the first iteration of OSGA ($s$). OSGA (s) chooses $s$ indices that corresponds to the largest (in magnitude) inner products. When can we conclude that at
least one of those indices is indeed in the support of \( x \)? The following lemma gives the answer. It was proved by Wang, Kwon, and Shim [15].

**Lemma 3.** For \( s \in \mathbb{N} \) and \( s \leq K \), suppose \( \Phi \) satisfies the RIP of order \( K + s \) with isometry constant \( \delta := \delta_{s+K} < \frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})} \). Define \( \Lambda := \{i_1, \ldots, i_l\} \) such that

\[
|\langle y, \phi_{i_1} \rangle| \geq \cdots \geq |\langle y, \phi_{i_l} \rangle| \geq \sup_{\phi \in \Phi, \phi \neq \phi_i, i \in \Lambda} |\langle y, \phi \rangle|.
\]  

(7)

Then \( \Lambda \cap T \neq \emptyset \).

### 3. Proof of Theorem 3

In this section, we shall prove the main result of the paper. First we prove a simple lemma, which will be used in the proof of Theorem 3.

**Lemma 4.** For \( s \in \mathbb{N} \) and \( s \leq K \), suppose \( 0 < \delta < \frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})} \), we have

\[
\frac{\delta}{\sqrt{(1-\delta)}} < \frac{1-\delta}{\sqrt{\delta}}.
\]  

(8)

Proof: A simple calculation shows that (8) is equivalent to

\[
\delta \sqrt{K} - \sqrt{s} (1-\delta)^2 < 0.
\]

Set

\[
f(\delta) = \delta \sqrt{K} - \sqrt{s} (1-\delta)^2 = -\sqrt{s} \delta^2 + (2\sqrt{s} + \sqrt{K}) \delta - \sqrt{s}.
\]

Since

\[
f\left(\frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})}\right) = -\left(\frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})}\right)^2 + \sqrt{s} - \sqrt{s} < 0,
\]

we have \( f(\delta) < 0 \), if \( 0 < \delta < \frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})} \). Hence, the lemma is proved.

We come now to the proof of Theorem 3. We will prove that the condition \( \delta := \delta_{sK} < \frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})} \) is sufficient to guarantee OSGA \((s)\) to select at least one correct index at each iteration.

Proof: For the first iteration, let \( \Gamma_1 := \{i_1, \ldots, i_l\} \) such that

\[
|\langle y, \phi_{i_1} \rangle| \geq \cdots \geq |\langle y, \phi_{i_l} \rangle| \geq \sup_{\phi \in \Phi, \phi \neq \phi_i, i \in \Gamma_1} |\langle y, \phi \rangle|.
\]

By Lemma 1 and Lemma 3, \( \delta_{s+K} < \delta_{sK} < \frac{\sqrt{s}}{(2\sqrt{s} + \sqrt{K})} \) guarantees \( \Gamma_1 \cap T \neq \emptyset \).

For the \((l+1)\)-th iteration, we have index set \( \Lambda^{l+1} \) with \( |\Lambda^{l+1}| = (l+1)s \). Assume OSGA\((s)\) selected at least one correct index at the \( k \)-th iteration, where \( k = 2, \ldots, l \), that is, \( \Gamma_k \cap T \neq \emptyset \), where \( \Gamma_k := \Lambda^{k} \setminus \Lambda^{k-1}, k = 2, \ldots, l \). We will prove that OSGA \((s)\) at the \((l+1)\)-th iteration selects at least one correct index. Denote:

\[
T_1 = \Lambda^l \cap T, T_2 = T \setminus \Lambda^l, |T_2| = K_2.
\]
We continue our proof by contradiction. Assume that at the \((I+1)\)-th iteration the OSGA \((s)\) does not select any correct index. This means \(\Gamma_{i+1} \cap T = \emptyset\). From the definition of OSGA \((s)\), it is obviously that \(|\Gamma_{i+1}| = s\).

The residual vector \(r' = y - \Phi_A^t\Phi_A^t y\) can be expressed as

\[
r' = y - \Phi_A^t\Phi_A^t y
= \Phi_{i_t}x_{i_t} + \Phi_{t_2}x_{t_2} - \Phi_A^t\Phi_A^t (\Phi_{i_t}x_{i_t} + \Phi_{t_2}x_{t_2}).
\]

(9)

By the definition of \(T_1\), using (2), we have

\[
\Phi_{i_t}x_{i_t} = \Phi_{i_t}x_{T_1}.
\]

(10)

Applying (10), let us continue (9) as

\[
r' = \Phi_{i_t}x_{i_t} + \Phi_{t_2}x_{t_2} - \Phi_A^t\Phi_A^t (\Phi_{i_t}x_{T_1} + \Phi_{t_2}x_{T_2})
= \Phi_{i_t}x_{i_t} + \Phi_{t_2}x_{t_2} - \Phi_A^t\Phi_A^t \Phi_{i_t}x_{i_t} - \Phi_A^t\Phi_A^t \Phi_{t_2}x_{t_2}
= \Phi_{t_2}x_{t_2} - \Phi_A^t\Phi_A^t \Phi_{t_2}x_{t_2}.
\]

(11)

Since \(A^t \cap T_2 = \emptyset\), we apply (4) and (3) of Lemma 1

\[
\|\Phi_A^t\Phi_T x_{T_2}\| \leq \|\Phi_A^t\Phi_T x_{T_2}\| \leq \frac{1}{(1-\delta)} \|\Phi_A^t\Phi_T x_{T_2}\| \leq \frac{\delta}{(1-\delta)} \|x_{T_2}\|.
\]

(12)

By (11), (12) and Lemma 1, we have

\[
\|\Phi_A^t\Phi_T x_{T_2}\| \leq \|\Phi_A^t\Phi_T x_{T_2}\| + \|\Phi_A^t\Phi_T x_{T_2}\| \leq \delta \|x_{T_2}\| + \delta \|x_{T_2}\| = \frac{\delta}{(1-\delta)} \|x_{T_2}\|.
\]

(13)

Thus

\[
\min_{i=1,\ldots,n} \langle r', \phi_i \rangle \leq \frac{1}{\delta^2} \|\Phi_A^t\Phi_T x_{T_2}\| \leq \frac{\delta}{(1-\delta)} \|x_{T_2}\|.
\]

(14)

Using (11) and Lemma 2, we obtain

\[
\|\Phi_A^t\Phi_T x_{T_2}\| \leq \|\Phi_A^t\Phi_T x_{T_2}\| = \| \Phi_A^t(I - \Phi_A^t\Phi_A^t)\Phi_{T_2}x_{T_2}\|
= \|\Phi_A^t(I - \Phi_A^t\Phi_A^t)\Phi_{T_2}x_{T_2}\|
= \|\Phi_A^t\Phi_{T\setminus A^t}x_{T\setminus A^t}\|
\]
\[ \geq (1 - \delta) \| x_{T \setminus N} \| = (1 - \delta) \| x_T \|. \]  

Hence
\[
\max_{j \in T} | \langle r_j, \phi_j \rangle | \geq \frac{1}{\sqrt{K}} \| \Phi r_j \| \geq \frac{(1 - \delta)}{\sqrt{K}} \| x_T \|. \quad (15)
\]

By the definition of \( \Gamma_{i+1} \), we get
\[
\min_{j \in \Gamma_{i+1}} | \langle r_j, \phi_j \rangle | \geq \max_{j \in T} | \langle r_j, \phi_j \rangle |.
\]

Using (14) and (16), we obtain
\[
\frac{\delta}{\sqrt{(1 - \delta)}} \| x_T \| \geq \frac{1}{\sqrt{K}} \| x_T \|. \quad (17)
\]

Since (17) is equivalent to (8), by Lemma 4, we must have \( \delta \geq \frac{\sqrt{K}}{2(1 - \delta + \sqrt{K})} \). This contradiction implies that we have \( \Gamma_{i+1} \cap T \neq \emptyset \). Therefore OSGA \((s)\) selects at least one correct index at every iteration. This means \( T \subset \Lambda^K \). Thus
\[
\hat{x}_{A^K} = x^K = \arg\min_z \| y - \Phi_{A^K} z \| = x_{A^K}
\]
and \( \hat{x}_{(A^K)^c} = x_{(A^K)^c} = 0 \). Then we conclude that \( x_{(A^K)^c} = x \). In other words, we can say OSGA \((s)\) exactly recover any \( K \)-sparse signal using at most \( K \) step.

Remark 1. As a consequence of Theorem 3, when \( s = 1 \), OSGA\((s)\) is OMP, then we have

**Theorem 4.** Suppose \( \Phi \) satisfies the RIP of order \( K + 1 \) with isometry constant
\[
\delta := \delta_{K + 1} < \frac{1}{2 \sqrt{K}}.
\]
Then OMP can exactly recover an arbitrary \( K \)-sparse signal \( x \in \mathbb{R}^N \) from \( y = \Phi x \) in \( K \) iterations.

Theorem 4 is similarly to Mo and Shen [12] Theorem 3.1. They proved an analog of Theorem 4 under the assumption \( \delta := \delta_{K + 1} < \frac{1}{1 + \sqrt{K}} \). For the noise case, that is, \( y = \Phi x + z \), where \( z \) is the noise term. Shen and Li [14] proved the following theorems

**Theorem 5.** Suppose \( \| z \| < B_z \) and \( \delta_{x, K} < \frac{1}{2 \sqrt{K}} \). Then OMP with the stopping rule
\[
\| r_k \| < B_z
\]
finds the support of \( x \in \mathbb{R}^N \) if all the nonzero coefficients \( x_i \) satisfy
\[
\| x_i \| \geq \frac{2(1 - \delta_{x, K}) B_z}{(1 - \delta_{x, K})^2 - \delta_{x, K} (1 + \sqrt{K})}.
\]

**Theorem 6.** Suppose \( \| A^* z \|_{\infty} < B_{\infty} \) and \( \delta_{x, K} < \frac{1}{2 \sqrt{K}} \). Then OMP with the stopping rule
\[
\| A^* r_k \|_{\infty} < B_{\infty}
\]
finds the support of \( x \in \mathbb{R}^N \) if all the nonzero coefficients \( x_i \) satisfy
\[
\| x_i \| \geq \frac{2(1 - \delta_{x, K}) B_{\infty}}{(1 - \delta_{x, K})^2 - \delta_{x, K} (1 + \sqrt{K})} (1 + \frac{\sqrt{K}}{\sqrt{1 - \delta_{x, K}}})
\]

4. Conclusion

OSGA \((s)\) is a natural extension of OMP in the sense that it selects \( s \) candidates per iteration whereas OMP selects only one. We analyzed the recovery performance of OSGA under RIP condition for the recovery sparse signal from incomplete measurements. As OSGA selects multi candidates at each step, it can recover the
sparse signal using fewer iterations, thus it can further reduce the computational complexity.

Our analysis provided a performance guarantee of the OSGA based on the RIP for its applications in CS. The emerging theory of CS has provided a new framework for signal acquisition [1, 4, 9]. CS may play an important role for design of measurement devices in diverse engineering fields, particularly where measurements are physically limited or data acquisition is very expensive.

Acknowledgments

The research was supported by the Natural Science Foundation of China (Grant No. 11271199 and 11101220). Corresponding author: Ye Peixin.

References


Authors

Ye Peixin, he received the M.S degree in mathematics from Xiamen University, Fujian, China, in 1998 and the Ph.D. degree in mathematics from Beijing Normal University, Beijing, China in 2001. From 2001 to 2003 he worked at the Institute of mathematics, Academy of mathematics and system sciences, Chinese academy of sciences as a post-doctor. Now he is a Full Professor at School of mathematical sciences, Nankai University, Tianjin, China. He has published more than 70 journal and conference papers. His current research interests include information-based complexity, approximation theory, machine learning and compressed sensing.
Wei Xiujie, she is currently a Ph.D. candidate in the Department of Basic Mathematics, School of Mathematical Science, Nankai University, Tianjin, China. She received her B.S. degree in mathematics from Qufu normal University, Shandong, China, in 2010 and the M.S. degree in mathematics from Nankai University, Tianjin, China, in 2013. Her research interests include approximation theory and compressed sensing.