

RESEARCH

Open Access



Efficiency of orthogonal super greedy algorithm under the restricted isometry property

Xiujie Wei^{1*}  and Peixin Ye²

*Correspondence:

weixiujie1986@163.com

¹School of Sciences, Tianjin Chengjian University, Tianjin, China
Full list of author information is available at the end of the article

Abstract

We investigate the efficiency of orthogonal super greedy algorithm (OSGA) for sparse recovery and approximation under the restricted isometry property (RIP). We first show that under the RIP conditions of the measurement matrix Φ and the minimum magnitude of the nonzero coordinates of the signal, for l_2 bounded or l_∞ bounded noise vector e , with explicit stopping rules, OSGA can recover the support of an arbitrary K -sparse signal x from $y = \Phi x + e$ in at most K steps. Then, we investigate the error performance of OSGA in m term approximation with regards to dictionaries satisfying the RIP in a separable Hilbert space. We establish a Lebesgue-type inequality for OSGA. Based on this inequality, we obtain the optimal rate of convergence for the sparse class induced by such dictionaries.

Keywords: Orthogonal super greedy algorithm; Orthogonal multi matching pursuit; Restricted isometry property; Compressed sensing; m -term approximation; Lebesgue-type inequality

1 Introduction

Recovery and approximation by sparse linear combination of elements from a fixed redundant family is frequently utilized in many application areas, such as image or signal processing, PDE solvers and statistical learning, see [1]. In general, these problems are N - P hard. It is well known that greedy type algorithms are efficient approaches to solve them, see [2–5]. Among others, the orthogonal greedy algorithm (OGA) has been widely used in practice. OGA is a simple yet powerful algorithm for highly nonlinear sparse approximation that has seen a large amount of research over its history, see [3, 6, 7] and the reference therein.

In this paper, we consider the orthogonal super greedy algorithm (OSGA), which is more efficient than OGA from the viewpoint of computational complexity. The performances of (OSGA) in greedy approximation and signal recovering were analyzed in [8–12]. We will make further study on the efficiency of the OSGA from two aspects.

In Sect. 2, we will study the efficiency of the OSGA in recovering N -dimensional sparse signal from linear measurements. This topic is also known in the literature as compressed sensing (CS), [13–15]. We show that OSGA can recover a K -sparse signal from noisy measurements in at most K steps. We remark that in the field of signal processing, orthogonal super greedy algorithm OSGA is also known as orthogonal multi matching pursuit

(OMMP) [4]. So for the reader’s convenience, we will use the term OMMP instead of OSGA in Sect. 2.

In Sect. 3, we study the error performance of the OSGA in the general context. We investigate the efficiency of OSGA for m -term approximation with regard to dictionaries in a real separable Hilbert space H . Assuming that the dictionary satisfies the RIP condition, we establish a Lebesgue inequality for OSGA which bounds the error of OSGA by the best m -term approximation error. Based on this inequality, we derive the sharp convergence rate of OSGA on the sparse class induced by dictionary satisfying the RIP condition.

2 The efficiency of the OMMP in CS

In this section, we analyze the efficiency of the OMMP in compressed sensing.

We consider the following model. Suppose that $x \in \mathbb{R}^N$ is an unknown N -dimensional signal and we wish to recover it from M noisy measurements y given by inner products with fixed vectors, that is,

$$y = \Phi x + e, \tag{2.1}$$

where Φ is a known $M \times N$ measurement matrix with $M \ll N$ and $e \in \mathbb{R}^M$ is a vector of measurement errors. The error vector can either be zero, bounded, or Gaussian noise.

For $x = (x_j)_{j=1}^N \in \mathbb{R}^N$, define

$$\|x\|_p := \left(\sum_{i=1}^N x_i^p \right)^{1/p}, \quad 0 < p < \infty,$$

and

$$\|x\|_\infty := \sup_{1 \leq i \leq N} \{|x_i|\}.$$

The signal $x \in \mathbb{R}^N$ is said to be K -sparse if $\|x\|_0 := |\text{supp}(x)| = |\{i : x_i \neq 0\}| \leq K < N$. We study the performance of OMMP for the recovery of the support of the K -sparse signal x under model (2.1).

These algorithms recover the K -sparse signal by iteratively constructing the support of it by some greedy principles.

The OMMP is a stepwise forward selection and is easy to implement, it has been used for signal recovery. Now let us recall the definition of OMMP(s) in an algorithmic way (see Algorithm 1).

OSGA(s) generalizes the orthogonal greedy algorithm (OGA) in the sense that it selects multiple s indices in each iteration, it can recover the sparse signal using fewer steps and further reduce the complexity. To investigate the efficiency of the OMMP in CS, we use the restricted isometry property (RIP) condition of Φ which ensures the stable recovery of x from noisy measurements. This property is introduced by Candes and Tao [16, 17] as follows.

A matrix Φ is said to satisfy RIP of order K if there exists a constant $\delta \in (0, 1)$ such that, for all K -sparse vectors x ,

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2. \tag{2.2}$$

In particular, the minimum of all δ satisfying (2.2) is referred to as an isometry constant δ_K .

Algorithm 1 Orthogonal Matching Multi Pursuit (OMMP(s))

Input: Measurement matrix Φ , vector y , and s , the stopping criterion.

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

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

$$|\langle r^l, \phi_{i_1} \rangle| \geq \dots \geq |\langle r^l, \phi_{i_s} \rangle| \geq \sup_{\phi \in \Phi, \phi \neq \phi_{i_k}, k=1, \dots, s} |\langle r^l, \phi \rangle|.$$

Then

$$x^{l+1} := \arg \min_{z: \text{supp}(z) \in \Lambda^{l+1}} \|y - \Phi z\|_2$$

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, output Λ^k and $\widehat{x}_{\Lambda^k} = x^k$.

We recall some results on the efficiency of OMMP based on RIP. These results concern the estimate of the upper bound of sK order RIP constant δ_{sK} . In the noiseless case, Liu and Temlyakov [8] proved that $\delta_{sK} < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}}$ is a sufficient condition for OMMP(s) to recover every K -sparse signal in at most K iterations successfully. Then, Wang, Kwon, and Shim [10] relaxed the condition to $\delta_{sK} < \frac{\sqrt{s}}{3\sqrt{s+\sqrt{K}}}$. The constant was further relaxed as $\delta_{sK} < \frac{\sqrt{s}}{2\sqrt{s+\sqrt{K}}}$ by Satpathi, Das, and Chakraborty [18]. In [18] the authors also pointed out the possibility of extending the results to the noisy case. On the other hand, Dan Wei [9] defined the restricted orthogonality constant $\theta_{K,K'}$ and proved that $\delta_{sK-s+1} + \sqrt{\frac{K}{s}}\theta_{sK-s+1,s} < 1$ is a sufficient condition guaranteeing the perfect recovery of K -sparse signals by OMMP. She shows that this result implies the result of [18]. Moreover in [9] the performance of OMMP for support recovery from noisy measurements was also discussed.

Motivated by the above works, we investigate the OMMP in the general setting where noise is presented under the RIP based conditions. We prove if the sampling matrix Φ satisfies the RIP of order sK with the isometry constant

$$\delta_{sK} < 1 + \frac{1}{2}\sqrt{\frac{K}{s}} - \frac{1}{2}\sqrt{\frac{K}{s}} + 4\sqrt{\frac{K}{s}},$$

then OMMP can recover every K -sparse signal in at most K steps with the l_2 bounded and l_∞ bounded noise. It is easy to check that

$$\frac{\sqrt{s}}{2\sqrt{s+\sqrt{K}}} < 1 + \frac{1}{2}\sqrt{\frac{K}{s}} - \frac{1}{2}\sqrt{\frac{K}{s}} + 4\sqrt{\frac{K}{s}}.$$

So our results improve and extend the results in [18].

When dealing with the noisy measurements, one of the key components in OMMP is the stopping rule which depends on the structure of the noise. In general, there are several natural stopping rules for OMMP.

- (1) Stop after a fixed number of iterations K .
- (2) For l_2 bounded noise case: The error vector e is bounded in l_2 norm with $\|e\|_2 \leq b_2$, we set the stopping rule as $\|r^k\|_2 \leq b_2$.
- (3) For l_∞ bounded noise case: The error vector e is bounded in l_∞ norm with $\|\Phi^*e\|_\infty \leq b_\infty$, we set the stopping rule as

$$\|\Phi^*r^k\|_\infty \leq \left(1 + \frac{\sqrt{sK}\delta_{sK}}{1 - \delta_{sK}}\right)b_\infty,$$

where Φ^* is the transpose of Φ .

Since we mainly consider l_2 bounded noise and l_∞ bounded noise, we use stopping rules (2) and (3) respectively. In Sect. 2.1, some notations are introduced and some consequences of restricted isometry properties are presented. In Sect. 2.2, we give the RIP based sufficient conditions for OMMP(s) with l_2 bounded noise case by using the stopping rule (2). In Sect. 2.3, we establish the results in the l_∞ bounded noise case by using the stopping rule (3). In Sect. 2.4, we give the RIP based sufficient conditions for OMMP(s) with high probability in the Gaussian noise case.

2.1 Preliminary

The following notations will be used in this section. For a signal $x \in \mathbb{R}^N$, let $\Omega = \{1, \dots, N\}$ denote the index set of x and $T = \text{supp}(x) = \{i : x_i \neq 0\}$ the support of it.

Suppose that $\phi_1, \phi_2, \dots, \phi_N$ are the columns of the matrix Φ . We always assume $\|\phi_i\|_2 = 1$, $1 \leq i \leq N$. Suppose that Λ is a subset of Ω , let Φ_Λ denote the matrix Φ restricted to the columns indexed by Λ . We use the same way to define x_Λ for the vector $x \in \mathbb{R}^N$. Thus

$$\Phi_\Lambda x = \Phi x_\Lambda = \Phi_\Lambda x_\Lambda. \tag{2.3}$$

We recall some useful consequences of the RIP.

Lemma 2.1 ([5]) *Suppose that Γ and Λ are two disjoint sets of indices. If the matrix Φ 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*

$$\|\Phi_\Gamma^* \Phi_\Lambda x\|_2 \leq \delta_{|\Gamma \cup \Lambda|} \|x\|_2 \tag{2.4}$$

and

$$\|(\Phi_\Lambda^* \Phi_\Lambda)^{-1} x\|_2 \leq \frac{1}{(1 - \delta_{|\Lambda|})} \|x\|_2. \tag{2.5}$$

Lemma 2.2 ([6]) *Let $x \in \mathbb{R}^N$ be a K -sparse vector. Suppose that the matrix Φ satisfies the RIP of order K with isometry constant δ_K . Then, for $\Lambda \subset \Omega$,*

$$(1 - \delta_K) \|x_{T \setminus \Lambda}\|_2 \leq \|\Phi_{T \setminus \Lambda}^* (I - \Phi_\Lambda \Phi_\Lambda^\dagger) \Phi_{T \setminus \Lambda} x_{T \setminus \Lambda}\|_2 \leq (1 + \delta_K) \|x_{T \setminus \Lambda}\|_2, \tag{2.6}$$

where $\Phi^\dagger = (\Phi^* \Phi)^{-1} \Phi^*$ is the Moore–Penrose pseudo-inverse of Φ .

The residual vector after k steps can be written as

$$\begin{aligned} r^k &= y - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger y = (I - P_k)y \\ &= (I - P_k)(\Phi x + e) \\ &= (I - P_k)\Phi x + (I - P_k)e \\ &= r_1^k + r_2^k, \end{aligned}$$

where $P_k = \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger$ denotes the projection onto the linear space spanned by $\{\phi_i, i \in \Lambda^k\}$, $r_1^k = (I - P_k)\Phi x$ is the signal part of the residual, and $r_2^k = (I - P_k)e$ is the noise part of the residual.

Denote:

$$T_k = \Lambda^k \cap T, \quad A_k = T \setminus \Lambda^k, \quad B_k = T \cup \Lambda^k.$$

Suppose that $S_k \subseteq \Omega \setminus B_k, |S_k| = s$, such that

$$\min_{i \in S_k} |\langle r^k, \phi_i \rangle| \geq \max_{i \in (\Omega \setminus B_k) \setminus S_k} |\langle r^k, \phi_i \rangle|.$$

The difference between OMMP(s) and the standard orthogonal matching pursuit (OMP) is that at each step of the OMMP(s), s elements are simultaneously selected from a dictionary instead of one. For a sufficient condition to guarantee OMMP(s) to select at least one correct index at each iteration, we need to give a lower bound of $M_{S_k} - m_{S_k}$. So we provide the following lemma which plays an important role in our analysis.

Lemma 2.3 *Let $m_{S_k} := \min_{i \in S_k} \{|\langle r_1^k, \phi_i \rangle|\}$ and $M_{A_k} := \max_{i \in A_k} \{|\langle r_1^k, \phi_i \rangle|\}$. Assume $A_k \neq \emptyset$. Then there hold*

$$m_{S_k} \leq \frac{\delta_{sK}}{\sqrt{s}(1 - \delta_{sK})} \|x_{A_k}\|_2 \tag{2.7}$$

and

$$M_{A_k} \geq \frac{1 - \delta_{sK}}{\sqrt{|A_k|}} \|x_{A_k}\|_2. \tag{2.8}$$

Proof Using the fact $(I - P_k)\Phi_{\Lambda^k} = \Phi_{\Lambda^k} - \Phi_{\Lambda^k}(\Phi_{\Lambda^k}^* \Phi_{\Lambda^k})^{-1} \Phi_{\Lambda^k}^* \Phi_{\Lambda^k} = 0$, we express

$$r_1^k = (I - P_k)\Phi x$$

as

$$\begin{aligned} r_1^k &= (I - P_k)\Phi_T x_T = (I - P_k)(\Phi_{T_k} x_{T_k} + \Phi_{A_k} x_{A_k}) \\ &= (I - P_k)\Phi_{A_k} x_{A_k} = (I - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger)\Phi_{A_k} x_{A_k} \\ &= \Phi_{A_k} x_{A_k} - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger \Phi_{A_k} x_{A_k}. \end{aligned} \tag{2.9}$$

Since $\Lambda^k \cap A_k = \emptyset$, applying (2.5) and (2.4) of Lemma 2.1, and the fact that for any two integers $K \leq K'$ it holds

$$\delta_K \leq \delta_{K'}, \tag{2.10}$$

it follows that

$$\begin{aligned} \|\Phi_{\Lambda^k}^\dagger \Phi_{A_k} x_{A_k}\|_2 &= \|(\Phi_{\Lambda^k}^* \Phi_{\Lambda^k})^{-1} \Phi_{\Lambda^k}^* \Phi_{A_k} x_{A_k}\|_2 \\ &\leq \frac{1}{(1 - \delta_{sK})} \|\Phi_{\Lambda^k}^* \Phi_{A_k} x_{A_k}\|_2 \\ &\leq \frac{\delta_{s(k+|A_k|)}}{(1 - \delta_{sK})} \|x_{A_k}\|_2 \leq \frac{\delta_{sK}}{(1 - \delta_{sK})} \|x_{A_k}\|_2. \end{aligned} \tag{2.11}$$

By (2.9), (2.11), and Lemma 2.1, we have

$$\begin{aligned} \|\Phi_{S_k}^* r_1^k\|_2 &= \|\Phi_{S_k}^* (\Phi_{A_k} x_{A_k} - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger \Phi_{A_k} x_{A_k})\|_2 \\ &\leq \|\Phi_{S_k}^* \Phi_{A_k} x_{A_k}\|_2 + \|\Phi_{S_k}^* \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger \Phi_{A_k} x_{A_k}\|_2 \\ &\leq \delta_{s+|A_k|} \|x_{A_k}\|_2 + \delta_{s(k+1)} \|\Phi_{\Lambda^k}^\dagger \Phi_{A_k} x_{A_k}\|_2 \\ &\leq \delta_{sK} \|x_{A_k}\|_2 + \delta_{sK} \frac{\delta_{sK}}{(1 - \delta_{sK})} \|x_{A_k}\|_2 \\ &= \frac{\delta_{sK}}{(1 - \delta_{sK})} \|x_{A_k}\|_2. \end{aligned} \tag{2.12}$$

Thus

$$m_{S_k} \leq \frac{1}{\sqrt{s}} \|\Phi_{S_k}^* r_1^k\|_2 \leq \frac{\delta_{sK}}{\sqrt{s}(1 - \delta_{sK})} \|x_{A_k}\|_2.$$

Using (2.9) and Lemma 2.2, we obtain

$$\begin{aligned} \|\Phi_{A_k}^* r_1^k\|_2 &= \|\Phi_{A_k}^* (I - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger) \Phi_{A_k} x_{A_k}\|_2 \\ &= \|\Phi_{T \setminus \Lambda^k}^* (I - \Phi_{\Lambda^k} \Phi_{\Lambda^k}^\dagger) \Phi_{T \setminus \Lambda^k} x_{T \setminus \Lambda^k}\|_2 \\ &\geq (1 - \delta_{sK}) \|x_{T \setminus \Lambda^k}\|_2 = (1 - \delta_{sK}) \|x_{A_k}\|_2. \end{aligned} \tag{2.13}$$

Hence

$$M_{A_k} \geq \frac{1}{\sqrt{|A_k|}} \|\Phi_{A_k}^* r_1^k\|_2 \geq \frac{(1 - \delta_{sK})}{\sqrt{|A_k|}} \|x_{A_k}\|_2.$$

Therefore Lemma 2.3 is proved. □

2.2 l_2 bounded noise

We shall discuss the performance of OMMP(s) for the l_2 bounded noise case. That is, the error vector e is bounded in l_2 norm with $\|e\|_2 \leq b_2$ for some constant $b_2 > 0$. With a suitable stopping rule and a reasonable condition on the minimum magnitude of the nonzero coordinates of the K -sparse signal of x , OMMP(s) can recover the support of x in at most K iterations.

Theorem 2.1 *Suppose that $\|e\|_2 \leq b_2$ and Φ satisfies the RIP of order sK with the isometry constant $\delta := \delta_{sK} < 1 + \frac{1}{2}\sqrt{\frac{K}{s}} - \frac{1}{2}\sqrt{\frac{K}{s} + 4\sqrt{\frac{K}{s}}}$. Then OMMP(s) with the stopping rule $\|r^k\|_s \leq b_2$ recovers the support of any K -sparse signal $x \in \mathbb{R}^N$ from $y = \Phi x + e$ provided that all the nonzero components x_i satisfy*

$$|x_i| > \frac{2b_2}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}}. \tag{2.14}$$

Proof Denote $E_k = \max_{i \in \Omega} \{|\langle r_2^k, \phi_i \rangle|\}$. Recalling the definition of m_{S_k} and M_{A_k} in Lemma 2.3, there hold

$$\min_{i \in S_k} \{|\langle r^k, \phi_i \rangle|\} \leq \min_{i \in S_k} \{|\langle r_1^k, \phi_i \rangle|\} + \max_{i \in \Omega} \{|\langle r_2^k, \phi_i \rangle|\} = m_{S_k} + E_k$$

and

$$\max_{i \in A_k} \{|\langle r^k, \phi_i \rangle|\} \geq \max_{i \in A_k} \{|\langle r_1^k, \phi_i \rangle|\} - \max_{i \in \Omega} \{|\langle r_2^k, \phi_i \rangle|\} = M_{A_k} - E_k.$$

It is clear that in order for OMMP(s) to select at least one correct index at this step, it is necessary to have

$$\max_{i \in A_k} \{|\langle r^k, \phi_i \rangle|\} > \min_{i \in S_k} \{|\langle r^k, \phi_i \rangle|\}.$$

A sufficient condition to guarantee OMMP(s) to select at least one correct index at each iteration until all correct indices are selected is

$$M_{A_k} - E_k > m_{S_k} + E_k.$$

That is,

$$2E_k < M_{A_k} - m_{S_k} \tag{2.15}$$

holds for $k = 0, 1, \dots, K - 1$. We first provide an upper bound for E_k . The assumption $\|e\| \leq b_2$ yields that

$$E_k = \max_{i \in \Omega} \{|\langle r_2^k, \phi_i \rangle|\} \leq \max_{i \in \Omega} \|\phi_i\|_2 \|r_2^k\|_2 = \|(I - P_k)e\|_2 \leq \|e\|_2 \leq b_2. \tag{2.16}$$

Next we give a lower bound of $M_{A_k} - m_{S_k}$. Using Lemma 2.3 and assumption (2.14), we have

$$\begin{aligned} M_{A_k} - m_{S_k} &\geq \left(\frac{1 - \delta}{\sqrt{|A_k|}} - \frac{\delta}{\sqrt{s}(1 - \delta)} \right) \|x_{A_k}\|_2 \\ &\geq \left(\frac{1 - \delta}{\sqrt{|A_k|}} - \frac{\delta}{\sqrt{s}(1 - \delta)} \right) \sqrt{|A_k|} \frac{2b_2}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}} \\ &> 2b_2. \end{aligned}$$

This means OMMP(s) will select at least one correct index at the $k + 1$ -th iteration, and OMMP(s) can recover the support of x in at most K iterations. It remains to show that under the stopping rule $\|r^k\| \leq b_2$, OMMP exactly stops when all the correct indices are selected.

First, assume that $A_k = \emptyset$, then $T \subseteq A^k$ and $(I - P_k)\Phi x = 0$. Thus

$$\|r^k\|_2 = \|(I - P_k)\Phi x + (I - P_k)e\|_2 = \|(I - P_k)e\|_2 \leq \|e\|_2 \leq b_2.$$

So when all the K correct indices are selected, the l_2 norm of the residual will be less than b_2 and hence OMMP(s) stops. Now we show that OMMP(s) does not stop early. Suppose that the OMMP(s) has run k steps for some $k < K$. We will verify that $\|r^k\|_2 > b_2$ and so OMMP(s) does not stop at the current step.

Secondly, assume that $A_k \neq \emptyset$, then

$$\begin{aligned} \|r^k\|_2 &= \|(I - P_k)\Phi x + (I - P_k)e\|_2 \\ &\geq \|(I - P_k)\Phi x\|_2 - \|(I - P_k)e\|_2 \\ &= \|\Phi_{A_k}x_{A_k} - \Phi_{A^k}^\dagger \Phi_{A^k} \Phi_{A_k} x_{A_k}\|_2 - \|(I - P_k)e\|_2 \\ &= \|\Phi_{B_k} \bar{x}^{(k)}\|_2 - \|(I - P_k)e\|_2 \\ &\geq \|\Phi_{B_k} \bar{x}^{(k)}\|_2 - \|e\|_2 \\ &\geq \sqrt{1 - \delta_{|B_k|}} \|\bar{x}^{(k)}\|_2 - b_2 \\ &\geq \sqrt{1 - \delta} \|x_{A_k}\|_2 - b_2 \\ &\geq \sqrt{1 - \delta} \sqrt{|A_k|} \frac{2b_2}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}} - b_2 \\ &> b_2, \end{aligned}$$

where $\bar{x}^{(k)} = (x_{A_k} - \Phi_{A^k}^\dagger \Phi_{A^k} x_{A_k})^*$, and to derive the third inequality, we use the RIP condition. Hence OMMP(s) does not stop early, and the theorem is proved. \square

2.3 l_∞ bounded noise

We now give the RIP based sufficient conditions for OMMP(s) with l_∞ bounded noise. Our result is the following theorem.

Theorem 2.2 *Suppose that $\|\Phi^*e\|_\infty \leq b_\infty$ and Φ satisfies the RIP of order sK with the same isometry constant as in Theorem 2.1. Then OMMP(s) with the stopping rule*

$$\|\Phi^*r^k\|_\infty \leq \left(1 + \frac{\sqrt{sK}\delta}{1 - \delta}\right) b_\infty$$

finds the support of any K -sparse signal x if all the nonzero coefficients x_i satisfy

$$|x_i| > \frac{2(1 + \frac{\sqrt{sK}\delta}{1 - \delta})b_\infty}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}}. \tag{2.17}$$

Proof Denote $E_k = \|\Phi^*(I - P_k)e\|_\infty = \max_{i \in \Omega} |\langle r_2^k, \phi_i \rangle|$. We first prove that

$$2E_k < M_{A_k} - m_{S_k} \tag{2.18}$$

holds for $k = 0, 1, \dots, K - 1$, where m_{S_k} and M_{A_k} are defined as in Lemma 2.3. It follows from the proof of Theorem 2.1 that inequality (2.18) can ensure OMMP recovers the true support of x in at most K steps.

We first provide an upper bound of E_k . The assumption $\|\Phi^*e\|_\infty \leq b_\infty$ and the fact $\Phi_{\Lambda^k}^*(I - P_k)e = 0$ yield that

$$\begin{aligned} E_k &= \max_{i \in \Omega} |\langle r_2^k, \phi_i \rangle| = \max_{i \in \Omega \setminus \Lambda^k} |\phi_i^*(I - P_k)e| \\ &\leq \|\Phi_{\Omega \setminus \Lambda^k}^*e\|_\infty + \|\Phi_{\Omega \setminus \Lambda^k}^*P_k e\|_\infty \\ &\leq b_\infty + \max_{i \in \Omega \setminus \Lambda^k} \|\phi_i^* \Phi_{\Lambda^k} (\Phi_{\Lambda^k}^* \Phi_{\Lambda^k})^{-1} \Phi_{\Lambda^k}^* e\|_2 \\ &\leq b_\infty + \delta \|\Phi_{\Lambda^k}^* \Phi_{\Lambda^k}^{-1} \Phi_{\Lambda^k}^* e\|_2 \\ &\leq b_\infty + \frac{\delta}{1 - \delta} \|\Phi_{\Lambda^k}^* e\|_2 \\ &\leq b_\infty + \frac{\delta}{1 - \delta} \sqrt{sK} b_\infty \\ &\leq \left(1 + \frac{\delta \sqrt{sK}}{1 - \delta}\right) b_\infty. \end{aligned} \tag{2.19}$$

Here, to derive the third and fourth inequalities, we have used (2.4) and (2.5) of Lemma 2.1 respectively.

On the other hand, we can obtain a lower bound of $M_{A_k} - m_{S_k}$. By using Lemma 2.3 and assumption (2.17), we have

$$\begin{aligned} M_{A_k} - m_{S_k} &\geq \left(\frac{1 - \delta}{\sqrt{|A_k|}} - \frac{\delta}{\sqrt{s(1 - \delta)}}\right) \|x_{A_k}\|_2 \\ &\geq \left(\frac{1 - \delta}{\sqrt{|A_k|}} - \frac{\delta}{\sqrt{s(1 - \delta)}}\right) \sqrt{|A_k|} \frac{2(1 + \frac{\sqrt{sK}\delta}{1 - \delta})b_\infty}{(1 - \delta) - \sqrt{\frac{K}{s}} \frac{\delta}{1 - \delta}} \\ &> 2 \left(1 + \frac{\sqrt{sK}\delta}{1 - \delta}\right) b_\infty. \end{aligned} \tag{2.20}$$

The bounds (2.19) and (2.20) imply that (2.18) holds for $k = 0, 1, \dots, K - 1$. It remains to show that OMMP (s) exactly stops when all the correct indices are selected under the stopping rule $\|\Phi^*r^k\|_\infty \leq (1 + \frac{\sqrt{sK}\delta_{sK}}{1 - \delta_{sK}})b_\infty$.

First, assume that $A_k = \emptyset$, then $T \subseteq \Lambda^k$ and $(I - P_k)\Phi x = 0$. Thus

$$\|\Phi^*r^k\|_\infty = \|\Phi^*(I - P_k)(\Phi x + e)\|_\infty = \|\Phi^*(I - P_k)e\|_\infty \leq \left(1 + \frac{\delta \sqrt{sK}}{1 - \delta}\right) b_\infty,$$

where the last inequality is deduced from (2.19). Hence, OMMP(s) stops when all the correct indices are selected.

Secondly, assume that $A_k \neq \emptyset$, then

$$\begin{aligned} \|\Phi^* r^k\|_\infty &= \|\Phi^*(I - P_k)(\Phi x + e)\|_\infty \\ &\geq \|\Phi^*(I - P_k)\Phi x\|_\infty - \|\Phi^*(I - P_k)e\|_\infty \\ &\geq \max_{i \in A_k} |(I - P_k)\Phi x, \phi_i| - \|\Phi^*(I - P_k)e\|_\infty \\ &\geq \frac{1 - \delta}{\sqrt{|A_k|}} \|x_{A_k}\| - \left(1 + \frac{\delta\sqrt{sK}}{1 - \delta}\right) b_\infty \\ &\geq \frac{1 - \delta}{\sqrt{|A_k|}} \sqrt{|A_k|} \frac{2(1 + \frac{\sqrt{sK}\delta}{1 - \delta})b_\infty}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}} - \left(1 + \frac{\delta\sqrt{sK}}{1 - \delta}\right) b_\infty \\ &= \frac{(1 - \delta)2(1 + \frac{\sqrt{sK}\delta}{1 - \delta})b_\infty}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}} - \left(1 + \frac{\delta\sqrt{sK}}{1 - \delta}\right) b_\infty \\ &> \left(1 + \frac{\delta\sqrt{sK}}{1 - \delta}\right) b_\infty. \end{aligned}$$

Here, to derive the third and fourth inequalities, we have used (2.8) of Lemma 2.3 and assumption (2.17), respectively. Hence OMMP(s) does not stop early, and the theorem is proved. \square

2.4 Gaussian noise

As an application of the above results on l_2 and l_∞ bounded noise cases, we shall discuss the performance of OMMP(s) on recovering K -sparse signals with the Gaussian noise. Gaussian noise is of particular interest in statistics since it is probably the best simulation of real noise when the noise source is particularly complex, cf. [19, 20].

The motivation of applying the results on the l_2 bounded and l_∞ bounded noise cases to the Gaussian noise case comes from the following results of Cai, Xu, and Zhang [21], which shows that the Gaussian noise e is essentially bounded in both l_2 and l_∞ norms. They have shown that if e is zero-mean white Gaussian noise with covariance $\sigma^2 I_{M \times M}$, that is, $e \sim N(0, \sigma^2 I_{M \times M})$, then there hold

$$P(e \in \{e : \|e\|_2 \leq \sigma\sqrt{M + 2\sqrt{M \log M}}\}) \geq 1 - \frac{1}{M}$$

and

$$P(e \in \{e : \|\Phi^* e\|_\infty \leq \sigma\sqrt{2 \log N}\}) \geq 1 - \frac{1}{2\sqrt{\pi \log N}}.$$

Combining the above results with those of Theorem 2.1 and Theorem 2.2, we immediately obtain the following theorems.

Theorem 2.3 *Suppose that $e \sim N(0, \sigma^2 I_{M \times M})$ and Φ satisfies the RIP of order sK with the same isometry constant as in Theorem 2.1 and nonzero components x_i satisfy*

$$|x_i| > \frac{2\sigma\sqrt{M + 2\sqrt{M \log M}}}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1 - \delta}}}.$$

Then OMMP(s) with the stopping rule $\|r^k\|_2 \leq \sigma \sqrt{M + 2\sqrt{M \log M}}$ finds the support of any K -sparse signal $x \in \mathbb{R}^N$ from $y = \Phi x + e$ with probability at least $1 - \frac{1}{M}$.

Theorem 2.4 Suppose that $e \sim N(0, \sigma^2 I_{M \times M})$ and Φ satisfies the RIP of order sK with the same isometry constant as in Theorem 2.1 and nonzero components x_i satisfy

$$|x_i| > \frac{2(1 + \frac{\sqrt{sK\delta}}{1-\delta})\sigma \sqrt{2 \log N}}{(1 - \delta) - \sqrt{\frac{K}{s} \frac{\delta}{1-\delta}}}.$$

Then OMMP(s) with the stopping rule $\|\Phi^* r^k\|_\infty \leq (1 + \frac{\sqrt{sK\delta}}{1-\delta})\sigma \sqrt{2 \log N}$ finds the support of any K -sparse signal $x \in \mathbb{R}^N$ from $y = \Phi x + e$ with probability at least $1 - \frac{1}{2\sqrt{\pi \log N}}$.

Theorem 2.3 and Theorem 2.4 show that, with a suitable stopping rule and a reasonable condition on the minimum magnitude of the nonzero coordinates of the K -sparse signal of x , OMMP(s) can recover the support of x in at most K iterations with high probability from measurements with Gaussian noise.

3 Error performance of the OSGA

In this section, we study the error performance of the OSGA in the general context. Let H be a real separable Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$ for all $x \in H$. Recall that a set \mathcal{D} from H is called a dictionary if

$$\mathcal{D} = \{\phi_i, i \in I\} \subset H, \quad \text{and} \quad \overline{\text{span}} \mathcal{D} = H.$$

Without loss of generality we shall assume that the dictionary \mathcal{D} is normalized, i.e.,

$$\|\phi_i\| = 1, \quad i \in I.$$

The standard measure of approximation power of a dictionary is the error of the best m -term approximation. Given a dictionary \mathcal{D} , we define the m -sparse class as

$$\Sigma_m := \Sigma_m(\mathcal{D}) := \left\{ \sum_{j \in \Lambda} c_j \phi_j : c_j \in \mathbb{R}, \phi_j \in \mathcal{D}, \#\Lambda = m \right\}.$$

The error of best m -term approximation to a function $f \in H$ from the dictionary \mathcal{D} is defined as

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{g \in \Sigma_m} \|f - g\|. \tag{3.1}$$

Now we will use OSGA(s) to generate a m term approximant f_m and estimate the error $\|f - f_m\|$ in terms of $\sigma_m(f)$.

So let us first recall the definition of the OSGA(s) (see Algorithm 2).

To investigate the error performance of SOGA, one needs to make some assumptions on the dictionary \mathcal{D} . A simple and useful assumption is the coherence μ of a dictionary \mathcal{D} defined by

$$\mu := \mu(\mathcal{D}) := \sup_{\phi, \psi \in \mathcal{D}, \phi \neq \psi} |\langle \phi, \psi \rangle|.$$

Algorithm 2 Orthogonal Super Greedy Algorithm (OSGA(s))

Suppose we are given an integer s and a target element $r_0 := f \in H$. For each $m \geq 1$, we inductively define

- (1) $\phi_{i_{(m-1)s+1}}, \dots, \phi_{i_{ms}} \in \mathcal{D}$ are the elements of the dictionary \mathcal{D} satisfying the following inequality. Denote $I_m := \{i_{(m-1)s+1}, \dots, i_{ms}\}$ and assume that

$$\min_{i \in I_m} |\langle r_{m-1}, \phi_i \rangle| \geq \sup_{\phi \in \mathcal{D}, \phi \neq \phi_i, i \in I_m} |\langle r_{m-1}, \phi \rangle|.$$

- (2) Let $H_m := H_m(f) := \text{span}(\phi_{i_1}, \dots, \phi_{i_{ms}})$, and let P_{H_m} denote an operator of orthogonal projection onto H_m . Define

$$f_m := G_m(f) := \text{OSGA}_m(f, \mathcal{D}) := P_{H_m}(f).$$

- (3) Define the residual after the m th iteration of the algorithm

$$r_m := r_m^s(f) := f - f_m.$$

Dictionaries with coherence μ are called μ -coherent. It was proved in [8] that for μ -coherent dictionaries the OSGA provides the same (in the sense of order) upper bound on the sparse class for the error as the OGA. In this paper, we will improve this results under a weaker assumption on a dictionary. This assumption is RIP which is formulated in a finite dimensional context in Sect. 2. Here we require the definition of RIP in terms of the dictionary instead of the measurement matrix in a infinite dimensional context:

A dictionary \mathcal{D} is said to satisfy RIP of order k if there exists a constant $\delta \in (0, 1)$ such that, for any subset $J \subset I$ with $\sharp(J) \leq k$ and any scalars $a_i, i \in J$, the following inequalities hold:

$$(1 - \delta) \left(\sum_{i \in J} |a_i|^2 \right) \leq \left\| \sum_{i \in J} a_i \phi_i \right\|^2 \leq (1 + \delta) \left(\sum_{i \in J} |a_i|^2 \right). \tag{3.2}$$

As before, the minimum of all δ satisfying (3.2) is referred to as an isometry constant δ_k .

It is known from [22] that if the dictionary \mathcal{D} has coherence μ , then it satisfies RIP of order k for $k \leq \mu^{-1} + 1$ with isometry constant $\delta_k \leq \mu(k - 1)$, but not vice versa.

Recently, under the RIP assumption, the authors in [7] proved the almost optimality of OGA by establishing the corresponding Lebesgue-type inequality. This inequality quantifies the efficiency of OGA for individual elements. It is natural to ask if we can establish Lebesgue-type inequality for OSGA(s). We give a positive answer to this question. In fact, we prove the following inequality for the OSGA(s) when the dictionary \mathcal{D} satisfies RIP. This is the first Lebesgue-type inequality for the super greedy type algorithms obtained so far as we know.

Theorem 3.1 *Given parameter $s \in \mathbb{N}$, there exist fixed constants $A := 88, \delta^* := 1/10, C := 8$ such that the following holds for all $n \geq 0$: if \mathcal{D} is a dictionary in a Hilbert space H satisfying RIP of order $(As + 1)n$ with isometry constant $\delta_{(As+1)n} \leq \delta^*$, then for any target function*

$f \in H$, we have

$$\|r_{An}\| = \|f - f_{An}\| \leq C\sigma_n(f, \mathcal{D}).$$

In Theorem 3.1, we remark that the values of A, δ^*, C for which the above result holds are coupled. For example, it is possible to obtain a smaller value of C at the price of a larger value of A or of a smaller value of δ^* .

Based on this inequality, we will derive the convergence rate of OSGA on the sparse class induced by dictionaries satisfying RIP conditions.

The rest part of this section is organized as follows. The Sects. 3.1 and 3.2 are devoted to the proof of Theorem 3.1. In Sect. 3.3, as an application of Theorem 3.1, we study the rate of convergence of OSGA on the sparse class induced by RIP dictionaries.

3.1 Reduction of the residual

To prove Theorem 3.1, we establish the following lemma which quantifies the reduction of the residuals generated by the OSGA(s) under the RIP condition. When $s = 1$, this lemma was proved in [7]. However, to deal with the case of $s > 1$, we need some new techniques. We need to establish a new lemma (Lemma 3.1). In what follows, we denote by

$$\Lambda_k := \bigcup_{i=1}^k I_i := \{i_1, \dots, i_{sk}\}$$

the set of indices selected after k iterations of the OSGA(s) applied to the target element $f \in H$.

Lemma 3.1 *Let $(f_k)_{k \geq 0}$ be the sequence of approximations generated by the OSGA(s) applied to f , and let $g = \sum_{i \in T} z_i \phi_i, \#(T) \leq n$. Then, if T is not contained in Λ_k , we have*

$$\|r_{k+1}\|^2 \leq \|r_k\|^2 - \frac{1 - \delta_{\#(T \cup \Lambda_k)}}{(1 + \delta_s)\#(T \setminus \Lambda_k)} (\|r_k\|^2 - \|f - g\|^2). \tag{3.3}$$

Lemma 3.1 quantifies the reduction of $\|r_k\|$ at each iteration provided that T is not contained in Λ_k and that $\|r_k\| \geq \|f - g\|$. Note that in the case when $T \subseteq \Lambda_k$, we have $\|r_k\| \leq \|f - g\|$.

Proof We may assume that $\|r_k\| \geq \|f - g\|$, otherwise inequality (3.3) is trivially satisfied. Denote

$$F_{k+1} := \text{span}(\phi_i, i \in I_{k+1}).$$

Then H_{k+1} is a direct sum of H_k and F_{k+1} . Therefore,

$$\begin{aligned} r_{k+1} &= f - f_{k+1} = r_k + f_k - f_{k+1} \\ &= r_k + f_k - P_{H_{k+1}}(f) \\ &= r_k + f_k - P_{H_{k+1}}(r_k + f_k) \\ &= r_k - P_{H_{k+1}}(r_k). \end{aligned}$$

It is clear that $F_{k+1} \subset H_{k+1}$ implies

$$\begin{aligned} \|r_{k+1}\|^2 &= \|r_k - P_{H_{k+1}}(r_k)\|^2 \\ &\leq \|r_k - P_{F_{k+1}}(r_k)\|^2 \\ &= \|r_k\|^2 - \|P_{F_{k+1}}(r_k)\|^2. \end{aligned} \tag{3.4}$$

Now we estimate $\|P_{F_{k+1}}(r_k)\|$. Since the dictionary \mathcal{D} satisfies RIP of order s with isometry constant δ_s , we have

$$\left\| \sum_{i \in I_{k+1}} a_i \phi_i \right\|^2 \leq (1 + \delta_s) \left(\sum_{i \in I_{k+1}} |a_i|^2 \right). \tag{3.5}$$

Thus

$$\begin{aligned} \|P_{F_{k+1}}(r_k)\| &= \sup_{\psi \in F_{k+1}, \|\psi\| \leq 1} |(P_{F_{k+1}}(r_k), \psi)| \\ &= \sup_{(c_i)_{i \in I_{k+1}}, \|\sum_{i \in I_{k+1}} c_i \phi_i\| \leq 1} \left| \sum_{i \in I_{k+1}} \langle r_k, \phi_i \rangle c_i \right| \\ &\geq \sup_{(c_i)_{i \in I_{k+1}}, \sum_{i \in I_{k+1}} |c_i|^2 \leq (1 + \delta_s)^{-1}} \left| \sum_{i \in I_{k+1}} \langle r_k, \phi_i \rangle c_i \right| \\ &= (1 + \delta_s)^{-1/2} \left(\sum_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \right)^{1/2}. \end{aligned} \tag{3.6}$$

Using inequality (3.6), we can continue to estimate (3.4) as

$$\begin{aligned} \|r_{k+1}\|^2 &\leq \|r_k\|^2 - (1 + \delta_s)^{-1} \left(\sum_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \right) \\ &\leq \|r_k\|^2 - (1 + \delta_s)^{-1} \left(\max_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \right). \end{aligned}$$

Therefore to prove inequality (3.3), it suffices to prove that

$$(1 + \delta_s)^{-1} \left(\max_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \right) \geq \frac{1 - \delta_{\sharp(T \cup \Lambda_k)}}{(1 + \delta_s) \sharp(T \setminus \Lambda_k)} (\|r_k\|^2 - \|f - g\|^2). \tag{3.7}$$

To prove this, we first note that

$$\begin{aligned} 2\sqrt{\|r_k\|^2 - \|f - g\|^2} \|g - f_k\| &\leq \|r_k\|^2 - \|f - g\|^2 + \|g - f_k\|^2 \\ &= \|r_k\|^2 - \|g - f_k - r_k\|^2 + \|g - f_k\|^2 \\ &\leq 2|\langle g - f_k, r_k \rangle| = 2|\langle g, r_k \rangle|. \end{aligned}$$

This inequality can be written as

$$\|r_k\|^2 - \|f - g\|^2 \leq \frac{|\langle g, r_k \rangle|}{\|g - f_k\|^2}. \tag{3.8}$$

If we write $f_k = \sum_{i \in \Lambda_k} c_i^k \phi_i$, $\delta := \delta_{\#(T \cup \Lambda_k)}$, then the denominator of the right-hand side of inequality (3.8) satisfies the RIP

$$\begin{aligned} \|g - f_k\|^2 &= \left\| \sum_{i \in T} z_i \phi_i - \sum_{i \in \Lambda_k} c_i^k \phi_i \right\|^2 \\ &= \left\| \sum_{i \in T \setminus \Lambda_k} z_i \phi_i + \sum_{i \in T \cap \Lambda_k} (z_i - c_i^k) \phi_i + \sum_{i \in \Lambda_k \setminus T} (-c_i^k) \phi_i \right\|^2 \\ &\geq (1 - \delta) \left(\sum_{i \in T \setminus \Lambda_k} |z_i|^2 + \sum_{i \in T \cap \Lambda_k} |z_i - c_i^k|^2 + \sum_{i \in \Lambda_k \setminus T} |c_i^k|^2 \right) \\ &\geq (1 - \delta) \left(\sum_{i \in T \setminus \Lambda_k} |z_i|^2 \right). \end{aligned}$$

On the other hand, the numerator of the right-hand side of inequality (3.8) satisfies

$$\begin{aligned} |\langle g, r_k \rangle|^2 &= \left| \left\langle \sum_{i \in T} z_i \phi_i, r_k \right\rangle \right|^2 \\ &= \left| \left\langle \sum_{i \in T \setminus \Lambda_k} z_i \phi_i, r_k \right\rangle \right|^2 \\ &\leq \left(\sum_{i \in T \setminus \Lambda_k} |z_i| |\langle \phi_i, r_k \rangle| \right)^2 \\ &\leq \max_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \left(\sum_{i \in T \setminus \Lambda_k} |z_i| \right)^2 \\ &\leq \max_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2 \#(T \setminus \Lambda_k) \sum_{i \in T \setminus \Lambda_k} |z_i|^2. \end{aligned}$$

Therefore we obtain

$$\|r_k\|^2 - \|f - g\|^2 \leq \frac{\#(T \setminus \Lambda_k) (\max_{i \in I_{k+1}} |\langle r_k, \phi_i \rangle|^2)}{1 - \delta_{\#(T \cup \Lambda_k)}},$$

which implies (3.7). Thus we complete the proof of Lemma 3.1. □

The following proposition is an immediate consequence of Lemma 3.1.

Proposition 3.1 *Assume that, for a given integer $A > 0$ and $\delta^* < 1$, a dictionary \mathcal{D} satisfies RIP of order $(As + 1)n$ with $\delta_{(As+1)n} \leq \delta^*$. If $g = \sum_{i \in T} z_i \phi_i$, $\#(T) \leq n$, then for any triple of non-negative integers (j, m, L) such that $\#(T \cup \Lambda_j) \leq m$ and $j + mL \leq (As + 1)n$, we have*

$$\|r_{j+mL}\|^2 \leq e^{-\frac{1-\delta^*}{1+\delta^*}L} \|r_j\|^2 + \|f - g\|^2. \tag{3.9}$$

Proof By Lemma 3.1, if $g = \sum_{i \in T} z_i \phi_i$, $\sharp(T) \leq n$, then for any triple of non-negative integers (j, m, L) such that $\sharp(T \cup \Lambda_j) \leq m$ and $j + mL \leq (As + 1)n$, we have

$$\begin{aligned} \|r_{j+mL}\|^2 - \|f - g\|^2 &\leq \left(1 - \frac{1 - \delta^*}{(1 + \delta^*)m}\right)^{mL} \max\{0, \|r_j\|^2 - \|f - g\|^2\} \\ &\leq e^{-\frac{1 - \delta^*}{1 + \delta^*}L} \max\{0, \|r_j\|^2 - \|f - g\|^2\}, \end{aligned}$$

where we have used the fact that $\sharp(T \cup \Lambda_l) \leq m$ for all $l \geq j$. This implies inequality (3.9). Thus we complete the proof of Proposition 3.1. \square

3.2 Lebesgue-type inequality for SOGA

We are in a position to prove Theorem 3.1. Fix $f \in H$, we first observe that the assertion of the theorem can be derived from the following lemma.

Lemma 3.2 *If $0 \leq k < n$ satisfies*

$$\|r_{Ak}\| \leq 2\sigma_k(f), \quad \sigma_k(f) > 4\sigma_n(f), \tag{3.10}$$

then there exists $k < \tilde{k} \leq n$ such that

$$\|r_{A\tilde{k}}\| \leq 2\sigma_{\tilde{k}}(f), \quad \sigma_{\tilde{k}}(f) \leq 4\sigma_n(f). \tag{3.11}$$

Indeed, assuming that Lemma 3.2 holds, we complete the proof of Theorem 3.1 as follows. We let \tilde{k} be the largest integer in $\{0, 1, \dots, n\}$ for which $\|r_{A\tilde{k}}\| \leq 2\sigma_{\tilde{k}}(f)$. Since

$$\|r_0\| = \sigma_0 = \|f\|,$$

such \tilde{k} exists. If $\tilde{k} < n$, then we must have $\sigma_{\tilde{k}}(f) \leq 4\sigma_n(f)$, and therefore

$$\|r_{An}\| \leq \|r_{A\tilde{k}}\| \leq 2\sigma_{\tilde{k}}(f) \leq 8\sigma_n(f). \tag{3.12}$$

We are therefore left with proving Lemma 3.2. For this, we fix

$$\delta^* = \frac{1}{10}, \tag{3.13}$$

and $0 \leq k < n$ such that (3.10) holds. Let $k < K \leq n$ be the first integer such that $\sigma_k > 4\sigma_K$. By the definition of $\sigma_K(f)$, for any $B > 1$, there exists $g = \sum_{j \in T} z_j \psi_j$, $\sharp(T) = K$ such that

$$\|f - g\| \leq B\sigma_K(f).$$

The significance of K is that on the one hand

$$\|f - g\| \leq B\sigma_K(f) < \frac{B}{4}\sigma_k(f), \tag{3.14}$$

while on the other hand

$$\sigma_k(f) \leq 4\sigma_{K-1}(f). \tag{3.15}$$

To apply Proposition 3.1 for the above g and $j = Ak$, we need to bound $\sharp(T \cup \Lambda_{Ak})$ with A yet to be specified. To this end, we write $K = k + M$, with $M > 0$, and notice that if $S \subset T$ is any set with $\sharp(S) = M$ and $g_S := \sum_{j \in S} z_j \psi_j$, then

$$\begin{aligned} \|g_S\| &\geq \|f - (g - g_S)\| - \|f - g\| \\ &\geq \sigma_k(f) - B\sigma_K(f) \\ &\geq \left(1 - \frac{B}{4}\right)\sigma_k(f), \end{aligned} \tag{3.16}$$

where we have used the fact that $g - g_S \in \Sigma_k$. Using the RIP, we obtain the following lower bound for the coefficients of g : for any set $S \subset T$ with $\sharp(S) = M$,

$$\left(1 - \frac{B}{4}\right)^2 \sigma_k^2(f) \leq \|g_S\|^2 \leq (1 + \delta^*) \sum_{j \in S} |z_j|^2 = \frac{11}{10} \sum_{j \in S} |z_j|^2. \tag{3.17}$$

Take for S the set S_g of the M smallest coefficients of g and note that then, for any more general $S \subset T$ with $\sharp(S) \geq M$, one has

$$\left(\sum_{j \in S} |z_j|^2\right) / \left(\sum_{j \in S_g} |z_j|^2\right) \geq \sharp(S)/M,$$

and hence

$$\frac{10}{11} \left(1 - \frac{B}{4}\right)^2 \frac{\sharp(S)}{M} \sigma_k^2(f) \leq \sum_{j \in S} |z_j|^2. \tag{3.18}$$

Now we consider the particular set $S := T \setminus \Lambda_{Ak}$ satisfying $\sharp(S) \geq M$. Combining the above bound with the RIP, we obtain

$$\begin{aligned} \frac{10}{11} \left(1 - \frac{B}{4}\right)^2 \frac{\sharp(S)}{M} \sigma_k^2(f) (1 - \delta^*) &\leq \|g_S\|^2 \leq \|g - f_{Ak}\|^2 \\ &\leq (\|g - f\| + \|r_{Ak}\|)^2 \\ &\leq (B\sigma_K(f) + 2\sigma_k(f))^2 \\ &\leq \left(\frac{B}{4} + 2\right)^2 \sigma_k^2(f). \end{aligned}$$

Since $\delta^* = \frac{1}{10}$ this gives the bound

$$\sharp(T \setminus \Lambda_{Ak}) \leq \frac{9}{11} \frac{(\frac{B}{4} + 2)^2}{(1 - \frac{B}{4})^2} M \leq 11M, \tag{3.19}$$

where the second inequality is obtained by taking B sufficiently close to 1.

We proceed now to verifying Lemma 3.2 with $\tilde{k} = K - 1$ when $K - 1 > k$ and with $\tilde{k} = k + 1$ otherwise. In the first we can use the estimate in Proposition 3.1 with $j = Ak$ in combination with (3.15) to deal with the term $\|r_{Ak}\|$ in (3.9). When $K = k + 1$, however, we cannot

bound $\|r_{Ak}\|$ directly in terms of $\sigma_l(f)$ for some $l > k$. Accordingly, we use Proposition 3.1 in different ways for the two cases.

In the case where $K - 1 > k$, i.e., $M \geq 2$, we apply (3.9) with $j = Ak$, $m = 11M$, and $L = 4$. Indeed $Ak + Lm = Ak + 44M \leq An$ holds for $k + M \leq n$ whenever $A \geq 44$. Moreover, notice that for such A

$$\begin{aligned} A(K - 1) &= Ak + A(M - 1) \\ &\geq Ak + \frac{1}{2}AM \\ &= Ak + \frac{Am}{22} \\ &= Ak + Lm, \end{aligned}$$

whenever

$$A \geq 88. \tag{3.20}$$

This gives

$$\begin{aligned} \|r_{A(K-1)}\|^2 &\leq \|r_{Ak+mL}\|^2 \\ &\leq e^{-\frac{36}{11}} \|r_{Ak}\|^2 + \|f - g\|^2 \\ &\leq e^{-\frac{36}{11}} 4\sigma_k^2(f) + B^2\sigma_k^2(f) \\ &\leq e^{-\frac{36}{11}} 64\sigma_{K-1}^2(f) + B^2\sigma_{K-1}^2(f) \\ &\leq 4\sigma_{K-1}^2(f), \end{aligned}$$

where we have used (3.15) in the fourth inequality, and the last inequality follows by taking B sufficiently close to 1. We thus obtain (3.11) for the value $\tilde{k} = K - 1 > k$.

In the case where $K - 1 = k$, i.e., $M = 1$, we apply (3.9) with $j = Ak$, $m = 11$, and $L = 8$. From (3.19), we know that $\sharp(T \cup \Lambda_{Ak}) \leq 11$ and $An \geq A(k + 1) \geq Ak + mL$ for A satisfying (3.20). This yields

$$\begin{aligned} \|r_{A(k+1)}\|^2 &\leq \|r_{Ak+mL}\|^2 \\ &\leq e^{-\frac{72}{11}} \|r_{Ak}\|^2 + \|f - g\|^2 \\ &\leq e^{-\frac{72}{11}} 4\sigma_k^2(f) + B^2\sigma_{k+1}^2(f) \\ &\leq \left(4e^{-\frac{72}{11}} + \frac{B^2}{16}\right)\sigma_k^2(f). \end{aligned}$$

This implies that $\Lambda_{A(k+1)}$ contains T . In fact, if it missed one of the indices $i \in T$, then we derive from the RIP condition

$$\begin{aligned} (1 - \delta^*)|z_i|^2 &\leq \|g - f_{\Lambda(k+1)}\|^2 \\ &\leq (\|f - g\| + \|r_{A(k+1)}\|)^2 \end{aligned}$$

$$\begin{aligned} &\leq \left(B\sigma_k(f) + \sqrt{4e^{-\frac{72}{11}} + \frac{B^2}{16}}\sigma_k(f) \right)^2 \\ &\leq \left(\frac{B}{4} + \sqrt{4e^{-\frac{72}{11}} + \frac{B^2}{16}} \right) \sigma_k^2(f). \end{aligned}$$

On the other hand, we know from (3.18) that

$$\frac{10}{11} \left(1 - \frac{B}{4} \right)^2 \sigma_k^2(f) \leq |z_i|^2, \tag{3.21}$$

which for B sufficiently close to 1 is a contradiction since

$$\frac{10}{11} \left(1 - \frac{B}{4} \right)^2 > \frac{10}{9} \left(\frac{B}{4} + \sqrt{4e^{-\frac{72}{11}} + \frac{B^2}{16}} \right).$$

This implies that $\|r_{A(k+1)}\| \leq \sigma_{k+1}(f)$, and therefore inequality (3.11) holds for the value $\tilde{k} = k + 1$. This verifies Lemma 3.2 and hence completes the proof of Theorem 3.1.

3.3 Convergence rate on sparse class

As an application of Theorem 3.1, we will derive the convergence rate of SOGA on the sparse class induced by a dictionary. Firstly, we recall the definitions of these classes. For a general dictionary \mathcal{D} in H , we define the class of functions

$$A_1^0(\mathcal{D}) := \left\{ f \in H : f = \sum_{i \in \Lambda} c_i(f) \phi_i, \phi_i \in \mathcal{D}, \#\Lambda < \infty, \sum_{i \in \Lambda} |c_i(f)| \leq 1 \right\}$$

and we define $A_1(\mathcal{D})$ to be the closure of $A_1^0(\mathcal{D})$ in H . It is well known that the class $A_1(\mathcal{D})$ plays an important role in the study of greedy approximation with respect to dictionaries. In [23], DeVore and Temlyakov proved that, for an arbitrary dictionary \mathcal{D} in H , the OGA provides, after m iterations, an approximation of $f \in A_1(\mathcal{D})$ with the following upper bound of the convergence rate:

$$r_m^{\text{OGA}}(f) \leq m^{-1/2}.$$

Note that the rate $m^{-1/2}$ is sharp since when \mathcal{D} is an ortho-normal basis of H , it is easy to find $f_0 \in A_1(\mathcal{D})$ such that

$$r_m^{\text{OGA}}(f_0) = c \cdot m^{-1/2}.$$

Unlike OGA, to get the same convergence rate on $A_1(\mathcal{D})$ for SOGA, one must make some extra assumptions on the dictionary \mathcal{D} . A specific feature of a dictionary, μ -coherence, allows us to build OSGA with the same rate of convergence. Assuming that the dictionary \mathcal{D} is μ -coherent, Liu and Temlyakov in [8] proved the following theorem.

Theorem 3.2 *If \mathcal{D} is a dictionary in a Hilbert space H with coherence parameter μ . Then, for $s \leq (2\mu)^{-1}$, OSGA(s) provides, after m iterations, an approximation of $f \in A_1(\mathcal{D})$ with the following upper bound on the error:*

$$\|r_m\|^2 = \|f - f_m\|^2 \leq 40.5(sm)^{-1}.$$

Now we present our results. By an immediate consequence of Theorem 3.1, we get the following theorem.

Theorem 3.3 *Given parameter $s \in \mathbb{N}$, for all $n \geq 0$, if \mathcal{D} is a dictionary in a Hilbert space H satisfying RIP of order $(88s + 1)n$ with isometry constant $\delta_{(88s+1)n} \leq 1/10$, then OSGA(s) provides, after m iterations, an approximation of $f \in A_1(\mathcal{D})$, with the following error bound:*

$$\|r_{88m}\| = \|f - f_{88m}\| \leq Cn^{-\frac{1}{2}}.$$

Proof As the discussion above, the results of DeVore and Temlyakov [8] imply the following estimate for the best m term approximation error of functions $f \in A_1(\mathcal{D})$:

$$\sigma_n(f, \mathcal{D}) \leq n^{-\frac{1}{2}}, \quad n = 1, 2, \dots \quad (3.22)$$

Combining (3.22) with Theorem 3.1, we complete the proof of Theorem 3.3. \square

Acknowledgements

Thanks for all the referenced authors.

Funding

This work is supported by the National Nature Science Foundation of China [Grant Nos. 11701411, 11271199, 11671213, 11401247].

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The authors contributed equally and significantly in writing this paper. All authors read and approved the final manuscript.

Author details

¹School of Sciences, Tianjin Chengjian University, Tianjin, China. ²School of Mathematical Sciences and LPMC, Nankai University, Tianjin, China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 January 2019 Accepted: 16 April 2019 Published online: 08 May 2019

References

- DeVore, R.A.: Nonlinear approximation. *Acta Numer.* **7**, 51–150 (1998)
- Pati, Y.C., Rezaifar, R., Krishnaprasad, P.S.: Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition (1993)
- Lin, J.H., Li, S.: Nonuniform support recovery from noisy measurements by orthogonal matching pursuit. *J. Approx. Theory* **165**, 20–40 (2013)
- Liu, E., Temlyakov, V.N.: Super greedy type algorithm. *Adv. Comput. Math.* **37**, 493–504 (2012)
- Needell, D., Tropp, J.A.: Cosamp: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **26**, 301–321 (2009)
- Cai, T., Wang, L.: Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* **57**, 4680–4688 (2011)
- Cohen, A., Dahmen, W., DeVore, R.A.: Orthogonal matching pursuit under the restricted isometry property. *Constr. Approx.* **45**, 113–127 (2017)
- Liu, E., Temlyakov, V.N.: The orthogonal super greedy algorithm and applications in compressed sensing. *IEEE Trans. Inf. Theory* **58**, 2040–2047 (2012)
- Wei, D.: Analysis of orthogonal multi-matching pursuit under restricted isometry property. *Sci. China Math.* **57**, 2179–2188 (2014)
- Wang, J., Kwon, S., Shim, B.: Generalized orthogonal matching pursuit. *IEEE Trans. Signal Process.* **60**, 6202–6216 (2012)

11. Wei, X.J., Ye, P.X.: Estimates of restricted isometry constant in super greedy algorithms. *Int. J. Future Gener. Commun. Netw.* **8**, 137–144 (2015)
12. Yi, S., Li, B., Pan, W.L., Li, J.: Analysis of generalised orthogonal matching pursuit using restricted isometry constant. *Electron. Lett.* **14**, 1020–1022 (2014)
13. Baraniuk, R.G.: Compressive sensing. *IEEE Signal Process. Mag.* **24**, 118–121 (2007)
14. Candes, E.J.: Compressive sampling. In: *Int. Congress of Mathematics*, vol. 3, pp. 1433–1452 (2006)
15. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006)
16. Candes, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005)
17. Candes, E.J., Tao, T.: Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory* **52**, 5406–5425 (2005)
18. Satpathi, S., Das, R., Chakraborty, M.: Improving the bound on the rip constant in generalized orthogonal matching pursuit. *IEEE Signal Process. Lett.* **20**, 1074–1077 (2013)
19. Candes, E.J., Tao, T.: The Dantzig selector: statistical estimation when p is much larger than n . *Ann. Stat.* **35**, 2313–2351 (2007)
20. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* **32**, 407–451 (2004)
21. Cai, T., Xu, G., Zhang, J.: On recovery of sparse signals via l_1 minimization. *IEEE Trans. Inf. Theory* **55**, 3388–3397 (2009)
22. Tropp, J.A.: Greedy is good: algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* **50**, 2231–2242 (2004)
23. DeVore, R.A., Temlyakov, V.N.: Some remarks on greedy algorithms. *Adv. Comput. Math.* **5**, 173–187 (1996)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
