IMI

Preprint Series

Interdisciplinary Mathematics Institute

2010:02

On performance of greedy algorithms

V. Temlyakov and P. Zheltov

College of Art and Science
University of South Carolina

ON PERFORMANCE OF GREEDY ALGORITHMS

VLADIMIR N. TEMLYAKOV AND PAVEL ZHELTOV

ABSTRACT. In this paper we show that for dictionaries with small coherence in a Hilbert space the Orthogonal Greedy Algorithm (OGA) performs almost as well as the best $m-$term approximation for all signals with sparsity almost as high as the best theoretically possible threshold $s = \frac{1}{2}(M^{-1}+1)$ by proving a Lebesgue-type inequality for arbitrary signals. On the other hand, we present an example of a dictionary with coherence M and an $s-$sparse signal for which OGA fails to pick up any atoms from the support, thus showing that the above threshold is sharp. Also, by proving a Lebesgue-type inequality for Pure Greedy Algorithm (PGA), we show that PGA matches the rate of convergence of the best $m-$term approximation, even beyond the saturation limit of $m^{-\frac{1}{2}}$.

1. Introduction

In this paper we mainly study the efficiency of the Orthogonal Greedy Algorithm (OGA), which is also known as the Orthogonal Matching Pursuit (OMP) in the compressed sensing community, when dealing mostly with finite-dimensional spaces. To preserve a more theoretical flavor of our result, we'll stay with a term more traditional in the field of approximation theory.

OGA is a simple yet powerful algorithm for highly nonlinear sparse approximation that enjoyed a long history of research, for example, see [1, 2, 4, 9, 10, 14, 16, 17], and [15] for a survey. Since conception, its performance served as a baseline of comparison for the other algorithms, such as Regularized OMP [13], Stagewise OMP [6] and others [12].

Previous work ([3, 4, 8, 16]) has shown that both OGA and convex relaxation known as Basis Pursuit recover sparse signals $f = \Phi x$ exactly if support size doesn't exceed a critical threshold $m = \frac{1}{2}(M^{-1}+1)$. In particular, Orthogonal Greedy Algorithm does so in exactly m steps.

2000 *Mathematics Subject Classification.* Primary 41A25; Secondary 41A17, 41A46.

Key words and phrases. Additive-type Lebesgue inequality, Greedy algorithm, Orthogonal Matching Pursuit, m-term approximation, incoherent dictionary, coherence, Orthogonal Greedy Algorithm, sparse representation.

In fact, OGA will recover one atom from the support of sparse f on every step.

The results in this paper are threefold. In the first section, we will show that OGA performs as well as the best $m-$term approximation (up to a factor of $\exp(\sqrt{\log m})$) for all signals with sparsity almost as high as the best theoretically possible threshold $m = O(\frac{1}{M})$ (up to the same factor). After that, we will show that this threshold is sharp by constructing an explicit example of a dictionary with small coherence and a signal with sparsity $m = \frac{1}{2}(M^{-1}+1)$, for which OGA fails to find its sparse approximation within m steps.

In a separate section we will explore OGA's sibling Pure Greedy Algorithm (PGA), which goes under the name of Projection Pursuit ([9, 10]) among statisticians. While even simpler to implement, as has been shown in [2, 11] to perform in the sense of rate of approximation of elements from special classes (namely, $\mathcal{A}_1(\mathcal{D})$) less efficiently than OGA.

We will work in a Hilbert space H. The *dictionary* $\mathcal{D}$ is an arbitrary collection of elements $\{\varphi_i, i \in \mathbb{N}\} \subset H$ such that $\operatorname{span}\mathcal{D}$ is dense in H. For convenience, we will assume all elements (or atoms) are normalized ($\|\varphi\| = 1$). We would be interested in the property of $\mathcal{D}$ called *coherence*:

$$M := \sup\{|\langle \varphi_i, \varphi_j\rangle| : \quad \varphi_i, \varphi_j \in \mathcal{D}, \varphi_i \neq \varphi_j\}.$$

We will commonly use $\Gamma \subset \mathbb{N}$ for a finite set of indices, and Φ_Γ for the collection of atoms from $\mathcal{D}$ indexed by Γ. Moreover, we will omit subscript Γ where it will be obvious from context. It would be convenient then to abuse finite-dimensional notation for linear combinations of elements from Φ,

$$\Phi = \{\varphi\}_{\gamma\in\Gamma}, \quad \Phi x = \sum_{\gamma\in\Gamma} x_\gamma \varphi_\gamma,$$

for scalar products of f with Φ (which can be seen as adjoint operator of Φ)

$$\Phi^* f = [\langle \varphi, f\rangle]_{\gamma\in\Gamma},$$

and for coefficients of projection of f onto $\operatorname{span}\Phi$ (pseudoinverse):

$$\Phi^\dagger f = \operatorname*{arg\,min}_{x\in\mathbb{R}^m} \|f - \Phi x\|, \quad \Phi\Phi^\dagger f = \operatorname{proj}_\Phi f.$$

Also, we will use notation log for the base-2 logarithm.

Finally, Pure and Orthogonal Greedy Algorithms construct sequences of approximations of a given signal $f \in H$ according to the following theoretical procedure:

Initialize residual $f_0 := f$ and the index set $\Gamma_0 := \varnothing$
Repeat for $s = 1, 2, \dots$:

Find the best atom in $\mathcal{D}$:	$\varphi_s = \arg\max_{\varphi\in\mathcal{D}} \lvert\langle f_{s-1}, \varphi\rangle\rvert$
Add it to the list:	$\Phi_s = \Phi_{s-1} \cup \{\varphi_s\}$
PGA: Subtract its contribution:	$f_s = f_{s-1} - \langle f_{s-1}, \varphi_s\rangle\varphi_s$
OGA: Project onto Φ_s:	$f_s = f - \operatorname{proj}_{\Phi_s} f$

2. Lebesgue-type inequalities for OGA

To reduce visual clutter, we will use the notation f_k for the residuals of both OGA and PGA throughout the paper. It will always be clear from context which algorithm is being used.

First, we will assume that the maximizer exists, otherwise some modifications are necessary.

For a function f from H we will define its best m-term approximation error

$$\sigma_m(f) := \inf_{\Gamma:|\operatorname{supp}\Gamma|=m} \inf_{x\in\mathbb{R}^m} \|f - \Phi_\Gamma x\| .$$

This quantity will serve as a benchmark for performance of the greedy algorithms. Following [5], we call such inequalities Lebesgue-type.

The first result of this kind was proven by Gilbert, Muthukrishnan and Strauss in [7]:

Theorem 1. *For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$*

$$\|f_m\| \le 8\sqrt{m}\sigma_m(f), \ \textit{if } m+1 \le \frac{1}{8\sqrt{2}M}.$$

The constants were improved by Tropp in [16];

Theorem 2. *For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$*

$$\|f_m\| \le \sqrt{1+6m}\sigma_m(f), \ \textit{if } m \le \frac{1}{3M}.$$

Of course, this provides a guarantee that for sparse signals OGA will recover its support exactly after at most m iterations, but on the other hand, the factor in front of σ is huge. This problem was solved by Donoho, Elad and Temlyakov in [5], where they have shown that

Theorem 3. *For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$*

$$\left\|f_{\lfloor m\log m\rfloor}\right\| \le 24\sigma_m(f), \ \textit{if } m \le \frac{1}{20M^{\frac{2}{3}}}.$$

This is much better in a sense that we have an absolute constant as an extra factor, but to obtain that they had to sacrifice depth of search ($m \log m$ now) and critical sparsity (only $M^{-\frac{2}{3}}$) to kill the square root. The method they used was based on the following fact:

Theorem 4. *For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$ and any k, s*

$$\|f_{k+s}\|^2 \leq 2\,\|f_k\|\,(3M(k+s)\,\|f_k\| + \sigma_s(f_k))\,, \;\; \textit{if}\; k+s \leq \frac{1}{2M}.$$

In other words, it is possible to estimate f_{k+s} in terms of the best $m-$term approximation of f_k. A clever recursive argument primed with Theorem 2 then establishes Theorem 3.

The approach used in this paper is a modification of their argument, replacing a crude triangle inequality in (2.6, [5]) by Parseval identity. This allows to essentially close the gap between $\frac{2}{3}$ and 1. For the sake of brevity, we refer you to their paper [5] for proofs of most of the important lemmas that we will need, initial setup and notation. We believe that current result (following) is more natural for this construction, and we suggest to call it an *additive-type Lebesgue inequality.*

Theorem 5. *(Additive-type Lebesgue inequality) For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$ and any $k < s$*

$$\|f_{k+s}\|^2 \leq 7Ms\,\|f_k\|^2 + \sigma_s(f_k)^2, \;\; \textit{if}\; k+s < \frac{1}{2M}.$$

Corollary 6. *For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$*

$$\left\|f_{m\lfloor 2^{\sqrt{\log m}}\rfloor}\right\| \leq 3\sigma_m(f), \;\; \textit{if} \quad m2^{\sqrt{2\log m}} \leq \frac{1}{26M}.$$

Note that the expression $m2^{\sqrt{\log m}}$ grows slower than any $m^{1+\varepsilon}$. Comparing to $m \log m$ in Corollary 2.1, [5], some sacrifices in the depth of search had to be made, but they don't offset the gains on the sparsity front, which is evident from another corollary:

Corollary 7. *For every $M-$coherent dictionary $\mathcal{D}$, any signal $f \in H$ and any fixed $\delta > 0$*

$$\left\|f_{m2^{\lceil \frac{1}{\delta}\rceil}}\right\| \leq 3\sigma_m(f), \;\; \textit{if} \quad m \leq \left(\frac{1}{14M}\right)^{\frac{1}{1+\delta}} 2^{-\lceil \frac{1}{\delta}\rceil}$$

3. Lebesque inequality for PGA

In a PGA setting we lack Theorem 2, and the fact that atoms chosen previously can still reappear in the expansion precludes a full force of Theorem 5. On the other hand this is an advantage over OGA in a sense that PGA outputs a greedy expansion, i.e. it is completely sequential, and therefore we can use $f_{m+n} = (f_m)_n$. This allows Theorem 8, an analogue of Theorem 5 with $k = 0$, to be a source of a surprising pair of corollaries. We will show that if the best $m-$term approximation rate is $O(m^{-r})$ for some fixed r, then PGA matches this rate up to a constant factor. This is a first result that breaks the saturation barrier. While a similar result was proven in [5] for a general class of dictionaries called $\lambda-$quasiorthogonal (which includes $M-$coherent dictionaries), it suffered from what is known as *saturation property*. That is, even imposing extremely tough restrictions on $\sigma_m(f)$, we cannot get better than $m^{-\frac{1}{2}}$ rate of approximation by PGA. In fact, DeVore and Temlyakov in [2] construct a signal and a dictionary such that $\sigma_2(f) = 0$, but $\|f - f_m\| \geq m^{-\frac{1}{2}}$ for $m \geq 4$. Note that the coherence of their dictionary is $M = \sqrt{33/89} = 0.61...$, which entails $m < \frac{1}{2M} < 1$, and therefore, the following theorems are, of course, useless.

Theorem 8. *(Additive-type Lebesgue inequality for PGA) For every $M-$coherent dictionary $\mathcal{D}$ and any signal $f \in H$*

$$\|f_s\|^2 \leq 9Ms\|f\|^2 + \sigma_s(f)^2, \text{ if } s \leq \frac{1}{2M}.$$

Corollary 9. *Let $\mathcal{D}$ have coherence M and signal $f \in H$ be such that for some fixed $r > 0$ and for all*

$$m2^{\sqrt{10r\log m}} \leq \frac{1}{18M}$$

it is true that

$$\sigma_m(f) \leq m^{-r}\|f\|.$$

Then for all such m

$$\|f_{m2^{\sqrt{10r\log m}}}\| \leq 2m^{-r}\|f\|.$$

Just as Corollary 7 is a "hard" realization of a "soft" Corollary 6 as far as power of m is concerned, Corollary 10 is a power of m version of a previous Corollary 9.

Corollary 10. *Let $\mathcal{D}$ have coherence M and signal $f \in H$ be such that for some fixed $\delta > 0$, $r > 0$ and for all*

$$m \leq N(\delta) := \left(\frac{1}{18M}\right)^{\frac{1}{1+\delta}} 2^{-\lceil\frac{1}{\delta}\rceil}$$

it is true that

$$\sigma_m(f) \leq m^{-r} \|f\|.$$

Then there exists a constant $C(\delta, r)$ such that m-th PGA residual is suboptimal:

$$\|f_m\| \leq C(\delta, r) m^{-r} \|f\| \text{ for all } m \leq N(\delta).$$

In other words, almost on the whole interval $m \in [0, O(\frac{1}{M})]$ we get PGA residuals matching the rate of best approximation. The proofs repeat for most part the corresponding proofs for OGA, so we will describe the necessary changes only.

4. A dictionary with small coherence that is difficult for OGA

From [16] (see also [4]) we know that OGA recovers $m-$sparse signal over $M-$coherent dictionary $\mathcal{D}$ exactly in m steps if

$$m < \frac{1}{2}\left(\frac{1}{M} + 1\right).$$

We will show that an above estimate is sharp.

Theorem 11. *For any $0 < M < 1$ there exists $M-$coherent dictionary $\mathcal{D}$ and an $m-$sparse signal f such that $m = \frac{1}{2}\left(\frac{1}{M}+1\right)$ but OGA will never recover x exactly.*

Proof. Let $\{e_j\}_{j=1}^{\infty}$ be the standard basis for $H = \ell^2$ and a signal $f = \sum_{i=1}^{m} e_i$ with norm $\|f\| = \sqrt{m}$. Let the dictionary $\mathcal{D}$ be a basis of H comprised of the following two kinds of atoms:

$$D_{\text{good}} = \{\varphi_i = \alpha e_i - \beta f, \quad i = 1, \ldots, m\}, \text{ and}$$
$$D_{\text{bad}} = \{\varphi_j = \eta e_j + \gamma f, \quad j = m+1, \ldots\}.$$

It is enough to consider $\alpha, \beta, \gamma > 0$. Also, let all φ's above to be norm-1: η is chosen in a way to normalize $\varphi_j, j = m+1, \ldots,$ and

$$(\alpha - \beta)^2 + (m-1)\beta^2 = 1. \tag{1}$$

The following are the scalar products of f with the dictionary:

$$\text{For } \varphi_i \in \mathcal{D}_{\text{good}} \quad \langle \varphi_i, f\rangle = \langle \alpha e_i - \beta f, f\rangle = \alpha - m\beta$$
$$\text{For } \varphi_j \in \mathcal{D}_{\text{bad}} \quad \langle \varphi_j, f\rangle = \langle \eta e_j + \gamma f, f\rangle = m\gamma.$$

Let's require the above dot products to be equal ($R := m\gamma = \alpha - m\beta$). This will allow some realization of OGA to select φ_{m+1} on the first step. Now the scalar products of the elements in $\mathcal{D}$ are as follows $(i \neq i', j \neq j', i, i' \leq m < j, j')$:

$$\begin{aligned}
\langle \varphi_i, \varphi_{i'} \rangle &= \langle \alpha e_i - \beta f, \alpha e_{i'} - \beta f \rangle = m\beta^2 - 2\alpha\beta = -m\beta(2\gamma + \beta), \\
\langle \varphi_j, \varphi_{j'} \rangle &= \langle \eta e_j + \gamma f, \eta e_{j'} + \gamma f \rangle = m\gamma^2, \\
\langle \varphi_i, \varphi_j \rangle &= \langle \alpha e_i - \beta f, \eta e_j + \gamma f \rangle = \gamma(-m\beta + \alpha) = \gamma R = m\gamma^2.
\end{aligned}$$

Then coherence of such a dictionary is

$$M := \max(m\gamma^2, m\beta(2\gamma + \beta)),$$

and it make sense to require $\gamma^2 = \beta(2\gamma + \beta)$. Solving this quadratic equation, we get $\gamma = (1 + \sqrt{2})\beta$. Now we can find $\alpha = m\gamma + m\beta = m(2 + \sqrt{2})\beta$, and plugging in (1) we can find β:

$$(m(2 + \sqrt{2}) - 1)^2\beta^2 + (m - 1)\beta^2 = 1$$

$$\beta^2 = \frac{1}{m^2(2 + \sqrt{2})^2 - m(3 + 2\sqrt{2})}.$$

The bottom line is that

$$M = m\gamma^2 = m(1 + \sqrt{2})^2\beta^2 = \frac{(1 + \sqrt{2})^2}{m(2 + \sqrt{2})^2 - (3 + 2\sqrt{2})}.$$

Denote $A := 1 + \sqrt{2}$ and notice that $2 + \sqrt{2} = \sqrt{2}A, \quad 3 + 2\sqrt{2} = A^2$. Now simplify:

$$M = \frac{A^2}{2A^2 \cdot m - A^2} = \frac{1}{2m - 1}, \qquad \text{or} \qquad m = \frac{1}{2}\left(\frac{1}{M} + 1\right).$$

Now remember OGA picked a wrong atom ψ from $\mathcal{D}_{\text{bad}}$ on the first step: $f' = f - \langle f, \psi \rangle \psi$. By induction, suppose that by the $n-$th step OGA has selected n atoms $\psi_1, \psi_2, \ldots, \psi_n$ from $\mathcal{D}_{\text{bad}}$. Due to projection, OGA will never select an atom twice, so let's see what happens for $\varphi \in \mathcal{D} \setminus \Psi$:

$$\langle \varphi, f - \sum_{j=1}^{n} c_j \psi_j \rangle = \langle \varphi, f \rangle - \sum_{j=1}^{n} c_j \langle \varphi, \psi_j \rangle = R - M \sum_{j=1}^{n} c_j.$$

Since all the scalar products are still the same (they do not depend on φ), some realization of OGA will select another atom from $\mathcal{D}_{\text{bad}}$, completing the induction. In fact, OGA will never select a correct atom from $\mathcal{D}_{\text{good}}$ ever, a disastrous failure. □

5. Proofs

We will need the following simple lemmas.

Lemma 12. *Let $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ be a finite collection of atoms from $M-$coherent dictionary $\mathcal{D}$. Then for any $x \in \mathbb{R}^n$*

$$(1 - Mn)\|x\|_2^2 \le \|\Phi x\|^2 \le (1 + Mn)\|x\|_2^2 .$$

Lemma 13. *Let $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ be a finite collection of atoms from $M-$coherent dictionary $\mathcal{D}$. Then for any $f \in H$*

$$\|\Phi^* f\|_2 \ge \frac{1 - Mn}{\sqrt{1 + Mn}} \|\operatorname{proj}_\Phi f\|_H .$$

Proof. For the proofs of the above lemmas see [5], Lemmas 2.1, 2.2. □

Proof of Proof of Theorem 5. Let's say OGA is on its $i+1$-st iteration. Let's denote its choice $\varphi_{i+1} := \arg\max_{\varphi \in \mathcal{D}} |\langle \varphi, f_i \rangle|$, and $d_{i+1} := |\langle f_i, \varphi_{i+1} \rangle|$. Also, let $G \subset \mathcal{D}$ be the collection of s distinct elements that have biggest scalar products with f_k:

$$G = \{g : |G| = s, |\langle g, f_k \rangle| \ge |\langle \varphi, f_k \rangle| \text{ for all } g \in G, \varphi \in \mathcal{D} \setminus G\}.$$

By the same construction as in [5], we have, denoting $p_{k+i} = \operatorname{proj}_{\Phi_{k+i}} f_k$, (2)

$$\|f_{k+s}\|^2 \le \|f_k\|^2 - \sum_{i=1}^{s} |\langle f_{k+i}, g_i \rangle|^2 = \|f_k\|^2 - \sum_{i=1}^{s} |\langle f_k, g_i \rangle - \langle p_{k+i}, g_i \rangle|^2$$

Replacing triangle inequality by a more appropriate Parseval identity in (2.6, [5]), using Lemma 13 we can estimate

$$\sum_{i=1}^{s} |\langle f_k, g_i \rangle|^2 \ge \frac{(1 - Ms)^2}{1 + Ms} \|\operatorname{proj}_\Psi f_k\|^2 = \frac{(1 - Ms)^2}{1 + Ms} (\|f_k\|^2 - \sigma_s(f_k)^2).$$

A slightly more delicate approach to (2.4, [5]) under the assumptions $k \le s, M(k+s) \le \frac{1}{2}$ gives an improved estimate on the second component of the sum in (2):

$$(3) \quad \sum_{i=1}^{s} |\langle p_{k+i}, g_i \rangle|^2 = \sum_{i=1}^{s} |\langle \Phi_{k+i} c_{k+i}, g_i \rangle|^2 \le \sum_{i=1}^{s} \left|\langle c_{k+i}, \Phi_{k+i}^* g_i \rangle\right|^2 \le$$

$$\le \sum_{i=1}^{s} \frac{M^2(k+i)}{1 - M(k+i)} \|f_k\|^2 \le M^2 \sum_{i=1}^{s} 2(k+i) \|f_k\|^2 \le 3(Ms)^2 \|f_k\|^2 .$$

To combine the two inequalities above back into (2), similarly to (2.5, [5]) we use triangle inequality:

$$(4)\quad \left(\sum_{i=1}^{s} |\langle f_k, g_i\rangle - \langle p_{k+i}, g_i\rangle|^2\right)^{\frac{1}{2}} \geq$$
$$\geq \left(\sum_{i=1}^{s} |\langle f_k, g_i\rangle|^2\right)^{\frac{1}{2}} - \left(\sum_{i=1}^{s} |\langle p_{k+i}, g_i\rangle|^2\right)^{\frac{1}{2}} \geq$$
$$\geq (1-\alpha x)\left(\|f_k\|^2 - \sigma_s(f_k)^2)\right)^{\frac{1}{2}} - \beta x \|f_k\|,$$

where we denoted $x := Ms$, $\beta = 2$, and used that on $x \in [0, \frac{1}{2}]$

$$\frac{1-x}{\sqrt{1+x}} \geq 1 - \alpha x \text{ for } \alpha = \frac{3}{2}.$$

For the sake of the presentation, $\beta = 2 > \sqrt{3}$ will suffice. A more careful treatment of (3) as a right Riemann sum of an increasing function yields an even better $\beta = 4\sqrt{\ln(\frac{3}{2}) - \frac{1}{4}} = 1.577...$ (see [18]).

The rest is a simple calculus exercise: observe that the convex quadratic in (4) is above its tangent line at $x = 0$ (for simplicity of presentation, let $a = \|f_k\|^2, c = \sigma_s(f_k)^2$):

$$((1-\alpha x)\sqrt{a-c} - \beta x\sqrt{a})^2 \geq a - c - 2x\sqrt{a-c}(\alpha\sqrt{a-c} + \beta\sqrt{a}) \geq$$
$$\geq a - c - 2x(\alpha+\beta)a.$$

Therefore,

$$\|f_{k+s}\|^2 \leq a - (a - c - 7xa) = 7xa + c = 7Ms\|f_k\|^2 + \sigma_s(f_k)^2.$$

□

From here, several analogues to [2.4, 2.5, [5]] can be established, although the nature of the estimate allows a purely iterative argument instead of a recursive one. We will need the following trivial lemma about sequences.

Lemma 14. *Let $\{a_l\}_{l=1}^{\infty}, \{b_l\}_{l=1}^{\infty}$ be nonnegative sequences of real numbers such that $b_l < \frac{1}{2}$ for all l, and c be a nonnegative real number. Also, let*

$$(5)\quad a_{l+1} \leq a_l b_l + c \text{ for all } l \in \mathbb{N}.$$

Then for all natural L

$$a_{L+1} \leq a_1 \prod_{l=1}^{L} b_l + 2c.$$

Proof. For $L = 1$, the statement is obvious. Suppose the desired inequality holds for some $L-1$. Then by (5) and by induction hypothesis,

$$a_{L+1} \leq a_L b_L + c \leq \left(a_1 \prod_{l=1}^{L-1} b_l + 2c \right) b_L + c \leq a_1 \prod_{l=1}^{L} b_l + 2c.$$

□

Proof of Corollaries 6,7. Fix $m \geq 1$. Let $k_l := m(2^l - 1)$ be a sequence of indices and $\{a_l\}_{l=1}^{\infty}$ be a sequence of squared norms $a_l := \|f_{k_l}\|^2$. Then by Theorem 5, while $Mm2^l \leq \frac{1}{2}$, we have

$$a_{l+1} \leq 7Mm2^l a_l + \sigma_{k_l+m}(f_{k_l})^2 \leq 7Mm2^l a_l + \sigma_m(f)^2, \tag{6}$$

where we used a degrees-of-freedom argument to estimate

$$\sigma_{k_l+m}(f_{k_l}) \leq \sigma_m(f) =: \sigma.$$

By Lemma 14 until $7Mm2^l > \frac{1}{2}$ we get

$$a_L \leq a_1 \prod_{l=1}^{L-1} 7Mm2^l + 2\sigma^2.$$

From Theorem 2 we can initialize $a_1 = \|f_m\|^2 \leq (6m+1)\sigma^2$ for the final estimate

$$a_L \leq \left(2 + (6m+1) \prod_{l=1}^{L-1} 7Mm2^l \right) \sigma^2.$$

For a meaningful conclusion, we need the product to overpower m in $6m+1$. If we require $7Mm2^{L-1} \leq \frac{1}{2} m^{-\delta}$ for some fixed $\delta \geq 0$, then

$$7Mm2^l = (7Mm2^{L-1})2^{l-L+1} \leq m^{-\delta} 2^{l-L},$$

and therefore

$$\prod_{l=1}^{L-1} 7Mm2^l \leq m^{-(L-1)\delta} \prod_{l=1}^{L-1} 2^{l-L} \leq m^{-(L-1)\delta} 2^{-\frac{1}{2}(L-1)^2}.$$

Each of the factors can do the job, thus we obtain two corollaries. Both conditions will then provide us with

$$\|f_{m2^L}\| \leq 3\sigma_m(f). \tag{7}$$

Proof of Corollary 6. If $\delta = 0$, we need $7Mm2^{L-1} \leq \frac{1}{2}$ and $m2^{-\frac{1}{2}(L-1)^2} < 1$. Stipulating $L = \lceil \sqrt{2 \log m} \rceil + 1$, if

$$m2^{\sqrt{2 \log m}} \leq \frac{1}{26M},$$

then after $m(2^{\lceil\sqrt{\log m}\rceil+1}-1)$ iterations we get (7). □

Proof of Corollary 7. Similarly, if there exists $\delta > 0$ such that

$$7Mm2^{\lceil\frac{1}{\delta}\rceil} \leq \frac{1}{2}m^{-\delta}, \text{ or, rewriting as } \quad m \leq \left(\frac{1}{14M}\right)^{\frac{1}{1+\delta}} 2^{-\lceil\frac{1}{\delta}\rceil},$$

we get the job done after $m2^{\lceil\frac{1}{\delta}\rceil+1}$ iterations. □

Proof of Theorem 8. In (2), we have an expansion $f_s = f - \Phi_s c_s$ instead, and using $\|\Phi_s c_s\| \leq \|f\| + \|f_s\| \leq 2\|f\|$, the following estimate holds:

$$|\langle\Phi_s c_s, g_i\rangle| \leq M\sqrt{s}\,\|c_s\|_2 \leq \frac{M\sqrt{s}}{\sqrt{1-Ms}}\|\Phi_s c_s\| \leq \frac{2M\sqrt{s}\,\|f\|}{\sqrt{1-Ms}},$$

and then (4) holds with $\alpha = \frac{3}{2}, \beta = 3$, and therefore, for all $s \leq \frac{1}{2M}$

$$\|f_s\|^2 \leq 9Ms\,\|f\|^2 + \sigma_s(f)^2.$$

□

Proof of Corollary 9. From above, by using the same argument as in proof of Theorem 5, we obtain (if $9Mm2^{L-1} \leq \frac{1}{2}$)

$$\|f_{m2^L}\|^2 \leq 2^{-\frac{1}{2}(L-1)^2}\|f\|^2 + 2\sigma_m(f)^2.$$

If we decree a certain rate of convergence on $\sigma_m(f)$, we can match it for the PGA residual at some price. Suppose $\sigma_m(f) \leq m^{-r}\|f\|$. Selecting $L = \sqrt{10r\log m} \geq \lceil\sqrt{4r\log m - 2} + 1\rceil$, we get that

$$\|f_{m2^L}\| \leq 2m^{-r}\|f\|$$

for all m such that

$$m2^L \leq \frac{1}{18M}.$$

□

Proof of Corollary 10. If we require a stronger condition $9Mm2^{L-1} \leq \frac{1}{2}m^{-\delta}$, we get

$$\|f_{m2^L}\|^2 \leq m^{-(L-1)\delta}\|f\|^2 + 2\sigma_m(f)^2. \tag{8}$$

Suppose now we have $m \leq N(\delta) := \left(\frac{1}{9M}\right)^{\frac{1}{1+\delta}} 2^{-\lceil\frac{2r}{\delta}\rceil}$, and $L = \lceil\frac{2r}{\delta}\rceil + 1$. Let $n = \lfloor m2^{-L+1}\rfloor$. Using that $\sigma_n(f) \leq n^{-r}\|f\|$, by (8) we have

$$\|f_m\|^2 \leq n^{-(L-1)\delta}\|f\|^2 + 2\sigma_n(f)^2 \leq \left(n^{-2r} + 2n^{-2r}\right)\|f\|^2 \leq$$
$$\leq C(\delta, r)m^{-2r}\|f\|^2.$$

□

6. References

References

[1] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13(1):57–98, 1997.

[2] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5(2-3):173–187, 1996.

[3] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[4] David L. Donoho, Michael Elad, and Vladimir N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. Technical Report 6, 2004.

[5] David L. Donoho, Moshe Elad, and Vladimir N. Temlyakov. On Lebesgue-type inequalities for greedy approximation. *Journal of Approximation Theory*, 147:185–195, 2007.

[6] David L. Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Preprint, 2007.

[7] Anna C. Gilbert, S. Muthukrishnan, and Martin J. Strauss. Approximation of functions over redundant dictionaries using coherence. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 243–252, New York, 2003. ACM.

[8] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[9] Peter J. Huber. Projection pursuit. *Ann. Statist.*, 13(2):435–525, 1985. With discussion.

[10] Lee K. Jones. On a conjecture of Huber concerning the convergence of projection pursuit regression. *Ann. Statist.*, 15(2):880–882, 1987.

[11] E. D. Livshitz and V. N. Temlyakov. Two lower estimates in greedy approximation. *Constr. Approx.*, 19(4):509–523, 2003.

[12] D. Needell and J. A. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2009.

[13] Deanna Needell and Roman Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 9(3):317–334, 2009.

[14] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. 27th Annu. Asilomar Conf. Signals, Systems and Computers*, November 1993.

[15] V. N. Temlyakov. Greedy approximation. *Acta Numer.*, 17:235–409, 2008.

[16] Joel A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[17] Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[18] Pavel Zheltov. *Lebesgue-type inequalities for greedy algorithms, in preparation.* PhD thesis, USC, 2010.

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

E-mail address: `V.N.Temlyakov: temlyak@math.sc.edu`

E-mail address: `P.Zheltov: zheltovp@sc.edu`