Interdisciplinary Mathematics Institute

2012:04

Greedy Expansions in Hilbert Spaces

J.L. Nelson and V.N. Temlyakov

IMI

Preprint Series

College of Arts and Sciences

University of South Carolina

GREEDY EXPANSIONS IN HILBERT SPACES

J.L. NELSON AND V.N.TEMLYAKOV

Dedicated to Boris Kashin on the occasion of his 60th birthday.

Abstract. We study rate of convergence of expansions of elements in a Hilbert space H into series with regard to a given dictionary $\mathcal{D}$. The primary goal of this paper is to study representations of an element $f \in H$ by a series

$$f \sim \sum_{j=1}^{\infty} c_j(f) g_j(f), \quad g_j(f) \in \mathcal{D}.$$

Such a representation involves two sequences: $\{g_j(f)\}_{j=1}^{\infty}$ and $\{c_j(f)\}_{j=1}^{\infty}$. In this paper the construction of $\{g_j(f)\}_{j=1}^{\infty}$ is based on ideas used in greedy-type nonlinear approximation, hence the use of the term *greedy expansion*.

An interesting open problem questions, "What is the best possible rate of convergence of greedy expansions for $f \in A_1(\mathcal{D})$?" Previously it was believed that the rate of convergence was slower than $m^{-\frac{1}{4}}$. The qualitative result of this paper is that the best possible rate of convergence of greedy expansions for $f \in A_1(\mathcal{D})$ is faster than $m^{-\frac{1}{4}}$. In fact, we prove it is faster than $m^{-\frac{2}{7}}$.

1. Introduction

Let H be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and norm $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$. A set $\mathcal{D} \subset H$ of functions (elements) is a dictionary if each $g \in \mathcal{D}$ is normalized ($\|g\| = 1$) and $\overline{\operatorname{span}}\, \mathcal{D} = H$. To have approximations with nonnegative coefficients, it is convenient to consider the symmetrized dictionary $\mathcal{D}^{\pm} := \{\pm g, g \in \mathcal{D}\}$ as well.

If $f \in H$, we assume the existence of $g = g(f) \in \mathcal{D}$, the element from $\mathcal{D}$ that maximizes $|\langle f, g \rangle|$, and define the greedy approximant as

$$G(f) := G(f, \mathcal{D}) := \langle f, g \rangle\, g \tag{1}$$

and the residual as

$$R(f) := R(f, \mathcal{D}) := f - G(f).$$

Built of these two bricks, the Pure Greedy Algorithm is an iterative process that chips away at the residual by creating greedy approximants for successive residuals. The resulting approximants of the residuals are combined to create an approximation for the original $f \in H$. The Pure Greedy Algorithm is given by Algorithm 1.

Date: December 2011.

1991 *Mathematics Subject Classification.* Primary: 41A65; secondary: 41A25, 41A46, 46B20.

This research was supported by the National Science Foundation Grant DMS-0906260.

Algorithm 1: Pure Greedy Algorithm (PGA)

Input: f, $\mathcal{D}$
begin
$f_0 := R_0(f) := R_0(f, \mathcal{D}) := f$;
$G_0(f) := G_0(f, \mathcal{D}) := 0$;
for $m \geq 1$ **do**
Greedy choice:

$$g(R_{m-1}(f)) \in \left\{ g \in \mathcal{D} : |\langle R_{m-1}(f), g \rangle| = \max_{g \in \mathcal{D}} |\langle R_{m-1}(f), g \rangle| \right\}$$

Greedy approximation:

$$G_m(f) := G_m(f, \mathcal{D}) := G_{m-1}(f) + G(R_{m-1}(f))$$

Calculate residual:

$$f_m := R_m(f) := R_m(f, \mathcal{D}) := f - G_m(f) = R(R_{m-1}(f))$$

end

The primary goal of this paper is to study representations of an element $f \in H$ by a series

$$f \sim \sum_{j=1}^{\infty} c_j(f) g_j(f), \quad g_j(f) \in \mathcal{D}. \tag{2}$$

where the coefficients $c_j(f)$ are created by design. We require that $\vec{\mathcal{D}} = \{g_j(f)\}_{j=1}^{\infty}$ is inductively constructed based on the greedy step from the PGA: $g_j(f) := g(f_{j-1})$ where

$$f_{j-1} := f - \sum_{i=1}^{j-1} c_i(f) g_i(f).$$

This choice is why such expansions are called *greedy expansions*.

After a dictionary element has been chosen, its corresponding coefficient must be made. There is freedom of choice in coefficients $c_j(f)$ of greedy expansions. Close study (see Temlyakov (2007b) for example) has shown that the obvious choice is not always the best, in terms of ushering along convergence. In the case of PGA the choice is $c_j(f) = \langle f_{j-1}, g_j(f) \rangle$. In the case of an orthonormal dictionary, this inner product boils down to $c_j(f) = \langle f, g_j(f) \rangle$, but with a general dictionary, the closed form of $c_j(f)$ is more complicated.

For now the rate of convergence of greedy expansions is the topic of interest. (Temlyakov, 2011, Chapter 6) presents results on greedy expansions, but specifically for the Pure Greedy Algorithm, DeVore and Temlyakov (1996) proved that for a general dictionary $\mathcal{D}$ and $f \in A_1(\mathcal{D})$ the estimate

$$\|f - G_m(f, \mathcal{D})\| \leq m^{-\frac{1}{6}}$$

holds, where $A_1(\mathcal{D})$ denotes the closure of the convex hull of the symmetric dictionary $\mathcal{D}^{\pm}$. Konyagin and Temlyakov (1999) improved the DeVore-Temlyakov estimate to

$$\|f - G_m(f, \mathcal{D})\| \leq 4m^{-\frac{11}{62}}.$$

These estimates brought up the following central theoretical open problem in greedy approximation in Hilbert spaces.

Problem 1.1. *Find the order of decay of the sequence*

$$\gamma(m) := \sup_{f \in A_1(\mathcal{D}), \mathcal{D}, \{G_m\}} \|f - G_m(f, \mathcal{D})\|,$$

where the supremum is taken over all dictionaries $\mathcal{D}$, all elements $f \in A_1(\mathcal{D})$ and all possible choices of $\{G_m\}$.

Sil′nichenko (2004) proved the upper estimate

$$\gamma(m) \leq Cm^{-\frac{s}{2(2+s)}}$$

where s is a solution from $[1, 1.5]$ of the equation

$$(1+x)^{\frac{1}{2+x}} \left(\frac{2+x}{1+x} \right) - \frac{1+x}{x} = 0.$$

Numerical calculations of s by Sil′nichenko (2004) give

$$\frac{s}{2(2+s)} = 0.182 \cdots > \frac{11}{62}.$$

The technique used by Sil′nichenko (2004) is a development of the method of Konyagin and Temlyakov (1999).

There is also some progress in lower estimates. The estimate

$$\gamma(m) \geq Cm^{-0.27},$$

with a positive constant C, has been proved in Livshitz and Temlyakov (2003). Previous lower estimates appear on page 59 of Temlyakov (2003). Recently, Livshitz (2009), developing the technique from Livshitz and Temlyakov (2003), proved the lower estimate

$$\gamma(m) \geq Cm^{-0.1898}.$$

Although the Pure Greedy Algorithm gives, for every element $f \in H$, a convergent expansion in a series with respect to a dictionary $\mathcal{D}$, alterations of the PGA can have their virtues. Temlyakov (2007a) developed such alteration that also provides a convergent expansion but generalizes the PGA with a weakness sequence and a tuning parameter. The *weakness sequence* is a sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \leq t_k \leq 1$. The k^{th} term in the weakness sequence prescribes that the greedy choice should be at least t_k times as good as an optimal greedy choice. In fact, when $\tau = \{1\}$, the algorithm is called the Pure Greedy Algorithm with parameter b (PGA(b)). The *tuning parameter* $b \in (0, 1]$ then attempts to ameliorate the shortcomings of the greedy choice by scaling the greedy approximant down–analogous to the way that someone who utters an insult might play it down by saying, "Just kidding!" More precisely, WGA(b) updates the approximant by adding an orthogonal projection of the residual $f_{m-1}^{\tau,b}$ onto $\varphi_m^{\tau,b}$ multiplied by b, so the greedy expansion for $f \in H$ is a series of the form

$$f \sim \sum_{j=1}^{\infty} c_j(f)\varphi_j^{\tau,b}, \quad c_j(f) := b\left\langle f_{j-1}^{\tau,b}, \varphi_j^{\tau,b}\right\rangle.$$

With these alterations, shown in Algorithm 2, the Weak Greedy Algorithm with parameter b arises.

Algorithm 2: Weak Greedy Algorithm with parameter b (WGA(b))

Input: f, $\mathcal{D}$, $\tau := \{t_m\}_{m=1}^{\infty}$ ($t_m \in [0,1]$), $b \in (0,1]$

begin

$f_0^{\tau,b} := f$;

$G_0(f) := G_0(f,\mathcal{D}) := 0$;

for $m \geq 1$ **do**

Greedy choice: $\varphi_m^{\tau,b} \in \mathcal{D}$ is any element satisfying

$$\left|\left\langle f_{m-1}^{\tau,b}, \varphi_m^{\tau,b}\right\rangle\right| \geq t_m \sup_{g\in\mathcal{D}}\left|\left\langle f_{m-1}^{\tau,b}, g\right\rangle\right|$$

Greedy approximation:

$$G_m^{\tau,b}(f,\mathcal{D}) := b\sum_{j=1}^{m}\left\langle f_{j-1}^{\tau,b}, \varphi_j^{\tau,b}\right\rangle \varphi_j^{\tau,b}$$

Calculate residual:

$$f_m^{\tau,b} := f_{m-1}^{\tau,b} - b\left\langle f_{m-1}^{\tau,b}, \varphi_m^{\tau,b}\right\rangle \varphi_m^{\tau,b}$$

end

Temlyakov (2007a) gives the following convergence rate of WGA(b).

Theorem 1.2. *Let $\mathcal{D}$ be an arbitrary dictionary in H. Assume $\tau := \{t_k\}_{k=1}^{\infty}$ is a nonincreasing sequence and $b \in (0,1]$. Then for $f \in A_1(\mathcal{D})$ we have*

$$\left\|f - G_m^{\tau,b}(f,\mathcal{D})\right\| \leq \left(1 + b(2-b)\sum_{k=1}^{m} t_k^2\right)^{\frac{-(2-b)t_m}{2(2+(2-b)t_m)}}.$$

In the particular case $t_k = 1$, $k = 1,2,\dots$, we get the following rate of convergence

$$\left\|f - G_m^{1,b}(f,\mathcal{D})\right\| \leq Cm^{-r(b)}, \quad r(b) := \frac{2-b}{2(4-b)}.$$

The fact that $r(1) = \frac{1}{6}$ and $r(b) \to \frac{1}{4}$ as $b \to 0$ means that at each step of the Pure Greedy Algorithm we can choose a fixed fraction of the optimal coefficient for that step instead of the optimal coefficient itself. Surprisingly, this leads to better upper estimates than those known for the Pure Greedy Algorithm, so it happens that pure greed is good, as Tropp (2004) says, but it is not as good as conservative greed. Try explaining that to Wall Street.

In the general setting of Banach spaces, Temlyakov (2007b) pushed the flexibility of the greedy coefficients to the extreme by making them independent of the input f. Assuming that $\mathcal{D}$ is symmetric, it is easy to formulate the analogous Hilbert space algorithm DGA($\tau,\mathcal{C}$), given by Algorithm 3. If $\tau = \{t\}$ with $t \in (0,1]$, the notation says t instead of τ. When $t = 1$, it is ignored, but the resulting algorithm, DGA($\mathcal{C}$), still provides a greedy expansion.

Algorithm 3: Dual Greedy Algorithm with weakness τ and coefficients $\mathcal{C}$ (DGA($\tau,\mathcal{C}$))

Input: f, $\mathcal{D}$, $\tau := \{t_m\}_{m=1}^{\infty}$ ($t_m \in [0,1]$), $\mathcal{C} = \{c_k\}_{k=1}^{\infty}$
begin
  $f_0 := f$;
  $G_0 := 0$;
  for $m \geq 1$ **do**
    Greedy choice: $\varphi_m \in \mathcal{D}$ is any element satisfying
$$\langle f_{m-1}, \varphi_m\rangle \geq t_m \sup_{g\in\mathcal{D}} \langle f_{m-1}, g\rangle$$
    Greedy approximation:
$$G_m := G_{m-1} + c_m\varphi_m.$$
    Calculate residual:
$$f_m := f_{m-1} - c_m\varphi_m$$
end

The Banach space result on the rate of convergence of DGA($\tau,\mathcal{C}$) from Temlyakov (2007a) also leads to a Hilbert space version.

Theorem 1.3. *Let $\mathcal{C} := \left\{k^{-\frac{3}{4}}\right\}_{k=1}^{\infty}$. Then the DGA($\mathcal{C}$) converges for $f \in A_1(\mathcal{D})$ at the following rate: For any $r \in \left(0, \frac{1}{4}\right)$*

$$\|f_m\| \leq C(r)m^{-r}.$$

Thus, both PGA(b) and DGA($\mathcal{C}$) provide a rate of convergence for $f \in A_1(\mathcal{D})$ close to, but slower than $m^{-\frac{1}{4}}$. So, if even this blind chicken of an algorithm can find some corn, what is the best possible rate of convergence of greedy expansions (2) for $f \in A_1(\mathcal{D})$? This is an interesting open problem. The qualitative result of this paper is that the best possible rate of convergence is faster than $m^{-\frac{1}{4}}$. Section 3 argues the following theorem where

$$h(x,w,b) := \left(1 - (2-b)x + \left(1 - \frac{b}{2}\right)wx^2\right)(1+x)^{w(2-b)}.$$

Theorem 1.4. *Let $b \in \left(0, \frac{1}{2}\right]$ be given and let $w > 1$ be such that*

$$\min_{0\leq x\leq 1} h(x,w,b) < 1.$$

Then for the residual of PGA(b) we have

$$\|f_m\| \leq C(b,w)m^{-\rho(b,w)}, \text{ where } \rho(b,w) := \frac{\left(1-\frac{b}{2}\right)w}{2\left(\left(1-\frac{b}{2}\right)w+1\right)}.$$

For the function $h(x,w,b)$ we have for $b \leq \frac{1}{2}$

$$h\left(\frac{1}{2},1,b\right) = \left(\frac{1}{4}+\frac{3b}{8}\right)\left(\frac{3}{2}\right)^{2-b} \leq \frac{63}{64} < 1.$$

This implies the following lemma.

Lemma 1.5. *There exists a number $w > 1$ such that for all $b \leq \frac{1}{2}$ we have*

$$\min_{0\leq x\leq 1} h(x,w,b) < 1.$$

Observing that for $w > 1$

$$\lim_{b\to 0} \rho(b,w) = \frac{w}{2(w+1)} > \frac{1}{4}$$

we obtain from Theorem 1.4 and Lemma 1.5 the fact that PGA(b), with appropriate b, converges faster than m^{-r} with $r > \frac{1}{4}$. At the end of Section 3 we present an elementary numerics showing that one can take $\rho(b,w) = \frac{2}{7}$ for appropriate b and w.

Techniques from Konyagin and Temlyakov (1999) and Temlyakov (2007a) are used in this proof. For completeness we give the proof of Theorem 1.2 presently.

2. Rate of convergence of WGA(b)

An alternative characterization of $A_1(\mathcal{D})$ for a general dictionary $\mathcal{D}$ begins by defining the class of functions

$$\mathcal{A}_1^0(\mathcal{D},M) := \left\{ f \in H : f = \sum_{k\in\Lambda} c_k w_k,\ w_k \in \mathcal{D},\ \#\Lambda < \infty \text{ and } \sum_{k\in\Lambda} |c_k| \leq M \right\}$$

then $\mathcal{A}_1(\mathcal{D},M)$ denotes the closure in H of $\mathcal{A}_1^o(\mathcal{D},M)$. The union over all $M > 0$ of the classes $\mathcal{A}_1(\mathcal{D},M)$ is then denoted $\mathcal{A}_1(\mathcal{D})$. For $f \in \mathcal{A}_1(\mathcal{D})$, we define the norm, $|f|_{\mathcal{A}_1(\mathcal{D})}$, as the smallest M such that $f \in \mathcal{A}_1(\mathcal{D},M)$. For $M = 1$ we denote $A_1(\mathcal{D}) := \mathcal{A}_1(\mathcal{D},1)$.

For ease of notation, let

$$\begin{aligned} a_m &:= \left\| f_m^{\tau,b} \right\|^2 \\ d_{m-1} &:= \left| \left\langle f_{m-1}^{\tau,b}, \varphi_m^{\tau,b} \right\rangle \right|, \quad m = 1,2,\ldots, \end{aligned}$$

where $f_{m-1}^{\tau,b}$ is the m^{th} residual of WGA(b) and $\varphi_m^{\tau,b}$ is the m^{th} greedy choice. Consider the sequence $\{B_n\}$ defined by

$$\begin{aligned} B_0 &:= 1, \\ B_m &:= B_{m-1} + bd_{m-1}, \quad m = 1,2,\ldots. \end{aligned} \tag{3}$$

Then obviously $f_m^{\tau,b} \in \mathcal{A}_1(\mathcal{D}, B_m)$.

Lemma 3.5 from DeVore and Temlyakov (1996) states that if $f \in \mathcal{A}_1(\mathcal{D}, M)$ and $g(f) := \arg\sup_{g\in\mathcal{D}} |\langle f, g\rangle|$ then

$$\frac{\langle f, g(f)\rangle}{\|f\|} \geq \frac{\|f\|}{M}.$$

Applying this lemma to $f_{m-1}^{\tau,b}$ results in

$$\sup_{g\in\mathcal{D}} \left|\left\langle f_{m-1}^{\tau,b}, g\right\rangle\right| \geq \frac{\left\|f_{m-1}^{\tau,b}\right\|^2}{B_{m-1}}. \tag{4}$$

Combining this with the equality

$$\left\|f_m^{\tau,b}\right\|^2 = \left\|f_{m-1}^{\tau,b}\right\|^2 - b(2-b)\left\langle f_{m-1}^{\tau,b}, \varphi_m^{\tau,b}\right\rangle^2$$

we obtain the relations

$$a_m = a_{m-1} - b(2-b)d_{m-1}^2, \tag{5}$$

$$d_{m-1} \geq \frac{t_m a_{m-1}}{B_{m-1}}. \tag{6}$$

Substituting (6) into (5), begets

$$a_m \leq a_{m-1}\left(1 - \frac{b(2-b)t_m^2 a_{m-1}}{B_{m-1}^2}\right).$$

Knowing that $B_{m-1} \leq B_m$, multiplication by inverse squares narrows the gap but does not close or reverse it, leading to

$$a_m B_m^{-2} \leq a_{m-1}B_{m-1}^{-2}\left(1 - \frac{b(2-b)t_m^2 a_{m-1}}{B_{m-1}^2}\right).$$

At this point, Lemma 3.1 from Temlyakov (2000) interjects. It says that when $\{a_m\}_{m=0}^{\infty}$ satisfies the inequalities

$$a_0 \leq A, \quad a_m \leq a_{m-1}\left(1 - \frac{t_m^2 a_{m-1}}{A}\right), \quad m = 1, 2, \ldots,$$

then we have for each m

$$a_m \leq A\left(1 + \sum_{k=0}^{m} t_k^2\right)^{-1}.$$

Applying this lemma with $A = 1$ gives

$$a_m B_m^{-2} \leq \left(1 + b(2-b)\sum_{k=1}^{m} t_k^2\right)^{-1}. \tag{7}$$

For the time being, set aside this relation. Plugging (6) into (5) again–but not completely replacing d_{m-1} this time–gives the slightly different relation,

$$a_m \leq a_{m-1} - \frac{b(2-b)d_{m-1}t_m a_{m-1}}{B_{m-1}} = a_{m-1}\left(1 - \frac{b(2-b)t_m d_{m-1}}{B_{m-1}}\right). \tag{8}$$

When the square roots of both sides of this relation are taken and followed by the application of the inequality $(1-x)^{\frac{1}{2}} \leq 1 - \frac{1}{2}x$ for $x \leq 1$, the result is

$$a_m^{\frac{1}{2}} \leq a_{m-1}^{\frac{1}{2}}\left(1 - \frac{b\left(1-\frac{b}{2}\right)t_m d_{m-1}}{B_{m-1}}\right). \tag{9}$$

Now this relation can step aside until it is needed. We can return to the definition of $\{B_m\}$ in (3) and rewrite it in the form

$$B_m = B_{m-1}\left(1 + \frac{bd_{m-1}}{B_{m-1}}\right),$$

so that applying the inequality

$$(1+x)^\alpha \leq 1 + \alpha x, \quad 0 \leq \alpha \leq 1, \quad x \geq 0,$$

reveals that

$$B_m^{\left(1-\frac{b}{2}\right)t_m} \leq B_{m-1}^{\left(1-\frac{b}{2}\right)t_m}\left(1 + \frac{b\left(1-\frac{b}{2}\right)t_m d_{m-1}}{B_{m-1}}\right). \tag{10}$$

Multiplying (9) and (10) allows us to trivialize the complicated multiplier in the latter so that we obtain

$$a_m^{\frac{1}{2}} B_m^{\left(1-\frac{b}{2}\right)t_m} \leq a_{m-1}^{\frac{1}{2}} B_{m-1}^{\left(1-\frac{b}{2}\right)t_m}.$$

In order to compare successive terms, note that since $B_{m-1} \geq 1$ and $t_m \leq t_{m-1}$,

$$B_{m-1}^{\left(1-\frac{b}{2}\right)t_m} \leq B_{m-1}^{\left(1-\frac{b}{2}\right)t_{m-1}}.$$

Substituting this fact into the previous one, it becomes clear that the sequence $\left\{a_k^{\frac{1}{2}} B_k^{\left(1-\frac{b}{2}\right)t_k}\right\}$ is non-increasing

$$a_m^{\frac{1}{2}} B_m^{\left(1-\frac{b}{2}\right)t_m} \leq a_{m-1}^{\frac{1}{2}} B_{m-1}^{\left(1-\frac{b}{2}\right)t_{m-1}} \leq \cdots \leq a_0^{\frac{1}{2}} \leq 1. \tag{11}$$

Raising both sides of (7) to the power $\left(1-\frac{b}{2}\right)t_m$, squaring the ends of the inequality chain (11), and then combining those two results, we obtain

$$a_m^{1+\left(1-\frac{b}{2}\right)t_m} \leq \left(1 + b(2-b)\sum_{k=1}^{m} t_k^2\right)^{-\left(1-\frac{b}{2}\right)t_m}.$$

Raising both sides of this final relation to the power $\left(1+\left(1-\frac{b}{2}\right)t_m\right)^{-1}$ completes the proof. $\square$

Remark 2.1. *If instead of (3) we define the sequence $\{B_n\}$ by*

$$B_0 \geq 1, \quad B_m := B_{m-1} + bd_{m-1}, \quad m = 1, 2, \ldots \tag{12}$$

then we still get (7) for $f \in A_1(\mathcal{D})$.

3. Improved rate of convergence

Since the Pure Greedy Algorithm with parameter b (PGA(b)) is essentially WGA(b) with the weakness sequence $t_k = 1$ for all k, we will use the structure of the proof given by Konyagin and Temlyakov (1999) and assumptions of Theorem 1.4 to improve the rate of convergence given in Theorem 1.2. Rather than (11), we get the following inequality

$$a_m^{\frac{1}{2}} B_m^{\left(1-\frac{b}{2}\right)w} \leq C(b, w) \tag{13}$$

with some constant $C(b, w)$. We define as before

$$d(f) := |\langle f, g(f)\rangle|, \quad d_m := d(f_m), \quad m = 0, 1, 2, \ldots$$

where $f_m := f_m^{\tau,b}$ for convenience. We note that (5) implies that $a_0 \geq a_1 \geq \ldots$ and, therefore, for $f \in A_1(\mathcal{D})$, we have, for all m, $d_m \leq 1$. For ease of use, we employ the rewritten definition of the sequence $\{B_m\}$

$$B_m = B_{m-1}\left(1 + \frac{bd_{m-1}}{B_{m-1}}\right),$$

but a different B_0 will be specified later on.

It is a fact that for any $f \in H$ and $h \in \mathcal{A}_1(\mathcal{D})$ we have

$$|\langle f, h\rangle| \leq |\langle f, g(f)\rangle| |h|_{\mathcal{A}_1(\mathcal{D})}. \tag{14}$$

To see why this is so, observe that if $h \in \mathcal{A}_1^0(\mathcal{D}, M)$, it has a representation $h = \sum_k c_k g_k$ in terms of the dictionary $\mathcal{D} = \{g_k\}$ with $\sum |c_k| \leq M$. Therefore we have

$$\begin{aligned} |\langle f, h\rangle| &= \left|\sum_k c_k \langle f, g_k\rangle\right| \\ &\leq \sum_k |c_k| |\langle f, g_k\rangle| \\ &\leq |\langle f, g(f)\rangle| \sum_k |c_k| \leq |\langle f, g(f)\rangle| M. \end{aligned}$$

Then (14) follows by a limiting argument.

With this fact in hand, consider $\langle f_\ell, f\rangle$ for some $\ell = 0, 1, 2, \ldots$, the inner product of the ℓ^{th} residual with the original input. On one hand, (14) implies that we have for $f \in A_1(\mathcal{D})$

$$\langle f_\ell, f\rangle \leq d(f_\ell)|f|_{\mathcal{A}_1(\mathcal{D})} \leq d_\ell B_0, \tag{15}$$

but on the other hand we can figure directly that

$$\langle f_\ell, f\rangle = \left\langle f - b\sum_{j=0}^{\ell-1} d_j g(f_j), f\right\rangle \geq a_0 - bd_0\sum_{j=0}^{\ell-1} d_j. \tag{16}$$

Mashing together the contents of our hands, by multiplying the respective sides together, we get a lower bound

$$d_\ell \geq \frac{a_0 - bd_0\sum_{j=0}^{\ell-1} d_j}{B_0}, \quad \ell = 1, 2, \ldots. \tag{17}$$

With a mind to keeping future usage of this relation clean, denote

$$D_\ell := \sum_{j=0}^{\ell} d_j.$$

To avoid a wreck of our train of thought during the proof of Theorem 1.4, we should address the following lemma before we proceed. This lemma gives us a bound on $a_{m+1} = \left\|f_{m+1}^b\right\|^2$ in terms of the parameter b, the sequence seed B_0, $d_0 = \left|\left\langle f_0^b, \varphi_1^b\right\rangle\right|$ and the newly defined D_m.

Lemma 3.1. *We have*

$$a_{m+1} \leq a_0\left(1 - (2-b)\frac{bD_m}{B_0}\right) + \frac{d_0(2-b)}{2B_0}(bD_m)^2.$$

Proof. Let

$$x_0^b := \frac{a_0}{bd_0}, \qquad y_\ell^b := x_0^b - D_\ell.$$

The superscript b is intended to distinguish this notation from that used by Konyagin and Temlyakov (1999), whose definitions of x_0 and y_ℓ lack the parameter b due to having been defined with PGA in mind, not PGA(b) as in our case.

The difference between y_ℓ^b and $y_{\ell-1}^b$ is d_ℓ and can be estimated by the inequality-mash made in (17), rewritten in terms of x_0^b and D_ℓ

$$y_{\ell-1}^b - y_\ell^b = d_\ell \geq \frac{bd_0}{B_0}(x_0^b - D_{\ell-1}) = \frac{bd_0}{B_0}y_{\ell-1}^b. \tag{18}$$

Rearranged, this relation entails

$$y_\ell^b \leq \left(1 - \frac{bd_0}{B_0}\right)y_{\ell-1}^b.$$

Denote the somewhat unruly multiplier by $r = 1 - \frac{bd_0}{B_0}$, and define a related quantity by $s = \frac{1-r}{1+r} = \frac{bd_0}{2B_0 - bd_0}$. The reason for s will be clear shortly.

Using the definition of y_0^b, we can write it in two different ways to add 0 to d_0^2 to find the identity

$$d_0^2 = -s(y_0^b)^2 + s(x_0^b - d_0)^2 + d_0^2. \tag{19}$$

At this point, let us reorient ourselves towards finding an estimate for a_{m+1}, an intermediate step that we left out of (5) gives the identity

$$a_{m+1} = a_0 - b(2-b)\sum_{j=0}^{m} d_j^2 \tag{20}$$

so it would be useful to estimate the sum of squares.

To this end, let us prove by induction the following inequality

$$d_0^2 + \cdots + d_n^2 \geq d_0^2 + s(x_0^b - d_0)^2 - s(y_n^b)^2. \tag{21}$$

Suppose that (21) holds for some $n = m-1$. Thus for $n = m$ we obtain

$$\begin{aligned} d_0^2 + \cdots + d_{m-1}^2 + d_m^2 &\geq d_0^2 + s(x_0^b - d_0)^2 - s(y_{m-1}^b)^2 + d_m^2 \\ &= d_0^2 + s(x_0^b - d_0)^2 - s(y_{m-1}^b)^2 + (y_{m-1}^b - y_m^b)^2 \end{aligned}$$

where the second line follows from (18). At this time, the following lemma by Konyagin and Temlyakov (1999) is helpful, using r and s as we have already defined them.

Lemma 3.2. *Let $0 < r < 1$ be given and $s := \frac{1-r}{1+r}$. Then for x_1, x_2 such that $x_1 - x_2 \geq 0$ and $x_2 \leq r x_1$ we have*

$$-s x_2^2 \leq -s x_1^2 + (x_1 - x_2)^2.$$

Applying the lemma with

$$\begin{aligned} r &= (1 - \frac{bd_0}{B_0}) & s &= \frac{bd_0}{2B_0 - bd_0} \\ x_2 &= y_m^b = x_0^b - D_m & x_1 &= y_{m-1}^b = x_0^b - D_{m-1}, \end{aligned}$$

we continue to estimate the sum of squares

$$\begin{aligned} d_0^2 + \cdots + d_{m-1}^2 + d_m^2 &\geq d_0^2 + s(x_0^b - d_0)^2 - s(y_m^b)^2 \\ &= d_0^2 + s(D_m - d_0)(2x_0^b - D_m - d_0) \\ &= d_0^2 + \frac{bd_0}{2B_0 - bd_0}(D_m - d_0)(2x_0^b - D_m - d_0). \end{aligned}$$

Now we require another lemma by Konyagin and Temlyakov (1999).

Lemma 3.3. *Let A, B, C be positive numbers such that $C \geq A + B$. Then for any $0 \leq x \leq \min(A, B)$ we have*

$$x + \frac{(A-x)(B-x)}{C-x} \geq \frac{AB}{C}.$$

This lemma works with $A = D_m$, $B = 2x_0^b - D_m$, $C = \frac{2B_0}{b}$, $x = d_0$ to bring us to

$$d_0^2 + \cdots + d_{m-1}^2 + d_m^2 \geq \frac{bd_0}{2B_0} D_m(2x_0^b - D_m) = \frac{a_0}{B_0} D_m - \frac{bd_0}{2B_0} D_m^2.$$

Inserting this estimate into the identity (20) expands the latter to

$$a_{m+1} = a_0 - b(2-b)\sum_{j=0}^{m} d_j^2 \le a_0\left(1-(2-b)\frac{bD_m}{B_0}\right) + \frac{(2-b)d_0}{2B_0}(bD_m)^2.$$

This completes the proof of Lemma 3.1. $\square$

With this lemma in hand, the proof of Theorem 1.4 follows. Remember that we are trying to show (13), that for all m

$$a_m^{\frac{1}{2}} B_m^{\left(1-\frac{b}{2}\right)w} \le C(b,w).$$

However, the proof is recursive. We analyze the first iteration.

Shortly, the form of $h(x,w,b)$ will find an explanation. For now, it suffices to say that its role is to give a bound on a product of powers of a_{m+1} and B_{m+1} in terms of a proportion of a product of powers of a_0 and B_0. If we want the residuals of PGA(b) to go to zero, we would like $h(x,w,b)$ to be less than 1. This assumption can be rephrased in terms of optimization as

$$\min_{0\le x\le 1} h(x,w,b) < 1$$

so that it implies that there exists an interval $[u,v]\subset[0,1]$, $u<v$, such that

$$h(x,w,b) \le 1, \qquad x\in[u,v]. \tag{22}$$

For reasons of fit, choose B_0 such that $B_0 > \frac{b}{(v-u)}$.

Recall that for $f\in A_1(\mathcal{D})$ we have defined $a_m = \left\|f_m^b\right\|^2$ and $d_m = \left|\left\langle f_m^b, \varphi_{m+1}^b\right\rangle\right|$. The recursive nature of the sequence B_m force any estimates to start at the beginning. During the first step of PGA(b) one of two cases may apply:

$$d_0 \ge \frac{wa_0}{B_0} \tag{A}$$

$$d_0 < \frac{wa_0}{B_0}. \tag{B}$$

The $w>1$ as mentioned in Theorem 1.4, stands in the place of t_m from the proof of Theorem 1.2 (see (6)).

The first case has short-term implications. The second case makes us wait for the other shoe to drop.

Case (A). In this case the proof closely resembles the early part of the proof of Theorem 1.2. We have for $m=1$ that

$$a_1 = a_0 - b(2-b)d_0^2 \le a_0\left(1-\frac{b(2-b)wd_0}{B_0}\right),$$
$$B_1 = B_0 + bd_0.$$

First, we treat the case $\left(1-\frac{b}{2}\right)w\le 1$. As before, the inequality $(1-x)^{\frac{1}{2}} \le 1-\frac{1}{2}x$ gives

$$a_1^{\frac{1}{2}} \le a_0^{\frac{1}{2}}\left(1-\frac{b\left(1-\frac{b}{2}\right)wd_0}{B_0}\right).$$

Using the inequality $(1+x)^\gamma \le 1+\gamma x$, $0\le\gamma\le 1$, $x\ge 0$, we obtain

$$B_1^{\left(1-\frac{b}{2}\right)w} = B_0^{\left(1-\frac{b}{2}\right)w}\left(1+\frac{bd_0}{B_0}\right)^{\left(1-\frac{b}{2}\right)w} \le B_0^{\left(1-\frac{b}{2}\right)w}\left(1+\frac{b\left(1-\frac{b}{2}\right)wd_0}{B_0}\right).$$

Second, we treat the case $\left(1-\frac{b}{2}\right)w\ge 1$. Using the inequality $1-\alpha x\le (1-x)^\alpha$ for $\alpha\ge 1$, $x\in[0,1]$ we get

$$a_1^{\frac{1}{2}} \le a_0^{\frac{1}{2}}\left(1-\frac{b\,(2-b)\,wd_0}{B_0}\right)^{\frac{1}{2}} \le a_0^{\frac{1}{2}}\left(1-\frac{bd_0}{B_0}\right)^{\left(1-\frac{b}{2}\right)w}.$$

Next,

$$B_1^{\left(1-\frac{b}{2}\right)w} = B_0^{\left(1-\frac{b}{2}\right)w}\left(1+\frac{bd_0}{B_0}\right)^{\left(1-\frac{b}{2}\right)w}.$$

Thus,

$$a_1^{\frac{1}{2}}B_1^{\left(1-\frac{b}{2}\right)w} \le a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}, \tag{23}$$

so we have that (13) is proved for $m=1$.

Case (B). When we look at this case, Lemma 3.1 tells us that for $m\ge 0$

$$\begin{aligned} a_{m+1}^{\frac{1}{2}}B_{m+1}^{\left(1-\frac{b}{2}\right)w} &\le \left(a_0\left(1-\frac{(2-b)bD_m}{B_0}\right)+\frac{(2-b)d_0}{2B_0}(bD_m)^2\right)^{\frac{1}{2}}(B_0+bD_m)^{\left(1-\frac{b}{2}\right)w} \\ &\le a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}h\left(b\frac{D_m}{B_0},w,b\right)^{\frac{1}{2}}. \end{aligned} \tag{24}$$

What we want now is for $b\frac{D_m}{B_0}\in[u,v]$, the interval from (22). This happens either sooner or later.

Subcase B1 (Sooner). Since $D_m=D_{m-1}+d_m$, $d_m\le 1$, if bD_m increases beyond $(1+u)B_0$ then there exists m_1 such that $\frac{bD_{m_1}}{B_0}\in[u,v]$ since by the choice of B_0 we have $\frac{b}{B_0}<v-u$. With this m_1 we obtain

$$a_{m_1+1}^{\frac{1}{2}}B_{m_1+1}^{\left(1-\frac{b}{2}\right)w} \le a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w},\quad B_{m_1}\le(1+v)B_0. \tag{25}$$

Therefore, for $m\le m_1$ we have

$$\begin{aligned} a_m^{\frac{1}{2}}B_m^{\left(1-\frac{b}{2}\right)w} &\le a_0^{\frac{1}{2}}B_{m_1}^{\left(1-\frac{b}{2}\right)w} \\ &\le a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}(1+v)^{\left(1-\frac{b}{2}\right)w} \le C_1a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}. \end{aligned} \tag{26}$$

Subcase B2 (Later). If $B_m\le(1+u)B_0$ for all m then by (24) we obtain

$$a_{m+1}^{\frac{1}{2}}B_{m+1}^{\left(1-\frac{b}{2}\right)w} \le a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}\max_{0\le x\le 1}h(x,w,b)^{\frac{1}{2}} \le C_2a_0^{\frac{1}{2}}B_0^{\left(1-\frac{b}{2}\right)w}. \tag{27}$$

Here, inequality (27) holds for all m and, therefore, implies (13). This ends consideration of the first iteration of the algorithm.

Combining all cases. In both Case (A) and Subcase B1, we begin the first iteration with f and end up with f_{n_1} ($f_{n_1} := f_1$ in case (A) and $f_{n_1} := f_{m_1+1}$ in Subcase B1) with the property

$$a_{n_1}^{\frac{1}{2}} B_{n_1}^{\left(1-\frac{b}{2}\right)w} \le a_0^{\frac{1}{2}} B_0^{\left(1-\frac{b}{2}\right)w}. \tag{28}$$

For the interim $n < n_1$ (in Subcase B1) we obtain

$$a_n^{\frac{1}{2}} B_n^{\left(1-\frac{b}{2}\right)w} \le C_1 a_0^{\frac{1}{2}} B_0^{\left(1-\frac{b}{2}\right)w}. \tag{29}$$

We now apply another iteration of PGA(b) to f_{n_1} instead of f with $B_{n_1}(f)$ playing the role of $B_0(f_{n_1})$. The condition $B_0(f_{n_1}) > \frac{b}{(v-u)}$ is clearly satisfied.

Therefore, at the second iteration, if Subcase B2 occurs we get for all n

$$a_n^{\frac{1}{2}} B_n^{\left(1-\frac{b}{2}\right)w} \le C_2 a_{n_1}^{\frac{1}{2}} B_{n_1}^{\left(1-\frac{b}{2}\right)w} \le C_2 a_0^{\frac{1}{2}} B_0^{\left(1-\frac{b}{2}\right)w}.$$

Else, if case (A) occurs then with $n_2 = n_1 + 1$ we have

$$a_{n_2}^{\frac{1}{2}} B_{n_2}^{\left(1-\frac{b}{2}\right)w} \le a_{n_1}^{\frac{1}{2}} B_{n_1}^{\left(1-\frac{b}{2}\right)w} \le a_0^{\frac{1}{2}} B_0^{\left(1-\frac{b}{2}\right)w}$$

which also occurs for some n_2 if we have Subcase B1. Moreover, for intermediate $n \in (n_1, n_2)$ we have

$$a_n^{\frac{1}{2}} B_n^{\left(1-\frac{b}{2}\right)w} \le C_1 a_{n_1}^{\frac{1}{2}} B_{n_1}^{\left(1-\frac{b}{2}\right)w} \le C_1 a_0^{\frac{1}{2}} B_0^{\left(1-\frac{b}{2}\right)w}.$$

We continue these iterations to complete the proof of (13).

Using Remark 2.1 and combining (7) and (13) we obtain

$$a_m \le C(b,w) m^{-\frac{\left(1-\frac{b}{2}\right)w}{\left(1-\frac{b}{2}\right)w+1}}.$$

This completes the proof of Theorem 1.4. □

3.1. Numerical calculations.

Denote $\theta := \left(1-\frac{b}{2}\right)w$, then

$$h(x,w,b) = (1-(2-b)x+\theta x^2)(1+x)^{2\theta}.$$

Specify $x = \frac{1}{2}$ and $b = 0$ and get

$$h\left(\frac{1}{2}, w, 0\right) = \frac{\theta}{4}\left(\frac{3}{2}\right)^{2\theta}.$$

Specify $\theta = \frac{4}{3}$. Then

$$h\left(\frac{1}{2}, w, b\right) = \frac{1}{3}\left(\frac{3}{2}\right)^{\frac{8}{3}}$$

and

$$h\left(\frac{1}{2}, w, b\right)^3 = \left(\frac{1}{3}\right)^3\left(\frac{3}{2}\right)^8 = \frac{3^5}{2^8} = \frac{243}{256} < 1.$$

Therefore, for sufficiently small b we can take w such that $\left(1-\frac{b}{2}\right)w=\frac{4}{3}$ and

$$h\left(\frac{1}{2}, w, b\right) < 1.$$

For these b and w we obtain $\rho(b,w)=\frac{2}{7}$ in Theorem 1.4.

References

R. A. DeVore and V. N. Temlyakov (1996). Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5(2-3):173–187. ISSN 1019-7168.

S. V. Konyagin and V. N. Temlyakov (1999). Rate of convergence of pure greedy algorithm. *East J. Approx.*, 5(4):493–499. ISSN 1310-6236.

E. D. Livshitz and V. N. Temlyakov (2003). Two lower estimates in greedy approximation. *Constr. Approx.*, 19(4):509–523. ISSN 0176-4276.

E. D. Livshitz (2009). On lower bounds for the rate of convergence of greedy algorithms. *Izv. Ross. Akad. Nauk Ser. Mat.*, 73(6):125–144. ISSN 0373-2436.

A. V. Sil′nichenko (2004). On the rate of convergence of greedy algorithms. *Mat. Zametki*, 76(4):628–632. ISSN 0025-567X.

V. N. Temlyakov (2000). Weak greedy algorithms. *Adv. Comput. Math.*, 12(2-3):213–227. ISSN 1019-7168.

V. N. Temlyakov (2003). Nonlinear methods of approximation. *Found. Comput. Math.*, 3 (1):33–107. ISSN 1615-3375.

V. N. Temlyakov (2007a). Greedy expansions in Banach spaces. *Adv. Comput. Math.*, 26 (4):431–449. ISSN 1019-7168.

V. N. Temlyakov (2007b). Greedy algorithms with prescribed coefficients. *J. Fourier Anal. Appl.*, 13(1):71–86. ISSN 1069-5869.

V. N. Temlyakov (2001). *Greedy Approximation*. Cambridge University Press, Cambridge, UK, 2011. ISBN 978-1-107-00337-8.

J. A. Tropp (2004). Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242. ISSN 0018-9448.

Mathematics Dept. of the University of South Carolina, 1523 Greene Street, Columbia, SC 29208, USA.

Mathematics Dept. of the University of South Carolina, 1523 Greene Street, Columbia, SC 29208, USA; Steklov Institute of Mathematics, Gubkin str. 8, Moscow 119991, Russia.