Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Rescaled pure greedy algorithm for Hilbert and Banach spaces ☆

Guergana Petrova

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form 27 September 2015
Accepted 20 October 2015
Available online xxxx
Communicated by Leslie F. Greengard

MSC:
41A25
41A46

Keywords:
Greedy algorithms
Rates of convergence

ABSTRACT

We show that a very simple modification of the Pure Greedy Algorithm for approximating functions by sparse sums from a dictionary in a Hilbert or more generally a Banach space has optimal convergence rates.

1. Introduction

Greedy algorithms have been used quite extensively as a tool for generating approximations from redundant families of functions, such as frames or more general dictionaries $\mathcal{D}$, see [8,10,3,2,15]. Given a Banach space X, a dictionary is any set $\mathcal{D}$ of norm one elements from X whose span is dense in X. The most natural greedy algorithm in a Hilbert space is the Pure Greedy Algorithm (**PGA**), which is also known as Matching Pursuit, see [3] for the description of this and other algorithms. The fact that the **PGA** lacks optimal convergence properties has led to a variety of modified greedy algorithms such as the Relaxed Greedy Algorithm (**RGA**), the Orthogonal Greedy Algorithm (**OGA**), and their weak versions. There are also analogues of these, developed for approximating functions in Banach spaces, see [15].

The central issues in the study of these algorithms is their ease of implementation and their approximation power, measured in terms of convergence rates. If f_m is the output of a greedy algorithm after m iterations, then f_m is a linear combination of at most m dictionary elements. Such linear combinations are said to be

☆ This research was supported by the Office of Naval Research Contract ONR N00014-11-1-0712, and by the NSF Grants DMS 1222715 and DMS 1521067.

E-mail address: gpetrova@math.tamu.edu.

http://dx.doi.org/10.1016/j.acha.2015.10.008
1063-5203/

sparse of order m. The quality of the approximation is measured by the decay of the error $\|f - f_m\|$ as $m \to \infty$, where $\|\cdot\|$ is the norm in the Hilbert or Banach space, respectively. Of course, the decay rate of this error is governed by properties of the target function f. The typical properties imposed on f are that it is sparse, or more generally, that it is in some way compressible. Here, compressible means that it can be written as a (generally speaking, infinite) linear combination of dictionary elements with some restrictions on the coefficients. The most frequently applied assumption on f is that it is in the unit ball of the class $\mathcal{A}_1(\mathcal{D})$, that is the set of all functions which are a convex combination of dictionary elements (provided we consider symmetric dictionaries). It is known that the elements in this class can be approximated by m sparse vectors to accuracy $\mathcal{O}(m^{-1/2})$, see Theorem 2.1, and so this rate of approximation serves as a benchmark for the performance of greedy algorithms.

It has been shown in [3] in the case of Hilbert space that whenever $f \in \mathcal{A}_1(\mathcal{D})$, the output f_m of the **PGA** satisfies

$$\|f - f_m\| = \mathcal{O}(m^{-1/6}), \quad m \to \infty.$$

Later results gave slight improvements of the above estimate. For example, in [9], the rate $\mathcal{O}(m^{-1/6})$ was improved to $\mathcal{O}(m^{-11/62})$. Based on the method from the latter paper, Sil'nichenko [14] then showed a rate of $\mathcal{O}(m^{-\frac{s}{2(s+2)}})$, where s solves a certain equation, and that $\frac{s}{2(s+2)} > 11/62$. Similar estimates for the weak versions of the **PGA** can be found in [15]. Estimates for the error from below have also been provided, see [12,11].

The fact that the **PGA** does not attain the optimal rate for approximating the elements in $\mathcal{A}_1(\mathcal{D})$ has led to various modifications of this algorithm. Two of these modifications, the Relaxed and the Orthogonal Greedy Algorithm were shown to achieve the optimal rate $\mathcal{O}(m^{-1/2})$, see [3].

The purpose of the present paper is to show that a very simple modification of the **PGA**, namely just rescaling f_m at each iteration, already leads to the improved convergence rate $\mathcal{O}(m^{-1/2})$ for functions in $\mathcal{A}_1(\mathcal{D})$ and the rate $\mathcal{O}(m^{-\theta/2})$ for functions from the interpolation space $[H, \mathcal{A}_1(\mathcal{D})]_{\theta,\infty}$, $0 < \theta < 1$. The rescaling we suggest is simply the orthogonal projection of f onto f_m. We call this modified algorithm a Rescaled Pure Greedy Algorithm (**RPGA**) and prove optimal convergence rates for its weak version in Hilbert and Banach spaces. In a subsequent paper, see [5], we show that this strategy can also be applied successfully for developing an algorithm for convex optimization.

The paper is organized as follows. In Section 2, we spell out our notation and recall some simple known facts related to greedy algorithms. In Section 3, we present the **RPGA** for a Hilbert space and prove the above convergence rates. The remaining parts of this paper consider a modification of this algorithm for Banach spaces and weak versions of this algorithm.

2. Notation and preliminaries

We denote by H a Hilbert space and by X a Banach space with $\|\cdot\|$ being the norm in these spaces, respectively. A set of functions $\mathcal{D} \subset H(\text{or } X)$ is called a dictionary if $\|\varphi\| = 1$ for every $\varphi \in \mathcal{D}$ and the closure of $\text{span}(\mathcal{D})$ is $H(\text{or } X)$. An example of a dictionary is any Shauder basis for $H(\text{or } X)$. However, the main idea behind dictionaries is to cover redundant families such as frames. A common example of dictionaries is the union of several Shauder bases.

We denote by $\Sigma_m(\mathcal{D})$ the set, consisting of all m-sparse elements with respect to the dictionary $\mathcal{D}$, namely

$$\Sigma_m := \Sigma_m(\mathcal{D}) = \{g : \ g = \sum_{\varphi \in \Lambda} c_\varphi \varphi, \ \Lambda \in \mathcal{D}, \ |\Lambda| \le m\}.$$

Here, we use the notation $|\Lambda|$ to denote the cardinality of the index set Λ. For a general element f from $H(\text{or } X)$, we define the error of approximation

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

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

of f by elements from Σ_m. The rate of decay of $\sigma_m(f)$ as $m\to\infty$ says how well f can be approximated by sparse elements.

For a general dictionary $\mathcal{D}\subset H$(or X), we define the class of functions

$$\mathcal{A}_1^o(\mathcal{D},M) := \{f=\sum_{k\in\Lambda} c_k(f)\varphi_k : \varphi_k\in\mathcal{D},\ |\Lambda|<\infty,\ \sum_{k\in\Lambda}|c_k(f)|\le M\},$$

and by $\mathcal{A}_1(\mathcal{D},M)$ its closure in H(or X). Then, $\mathcal{A}_1(\mathcal{D})$ is defined to be the union of the classes $\mathcal{A}_1(\mathcal{D},M)$ over all $M>0$. For $f\in\mathcal{A}_1(\mathcal{D})$, we define the norm of f as

$$\|f\|_{\mathcal{A}_1(\mathcal{D})} := \inf\{M :\ f\in\mathcal{A}_1(\mathcal{D},M)\}.$$

A fundamental result for approximating $\mathcal{A}_1(\mathcal{D})$ is the following, see [3].

Theorem 2.1. *For a general dictionary $\mathcal{D}\subset H$ and $f\in\mathcal{A}_1(\mathcal{D})\subset H$, we have*

$$\sigma_m(f,\mathcal{D})\le c\|f\|_{\mathcal{A}_1(\mathcal{D})}m^{-1/2},\quad m=1,2,\dots,$$

where c is a constant.

When analyzing the convergence of greedy algorithms, we will use the following lemma, proved in [13]. Several similar versions of this lemma have been proved and utilized in the analysis of greedy algorithms, see [15]. For example, Lemma 3.1 in [16] is the same as Lemma 2.2 in the case $\ell=1$, $r_m=t_m^2$, $J=0$, and $r=B$.

Lemma 2.2. *Let $\ell>0$, $r>0$, $B>0$, and $\{a_m\}$ and $\{r_m\}$ be finite or infinite sequences of non-negative numbers satisfying the inequalities*

$$a_J\le B,\quad a_m\le a_{m-1}(1-\frac{r_m}{r}a_{m-1}^{\ell}),\quad m=J+1,J+2,\dots.$$

Then, we have

$$a_m\le \max\{1,\ell^{-1/\ell}\}r^{1/\ell}(rB^{-\ell}+\Sigma_{k=J+1}^m r_k)^{-1/\ell},\quad m=J+1,J+2,\dots. \tag{2.1}$$

3. The Hilbert space case

In order to show the simplicity of our results, we begin with the standard case of the **RPGA** in a Hilbert space. Later, we treat the case of Banach spaces and weak algorithms, but the reader familiar with this topic will see that the results in these more general settings follow by standard modifications of the results from this section. We denote the inner product in the Hilbert space H by $\langle\cdot,\cdot\rangle$, and so the norm of $f\in H$ is $\|f\|=\langle f,f\rangle^{1/2}$.

The **RPGA($\mathcal{D}$)** with a dictionary $\mathcal{D}$ is defined by the following simple steps.

RPGA($\mathcal{D}$):

- **Step** 0: Define $f_0:=0$.
- **Step** m:

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

– If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$, for $k \geq m$.
– If $f \neq f_{m-1}$, choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

With

$$\lambda_m := \langle f - f_{m-1}, \varphi_m \rangle, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m, \quad s_m := \frac{\langle f, \hat{f}_m \rangle}{\|\hat{f}_m\|^2},$$

define the next approximant to be

$$f_m = s_m \hat{f}_m,$$

and proceed to Step $m+1$.

Note that if the output at each Step m were $\hat{f}_m$ and not $f_m = s_m \hat{f}_m$, this would be the **PGA**. However, the new algorithm uses not $\hat{f}_m$, but the best approximation to f from the one dimensional space $\text{span}\{\hat{f}_m\}$, that is $s_m \hat{f}_m$. Adding this step, which is just appropriate scaling of the output of the **PGA**, allows us to prove optimal convergence rate of $m^{-1/2}$ for the proposed algorithm for functions from $\mathcal{A}_1(\mathcal{D})$.

Next, we show that the **RPGA** and the Relaxed Greedy Algorithm (**RGA**) provide different sequences of approximants $\{f_m\}$ and $\{f_m^r\}$, respectively, and thus **RPGA** is different from the known so far greedy algorithms. For both algorithms

$$f_0 = f_0^r = 0, \quad f_1 = f_1^r = \langle f, \varphi_1 \rangle \varphi_1,$$

where $\varphi_1 \in \mathcal{D}$ is such that $|\langle f, \varphi_1 \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f, \varphi \rangle|$. For both **RPGA** and **RGA**, the next element $\varphi_2 \in \mathcal{D}$ is chosen as $|\langle f - f_1, \varphi_2 \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_1, \varphi \rangle|$. One can easily compute that the next approximant, generated by the **RPGA** is

$$f_2 = s_2 f_1 + s_2 \langle f - f_1, \varphi_2 \rangle \varphi_2, \quad s_2 = \frac{\langle f, \varphi_1 \rangle^2 + \langle f, \varphi_2 \rangle^2 - \langle f, \varphi_1 \rangle \langle f, \varphi_2 \rangle \langle \varphi_1, \varphi_2 \rangle}{\langle f, \varphi_1 \rangle^2 + \langle f, \varphi_2 \rangle^2 - \langle f, \varphi_1 \rangle^2 \langle \varphi_1, \varphi_2 \rangle^2},$$

while the classical **RGA** would give

$$f_2^r = \frac{1}{2} f_1 + \frac{1}{2} \varphi_2.$$

There are some modifications of the **RGA**, see [1], where the approximant at Step m is determined not as

$$f_m^r = (1 - \frac{1}{m}) f_{m-1}^r + \frac{1}{m} \varphi_m, \quad \text{where } |\langle f - f_{m-1}^r, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}^r, \varphi \rangle|,$$

but as

$$f_m^r = (1 - a_m) f_{m-1}^r + a_m \varphi_m, \tag{3.1}$$

where a_m and φ_m are the solutions of the minimization problem

$$\min_{a \in [0,1], \varphi \in \mathcal{D}} \|f - \big((1-a) f_{m-1}^r + a\varphi\big)\|.$$

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

In other modifications, such as the Weak Relaxed Greedy Algorithm (**WRGA**) and versions of it, see [15], the direction φ_m at step m is chosen, such that

$$\langle f - f_{m-1}^r, \varphi_m - f_{m-1}^r \rangle \geq t_m \|f - f_{m-1}^r\|^2.$$

Then the next approximation f_m^r is computed according to

$$f_m^r = (1 - \beta_m) f_{m-1}^r + \beta_m \varphi_m,$$

where the weights β_m are given by a specific formula using the weakness sequence $\{t_k\}$.

Note that the sequence, generated by the **RPGA** is a linear combination of f_{m-1} and φ_m, that is

$$f_m = s_m f_{m-1} + \lambda_m s_m \varphi_m,$$

but it is different from the convex combinations (3.1), from other variations of the **RGA**, for example, the **WRGA**, and from the best approximation to f from $\operatorname{span}\{f_{m-1}^r, \varphi_m\}$. One can compute that the best approximation to f from $\operatorname{span}\{f_1^r, \varphi_2\}$ is

$$f_2^r = \frac{\langle f, \varphi_1 \rangle^2 - \langle f, \varphi_1 \rangle \langle f, \varphi_2 \rangle \langle \varphi_1, \varphi_2 \rangle}{\langle f, \varphi_1 \rangle^2 (1 - \langle \varphi_1, \varphi_2 \rangle^2)} f_1 + \frac{\langle f, \varphi_2 \rangle - \langle f, \varphi_1 \rangle \langle \varphi_1, \varphi_2 \rangle}{1 - \langle \varphi_1, \varphi_2 \rangle^2} \varphi_2,$$

and again $f_2 \neq f_2^r$.

While the **RPGA** and its versions can be viewed as a new modification of the **RGA**, we feel that it is closer to the Weak Greedy Algorithm (**WGA**), described in [15] and introduced in [7]. The reason is that even thought the suggested algorithm does not provide a greedy expansion for $f \in H$ as the **WGA** does, in both methods (the **RPGA** and **WGA**) the direction at the next step is determined as the supremum of the inner product of dictionary elements with the residual of the previous approximation. Moreover, both algorithms use at each step orthogonal projections onto one dimensional spaces. On the other hand, in most versions of the **RGA**, see for example [1] and [15], the next direction at every step is chosen so that it does not optimize the inner product, but rather the squared error in some form of linear combination.

We continue with the following theorem.

Theorem 3.1. *If $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, then the output $(f_m)_{m \geq 0}$ of the* **RPGA***$(\mathcal{D})$ satisfies the error estimate*

$$e_m := \|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} (m+1)^{-1/2}, \quad m = 0, 1, 2 \ldots. \tag{3.2}$$

Proof. Since f_m is the orthogonal projection of f onto the one dimensional space spanned by $\hat{f}_m$, we have

$$\langle f - f_m, f_m \rangle = 0, \quad m \geq 0. \tag{3.3}$$

Note that the definition of $\hat{f}_m$ and the choice of λ_m give

$$\begin{aligned} \|f - \hat{f}_m\|^2 &= \langle f - f_{m-1} - \lambda_m \varphi_m, f - f_{m-1} - \lambda_m \varphi_m \rangle \\ &= \|f - f_{m-1}\|^2 - 2\lambda_m \langle f - f_{m-1}, \varphi_m \rangle + \lambda_m^2 \|\varphi_m\|^2 \\ &= \|f - f_{m-1}\|^2 - \langle f - f_{m-1}, \varphi_m \rangle^2, \end{aligned} \tag{3.4}$$

where we have used that $\|\varphi_m\| = 1$. Now, assume $f \neq f_{m-1}$. Since f_m is the orthogonal projection of f onto $\operatorname{span}\{\hat{f}_m\}$, we have

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

$$e_m^2 = \|f - f_m\|^2 = \|f - s_m \hat{f}_m\|^2 \le \|f - \hat{f}_m\|^2.$$

We combine the latter inequality and (3.4) to derive that for any $f \in H$,

$$e_m^2 \le e_{m-1}^2 - \langle f - f_{m-1}, \varphi_m \rangle^2, \quad m = 1, 2, \dots. \tag{3.5}$$

We proceed with a lower estimate for $|\langle f - f_{m-1}, \varphi_m \rangle|$. Observe that

$$e_{m-1}^2 = \|f - f_{m-1}\|^2 = \langle f - f_{m-1}, f - f_{m-1} \rangle = \langle f - f_{m-1}, f \rangle, \tag{3.6}$$

where we have used (3.3).

It is enough to prove (3.2) for functions f that are finite sums $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j| \le M$, for $M > 0$, since these functions are dense in $\mathcal{A}_1(\mathcal{D}, M)$. Let us fix $\varepsilon > 0$ and choose a representation for $f = \sum_{\varphi \in \mathcal{D}} c_\varphi \varphi$, such that

$$\sum_{\varphi \in \mathcal{D}} |c_\varphi| < M + \varepsilon.$$

It follows from (3.6) that

$$\begin{aligned} e_{m-1}^2 &= \sum_{\varphi \in \mathcal{D}} c_\varphi \langle f - f_{m-1}, \varphi \rangle \\ &\le |\langle f - f_{m-1}, \varphi_m \rangle| \sum_{\varphi \in \mathcal{D}} |c_\varphi| \\ &< |\langle f - f_{m-1}, \varphi_m \rangle| (M + \varepsilon), \end{aligned}$$

where we have used the choice of φ_m. We let $\varepsilon \to 0$ and obtain the inequality

$$M^{-1} e_{m-1}^2 \le |\langle f - f_{m-1}, \varphi_m \rangle|. \tag{3.7}$$

In particular, when $m = 1$ we have

$$M^{-1} \|f\|^2 = M^{-1} e_0^2 \le |\langle f, \varphi_1 \rangle| \le \|f\|,$$

and therefore

$$e_0^2 = \|f\|^2 \le M^2.$$

We combine (3.5) and (3.7) to obtain

$$e_m^2 \le e_{m-1}^2 - M^{-2} e_{m-1}^4 = e_{m-1}^2 (1 - M^{-2} e_{m-1}^2),$$

and apply Lemma 2.2 with $a_m = e_m^2$, $B = M^2$, $r_m := 1$, $r = M^2$, $J = 0$ and $\ell = 1$. Then, (2.1) gives

$$e_m^2 \le M^2 (m + 1)^{-1}, \quad m \ge 1,$$

and the theorem follows. □

Further, we investigate the dependence of the algorithm on errors in the computation of the inner product $\langle f - f_{m-1}, \varphi_m \rangle$ along the lines of [6], where similar analysis have been done for another greedy algorithm. One can prove the following theorem.

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

Theorem 3.2. *If at Step m of the* **RPGA**$(\mathcal{D})$ *the coefficient λ_m is computed according to*

$$\lambda_m := (1+\epsilon_m)\langle f - f_{m-1}, \varphi_m\rangle, \quad |\epsilon_m| < 1,$$

then for any $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, the output $(f_m)_{m\geq 0}$ of the **RPGA**$(\mathcal{D})$ *satisfies the error estimate*

$$e_m := \|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} \left(1 + \sum_{j=1}^{m}(1-\epsilon_j^2)\right)^{-1/2}, \quad m = 1, 2 \ldots . \tag{3.8}$$

Proof. The proof follows the arguments from Theorem 3.1. The presence of $(1+\epsilon_m)$ is accounted for in the estimate

$$\|f - f_m\|^2 \leq \|f - f_{m-1}\|^2 \left(1 - \frac{(1-\epsilon_m^2)}{M^2}\|f - f_{m-1}\|^2\right),$$

where $f = \sum_j c_j\varphi_j$ with $\sum_j |c_j| \leq M$. Application of Lemma 2.2 with appropriately chosen parameters gives (3.8), and the proof is completed. □

Next, we study the approximation properties of the **RPGA** for functions $f \in H$ that are less sparse than being in $\mathcal{A}_1(\mathcal{D})$. Following similar arguments as those in [2], we prove that the convergence rate of the **RPGA** is the same as the rate of the **OGA** and **RGA** for functions from the interpolation spaces $\mathcal{B}_p$, defined by

$$\mathcal{B}_p := [H, \mathcal{A}_1(\mathcal{D})]_{\theta,\infty} = \{f \in H : \ K(f,t) \leq Ct^\theta, \ t > 0\}, \quad 0 < \theta < 1,$$

where $\frac{1}{p} = \frac{1+\theta}{2}$ and

$$K(f,t) := K(f,t; H, \mathcal{A}_1(\mathcal{D})) = \inf_{h\in\mathcal{A}_1(\mathcal{D})}\{\|f-h\|_H + t\|h\|_{\mathcal{A}_1(\mathcal{D})}\}$$

is the K-functional for the pair $(H, \mathcal{A}_1(\mathcal{D}))$. The norm on $\mathcal{B}_p$ is defined as the smallest C for which $K(f,t) \leq Ct^\theta$. The following theorem holds.

Theorem 3.3. *If $f \in \mathcal{B}_p$, with $p = \frac{2}{1+\theta}$, then the output $(f_m)_{m\geq 0}$ of the* **RPGA**$(\mathcal{D})$ *satisfies the error inequality*

$$\|f - f_m\| \leq 2\|f\|_{\mathcal{B}_p}(m+1)^{-\theta/2}, \quad m = 1, 2, \ldots .$$

Proof. First, we show that for any $f \in H$ and any $h \in \mathcal{A}_1(\mathcal{D})$, the error $e_m = \|f - f_m\|$ of the **RPGA**$(\mathcal{D})$ satisfies the inequality

$$e_m^2 \leq \|f-h\|^2 + \frac{4}{m+1}\|h\|^2_{\mathcal{A}_1(\mathcal{D})}. \tag{3.9}$$

It is enough to prove (3.9) for functions h that are finite sums $h = \sum_j c_j\varphi_j$ with $\sum_j |c_j| \leq M$, since these functions are dense in $\mathcal{A}_1(\mathcal{D}, M)$. Let us fix $\varepsilon > 0$ and choose a representation for $h = \sum_{\varphi\in\mathcal{D}} c_\varphi\varphi$, such that

$$\sum_{\varphi\in\mathcal{D}} |c_\varphi| < M + \varepsilon.$$

We denote by a_m the sequence of numbers

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

$$a_m := e_m^2 - \|f-h\|^2, \quad m = 1, 2, \dots,$$

and consider the following cases.

Case 1: $a_0 := \|f\|^2 - \|f-h\|^2 \le 0$. It follows from the monotonicity of the error sequence $\{e_m\}$ that the sequence $\{a_m\}$ is monotone as well. Therefore $a_m \le 0$ for all $m \ge 1$, and thus estimate (3.9) holds.

Case 2: $0 < a_0 < 4M^2$. Assume that $0 < a_{m-1} < 4M^2$, $m \ge 1$. We notice that because of (3.3) and the choice of φ_m, we have

$$\begin{aligned} e_{m-1}^2 &= \langle f - f_{m-1}, f - f_{m-1}\rangle = \langle f - f_{m-1}, f\rangle = \langle f - f_{m-1}, f-h\rangle + \langle f - f_{m-1}, h\rangle \\ &\le e_{m-1}\|f-h\| + \sum_{\varphi\in\mathcal{D}} c_\varphi \langle f - f_{m-1}, \varphi\rangle \le e_{m-1}\|f-h\| + |\langle f - f_{m-1}, \varphi_m\rangle| \sum_{\varphi\in\mathcal{D}} |c_\varphi| \\ &\le \frac{1}{2}\left(e_{m-1}^2 + \|f-h\|^2\right) + |\langle f - f_{m-1}, \varphi_m\rangle|(M+\varepsilon), \end{aligned}$$

and therefore

$$|\langle f - f_{m-1}, \varphi_m\rangle| \ge \frac{e_{m-1}^2 - \|f-h\|^2}{2M} = \frac{a_{m-1}}{2M}.$$

It follows from (3.5) and the above lower estimate of $|\langle f - f_{m-1}, \varphi_m\rangle|$, using the fact that $a_{m-1} > 0$, that

$$e_m^2 \le e_{m-1}^2 - \frac{a_{m-1}^2}{4M^2}.$$

Subtracting $\|f-h\|^2$ from both sides gives

$$a_m \le a_{m-1}\left(1 - \frac{a_{m-1}}{4M^2}\right). \tag{3.10}$$

Since the function $\psi(t) := t(1 - \frac{t}{4M^2})$ on $(0, 4M^2)$ has maximum M^2, it follows that

$$a_m \le M^2 < 4M^2.$$

Therefore, either all elements of the sequence $\{a_m\}$ belong to the interval $(0, 4M^2)$ and then satisfy relation (3.10), or for some $m^* \ge 1$ we have that $a_{m^*} \le 0$. Then for $m \ge m^*$ the arguments are as in Case 1. We now use Lemma 2.2 for the positive elements in the sequence $\{a_m\}$ with $\ell = 1$, $r_m = 1$, $B = 4M^2$, $J = 0$ and $r = 4M^2$ to derive that

$$a_m \le 4M^2(m+1)^{-1},$$

which gives (3.9).

Case 3: $a_0 \ge 4M^2$. In this case, as we did in Case 2, we show that

$$a_1 \le a_0\left(1 - \frac{a_0}{4M^2}\right).$$

Therefore $a_1 < 0$, that is $e_1^2 < \|f-h\|^2$, which gives (3.9) because of monotonicity.

Now that we had proved (3.9), we have that

$$\|f - f_m\| \le K(f, 2(m+1)^{-1/2}) \le 2K(f, (m+1)^{-1/2}),$$

and therefore for $f \in \mathcal{B}_p$,

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

$$\|f - f_m\| \le K(f, 2(m+1)^{-1/2}) \le 2\|f\|_{\mathcal{B}_p}(m+1)^{-\theta/2},$$

which completes the proof of the theorem. □

We continue our discussion with the observation that $\lim_{t\to 0} K(f,t;H,\mathcal{A}_1(\mathcal{D})) = 0$, see [2], and thus the following corollary holds.

Corollary 3.4. *If $f \in H$ and $(f_m)_{m\ge 0}$ is the output of the* **RPGA***$(\mathcal{D})$ with a dictionary $\mathcal{D}$, then the error*

$$e_m = \|f - f_m\| \to 0, \quad as \quad m \to 0.$$

In the sections that follow, we introduce variants of the **RPGA** and prove convergence results similar to Theorem 3.1 and Theorem 3.3.

4. The weak rescaled pure greedy algorithm for Hilbert spaces

In this section, we describe the Weak Rescaled Pure Greedy Algorithm (**WRPGA**). It is determined by a weakness sequence $\{t_k\}_{k=1}^{\infty}$, where all $t_k \in (0,1]$, and a dictionary $\mathcal{D}$. We denote it by **WRPGA**$(\{t_k\},\mathcal{D})$.

WRPGA$(\{t_k\},\mathcal{D})$:

- **Step** 0: Define $f_0 = 0$.
- **Step** m:
  - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \ge m$.
  - If $f \ne f_{m-1}$, choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m\rangle| \ge t_m \sup_{\varphi\in\mathcal{D}} |\langle f - f_{m-1}, \varphi\rangle|.$$

With

$$\lambda_m = \langle f - f_{m-1}, \varphi_m\rangle, \quad \hat{f}_m := f_{m-1} + \lambda_m\varphi_m, \quad s_m = \frac{\langle f, \hat{f}_m\rangle}{\|\hat{f}_m\|^2},$$

define the next approximant to be

$$f_m = s_m\hat{f}_m,$$

and proceed to Step $m+1$.

In the case when all elements t_k of the weakness sequence are $t_k = 1$, this algorithm is the **RPGA**$(\mathcal{D})$. The following theorem holds.

Theorem 4.1. *If $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, then the output $(f_m)_{m\ge 0}$ of the* **WRPGA***$(\{t_k\},\mathcal{D})$ satisfies the inequality*

$$e_m := \|f - f_m\| \le \|f\|_{\mathcal{A}_1(\mathcal{D})}\left(\sum_{k=1}^{m} t_k^2\right)^{-1/2}, \quad m \ge 1.$$

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

Proof. The proof is similar to the one of Theorem 3.1, where we show that for any $f \in H$, the error $e_m^2 = \|f - f_m\|^2$ satisfies the recursive inequality,

$$e_m^2 \le e_{m-1}^2 - \langle f - f_{m-1}, \varphi_m\rangle^2, \quad m = 1, 2, \dots. \tag{4.1}$$

The lower estimate for $|\langle f - f_{m-1}, \varphi_m\rangle|$ when $f \in \mathcal{A}_1(\mathcal{D}, M)$, is derived similarly as

$$M^{-1} t_m e_{m-1}^2 \le |\langle f - f_{m-1}, \varphi_m\rangle|, \tag{4.2}$$

where we have used the definition of φ_m. In particular, for $m = 1$, we have

$$M^{-1} t_1 \|f\|^2 = M^{-1} t_1 e_0^2 \le |\langle f, \varphi_1\rangle| \le \|f\|,$$

and thus $e_1^2 \le e_0^2 = \|f\|^2 \le M^2 t_1^{-2}$. It follows from (4.1) and (4.2) that

$$e_m^2 \le e_{m-1}^2 - M^{-2} t_m^2 e_{m-1}^4 = e_{m-1}^2 (1 - M^{-2} t_m^2 e_{m-1}^2), \quad m \ge 1.$$

We apply Lemma 2.2 with $a_m = e_m^2$, $B = M^2 t_1^{-2}$, $r_m := t_m^2$, $r = M^2$, $J = 1$ and $\ell = 1$ to obtain

$$e_m^2 \le M^2 \left(\sum_{k=1}^m t_k^2 \right)^{-1}, \quad m \ge 1,$$

and the proof is completed. □

Remark 4.2. Analogously to the **RPGA**($\mathcal{D}$), one may discuss stability issues of the above weak algorithm. We assume that at Step m, the coefficient λ_m is computed according to

$$\lambda_m := (1 + \epsilon_m)\langle f - f_{m-1}, \varphi_m\rangle, \quad |\epsilon_m| < 1.$$

Then, one can show that for $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, the output $(f_m)_{m\ge 0}$ of the **WRPGA**($\{t_k\}$, $\mathcal{D}$) with a weakness sequence $\{t_k\}$, satisfies the error estimate for $m = 1, 2 \dots$,

$$e_m := \|f - f_m\| \le \|f\|_{\mathcal{A}_1(\mathcal{D})} \left(t_1^2 + \sum_{j=2}^m (1 - \epsilon_j^2) t_j^2 \right)^{-1/2} \le \|f\|_{\mathcal{A}_1(\mathcal{D})} \left(\sum_{j=1}^m (1 - \epsilon_j^2) t_j^2 \right)^{-1/2}.$$

5. The Banach space case

In this section, we will state the **RPGA**($\mathcal{D}$) algorithm for Banach spaces X with norm $\|\cdot\|$ and dictionary $\mathcal{D}$, and prove convergence results for certain Banach spaces. Let us first start with the introduction of the modulus of smoothness ρ of a Banach space X, which is defined as

$$\rho(u) := \sup_{f,g \in X, \|f\| = \|g\| = 1} \left\{ \frac{1}{2} (\|f + ug\| + \|f - ug\|) - 1 \right\}, \quad u > 0.$$

In this paper, we shall consider only Banach spaces X whose modulus of smoothness satisfies the inequality

$$\rho(u) \le \gamma u^q, \quad 1 < q \le 2, \quad \gamma\text{-constant}.$$

This is a natural assumption, since the modulus of smoothness of the spaces $X = L_p$, $1 < p < \infty$, for example, is known to satisfy such inequality. Recall that, see [4], for $X = L_p$,

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

$$\rho(u) \le \begin{cases} \frac{1}{p}u^p, & \text{if } 1 \le p \le 2, \\ \frac{p-1}{2}u^2, & \text{if } 2 \le p < \infty. \end{cases}$$

For every element $f \in X$, $f \neq 0$, we consider its norming functional $F_f \in X^*$ with the properties $\|F_f\| = 1$, $F_f(f) = \|f\|$. Note that if $X = H$ is a Hilbert space, the norming functional for any $f \in H$, $f \neq 0$, is

$$F_f(\cdot) = \frac{< f, \cdot >}{\|f\|}.$$

There is a relationship between the norming functional F_g for any $g \in X$, $g \neq 0$, and the modulus of smoothness of X, given by the following lemma.

Lemma 5.1. *Let X be a Banach space with modulus of smoothness ρ, where $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let $g \in X$, $g \neq 0$ with norming functional F_g. Then, for every $h \in X$, we have*

$$\|g + uh\| \le \|g\| + uF_g(h) + 2\gamma u^q \|g\|^{1-q} \|h\|^q, \quad u > 0.$$

Proof. The proof follows from Lemma 6.1 in [15] and the property of the modulus of smoothness. □

We now present the **RPGA**($\mathcal{D}$) for the Banach space X with dictionary $\mathcal{D}$.

RPGA($\mathcal{D}$):

- **Step** 0: Define $f_0 = 0$.
- **Step** m:
  - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \ge m$.
  - If $f \neq f_{m-1}$, choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|F_{f-f_{m-1}}(\varphi_m)| = \sup_{\varphi \in \mathcal{D}} |F_{f-f_{m-1}}(\varphi)|.$$

With

$$\lambda_m = \text{sign}\{F_{f-f_{m-1}}(\varphi_m)\} \|f - f_{m-1}\| (2\gamma q)^{\frac{1}{1-q}} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{1}{q-1}}, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m,$$

choose s_m such that

$$\|f - s_m \hat{f}_m\| = \min_{s \in \mathbb{R}} \|f - s\hat{f}_m\|,$$

define the next approximant to be

$$f_m = s_m \hat{f}_m,$$

and proceed to Step $m + 1$.

The following lemma holds.

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

Lemma 5.2. *Let X be a Banach space with modulus of smoothness ρ, such that $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let f_{m-1} be the output of the* **RPGA**$(\mathcal{D})$ *at Step $m-1$. Then, if $f \ne f_{m-1}$, we have*

$$F_{f-f_{m-1}}(f_{m-1}) = 0.$$

Proof. Let us denote by $L := \operatorname{span}\{\hat{f}_{m-1}\} \subset X$. Clearly, $f_{m-1} \in L$, and moreover, f_{m-1} is the best approximation to f from L. We apply Lemma 6.9 from [15] to the linear space L and the vector f_{m-1}, and derive the lemma. □

The next theorem provides the convergence rate for the new algorithm in Banach spaces.

Theorem 5.3. *Let X be a Banach space with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. If $f \in \mathcal{A}_1(\mathcal{D}) \subset X$, then the output $(f_m)_{m\ge 0}$ of the* **RPGA**$(\mathcal{D})$ *satisfies the estimate*

$$e_m := \|f - f_m\| \le c\|f\|_{\mathcal{A}_1(\mathcal{D})} m^{1/q-1}, \quad m \ge 1,$$

where $c = c(\gamma, q)$.

Proof. Clearly, we have $e_0 = \|f - f_0\| = \|f\|$. At Step m, $m = 1, 2, \dots$ of the algorithm, either $f = f_{m-1}$, in which case $f_k = f_{m-1}$, $k \ge m$, and therefore $e_m = 0$, or we have

$$e_m = \|f - f_m\| = \|f - s_m \hat{f}_m\| \le \|f - \hat{f}_m\| = \|(f - f_{m-1}) - \lambda_m \varphi_m\|.$$

We apply Lemma 5.1 to the latter inequality with $g = f - f_{m-1} \ne 0$, $u = |\lambda_m| > 0$, $h = -\operatorname{sign}\{\lambda_m\}\varphi_m$, and derive

$$\begin{aligned} e_m &\le \|f - f_{m-1}\| - \lambda_m F_{f-f_{m-1}}(\varphi_m) + 2\gamma|\lambda_m|^q \|f - f_{m-1}\|^{1-q} \|\varphi_m\|^q \\ &= e_{m-1} - \lambda_m F_{f-f_{m-1}}(\varphi_m) + 2\gamma|\lambda_m|^q e_{m-1}^{1-q} \\ &= e_{m-1} - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} e_{m-1} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{q}{q-1}}, \end{aligned} \tag{5.1}$$

where we have used that $\|\varphi_m\| = 1$ and the choice of λ_m. Now, we need an estimate from below for $|F_{f-f_{m-1}}(\varphi_m)|$. Using Lemma 5.2, we obtain that

$$e_{m-1} = \|f - f_{m-1}\| = F_{f-f_{m-1}}(f - f_{m-1}) = F_{f-f_{m-1}}(f). \tag{5.2}$$

As in the Hilbert space case, it is enough to consider functions f that are finite sums $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j| \le M$, since these functions are dense in $\mathcal{A}_1(\mathcal{D}, M)$. Let us fix $\varepsilon > 0$ and choose a representation for $f = \sum_{\varphi\in\mathcal{D}} c_\varphi \varphi$, such that

$$\sum_{\varphi\in\mathcal{D}} |c_\varphi| < M + \varepsilon.$$

It follows that

$$\begin{aligned} F_{f-f_{m-1}}(f) &= \sum_{\varphi\in\mathcal{D}} c_\varphi F_{f-f_{m-1}}(\varphi) \le \sum_{\varphi\in\mathcal{D}} |c_\varphi| |F_{f-f_{m-1}}(\varphi)| \\ &\le |F_{f-f_{m-1}}(\varphi_m)| \sum_{\varphi\in\mathcal{D}} |c_\varphi| < |F_{f-f_{m-1}}(\varphi_m)|(M + \varepsilon). \end{aligned}$$

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

We take $\epsilon \to 0$ and derive

$$F_{f-f_{m-1}}(f) \le |F_{f-f_{m-1}}(\varphi_m)|M. \tag{5.3}$$

In particular, when $m = 1$, we get

$$e_0 = \|f\| = F_f(f) \le |F_f(\varphi_1)|M \le \|F_f\|\|\varphi_1\|M = M.$$

Inequality (5.3) and relation (5.2) provide the lower estimate

$$M^{-1}e_{m-1} \le |F_{f-f_{m-1}}(\varphi_m)|,$$

which together with (5.1) result in

$$e_m \le e_{m-1}\left(1 - \frac{q-1}{q}\,(2\gamma q)^{\frac{1}{1-q}}\,M^{-\frac{q}{q-1}}e_{m-1}^{\frac{q}{q-1}}\right).$$

We now use Lemma 2.2 with $a_m = e_m$, $B = M$, $r_m := \frac{q-1}{q}\,(2\gamma q)^{\frac{1}{1-q}}$, $r = M^{\frac{q}{q-1}}$, $J = 0$ and $\ell = \frac{q}{q-1}$ to obtain

$$e_m \le M\left(1 + \frac{q-1}{q}\,(2\gamma q)^{\frac{1}{1-q}}\,m\right)^{1/q-1}, \quad m \ge 1,$$

and the theorem follows. □

6. The weak rescaled pure greedy algorithm for Banach spaces

In this section, we describe the Weak Rescaled Pure Greedy Algorithm for Banach spaces. It is determined by a weakness sequence $\{t_k\}_{k=1}^{\infty}$, where all $t_k \in (0, 1]$, and a dictionary $\mathcal{D}$. As in the Hilbert case, we denote it by **WRPGA**$(\{t_k\}, \mathcal{D})$.

WRPGA$(\{t_k\}, \mathcal{D})$:

- **Step** 0: Define $f_0 = 0$.
- **Step** m:
  - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \ge m$.
  - If $f \ne f_{m-1}$, choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|F_{f-f_{m-1}}(\varphi_m)| \ge t_m \sup_{\varphi\in\mathcal{D}} |F_{f-f_{m-1}}(\varphi)|.$$

With

$$\lambda_m = \operatorname{sign}\{F_{f-f_{m-1}}(\varphi_m)\}\|f - f_{m-1}\|(2\gamma q)^{\frac{1}{1-q}}|F_{f-f_{m-1}}(\varphi_m)|^{\frac{1}{q-1}}, \quad \hat{f}_m := f_{m-1} + \lambda_m\varphi_m,$$

choose s_m such that

$$\|f - s_m\hat{f}_m\| = \min_{s\in\mathbb{R}} \|f - s\hat{f}_m\|,$$

define the next approximant to be

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

$$f_m = s_m \hat{f}_m,$$

and proceed to Step $m+1$.

Next, we present the convergence rates for the **WRPGA**$(\{t_k\}, \mathcal{D})$ in Banach Spaces.

Theorem 6.1. *Let X be a Banach space with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. If $f \in \mathcal{A}_1(\mathcal{D}) \subset X$, then the output $(f_m)_{m\ge 0}$ of the* **WRPGA**$(\{t_k\}, \mathcal{D})$ *satisfies the error estimate*

$$e_m := \|f - f_m\| \le c\|f\|_{\mathcal{A}_1(\mathcal{D})} \left(\sum_{k=1}^{m} t_k^{\frac{q}{q-1}} \right)^{1/q-1}, \quad m \ge 1,$$

where $c = c(\gamma, q)$.

Proof. As in the proof of Theorem 5.3, we show that for any $f \in X$,

$$e_m \le e_{m-1} - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} e_{m-1} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{q}{q-1}}. \tag{6.1}$$

Next, similarly to Theorem 5.3, we prove a lower estimate for $|F_{f-f_{m-1}}(\varphi_m)|$ when $f \in \mathcal{A}_1(\mathcal{D})$, which is

$$M^{-1} t_m e_{m-1} \le |F_{f-f_{m-1}}(\varphi_m)|,$$

which together with (6.1) result in

$$e_m \le e_{m-1} \left(1 - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} t_m^{\frac{q}{q-1}} M^{-\frac{q}{q-1}} e_{m-1}^{\frac{q}{q-1}} \right).$$

Again, since $e_1 \le e_0 = \|f\| \le M t_1^{-1}$, we use Lemma 2.2 with $a_m = e_m$, $B = M t_1^{-1}$, $r_m := \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} t_m^{\frac{q}{q-1}}$, $r = M^{\frac{q}{q-1}}$, $J = 1$ and $\ell = \frac{q}{q-1}$. Then, (2.1) gives

$$e_m \le M \left(t_1^{\frac{q}{q-1}} + \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} \sum_{k=2}^{m} t_k^{\frac{q}{q-1}} \right)^{1/q-1}, \quad m \ge 2,$$

and the theorem follows. □

Remark 6.2. There are many weak greedy algorithms for Banach spaces, for example, the Weak Relaxed Greedy Algorithm **(WRGA)**, or the Weak Greedy Algorithm with fixed Relaxation **(WGAFR)**, see [15], Chapter 6. These algorithms are proven to provide the same convergence rate as Theorem 6.1, see Theorems 6.17 and 6.23 from [15], but they differ either in the way the next direction φ_m is chosen or in the way the next approximant f_m is determined. The **WRPGA**$(\{t_k\}, \mathcal{D})$ from this paper is closer to the **WGAFR** in the sense that the next direction φ_m is determined in the same way. However, a significant advantage of the algorithm, presented here, is that it requires at each step the solution of one dimensional optimization problem, rather than the two dimensional optimization problem that is used in **WGAFR**. While the **WGAFR** can be applied to functions f that can be approximated by elements from $\mathcal{A}_1(\mathcal{D})$, as shown in Theorem 6.23 from [15], we have restricted our discussion only to functions $f \in \mathcal{A}_1(\mathcal{D})$, since a perturbation analysis of the algorithms, presented here will be given in a subsequent paper.

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008

References

[1] A. Barron, Universal approximation bounds for superposition of n sigmoidal functions, IEEE Trans. Inform. Theory 39 (1993) 930–945.
[2] A. Baron, A. Cohen, W. Dahmen, R. DeVore, Approximation and learning by greedy algorithms, Ann. Statist. 36 (2008) 64–94.
[3] R. DeVore, V. Temlyakov, Some remarks on greedy algorithms, Adv. Comput. Math. 5 (1996) 173–187.
[4] M. Donahue, L. Gurvits, C. Darken, E. Sontag, Rate of convex approximation in non-Hilbert spaces, Constr. Approx. 13 (1997) 187–220.
[5] Z. Gao, G. Petrova, Rescaled pure greedy algorithm for convex optimization, http://www.math.tamu.edu/~gpetrova/Gao_Petrova.pdf, submitted for publication.
[6] R. Gribonval, M. Nielsen, Approximate weak greedy algorithms, Adv. Comput. Math. 14 (2001) 361–378.
[7] L. Jones, On a conjecture of Huber concerning the convergence of projection pursuit regression, Ann. Statist. 15 (1987) 880–882.
[8] L.K. Jones, A simple lemma on greedy approximation in Hilbert spaces and convergence rates for projection pursuit regression and neural network training, Ann. Statist. 20 (1992) 608–613.
[9] S. Konyagin, V. Temlyakov, Rate of convergence of pure greedy algorithm, East J. Approx. 5 (1999) 493–499.
[10] W. Lee, P. Bartlett, R. Williamson, Efficient agnostic learning of neural networks with bounded fan-in, IEEE Trans. Inform. Theory 42 (1996) 2118–2132.
[11] E. Livshitz, On lower estimates of rate of convergence of greedy algorithms, Izv. RAN, Ser. Mat. 73 (2009) 125–144.
[12] E. Livshitz, V. Temlyakov, Two lower estimates in greedy approximation, Constr. Approx. 19 (2003) 509–523.
[13] H. Nguyen, G. Petrova, Greedy strategies for convex optimization, http://www.math.tamu.edu/~gpetrova/Convex.pdf, submitted for publication.
[14] A.V. Sil'nichenko, Rates of convergence of greedy algorithms, Mat. Zametki 76 (2004) 628–632.
[15] V. Temlyakov, Greedy Approximation, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2011.
[16] V. Temlyakov, Weak greedy algorithms, Adv. Comput. Math. 12 (2000) 213–227.

Please cite this article in press as: G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.10.008