Industrial Mathematics Institute

2006:03

Relaxation in greedy approximation

V. Temlyakov

IMI
Preprint Series

Department of Mathematics
University of South Carolina

RELAXATION IN GREEDY APPROXIMATION

V. Temlyakov

University of South Carolina

Abstract. We study greedy algorithms in a Banach space from the point of view of convergence and rate of convergence. There are two well studied approximation methods: the Weak Chebyshev Greedy Algorithm (WCGA) and the Weak Relaxed Greedy Algorithm (WRGA). The WRGA is simpler than the WCGA in the sense of computational complexity. However, the WRGA has limited applicability. It converges only for elements of the closure of the convex hull of a dictionary. In this paper we study algorithms that combine good features of both algorithms the WRGA and the WCGA. In construction of such algorithms we use different forms of relaxation. First results on such algorithms have been obtained in a Hilbert space by A. Barron, A. Cohen, W. Dahmen, and R. DeVore. Their paper was a motivation for the research reported here.

1. Introduction

Let X be a Banach space with norm $\|\cdot\|$. We say that a set of elements (functions) $\mathcal{D}$ from X is a dictionary (symmetric dictionary) if each $g \in \mathcal{D}$ has norm bounded by one ($\|g\| \le 1$),

$$g \in \mathcal{D} \quad \text{implies} \quad -g \in \mathcal{D},$$

and $\overline{\operatorname{span}}\mathcal{D} = X$. We denote the closure (in X) of the convex hull of $\mathcal{D}$ by $A_1(\mathcal{D})$. We introduce a new norm, associated with a dictionary $\mathcal{D}$, in the dual space X' by the formula

$$\|F\|_{\mathcal{D}} := \sup_{g\in\mathcal{D}} F(g), \quad F \in X'.$$

We will study in this paper greedy algorithms with regard to $\mathcal{D}$. For a nonzero element $f \in X$ we denote by F_f a norming (peak) functional for f:

$$\|F_f\| = 1, \qquad F_f(f) = \|f\|.$$

The existence of such a functional is guaranteed by Hahn-Banach theorem. Let $\tau := \{t_k\}_{k=1}^{\infty}$ be a given sequence of nonnegative numbers $t_k \le 1$, $k = 1, \dots$. We define first the Weak Chebyshev Greedy Algorithm (WCGA) (see [T3]) that is a generalization for Banach spaces of Weak Orthogonal Greedy Algorithm defined and studied in [T2] (see also [DT] for Orthogonal Greedy Algorithm).

Weak Chebyshev Greedy Algorithm (WCGA). We define $f_0^c := f_0^{c,\tau} := f$. Then for each $m \ge 1$ we inductively define

1). $\varphi_m^c := \varphi_m^{c,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^c}(\varphi_m^c) \ge t_m \|F_{f_{m-1}^c}\|_{\mathcal{D}}. \tag{1.1}$$

2). Define

$$\Phi_m := \Phi_m^\tau := \operatorname{span}\{\varphi_j^c\}_{j=1}^m,$$

and define $G_m^c := G_m^{c,\tau}$ to be the best approximant to f from Φ_m.

3). Denote

$$f_m^c := f_m^{c,\tau} := f - G_m^c.$$

We define now the generalization for Banach spaces of the Weak Relaxed Greedy Algorithm studied in [T3] (see [T2] for the case of a Hilbert space).

Weak Relaxed Greedy Algorithm (WRGA). We define $f_0^r := f_0^{r,\tau} := f$ and $G_0^r := G_0^{r,\tau} := 0$. Then for each $m \ge 1$ we inductively define

1). $\varphi_m^r := \varphi_m^{r,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^r}(\varphi_m^r - G_{m-1}^r) \ge t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}^r}(g - G_{m-1}^r).$$

2). Find $0 \le \lambda_m \le 1$ such that

$$\|f - ((1-\lambda_m)G_{m-1}^r + \lambda_m \varphi_m^r)\| = \inf_{0 \le \lambda \le 1} \|f - ((1-\lambda)G_{m-1}^r + \lambda \varphi_m^r)\|$$

and define

$$G_m^r := G_m^{r,\tau} := (1-\lambda_m)G_{m-1}^r + \lambda_m \varphi_m^r.$$

3). Denote

$$f_m^r := f_m^{r,\tau} := f - G_m^r.$$

Remark 1.1. *It follows from the definition of WCGA and WRGA that the sequences $\{\|f_m^c\|\}$ and $\{\|f_m^r\|\}$ are nonincreasing sequences.*

Both of the above algorithms use the functional $F_{f_{m-1}}$ in a search for the mth element φ_m from the dictionary to be used in approximation. The construction of the approximant in the WRGA is different from the construction in the WCGA. In the WCGA we build the approximant G_m^c in a way to maximally use the approximation power of the elements $\varphi_1, \dots, \varphi_m$. The WRGA by its definition is designed for approximation of functions from $A_1(\mathcal{D})$. In building the approximant in the WRGA we keep the property $G_m^r \in A_1(\mathcal{D})$. We call the WRGA *relaxed* because at the mth step of the algorithm we use a linear combination (convex combination) of the previous approximant G_{m-1}^r and a new element φ_m^r. The relaxation parameter λ_m in the WRGA is chosen at the mth step depending on f. Recently, the following modification of the above idea of relaxation in greedy approximation has been studied in [BCDD]. Let a sequence $\mathbf{r} := \{r_k\}_{k=1}^\infty$, $r_k \in [0,1)$, of relaxation parameters be given. Then at each step of our new algorithm we build the mth approximant of the form $G_m = (1-r_m)G_{m-1} + \lambda\varphi_m$. With an approximant of this form we are not limited to approximation of functions from $A_1(\mathcal{D})$ as in the WRGA. Remarkable results on the approximation properties of such an algorithm in a Hilbert space have been obtained in [BCDD] (see Section 5 below). We will study here a realization of the above new idea of relaxation in the case of Banach spaces. In Section 2 we study the Greedy Algorithm with Weakness parameter t and Relaxation $\mathbf{r}$ (GAWR($t,\mathbf{r}$)). In addition to the acronym GAWR($t,\mathbf{r}$) we will use the abbreviated acronym GAWR for the name of this algorithm. We give a general definition of the algorithm in the case of a weakness sequence τ.

GAWR($\tau,\mathbf{r}$). Let $\tau := \{t_m\}_{m=1}^{\infty}$, $t_m \in [0,1]$, be a weakness sequence. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \geq t_m \|F_{f_{m-1}}\|_{\mathcal{D}}.$$

2). Find $\lambda_m \geq 0$ such that

$$\|f - ((1-r_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{\lambda\geq 0} \|f - ((1-r_m)G_{m-1} + \lambda\varphi_m)\|$$

and define

$$G_m := (1-r_m)G_{m-1} + \lambda_m\varphi_m.$$

3). Denote

$$f_m := f - G_m.$$

In the case $\tau = \{t\}$, $t \in (0,1]$, we write t instead of τ in the notation. We note that in the case $r_k = 0$, $k = 1,\dots,$ when there is no relaxation, the GAWR($\tau,\mathbf{0}$) coincides with the Weak Dual Greedy Algorithm [T4, p.66]. We will also consider here a relaxation of the X-greedy algorithm (see [T4, p.39]) that corresponds to $\mathbf{r} = \mathbf{0}$ in the definition that follows.

X-Greedy Algorithm with Relaxation r (XGAR(r)). We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ and $\lambda_m \geq 0$ are such that

$$\|f - ((1-r_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{g\in\mathcal{D},\lambda\geq 0} \|f - ((1-r_m)G_{m-1} + \lambda g)\|$$

and

$$G_m := (1-r_m)G_{m-1} + \lambda_m\varphi_m.$$

2). Denote

$$f_m := f - G_m.$$

We note that, practically, nothing is known about convergence and rate of convergence of the X-greedy algorithm. It will be seen from the results of Section 2 that relaxation helps to prove convergence results for the XGAR($\mathbf{r}$).

The following version of relaxed greedy algorithm will be studied in Section 3.

Weak Greedy Algorithm with Free Relaxation (WGAFR). Let $\tau := \{t_m\}_{m=1}^\infty$, $t_m \in [0,1]$, be a weakness sequence. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \ge 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \ge t_m \|F_{f_{m-1}}\|_{\mathcal{D}}.$$

2). Find w_m and $\lambda_m \ge 0$ such that

$$\|f - ((1-w_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{\lambda\ge 0, w} \|f - ((1-w)G_{m-1} + \lambda\varphi_m)\|$$

and define

$$G_m := (1-w_m)G_{m-1} + \lambda_m\varphi_m.$$

3). Denote

$$f_m := f - G_m.$$

We consider here approximation in uniformly smooth Banach spaces. For a Banach space X we define the modulus of smoothness

$$\rho(u) := \sup_{\|x\|=\|y\|=1} \left(\frac{1}{2}(\|x+uy\| + \|x-uy\|) - 1\right).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u\to 0} \rho(u)/u = 0.$$

It is easy to see that for any Banach space X its modulus of smoothness $\rho(u)$ is an even convex function satisfying the inequalities

$$\max(0, u-1) \le \rho(u) \le u, \quad u \in (0,\infty).$$

It is well known (see for instance [DGDS, Lemma B.1]) that in the case $X = L_p$, $1 \le p < \infty$ we have

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

It is also known (see [LT], p.63) that for any X with $\dim X = \infty$ one has

$$\rho(u) \ge (1+u^2)^{1/2} - 1$$

and for every X, $\dim X \ge 2$,

$$\rho(u) \ge Cu^2, \quad C > 0.$$

This limits power type modulus of smoothness of nontrivial Banach spaces to the case $1 \le q \le 2$.

We formulate in the Introduction two typical results of the paper. The first theorem is proved in Section 2 and the second theorem is proved in Section 3.

Theorem 1.1. *Let a sequence* $\mathbf{r} := \{r_k\}_{k=1}^\infty$, $r_k \in [0,1)$, *satisfy the conditions*

$$\sum_{k=1}^{\infty} r_k = \infty, \quad r_k \to 0 \quad as \quad k \to \infty.$$

Then the GAWR($t, \mathbf{r}$) *and the XGAR*($\mathbf{r}$) *converge in any uniformly smooth Banach space for each* $f \in X$ *and for all dictionaries* $\mathcal{D}$.

Theorem 1.2. *Let* X *be a uniformly smooth Banach space with modulus of smoothness* $\rho(u) \le \gamma u^q$, $1 < q \le 2$. *Take a number* $\epsilon \ge 0$ *and two elements* f, f^ϵ *from* X *such that*

$$\|f - f^\epsilon\| \le \epsilon, \quad f^\epsilon / A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. *Then we have for the WGAFR*

$$\|f_m\| \le \max\left(2\epsilon, C(q,\gamma)(A(\epsilon)+\epsilon)(1+\sum_{k=1}^m t_k^p)^{-1/p}\right), \quad p := q/(q-1).$$

Remark 1.2. *The setting in Theorem 1.2 with two functions* f *and* f^ϵ *covers the following noisy data setting. Let* $A(\epsilon) = A$ *and the target function* f^ϵ *is such that* $f^\epsilon/A \in A_1(\mathcal{D})$. *The task is to approximate* f^ϵ *from the noisy data* f *of it.*

In Section 4 we present some remarks on computational complexity of greedy algorithms and introduce two thresholding type greedy algorithms. We also demonstrate in Section 4 how a special structure of Hilbert spaces can be used in improving (in the sense of numerical constants) our error bounds.

Section 5 contains a discussion of the results of the paper. It also contains some historical remarks.

2. Convergence and rate of convergence of the GAWR

We begin with known results on the behavior of the WCGA. There are results in [T3] that give sufficient conditions for convergence of the WCGA in terms of the weakness sequence τ and the modulus of smoothness ρ. In particular, it is proved in [T3] that the WCGA with $\tau = \{t\}$, $t \in (0,1]$ converges in any uniformly smooth Banach space for each $f \in X$ and any dictionary $\mathcal{D}$. The following theorem from [T3] provides the rate of convergence of the WCGA.

Theorem 2.1. *Let* X *be a uniformly smooth Banach space with modulus of smoothness* $\rho(u) \le \gamma u^q$, $1 < q \le 2$. *Then for a sequence* $\tau := \{t_k\}_{k=1}^\infty$, $t_k \le 1$, $k = 1, 2, \dots$, *we have for any* $f \in A_1(\mathcal{D})$ *that*

$$\|f_m^{c,\tau}\| \le C(q,\gamma)(1+\sum_{k=1}^m t_k^p)^{-1/p}, \quad p := \frac{q}{q-1}, \tag{2.1}$$

with a constant $C(q,\gamma)$ which may depend only on q and γ.

This theorem gives the rate of convergence of WCGA for f in $A_1(\mathcal{D})$. It was pointed out in [BCDD] that it is important to have estimates for the rate of approximation of greedy algorithms for more general functions. We will now formulate the corresponding variant of Theorem 2.1. Theorem 2.1 was derived in [T3] from the following lemma (see [T3, Lemma 2.3])

Lemma 2.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have

$$\|f_m^{c,\tau}\| \leq \|f_{m-1}^{c,\tau}\| \inf_\lambda \left(1 - \lambda t_m A(\epsilon)^{-1}\left(1 - \frac{\epsilon}{\|f_{m-1}^{c,\tau}\|}\right) + 2\rho\left(\frac{\lambda}{\|f_{m-1}^{c,\tau}\|}\right)\right), \quad m = 1, 2, \dots.$$

In the same way as Theorem 2.1 was derived from Lemma 2.1 in [T3] we obtain the following result.

Theorem 2.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have

$$\|f_m^{c,\tau}\| \leq \max\left(2\epsilon, C(q,\gamma)(A(\epsilon)+\epsilon)(1 + \sum_{k=1}^m t_k^p)^{-1/p}\right), \quad p := q/(q-1). \tag{2.2}$$

We now proceed to the GAWR. We begin with an analogue of Lemma 2.1.

Lemma 2.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have for the GAWR($t, \mathbf{r}$) and for the XGAR($\mathbf{r}$) (for this algorithm $t = 1$)

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - r_m\left(1 - \frac{\epsilon}{\|f_{m-1}\|}\right) + 2\rho\left(\frac{r_m(\|f\| + A(\epsilon)/t)}{(1 - r_m)\|f_{m-1}\|}\right)\right), \quad m = 1, 2, \dots.$$

Proof. It is clear that it suffices to prove Lemma 2.2 in the case of the GAWR($t,\mathbf{r}$). By the definition of f_m

$$\|f_m\| = \inf_{\lambda\ge 0}\|f-((1-r_m)G_{m-1}+\lambda\varphi_m)\|.$$

We have for any λ

$$f-((1-r_m)G_{m-1}+\lambda\varphi_m) = (1-r_m)f_{m-1}+r_mf-\lambda\varphi_m \tag{2.3}$$

and

$$\|(1-r_m)f_{m-1}+r_mf-\lambda\varphi_m\| + \|(1-r_m)f_{m-1}-r_mf+\lambda\varphi_m\| \le \tag{2.4}$$

$$2(1-r_m)\|f_{m-1}\|\left(1+\rho\left(\frac{\|r_mf-\lambda\varphi_m\|}{(1-r_m)\|f_{m-1}\|}\right)\right).$$

We have

$$\|(1-r_m)f_{m-1}-r_mf+\lambda\varphi_m\| \ge F_{f_{m-1}}((1-r_m)f_{m-1}-r_mf+\lambda\varphi_m) = \tag{2.5}$$

$$(1-r_m)\|f_{m-1}\| - r_mF_{f_{m-1}}(f)+\lambda F_{f_{m-1}}(\varphi_m).$$

From the definition of φ_m we get

$$F_{f_{m-1}}(\varphi_m) \ge t\sup_{g\in\mathcal{D}} F_{f_{m-1}}(g). \tag{2.6}$$

By Lemma 2.2 from [T3] we obtain

$$\sup_{g\in\mathcal{D}} F_{f_{m-1}}(g) = \sup_{\phi\in A_1(\mathcal{D})} F_{f_{m-1}}(\phi) \ge A(\epsilon)^{-1}F_{f_{m-1}}(f^\epsilon) \ge A(\epsilon)^{-1}(F_{f_{m-1}}(f)-\epsilon). \tag{2.7}$$

Combining (2.6) and (2.7) we get

$$F_{f_{m-1}}(\varphi_m) \ge tA(\epsilon)^{-1}(F_{f_{m-1}}(f)-\epsilon). \tag{2.8}$$

We now choose $\lambda := \lambda^* := r_mA(\epsilon)/t$. Then for this λ we derive from (2.5) and (2.8)

$$\|(1-r_m)f_{m-1}-r_mf+\lambda^*\varphi_m\| \ge (1-r_m)\|f_{m-1}\| - r_m\epsilon. \tag{2.9}$$

The relations (2.4) and (2.9) imply

$$\|(1-r_m)f_{m-1}+r_mf-\lambda^*\varphi_m\| \le$$

$$(1-r_m)\|f_{m-1}\| + r_m\epsilon + 2(1-r_m)\|f_{m-1}\|\rho\left(\frac{r_m(\|f\|+A(\epsilon)/t)}{(1-r_m)\|f_{m-1}\|}\right). \quad \square$$

Proof of Theorem 1.1. We prove this theorem in two steps.

I. First, we prove that $\liminf_{m\to\infty}\|f_m\|=0$. The proof goes by contradiction. We want to prove that $\liminf_{m\to\infty}\|f_m\|=0$. Assume the contrary. Then there exists K and $\beta>0$ such that we have for all $k\ge K$ that $\|f_k\|\ge\beta$. By Lemma 2.2 for $m>K$

$$\|f_m\| \le \|f_{m-1}\|\left(1-r_m\left(1-\frac{\epsilon}{\beta}\right)+2\rho\left(\frac{r_m(\|f\|+A(\epsilon)/t)}{(1-r_m)\beta}\right)\right), \quad m=1,2,\dots.$$

We choose $\epsilon:=\beta/2$. Using the assumption that X is uniformly smooth and the assumption $r_k\to 0$ as $k\to\infty$, we find $N\ge K$ such that for $m\ge N$ we have

$$2\rho\left(\frac{r_m(\|f\|+A(\epsilon)/t)}{(1-r_m)\beta}\right)\le r_m/4.$$

Then for $m>N$

$$\|f_m\|\le\|f_{m-1}\|(1-r_m/4).$$

The assumption $\sum_{m=1}^\infty r_m=\infty$ implies that $\|f_m\|\to 0$ as $m\to\infty$. The obtained contradiction to the assumption $\beta>0$ completes the proof of part I.

II. Secondly, we prove that $\lim_{m\to\infty}\|f_m\|=0$. Using the assumption $r_k\to 0$ as $k\to\infty$ we find N_1 such that for $k\ge N_1$ we have $r_k\le 1/2$. For such k we obtain from Lemma 2.2

$$\|f_k\|-\epsilon\le(1-r_k)(\|f_{k-1}\|-\epsilon)+2\|f_{k-1}\|\rho\left(\frac{Br_k}{\|f_{k-1}\|}\right) \tag{2.10}$$

with $B:=2(\|f\|+A(\epsilon)/t)$. Denote $a_k:=\|f_{k-1}\|-\epsilon$. We note that from the definition of f_k and the representation (2.3) follows that

$$a_{k+1}\le a_k+r_k\|f\|. \tag{2.11}$$

Using the fact that the function $\rho(u)/u$ is monotone increasing on $[0,\infty)$ we obtain from (2.10) for $a_k>0$

$$a_{k+1}\le a_k\left(1-r_k+2\frac{\|f_{k-1}\|}{a_k}\rho\left(\frac{Br_k}{\|f_{k-1}\|}\right)\right)\le a_k\left(1-r_k+2\rho\left(\frac{Br_k}{a_k}\right)\right). \tag{2.12}$$

We now introduce an auxiliary sequence $\{b_k\}$ of positive numbers that is defined by the equation

$$2\rho(Br_k/b_k)=r_k.$$

The property $\rho(u)/u\to 0$ as $u\to 0$ implies $b_k\to 0$ as $k\to\infty$. The inequality (2.12) guarantees that for $k\ge N_1$ such that $a_k\ge b_k$ we have $a_{k+1}\le a_k$.

Let

$$U:=\{k:\quad k\ge N_1,\quad a_k\ge b_k\}.$$

If the set U is finite then we get

$$\limsup_{k\to\infty}a_k\le\lim_{k\to\infty}b_k=0.$$

This implies

$$\limsup_{m\to\infty} \|f_m\| \le \epsilon.$$

Consider the case when U is infinite. We note that part I of the proof implies that there is a subsequence $\{k_j\}$ such that $a_{k_j} \le 0$, $j = 1, 2, \dots$. This means that

$$U = \cup_{j=1}^{\infty}[l_j, n_j]$$

with the property $n_{j-1} < l_j - 1$. For $k \notin U$, $k \ge N_1$ we have

$$a_k < b_k. \tag{2.13}$$

For $k \in [l_j, n_j]$ we have by (2.11) and the monotonicity property of a_k, when $k \in [l_j, n_j]$, that

$$a_k \le a_{l_j} \le a_{l_j-1} + r_{l_j-1}\|f\| \le b_{l_j-1} + r_{l_j-1}\|f\|. \tag{2.14}$$

By (2.13) and (2.14) we obtain

$$\limsup_{k\to\infty} a_k \le 0 \quad \Rightarrow \quad \limsup_{m\to\infty} \|f_m\| \le \epsilon.$$

Taking into account that $\epsilon > 0$ is arbitrary we complete the proof. $\square$

We now proceed to results on the rate of approximation. We will need the following technical lemma. This lemma is a more general version of Lemma 2.1 from [T1] (see also Remark 5.1 in [T6]).

Lemma 2.3. *Let a sequence $\{a_n\}_{n=1}^{\infty}$ have the following property. For given positive numbers $\alpha < \beta \le 1$, $A > a_1$ we have for all $n \ge 2$*

$$a_n \le a_{n-1} + A(n-1)^{-\alpha}; \tag{2.15}$$

if for some $\nu \ge 2$ we have

$$a_\nu \ge A\nu^{-\alpha}$$

then

$$a_{\nu+1} \le a_\nu(1 - \beta/\nu). \tag{2.16}$$

Then there exists a constant $C(\alpha, \beta)$ such that for all $n = 1, 2, \dots$ we have

$$a_n \le C(\alpha, \beta)An^{-\alpha}.$$

Proof. We have $a_1 < A$ which implies that the set

$$V := \{\nu : a_\nu \ge A\nu^{-\alpha}\}$$

does not contain $\nu = 1$. We prove now that for any segment $[n, n+k] \subset V$ we have $k \le C_1(\alpha,\beta)n$. Indeed, let $n \ge 2$ be such that $n-1 \notin V$, which means

$$a_{n-1} < A(n-1)^{-\alpha}, \tag{2.17}$$

and $[n, n+k] \subset V$, which in turn means

$$a_{n+j} \ge A(n+j)^{-\alpha}, \quad j = 0, 1, \dots, k. \tag{2.18}$$

Then by the conditions (2.15) and (2.16) of the lemma we get

$$a_{n+k} \le a_n \prod_{\nu=n}^{n+k-1} (1-\beta/\nu) \le (a_{n-1} + A(n-1)^{-\alpha}) \prod_{\nu=n}^{n+k-1} (1-\beta/\nu). \tag{2.19}$$

Combining (2.17) – (2.19) we obtain

$$(n+k)^{-\alpha} \le 2(n-1)^{-\alpha} \prod_{\nu=n}^{n+k-1} (1-\beta/\nu). \tag{2.20}$$

Taking logarithms and using the inequalities

$$\ln(1-x) \le -x, \quad x \in [0,1);$$

$$\sum_{\nu=n}^{m-1} \nu^{-1} \ge \int_n^m x^{-1} dx = \ln(m/n),$$

we get from (2.20)

$$-\alpha \ln \frac{n+k}{n-1} \le \ln 2 + \sum_{\nu=n}^{n+k-1} \ln(1-\beta/\nu) \le \ln 2 - \sum_{\nu=n}^{n+k-1} \beta/\nu \le \ln 2 - \beta \ln \frac{n+k}{n}.$$

Hence

$$(\beta-\alpha)\ln(n+k) \le \ln 2 + (\beta-\alpha)\ln n + \alpha \ln \frac{n}{n-1},$$

which implies

$$n + k \le 2^{\frac{\alpha+1}{\beta-\alpha}} n$$

and

$$k \le C_1(\alpha,\beta)n.$$

Let us take any $m \in \mathbb{N}$. If $m \notin V$ we have the desired inequality with $C(\alpha,\beta) = 1$. Assume $m \in V$, and let $[n, n+k]$ be the maximal segment in V containing m. Then similarly to (2.19)

$$a_m \le a_n \le a_{n-1} + A(n-1)^{-\alpha} \le 2A(n-1)^{-\alpha} \le 2Am^{-\alpha} \left(\frac{n-1}{m}\right)^{-\alpha}. \tag{2.21}$$

Using the inequality $k \le C_1(\alpha,\beta)n$ proved above we get

$$\frac{m}{n-1} \le \frac{n+k}{n-1} \le C_2(\alpha,\beta). \tag{2.22}$$

Substituting (2.22) into (2.21) we complete the proof of Lemma 2.3. $\Box$

Theorem 2.3. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let $\mathbf{r} := \{2/(k+2)\}_{k=1}^{\infty}$. Consider the GAWR$(t,\mathbf{r})$ and the XGAR$(\mathbf{r})$ (for this algorithm $t = 1$). For a pair of functions f, f^ϵ, satisfying*

$$\|f - f^\epsilon\| \le \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D})$$

we have

$$\|f_m\| \le \epsilon + C(q,\gamma)(\|f\| + A(\epsilon)/t)m^{-1+1/q}.$$

Proof. By Lemma 2.2 we obtain

$$\|f_k\| - \epsilon \le (1 - r_k)(\|f_{k-1}\| - \epsilon) + C\gamma\|f_{k-1}\| \left(\frac{r_k(\|f\| + A(\epsilon)/t)}{\|f_{k-1}\|}\right)^q. \tag{2.23}$$

Consider, as in the proof of Theorem 1.1, the sequence $a_n := \|f_{n-1}\| - \epsilon$. We plan to apply Lemma 2.3 to the sequence $\{a_n\}$. We set $\alpha := 1 - 1/q \le 1/2$. The parameters $\beta \in (\alpha, 1]$ and A will be chosen later. We note that

$$\|f_m\| \le \|f_{m-1}\| + r_m\|f\|.$$

Therefore, the condition (2.15) of Lemma 2.3 is satisfied with $A \ge 2\|f\|$. Let $a_k \ge Ak^{-\alpha}$. Then by (2.23) we get

$$a_{k+1} \le a_k(1 - r_k + C\gamma(r_k(\|f\| + A(\epsilon)/t)/a_k)^q \le$$

$$a_k\left(1 - \frac{2}{k+2} + \frac{C\gamma(\|f\| + A(\epsilon)/t)^q 2^q}{A^q}\frac{k^{\alpha q}}{(k+2)^q}\right).$$

Setting $A := \max(2\|f\|, 2(2C\gamma)^{1/q}(\|f\| + A(\epsilon)/t))$ we continue

$$\le a_k\left(1 - \frac{3}{2(k+2)}\right).$$

Thus the condition (2.16) of Lemma 2.3 is satisfied with $\beta = 3/4$. Applying Lemma 2.3 we obtain

$$\|f_m\| \le \epsilon + C(q,\gamma)(\|f\| + A(\epsilon)/t)m^{-1+1/q}. \quad \square$$

3. Convergence and rate of convergence of the WGAFR

We begin with the proof of Theorem 1.2. The proof of this theorem is based on the following analogue of Lemma 2.1.

Lemma 3.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) \geq \epsilon$. Then we have for the WGAFR

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda \geq 0} \left(1 - \lambda t_m A(\epsilon)^{-1}\left(1 - \frac{\epsilon}{\|f_{m-1}\|}\right) + 2\rho\left(\frac{5\lambda}{\|f_{m-1}\|}\right)\right), \quad m = 1, 2, \dots.$$

Proof. By the definition of f_m

$$\|f_m\| = \inf_{\lambda \geq 0, w} \|f_{m-1} + wG_{m-1} - \lambda\varphi_m\|.$$

Similarly to the arguments in the proof of Lemma 2.2 we write the inequality

$$\|f_{m-1} + wG_{m-1} - \lambda\varphi_m\| + \|f_{m-1} - wG_{m-1} + \lambda\varphi_m\| \leq$$
$$2\|f_{m-1}\|(1 + \rho(\|wG_{m-1} - \lambda\varphi_m\|/\|f_{m-1}\|)) \tag{3.1}$$

and estimate for $\lambda \geq 0$

$$\|f_{m-1} - wG_{m-1} + \lambda\varphi_m\| \geq F_{f_{m-1}}(f_{m-1} - wG_{m-1} + \lambda\varphi_m) \geq$$
$$\|f_{m-1}\| - F_{f_{m-1}}(wG_{m-1}) + \lambda t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}}(g) =$$

By Lemma 2.2 of [T3]

$$\|f_{m-1}\| - F_{f_{m-1}}(wG_{m-1}) + \lambda t_m \sup_{\phi \in A_1(\mathcal{D})} F_{f_{m-1}}(\phi) \geq$$
$$\|f_{m-1}\| - F_{f_{m-1}}(wG_{m-1}) + \lambda t_m A(\epsilon)^{-1} F_{f_{m-1}}(f^\epsilon) \geq$$
$$\|f_{m-1}\| - F_{f_{m-1}}(wG_{m-1}) + \lambda t_m A(\epsilon)^{-1}(F_{f_{m-1}}(f) - \epsilon).$$

We set $w^* := \lambda t_m A(\epsilon)^{-1}$ and obtain

$$\|f_{m-1} - w^*G_{m-1} + \lambda\varphi_m\| \geq \|f_{m-1}\| + \lambda t_m A(\epsilon)^{-1}(\|f_{m-1}\| - \epsilon). \tag{3.2}$$

Combining (3.1) and (3.2) we get

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda \geq 0} \left(1 - \lambda t_m A(\epsilon)^{-1}\left(1 - \frac{\epsilon}{\|f_{m-1}\|}\right) + 2\rho\left(\frac{\|w^*G_{m-1} - \lambda\varphi_m\|}{\|f_{m-1}\|}\right)\right).$$

We now estimate

$$\|w^*G_{m-1} - \lambda\varphi_m\| \leq w^*\|G_{m-1}\| + \lambda.$$

Next,

$$\|G_{m-1}\| = \|f - f_{m-1}\| \leq 2\|f\| \leq 2(\|f^\epsilon\| + \epsilon) \leq 2(A(\epsilon) + \epsilon).$$

Thus, under assumption $A(\epsilon) \geq \epsilon$ we get

$$w^*\|G_{m-1}\| \leq 2\lambda t_m (A(\epsilon) + \epsilon)/A(\epsilon) \leq 4\lambda.$$

Finally,

$$\|w^*G_{m-1} - \lambda\varphi_m\| \leq 5\lambda.$$

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

Remark 3.1. *It follows from the definition of the WGAFR that the sequence* $\{\|f_m\|\}$ *is a nonicreasing sequence.*

Proof of Theorem 1.2. It is clear that it suffices to consider the case $A(\epsilon) \geq \epsilon$. Otherwise, $\|f_m\| \leq \|f\| \leq \|f^\epsilon\| + \epsilon \leq 2\epsilon$. Also, assume $\|f_m\| > 2\epsilon$ (otherwise, Theorem 1.2 trivially holds). Then by Remark 3.1 we have for all $k = 0, 1, \dots, m$ that $\|f_k\| > 2\epsilon$. By Lemma 3.1 we obtain

$$\|f_k\| \leq \|f_{k-1}\| \inf_{\lambda \geq 0} \left(1 - \lambda t_k A(\epsilon)^{-1}/2 + 2\gamma \left(\frac{5\lambda}{\|f_{k-1}\|}\right)^q\right). \tag{3.3}$$

Choose λ from the equation

$$\frac{\lambda t_k}{4A(\epsilon)} = 2\gamma \left(\frac{5\lambda}{\|f_{k-1}\|}\right)^q$$

what implies that

$$\lambda = \|f_{k-1}\|^{\frac{q}{q-1}} 5^{-\frac{q}{q-1}} (8\gamma A(\epsilon))^{-\frac{1}{q-1}} t_k^{\frac{1}{q-1}}.$$

Denote

$$A_q := 4(8\gamma)^{\frac{1}{q-1}} 5^{\frac{q}{q-1}}.$$

Using notation $p := \frac{q}{q-1}$ we get from (3.3)

$$\|f_k\| \leq \|f_{k-1}\| \left(1 - \frac{1}{4}\frac{\lambda t_k}{A(\epsilon)}\right) = \|f_{k-1}\| \left(1 - \frac{t_k^p \|f_{k-1}\|^p}{A_q A(\epsilon)^p}\right).$$

Raising both sides of this inequality to the power p and taking into account the inequality $x^r \leq x$ for $r \geq 1$, $0 \leq x \leq 1$, we obtain

$$\|f_k\|^p \leq \|f_{k-1}\|^p \left(1 - \frac{t_k^p \|f_{k-1}\|^p}{A_q A(\epsilon)^p}\right).$$

By Lemma 3.1 from [T2], using the estimates $\|f\| \leq A(\epsilon) + \epsilon$ and $A_q > 1$, we get

$$\|f_m\|^p \leq A_q (A(\epsilon) + \epsilon)^p \left(1 + \sum_{k=1}^m t_k^p\right)^{-1}$$

which implies

$$\|f_m\| \leq C(q, \gamma)(A(\epsilon) + \epsilon) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p}.$$

Theorem 1.2 is proved. □

We now prove a convergence theorem for an arbitrary uniformly smooth Banach space. Modulus of smoothness $\rho(u)$ of a uniformly smooth Banach space is an even convex function such that $\rho(0) = 0$ and $\lim_{u \to 0} \rho(u)/u = 0$. The following function $s(u) := \rho(u)/u$, $s(0) := 0$, associated with $\rho(u)$ is a continuous increasing on $[0, \infty)$ function. Therefore, the inverse function $s^{-1}(\cdot)$ is well defined.

Theorem 3.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Assume that a sequence $\tau := \{t_k\}_{k=1}^\infty$ satisfies the condition: for any $\theta > 0$ we have*

$$\sum_{m=1}^{\infty} t_m s^{-1}(\theta t_m) = \infty. \tag{3.4}$$

Then for any $f \in X$ we have for the WGAFR

$$\lim_{m\to\infty} \|f_m\| = 0.$$

Proof. By Remark 3.1 $\{\|f_m\|\}$ is a nonincreasing sequence. Therefore we have

$$\lim_{m\to\infty} \|f_m\| = \beta.$$

We prove that $\beta = 0$ by contradiction. Assume the contrary that $\beta > 0$. Then for any m we have

$$\|f_m\| \geq \beta.$$

We set $\epsilon = \beta/2$ and find f^ϵ such that

$$\|f - f^\epsilon\| \leq \epsilon \quad \text{and} \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D})$$

with some $A(\epsilon) \geq \epsilon$. Then by Lemma 3.1 we get

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda\geq 0}(1 - \lambda t_m A(\epsilon)^{-1}/2 + 2\rho(5\lambda/\beta)).$$

Let us specify $\theta := \beta/(40A(\epsilon))$ and take $\lambda = \beta s^{-1}(\theta t_m)/5$. Then we obtain

$$\|f_m\| \leq \|f_{m-1}\|(1 - 2\theta t_m s^{-1}(\theta t_m)).$$

The assumption

$$\sum_{m=1}^{\infty} t_m s^{-1}(\theta t_m) = \infty$$

implies that

$$\|f_m\| \to 0 \quad \text{as} \quad m \to \infty.$$

We got a contradiction which proves the theorem. $\square$

We consider one more variant of the Weak Relaxed Greedy Algorithm. We first give a remark explaining why we consider this variant. At the second step of the WRGA we are finding the parameter λ_m from the optimization step

$$\|f_{m-1} - \lambda_m(\varphi_m - G_{m-1})\| = \inf_{\lambda\in[0,1]} \|f_{m-1} - \lambda(\varphi_m - G_{m-1})\|.$$

In this case we can interpret relaxation as the replacement of an element φ_m by an element $\varphi_m - G_{m-1}$ in the search for best approximation. This leads (for $\lambda \in [0,1]$) to a limitation of approximation of functions from $A_1(\mathcal{D})$. We want to get rid of this limitation and replace the element $\varphi_m - G_{m-1}$ by the element $\varphi_m - \alpha_m G_{m-1}$ with a prescribed in advance sequence $\{\alpha_m\}$ of nonnegative numbers. We note that in the case $\alpha_m = 0$ there is no relaxation and we get the Weak Dual Greedy Algorithm (see [T4, p.66]).

Weak α-Relaxed Greedy Algorithm (Wα-RGA). We define $f_0^\alpha := f_0^{\alpha,\tau} := f$ and $G_0^\alpha := G_0^{\alpha,\tau} := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m^\alpha := \varphi_m^{\alpha,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^\alpha}(\varphi_m^\alpha - \alpha_m G_{m-1}^\alpha) \geq t_m \sup_{g\in\mathcal{D}} F_{f_{m-1}^\alpha}(g - \alpha_m G_{m-1}^\alpha).$$

2). Find $\lambda_m \geq 0$ such that

$$\|f - ((1-\lambda_m\alpha_m)G_{m-1}^\alpha + \lambda_m\varphi_m^\alpha)\| = \inf_{\lambda\geq 0} \|f - ((1-\lambda\alpha_m)G_{m-1}^\alpha + \lambda\varphi_m^\alpha)\|$$

and define

$$G_m^\alpha := G_m^{\alpha,\tau} := (1-\lambda_m\alpha_m)G_{m-1}^\alpha + \lambda_m\varphi_m^\alpha.$$

3). Denote

$$f_m^\alpha := f_m^{\alpha,\tau} := f - G_m^\alpha.$$

Theorem 3.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Assume that sequences $\tau := \{t_k\}_{k=1}^\infty$ and $\alpha := \{\alpha_k\}_{k=1}^\infty$ satisfy the conditions: $\alpha_m \in [0,1]$, $\alpha_m \to 0$ as $m \to \infty$,*

$$\sum_{m=1}^\infty t_m\alpha_m s^{-1}(\theta t_m\alpha_m) = \infty \quad \textit{for any} \quad \theta > 0.$$

Then the Wα-RGA converges for each $f \in X$ and for all dictionaries $\mathcal{D}$.

Proof. In the proof we will drop α from the notations. We begin with the inequality

$$\|f_{m-1} - \lambda(\varphi_m - \alpha_m G_{m-1})\| + \|f_{m-1} + \lambda(\varphi_m - \alpha_m G_{m-1})\| \leq \tag{3.5}$$

$$2\|f_{m-1}\|(1 + \rho(\lambda\|\varphi_m - \alpha_m G_{m-1}\|/\|f_{m-1}\|)).$$

Next, for $\lambda \geq 0$

$$\|f_{m-1} + \lambda(\varphi_m - \alpha_m G_{m-1})\| \geq F_{f_{m-1}}(f_{m-1} + \lambda(\varphi_m - \alpha_m G_{m-1})) =$$

$$\|f_{m-1}\| + \lambda F_{f_{m-1}}(\varphi_m - \alpha_m G_{m-1}) \geq \|f_{m-1}\| + \lambda t_m\left(\sup_{g\in\mathcal{D}} F_{f_{m-1}}(g) - F_{f_{m-1}}(\alpha_m G_{m-1})\right) =$$

By Lemma 2.2 from [T3]

$$\|f_{m-1}\| + \lambda t_m\left(\sup_{\phi\in A_1(\mathcal{D})} F_{f_{m-1}}(\phi) - F_{f_{m-1}}(\alpha_m G_{m-1})\right).$$

Assume that $\phi_m \in A_1(\mathcal{D})$ is such that

$$\|f - \phi_m/\alpha_m\| \leq \epsilon_m \quad \text{with} \quad \epsilon_m \to 0.$$

Then we continue

$$\geq \|f_{m-1}\| + \lambda t_m(F_{f_{m-1}}(\phi_m) - F_{f_{m-1}}(\alpha_m G_{m-1})) \geq$$

$$\|f_{m-1}\| + \lambda t_m(F_{f_{m-1}}(\alpha_m f) - \alpha_m\epsilon_m - F_{f_{m-1}}(\alpha_m G_{m-1})) =$$

$$\|f_{m-1}\| + \lambda t_m(F_{f_{m-1}}(\alpha_m f_{m-1}) - \alpha_m\epsilon_m) = \|f_{m-1}\|(1 + \lambda t_m\alpha_m(1 - \epsilon_m/\|f_{m-1}\|)).$$

This and (3.5) imply

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda\geq 0}\left(1 - \lambda t_m\alpha_m(1 - \epsilon_m/\|f_{m-1}\|) + 2\rho(\lambda(1+2\|f\|)/\|f_{m-1}\|)\right). \tag{3.6}$$

We complete the proof by the contradiction argument in the same way as in the proof of Theorem 3.1. The sequence $\{\|f_m\|\}$ is nonincreasing. Suppose

$$\lim_{k\to\infty}\|f_m\| = \beta > 0.$$

Then (3.6) implies for big enough m

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda\geq 0}(1 - \lambda t_m\alpha_m/2 + 2\rho(\lambda(1+2\|f\|)/\beta)). \tag{3.7}$$

Our assumption on τ and α

$$\sum_{m=1}^{\infty} t_m\alpha_m s^{-1}(\theta t_m\alpha_m) = \infty$$

implies

$$\prod_{m=1}^{\infty}(1 - t_m\alpha_m s^{-1}(\theta t_m\alpha_m)) = 0.$$

From this and (3.7) one can derive that $\|f_m\| \to 0$ as $m \to \infty$ which is in contradiction with $\beta > 0$. $\square$

We conclude this section by the following remark. The algorithms GAWR and WGAFR are both of dual type greedy algorithms. The first steps are similar for both algorithms: we use the norming functional $F_{f_{m-1}}$ in the search for an element φ_m. The WGAFR provides more freedom than the GAWR does in choosing good coefficients w_m and λ_m. This results in more flexibility in choosing the weakness sequence $\tau = \{t_m\}$. For instance, the condition (3.4) of Theorem 3.1 is satisfied if $\tau = \{t\}$, $t \in (0,1]$ for any uniformly smooth Banach space. In the case $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$, the condition (3.4) is satisfied if

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p := q/(q-1).$$

In Section 2 in parallel with consideration of the GAWR we studied the XGAR($\mathbf{r}$). In the same way we can consider in parallel with the WGAFR the following analogue of the XGAR($\mathbf{r}$).

X-Greedy Algorithm with Free Relaxation (XGAFR). We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ and $\lambda_m \geq 0$, w_m are such that

$$\|f - ((1-w_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{g\in\mathcal{D},\lambda\geq 0,w} \|f - ((1-w)G_{m-1} + \lambda g)\|$$

and

$$G_m := (1-w_m)G_{m-1} + \lambda_m\varphi_m.$$

2). Denote

$$f_m := f - G_m.$$

The technique developed in this section implies the following theorem.

Theorem 3.3. *Theorems 1.2, 3.1 and Lemma 3.1 hold for the XGAFR with $\tau = \{1\}$.*

4. Some remarks

We begin with a remark on computational complexity of greedy algorithms. The main point of this paper is in proving that relaxation allows us to build greedy algorithms (see the WGAFR) that are computationally simpler than the WCGA and perform as well as the WCGA. We note that the WCGA and the WGAFR differ in the second step of the algorithm. However, the most computationally involved step of all greedy algorithms is the greedy step (the first step of the algorithm). One of the goals of this paper was to get rid of the assumption $f \in A_1(\mathcal{D})$ (as in the WRGA). All new relaxed greedy algorithms of the paper are applicable to (and converge for) any $f \in X$. We want to point out that the information $f \in A_1(\mathcal{D})$ allows us to simplify substantially the greedy step of the algorithm. It is remarked in [T3, Remark 2.2] that we can replace the first step (1.1) of the WCGA by the following search criterion

$$F_{f_{m-1}}(\varphi_m) \geq t_m\|f_{m-1}\|. \tag{4.1}$$

The requirement (4.1) is weaker than the requirement of the greedy step of the WCGA. However, Theorem 2.1 holds for this modification of the WCGA. The relation (4.1) is a threshold type inequality and can be checked easier than (1.1).

We now consider two algorithms with a different type of thresholding. These algorithms work for any $f \in X$. We begin with the Dual Greedy Algorithm with Relaxation and Thresholding (DGART).

DGART. We define $f_0 := f$ and $G_0 := 0$. Then for a given parameter $\delta \in (0, 1/2]$ we inductively define for $m \geq 1$

1). $\varphi_m \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \geq \delta. \tag{4.2}$$

If there is no $\varphi_m \in \mathcal{D}$ satisfying (4.2) then we stop.

2). Find w_m and λ_m such that

$$\|f-((1-w_m)G_{m-1}+\lambda_m\varphi_m)\| = \inf_{\lambda,w}\|f-((1-w)G_{m-1}+\lambda\varphi_m)\|$$

and define

$$G_m := (1-w_m)G_{m-1}+\lambda_m\varphi_m.$$

3). Denote

$$f_m := f-G_m.$$

If $\|f_m\|\le\delta\|f\|$ then we stop, otherwise we proceed to the $(m+1)$th iteration.

The following algorithm is a thresholding type modification of the WCGA. This modification can be applied to any $f\in X$.

Chebyshev Greedy Algorithm with Thresholding (CGAT). For a given parameter $\delta\in(0,1/2]$ we conduct instead of the greedy step (1.1) of the WCGA the following thresholding step: find $\varphi_m\in\mathcal{D}$ such that $F_{f_{m-1}}(\varphi_m)\ge\delta$. If such φ_m exists then we pick it and apply steps 2 and 3 of the WCGA. If such φ_m does not exist then we stop. We also stop, if $\|f_m\|\le\delta\|f\|$.

Theorem 4.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)\le\gamma u^q$, $1<q\le 2$. Take a number $\epsilon\ge 0$ and two elements f, f^ϵ from X such that*

$$\|f-f^\epsilon\|\le\epsilon,\quad f^\epsilon/A(\epsilon)\in A_1(\mathcal{D}),$$

with some number $A(\epsilon)>0$. Then the DGART (CGAT) will stop after $m\le C(\gamma)\delta^{-p}\ln(1/\delta)$, $p:=q/(q-1)$, iterations with

$$\|f_m\|\le\epsilon+\delta A(\epsilon).$$

Proof. We begin with the error bound. For both algorithms the DGART and the CGAT our stopping criterion guarantees that either $\|F_{f_m}\|_{\mathcal{D}}\le\delta$ or $\|f_m\|\le\delta\|f\|$. In the latter case the required bound follows from simple inequalities

$$\|f\|\le\epsilon+\|f^\epsilon\|\le\epsilon+A(\epsilon).$$

Thus, assume that $\|F_{f_m}\|_{\mathcal{D}}\le\delta$ holds. We will use the following well known lemma (see, for instance, [T3, Lemma 2.1]).

Lemma 4.1. *Let X be a uniformly smooth Banach space and L be a finite-dimensional subspace of X. For any $\psi\in X\setminus L$ denote by ψ_L the best approximant of ψ from L. Then we have*

$$F_{\psi-\psi_L}(\phi)=0$$

for any $\phi\in L$.

In the case of CGAT we apply Lemma 4.1 with $\psi=f$ and $L=\operatorname{span}(\varphi_1,\dots,\varphi_m)$ and obtain

$$\|f_m\|=F_{f_m}(f_m)=F_{f_m}(f)\le\epsilon+F_{f_m}(f^\epsilon)\le\epsilon+\|F_{f_m}\|_{\mathcal{D}}A(\epsilon)\le\epsilon+\delta A(\epsilon).$$

In the case of DGART we apply Lemma 4.1 with $\psi = f_{m-1}$ and $L = \operatorname{span}(G_{m-1}, \varphi_m)$ and get

$$\|f_m\| = F_{f_m}(f_m) = F_{f_m}(f_{m-1}) = F_{f_m}(f) \le \epsilon + F_{f_m}(f^\epsilon) \le \epsilon + \|F_{f_m}\|_{\mathcal{D}} A(\epsilon) \le \epsilon + \delta A(\epsilon).$$

This proves the required bound.

We now proceed to the bound of m. We prove the bound for both algorithms simultaneously. We note that for the DGART

$$\|f_k\| = \inf_{\lambda,w} \|f_{k-1} + wG_{k-1} - \lambda\varphi_k\| \le \inf_{\lambda\ge 0} \|f_{k-1} - \lambda\varphi_k\|.$$

We write for all $k \le m$, $\lambda \ge 0$

$$\|f_{k-1} - \lambda\varphi_k\| + \|f_{k-1} + \lambda\varphi_k\| \le 2\|f_{k-1}\|(1 + \rho(\lambda/\|f_{k-1}\|)). \tag{4.3}$$

Next,

$$\|f_{k-1} + \lambda\varphi_k\| \ge F_{f_{k-1}}(f_{k-1} + \lambda\varphi_k) \ge \|f_{k-1}\| + \lambda\delta. \tag{4.4}$$

Combining (4.3) with (4.4) we obtain

$$\|f_k\| \le \inf_{\lambda\ge 0} \|f_{k-1} - \lambda\varphi_k\| \le \inf_{\lambda\ge 0} \left(\|f_{k-1}\| - \lambda\delta + 2\|f_{k-1}\|\gamma(\lambda/\|f_{k-1}\|)^q\right).$$

Solving the equation $\delta x/2 = 2\gamma x^q$ we get $x_1 = (\delta/(4\gamma))^{1/(q-1)}$. Setting $\lambda := x_1\|f_{k-1}\|$ we obtain

$$\|f_k\| \le \|f_{k-1}\|(1 - \delta x_1/2) = \|f_{k-1}\|(1 - c(\gamma)\delta^p).$$

Thus,

$$\|f_k\| \le \|f\|(1 - c(\gamma)\delta^p)^k.$$

By the stopping condition $\|f_m\| \le \delta\|f\|$ we get that $m \le n$ where n is the smallest such that

$$(1 - c(\gamma)\delta^p)^n \le \delta.$$

This implies

$$m \le C(\gamma)\delta^{-p}\ln(1/\delta). \quad \Box$$

We now make a remark on the case of a Hilbert space. It is known and easy to check that for a Hilbert space one has

$$\rho(u) \le (1 + u^2)^{1/2} - 1 \le u^2/2.$$

Thus, a Hilbert space is a uniformly smooth Banach space with $q = 2$ and $\gamma = 1/2$. In this case one can find numerical values for the constants in the error bounds for the $\|f_m\|$. For instance, we get in Theorem 2.2 $C(2, 1/2) = 4$. We will show how the use of a special structure of Hilbert spaces allows us to get better numerical constants in the error bounds. We will consider the Weak Orthogonal Greedy Algorithm (WOGA) and the WGAFR. The WOGA is the WCGA for a Hilbert space.

Theorem 4.2. *Let H be a Hilbert space. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from H such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have for the WOGA and for the WGAFR

$$\|f_m\| \leq \epsilon + A(\epsilon)(1 + \sum_{k=1}^m t_k^2)^{-1/2}.$$

Proof. The proof is the same for both algorithms. We will carry it out only for the WOGA. In the case $\epsilon = 0$, $A(\epsilon) = 1$, Theorem 4.2 for the WOGA has been proved in [T2]. The proof of Theorem 4.2 repeats the arguments from [T2]. Assume $\|f_m\| \geq \epsilon$ (otherwise the statement is trivial). Then for all $k \leq m$ we have $\|f_k\| \geq \epsilon$. Next,

$$\|f_{k-1}\|^2 = \langle f_{k-1}, f_{k-1}\rangle = \langle f_{k-1}, f\rangle \leq \|f_{k-1}\|\epsilon + \langle f_{k-1}, f^\epsilon\rangle \leq \|f_{k-1}\|\epsilon + A(\epsilon) \sup_{g\in\mathcal{D}}\langle f_{k-1}, g\rangle.$$

Therefore,

$$\langle f_{k-1}, \varphi_k\rangle \geq t_k A(\epsilon)^{-1}(\|f_{k-1}\| - \epsilon)\|f_{k-1}\|,$$

and

$$\|f_k\|^2 \leq \|f_{k-1}\|^2 - \langle f_{k-1}, \varphi_k\rangle^2 \leq \|f_{k-1}\|^2 \left(1 - \frac{t_k^2(\|f_{k-1}\| - \epsilon)^2}{A(\epsilon)^2}\right). \tag{4.5}$$

Using the inequality

$$(\|f_k\| - \epsilon)/\|f_k\| \leq (\|f_{k-1}\| - \epsilon)/\|f_{k-1}\|$$

we get from (4.5)

$$(\|f_k\| - \epsilon)^2 \leq (\|f_{k-1}\| - \epsilon)^2 \left(1 - \frac{t_k^2(\|f_{k-1}\| - \epsilon)^2}{A(\epsilon)^2}\right). \tag{4.6}$$

Applying Lemma 3.1 from [T2] we obtain

$$(\|f_m\| - \epsilon)^2 \leq A(\epsilon)^2 \left(1 + \sum_{k=1}^m t_k^2\right)^{-1}. \quad \square$$

We conclude this section by a technical remark. An important part of the technique, presented in the paper, consists in deriving desired upper bounds from some recurrent inequalities. Typically, in order to avoid too much technical details, we prove the upper bounds for the rate of convergence under an assumption that modulus of smoothness has a power type. We will now demonstrate how one can handle a case of general modulus of smoothness.

Lemma 4.2. *Let $\rho(u)$ be modulus of smoothness of a uniformly smooth Banach space. Assume a sequence $a_0 \ge a_1 \ge a_2 \ge \dots$ of positive numbers satisfies the inequality*

$$a_k \le a_{k-1} \inf_\lambda (1 - \lambda t A^{-1} + 2\rho(B\lambda/a_{k-1})), \quad a_0 \le A, \tag{4.7}$$

with positive constants A, $B \ge 1$, $t \in (0,1]$. Then

$$a_m \le \frac{2AB}{tm\rho^{-1}(1/(2m))}. \tag{4.8}$$

Proof. We specify $\lambda := a_{k-1}s^{-1}(\theta t)/B$ with $\theta := \frac{a_{k-1}}{4AB}$. This implies

$$a_k \le a_{k-1}\left(1 - 2\frac{a_{k-1}t}{4AB}s^{-1}\left(\frac{a_{k-1}t}{4AB}\right)\right). \tag{4.9}$$

Setting $b_k := \frac{a_k t}{4AB}$ and $w(u) := us^{-1}(u)$ we get from (4.9)

$$b_k \le b_{k-1}(1 - 2w(b_{k-1})). \tag{4.10}$$

Using the property $w(\alpha u) \le \alpha w(u)$, $\alpha \in (0,1]$, we obtain from (4.10)

$$2w(b_k) \le 2w(b_{k-1})(1 - 2w(b_{k-1})).$$

It is easy to check that $2w(b_0) \le 1$. Applying Lemma 3.4 from [DT], we conclude

$$2w(b_m) \le 1/m.$$

It remains to note that $w^{-1}(x) = x/\rho^{-1}(x)$. $\square$

With a help of Lemma 4.2 one can prove the following version of Theorem 1.2 in the case of general modulus of smoothness.

Theorem 4.3. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Take a number $\epsilon \ge 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \le \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have for the WGAFR with $\tau = \{t\}$, $t \in (0,1]$,

$$\|f_m\| \le \max\left(2\epsilon, CA(\epsilon)(tm\rho^{-1}(1/(2m))^{-1}\right).$$

5. Discussion

We begin with brief discussion of the development of the theory of greedy approximation with regard to a general dictionary in order to show the place of results of this paper. The first greedy algorithm with regard to a general dictionary was studied in a Hilbert space. This algorithm is the Pure Greedy Algorithm (PGA) (see [FS], [H], [J1], [DT], [T4]). It coincides with the XGAR($\mathbf{0}$) (X-greedy algorithm), defined in the Introduction, in the case $X = H$ is a Hilbert space and $\mathbf{r} = \mathbf{0}$, which means there is no relaxation. Let us right away point out one good feature of the X-greedy algorithm. For an element $f \in X$ it provides an expansion into a series

$$f \sim \sum_{j=1}^{\infty} c_j(f) g_j(f), \quad g_j(f) \in \mathcal{D}, \quad c_j(f) > 0, \quad j = 1, 2, \dots \tag{5.1}$$

such that

$$G_m = \sum_{j=1}^{m} c_j(f) g_j(f).$$

The first steps in the theory of greedy approximation were devoted to the study of convergence of the expansion (5.1) and were done in a Hilbert space. P.J. Huber [H] proved convergence of the PGA in the weak topology and conjectured that the PGA converges in the strong sense (in the norm of H). L. Jones [J1] proved this conjecture. It is a fundamental result in the theory of greedy approximation that guarantees convergence of (5.1), obtained by the PGA, for any $f \in H$ and any dictionary $\mathcal{D}$. The reader can find results on greedy expansions in Banach spaces in [T5]. The next step was to understand efficiency of the PGA in terms of rate of convergence. There is some progress in this direction. However, the problem is still open. We formulate only two results in this direction (see [T4] for more details). It was proved in [DT] that for a general dictionary $\mathcal{D}$ the PGA provides for $f \in A_1(\mathcal{D})$ the estimate

$$\|f_m\| \le m^{-1/6}.$$

A lower estimate has been proved in [LiT]. It was shown that there exist a dictionary $\mathcal{D}$, a positive constant C, and an element $f \in A_1(\mathcal{D})$ such that for the PGA

$$\|f_m\| \ge C m^{-0.27}. \tag{5.2}$$

We note that even before the lower estimate (5.2) was proved, people began looking for other greedy algorithms that provide good rate of approximation of functions from $A_1(\mathcal{D})$. Two different ideas have been used at this step. The first idea was the idea of relaxation (see [J2], [B], [DT], [T2]). The corresponding algorithms (for example, the WRGA, defined in the Introduction) were designed for approximation of functions from $A_1(\mathcal{D})$. These algorithms do not provide an expansion into a series but they have other good features. It was established (see [J2], [B]) for the WRGA with $\tau = \{1\}$ in a Hilbert space that for $f \in A_1(\mathcal{D})$

$$\|f_m\| \le C m^{-1/2}.$$

Also, for the WRGA we always have $G_m \in A_1(\mathcal{D})$. The latter property, clearly, limits the applicability of the WRGA to the $A_1(\mathcal{D})$.

The second idea was the idea of building the best approximant from the span$(\varphi_1,\dots,\varphi_m)$ instead of the use of only one element φ_m for an update of the approximant. This idea was realized in the Weak Orthogonal Greedy Algorithm (see [DT], [T2]) in the case of a Hilbert space and in the Weak Chebyshev Greedy Algorithm (see [T3]) in the case of a Banach space.

The realization of both ideas resulted in construction of algorithms (WRGA and WCGA) that are good for approximation of functions from $A_1(\mathcal{D})$. Both algorithms do not provide expansions into series. The WCGA has the following advantage over the WRGA. It has been proved that the WCGA (under some assumptions on the weakness sequence τ) converges for each $f \in X$ in any uniformly smooth Banach space [T3]. Moreover, the behavior of the WCGA has been studied well enough to provide the upper estimates of approximation in terms of the intermediate approximation of f by an element f^ϵ and the $A_1(\mathcal{D})$ norm of f^ϵ (see Theorem 2.2 from the Introduction).

Recently, the following important question has been raised in [BCDD]. The authors pointed out that the Orthogonal Greedy Algorithm (OGA) has a defect comparing to the PGA: the OGA is computationally more complex than the PGA since each step of the OGA requires the evaluation of the orthogonal projection. The WRGA does not have this defect but it works only for elements from $A_1(\mathcal{D})$. The following remarkable result has been obtained in [BCDD]. The authors proved the following error estimate for a proper modification of the relaxed greedy algorithm: for $f \in H$ with f^ϵ such that $\|f-f^\epsilon\| \le \epsilon$, $f^\epsilon/A(\epsilon) \in A_1(\mathcal{D})$

$$\|f_m\|^2 \le \epsilon^2 + 4(A(\epsilon)^2 - \|f^\epsilon\|^2)m^{-1}. \tag{5.3}$$

They worked in a Hilbert space and used the following algorithm that is slightly different from our GAWR. At the first step of the algorithm they define $\varphi_m \in \mathcal{D}$ as the one that satisfies

$$\langle f-(1-r_m)G_{m-1},\varphi_m\rangle = \sup_{g\in\mathcal{D}}\langle f-(1-r_m)G_{m-1},g\rangle.$$

At the first step of the GAWR$(1,\mathbf{r})$ in the case of a Hilbert space we look for $\varphi_m \in \mathcal{D}$ that satisfies

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

The steps 2 and 3 in their algorithm are the same as in the GAWR. They proved the inequality (5.3) in the case $r_1 = 1$, $r_k = 2/k$, $k \ge 2$.

The above result (5.3) from [BCDD] was a motivation for the research reported in this paper. We considered different versions of the weak relaxed greedy algorithms. The WGAFR and XGAFR, studied in Section 3, are the most powerful ones out of the versions considered here. We proved convergence of the WGAFR in Theorem 3.1. This theorem is the same as the corresponding convergence result for the WCGA (see [T3, Theorem 2.1]). The results on the rate of convergence for the WGAFR and the WCGA are also the same (see Theorem 1.2 and Theorem 2.2). Thus, the WGAFR performs in the same way as the WCGA from the point of view of convergence and rate of convergence and outperforms the WCGA in terms of computational complexity.

In the WGAFR we are optimizing over two parameters w and λ at each step of the algorithm. In other words we are looking for the best approximation from a 2-dimensional linear subspace at each step. In the two other versions of the weak relaxed greedy algorithms (see GAWR and Wα-RGA), considered here, we approximate from a 1-dimensional linear subspace at each step of the algorithm. This makes computational complexity of these algorithms very close to that of the PGA. The analysis of these two versions turns out to be more complicated than the analysis of the WGAFR. Also, the results, obtained for these two versions, are not as general as in the case of the WGAFR. For instance, we present results on the GAWR only in the case $\tau = \{t\}$, when the weakness parameter t is the same for all steps.

We introduce a new norm generated by a given dictionary $\mathcal{D}$ in a Banach space X:

$$\|f\|_{A_1(\mathcal{D})} := \min\{a > 0 : \quad f/a \in A_1(\mathcal{D})\}.$$

If there is no a such that $f/a \in A_1(\mathcal{D})$ then we set $\|f\|_{A_1(\mathcal{D})} = \infty$. Denote

$$\mathcal{A}_1(\mathcal{D}) := \{f \in X : \|f\|_{A_1(\mathcal{D})} < \infty\}.$$

The results of this paper allow us to express the upper bounds of approximation by greedy algorithms in terms of K-functional. This type of bounds have been obtained in [BCDD] in the case of a Hilbert space. We remind a definition of the K-functional. Let two Banach spaces $Y \subset X$ with norms $\|\cdot\|_Y$ and $\|\cdot\|_X$ be given. Define for $f \in X$ the K-functional

$$K(f,u) := K(f,u,X,Y) := \inf_{y \in Y}(\|f-y\|_X + u\|y\|_Y).$$

Then Theorem 2.2 implies for the WCGA and X with $\rho(u) \le \gamma u^q$, $1 < q \le 2$, that

$$\|f_m^{c,\tau}\| \le C(\gamma,q)K(f,(1+\sum_{k=1}^m t_k^p)^{-1/p}, X, \mathcal{A}_1(\mathcal{D})).$$

Similarly, for the GAWR$(t,\mathbf{r})$ and the XGAR$(\mathbf{r})$, under conditions of Theorem 2.3 we obtain

$$\|f_m\| \le C(\gamma,q,t)K(f, m^{-1+1/q}, X, \mathcal{A}_1(\mathcal{D})).$$

Theorem 1.2 gives for the WGAFR

$$\|f_m\| \le C(\gamma,q)K(f,(1+\sum_{k=1}^m t_k^p)^{-1/p}, X, \mathcal{A}_1(\mathcal{D})).$$

Theorem 4.1 gives for the CGAT and for the DGART

$$\|f_m\| \le K(f,\delta,X,\mathcal{A}_1(\mathcal{D})).$$

Theorem 4.2 gives in the case of a Hilbert space for the WOGA and for the WGAFR

$$\|f_m\| \le K(f,(1+\sum_{k=1}^m t_k^2)^{-1/2}, H, \mathcal{A}_1(\mathcal{D})).$$

We note that in the case of OGA, i.e. WOGA with $\tau = \{1\}$, the following error bound has been obtained in [BCDD]

$$\|f_m\| \le K(f, 2m^{-1/2}, H, \mathcal{A}_1(\mathcal{D})).$$

References

[B] Andrew R. Barron, *Universal approximation bounds for superposition of n sigmoidal functions*, IEEE Transactions on Information Theory **39** (1993), 930– 945.

[BCDD] A. Barron, A. Cohen, W. Dahmen, and R. DeVore, *Approximation and learning by greedy algorithms*, Manuscript (2005), 1–27.

[DT] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in Computational Mathematics **5** (1996), 173–187.

[DGDS] M. Donahue, L. Gurvits, C. Darken, E. Sontag, *Rate of convex approximation in non-Hilbert spaces*, Constr. Approx. **13** (1997), 187–220.

[FS] J.H. Friedman and W. Stuetzle, *Projection pursuit regression*, J. Amer. Statist. Assoc. **76** (1981), 817–823.

[H] P.J. Huber, *Projection Pursuit*, The Annals of Statistics **13** (1985), 435–475.

[J1] L. Jones, *On a conjecture of Huber concerning the convergence of projection pursuit regression*, Annals of Stat. **15** (1987), 880–882.

[J2] L. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, The Annals of Statistics **20** (1992), 608–613.

[LT] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces I*, Springer-Verlag, Berlin, 1977.

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

[T1] V.N. Temlyakov, *Greedy Algorithms and m-term Approximation With Regard to Redundant Dictionaries*, J. Approx. Theory **98** (1999), 117–145.

[T2] V.N. Temlyakov, *Weak greedy algorithms*, Advances in Comp. Math. **12** (2000), 213–227.

[T3] V.N. Temlyakov, *Greedy algorithms in Banach spaces*, Advances in Comp. Math. **14** (2001), 277–292.

[T4] V.N. Temlyakov, *Nonlinear Methods of Approximation*, Found. Comput. Math. **3** (2003), 33–107.

[T5] V.N. Temlyakov, *Greedy expansions in Banach spaces*, IMI Preprints Series **6** (2003), 1–21.

[T6] V.N. Temlyakov, *Greedy-Type Approximation in Banach Spaces and Applications*, Constr. Approx. **21** (2005), 257–292.