INDUSTRIAL
MATHEMATICS
INSTITUTE

2000:10

Greedy algorithms in Banach spaces

V.N. Temlyakov

IMI

Preprint Series

Department of Mathematics
University of South Carolina

Greedy Algorithms in Banach Spaces[1]

V.N.Temlyakov

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

Abstract. We study efficiency of approximation and convergence of two greedy type algorithms in uniformly smooth Banach spaces. The Weak Chebyshev Greedy Algorithm (WCGA) is defined for an arbitrary dictionary $\mathcal{D}$ and provides nonlinear m-term approximation with regard to $\mathcal{D}$. This algorithm is defined inductively with the m-th step consisting of two basic substeps: 1) selection of an m-th element φ^c_m from $\mathcal{D}$ and 2) constructing an m-term approximant G^c_m. We include the name of Chebyshev in the name of this algorithm because at the substep 2) the approximant G^c_m is chosen as the best approximant from $\operatorname{span}(\varphi^c_1, \dots, \varphi^c_m)$. The term Weak Greedy Algorithm indicates that at each substep 1) we choose φ^c_m as an element of $\mathcal{D}$ that satisfies some condition which is "t_m-times weaker" than the condition for φ^c_m to be optimal ($t_m = 1$). We got error estimates for Banach spaces with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q < \infty$. We proved that for any f from the closure of the convex hull of $\mathcal{D} \cup 0$ the error of m-term approximation by WCGA is of order $(1 + t^p_1 + \dots + t^p_m)^{-1/p}$, $1/p + 1/q = 1$. Similar results are obtained for Weak Relaxed Greedy Algorithm (WRGA) and its modification. In this case an approximant G^r_m is a convex linear combination of $0, \varphi^r_1, \dots, \varphi^r_m$. We also proved some convergence results for WCGA and WRGA.

1. Introduction

The core problem of approximation continues to be the development of efficient methods for replacing general functions by simpler functions. Some methods were invented long ago (Fourier sums, Taylor polynomials, best approximation by trigonometric or algebraic polynomials etc.). More recently however, driven by several numerical applications, the directions of approximation theory have moved toward nonlinear approximation. This includes the comparatively new subject of nonlinear m-term approximation. It has found applications in numerical solution of integral equations, image compression, design of neural networks, and so on.

The purpose of this paper is to continue investigations of nonlinear m-term approximation. We concentrate here on studying m-term approximation with regard to redundant dictionaries in Banach spaces. This paper is based on a combination of ideas and methods developed for Banach spaces in [DDGS] with the approach used in [T] in the case of Hilbert spaces. The papers [DDGS] and [T] contain detailed historical remarks and we refer the reader to those papers. In this paper we will mention only those results which are directly connected with the results presented here.

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

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

[1]This research was supported by the National Science Foundation Grant DMS 9970326 and by ONR Grant N00014-91-J1343

and $\overline{\operatorname{span}}\mathcal{D} = X$.

We will study in this paper two types of greedy algorithms with regard to $\mathcal{D}$. For an element $f \in X$ we denote by F_f a peack 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 positive numbers $t_k \le 1$, $k = 1, \dots$. We define first the Weak Chebyshev Greedy Algorithm (WCGA) that is a generalization for Banach spaces of Weak Orthogonal Greedy Algorithm defined and studied in [T] (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

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

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 [T] in the case of 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

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

The term "weak" in both definitions means that at the step 1). we do not shoot for the optimal element of the dictionary which realizes the corresponding sup but are satisfied with weaker property than being optimal. The obvious reason for this is that we don't know in general that the optimal one exists. Another, practical reason is that the weaker the assumption the easier to satisfy it and therefore easier to realize in practice. The Weak Relaxed Greedy Algorithm provides incremental approximants discussed in [DDGS]. In [DDGS] they also impose weaker assumptions

(ϵ-greedy) on an element of the dictionary than being optimal. For instance, for a given sequence $\{\epsilon_n\}_{n=1}^{\infty}$, $\epsilon_n > 0$, $n = 1, 2, \dots$, they take $0 \le \alpha_m \le 1$ and $g_m \in \mathcal{D}$ satisfying

$$\|f - ((1-\alpha_m)G_{m-1} + \alpha_m g_m)\| \le \inf_{0\le\alpha\le 1, g\in\mathcal{D}} \|f - ((1-\alpha)G_{m-1} + \alpha g)\| + \epsilon_m$$

instead of trying to find optimal ones. Their approach is different from ours.

We study in this paper the questions of convergence and the rate of convergence for the two above defined methods of approximation. It is clear that in the case of WRGA the assumption that f belongs to the closure of convex hull of $\mathcal{D}$ is natural. We denote the closure of convex hull of $\mathcal{D}$ by $\mathcal{A}_1(\mathcal{D})$. It has been proven in [T] that in the case of Hilbert space both algorithms WCGA and WRGA give the approximation error for the class $\mathcal{A}_1(\mathcal{D})$ of the order

$$(1 + \sum_{k=1}^{m} t_k^2)^{-1/2}.$$

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

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

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 a convex function on $(0,\infty)$ satisfying the inequalities

$$\max(0, u-1) \le \rho(u) \le u. \tag{1.1}$$

It has been established in [DDGS] that the approximation error of an algorithm analogous to our WRGA with $t_k = 1$, $k = 1, 2, \dots$, for the class $\mathcal{A}_1(\mathcal{D})$ can be expressed in terms of modulus of smoothness of Banach space. Namely, if modulus of smoothness ρ of X satisfies the inequality $\rho(u) \le \gamma u^q$, $q > 1$, then the error is of $O(m^{1/q-1})$. We prove here that both algorithms WCGA and WRGA provide approximation for the class $\mathcal{A}_1(\mathcal{D})$ in a Banach space X with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q < \infty$, of order

$$(1 + \sum_{k=1}^{m} t_k^p)^{-1/p}, \quad p := \frac{q}{q-1}. \tag{1.2}$$

We prove also that both algorithms WCGA and WRGA converge for $f \in \mathcal{A}_1(\mathcal{D})$ if

$$\lim_{n\to\infty} n\rho(\frac{A}{nt_n}) = 0,$$

for any positive A.

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

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

In Section 4 we study a modification WRGA(2) of WRGA. The basic goal of this modification is to get rid of solving the optimization problem at the step 2). It is known from [DDGS], [DT], [T] that in many cases this can be done. In our modification WRGA(2) we use at the step 2) the following numbers

$$\beta_m := \frac{2t_m^{\frac{1}{q-1}}}{1 + t_1^p + \dots + t_{m-1}^p}, \quad p := \frac{q}{q-1}, \quad m > N,$$

instead of λ_m and prove that this does not effect the error estimate (1.2).

2. Convergence and rate of approximation of WCGA

We prove here the following two theorems.

Theorem 2.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Assume that a sequence $\tau := \{t_k\}_{k=1}^\infty$ satisfies the conditions $1 \ge t_1 \ge t_2 \ge \dots$ and*

$$\lim_{n \to \infty} n\rho(\frac{A}{nt_n}) = 0,$$

for any positive A. Then for any $f \in \mathcal{A}_1(\mathcal{D})$ we have

$$\lim_{m \to \infty} \|f_m^{c,\tau}\| = 0.$$

Theorem 2.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q < \infty$. Then for a sequence $\tau := \{t_k\}_{k=1}^\infty$, $t_k \le 1$, $k = 1, 2, \dots,$ we have for any $f \in \mathcal{A}_1(\mathcal{D})$ that*

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

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

We will use the following two simple and well-known lemmas in the proof of the above two theorems.

Lemma 2.1. *Let X be a uniformly smooth Banach space and L be a finite-dimensional subspace of X. For any $f \in X \setminus L$ denote by f_L the best approximant of f from L. Then we have*

$$F_{f-f_L}(\phi) = 0$$

for any $\phi \in L$.

Proof. Let us assume the contrary: there is a $\phi \in L$ such that $\|\phi\| = 1$ and

$$F_{f-f_L}(\phi) = \beta > 0.$$

For any λ we have from the definition of $\rho(u)$ that

$$\|f - f_L - \lambda\phi\| + \|f - f_L + \lambda\phi\| \leq 2\|f - f_L\|(1 + \rho(\frac{\lambda}{\|f - f_L\|})). \tag{2.1}$$

Next

$$\|f - f_L + \lambda\phi\| \geq F_{f-f_L}(f - f_L + \lambda\phi) = \|f - f_L\| + \lambda\beta. \tag{2.2}$$

Combining (2.1) and (2.2) we get

$$\|f - f_L - \lambda\phi\| \leq \|f - f_L\|(1 - \frac{\lambda\beta}{\|f - f_L\|} + 2\rho(\frac{\lambda}{\|f - f_L\|})). \tag{2.3}$$

Taking into account that $\rho(u) = o(u)$ we find $\lambda' > 0$ such that

$$(1 - \frac{\lambda'\beta}{\|f - f_L\|} + 2\rho(\frac{\lambda'}{\|f - f_L\|})) < 1.$$

Then (2.3) gives

$$\|f - f_L - \lambda'\phi\| < \|f - f_L\|$$

what contradicts the assumption that $f_L \in L$ is the best approximant of f.

Lemma 2.2. *For any bounded linear functional F and any dictionary $\mathcal{D}$ we have*

$$\sup_{g \in \mathcal{D}} F(g) = \sup_{f \in \mathcal{A}_1(\mathcal{D})} F(f).$$

Proof. The inequality

$$\sup_{g \in \mathcal{D}} F(g) \leq \sup_{f \in \mathcal{A}_1(\mathcal{D})} F(f)$$

is obvious. We prove the opposite inequality. Take any $f \in \mathcal{A}_1(\mathcal{D})$. Then for any $\epsilon > 0$ there exist $g_1^\epsilon, \dots, g_N^\epsilon \in \mathcal{D}$ and numbers $a_1^\epsilon, \dots, a_N^\epsilon$ such that $a_i^\epsilon > 0$, $a_1^\epsilon + \dots + a_N^\epsilon \leq 1$ and

$$\|f - \sum_{i=1}^N a_i^\epsilon g_i^\epsilon\| \leq \epsilon.$$

Thus

$$F(f) \leq \|F\|\epsilon + F(\sum_{i=1}^N a_i^\epsilon g_i^\epsilon) \leq \epsilon\|F\| + \sup_{g \in \mathcal{D}} F(g)$$

what proves Lemma 2.2.

We will also need one more lemma.

Lemma 2.3. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Then for any $f \in \mathcal{A}_1(\mathcal{D})$ we have*

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

Proof. We have for any λ

$$\|f^c_{m-1}-\lambda\varphi^c_{m-1}\|+\|f^c_{m-1}+\lambda\varphi^c_{m-1}\|\le 2\|f^c_{m-1}\|(1+\rho(\frac{\lambda}{\|f^c_{m-1}\|})) \tag{2.4}$$

and by 1) from the definition of WCGA and Lemma 2.2 we get

$$F_{f^c_{m-1}}(\varphi^c_m)\ge t_m\sup_{g\in\mathcal{D}}F_{f^c_{m-1}}(g)= \tag{2.5}$$

$$t_m\sup_{\phi\in\mathcal{A}_1(\mathcal{D})}F_{f^c_{m-1}}(\phi)\ge t_mF_{f^c_{m-1}}(f).$$

By Lemma 2.1 we obtain

$$F_{f^c_{m-1}}(f)=F_{f^c_{m-1}}(f^c_{m-1})=\|f^c_{m-1}\|.$$

Thus similarly to (2.3) we get from (2.4)

$$\|f^c_m\|\le\inf_\lambda\|f^c_{m-1}-\lambda\varphi^c_m\|\le\|f^c_{m-1}\|\inf_\lambda(1-\lambda t_m+2\rho(\frac{\lambda}{\|f^c_{m-1}\|})) \tag{2.6}$$

what proves the lemma.

Proof of Theorem 2.1. Denote $\epsilon(u):=\rho(u)/u$. Then we have

$$\lim_{u\to 0^+}\epsilon(u)=0.$$

It is easy to derive from the definition of $\rho(u)$ that $\rho(u)$ is a convex function. This implies that $\epsilon(u)$ is an increasing continuous function on $(0,\infty)$. The relation (1.1) implies that $\rho(2)\ge 1$ or $\epsilon(2)\ge 1/2$. We define inductively a sequence $2\ge\delta_1\ge\delta_2\ge\dots\ge 0$. Assume $\delta_1,\dots,\delta_{m-1}$ have been chosen. We choose $\delta_m\le\delta_{m-1}$ such that the $\lambda_m:=\delta_m\|f^c_{m-1}\|$ satisfies the equation (see Lemma 2.3)

$$2\rho(\frac{\lambda_m}{\|f^c_{m-1}\|})=\frac{1}{2}\lambda_mt_m$$

what is equivalent to

$$\epsilon(\delta_m)=\frac{1}{4}t_m\|f^c_{m-1}\|.$$

Then (2.6) implies that

$$\|f^c_m\|\le\|f^c_{m-1}\|(1-\frac{1}{2}\delta_mt_m\|f^c_{m-1}\|). \tag{2.7}$$

Denote

$$a_{m-1}:=\frac{1}{2}\delta_mt_m\|f^c_{m-1}\|.$$

Then using that $\delta_{m+1}\le\delta_m$ and $t_{m+1}\le t_m$ we get from (2.7)

$$a_m\le a_{m-1}(1-a_{m-1}).$$

Thus by Lemma 3.4 from [DT] we get

$$a_{m-1} \le \frac{1}{m}$$

and

$$\frac{1}{2}\delta_m t_m \|f^c_{m-1}\| \le \frac{1}{m},$$

$$\delta_m \le \frac{2}{m t_m \|f^c_{m-1}\|}.$$

Assume the contrary that

$$\lim_{m\to\infty} \|f^c_{m-1}\| = \alpha > 0.$$

Then

$$\delta_m \le \frac{2}{m t_m \alpha}.$$

From the definition of δ_m and the monotonicity of $\epsilon(u)$ we get

$$\frac{1}{4} t_m \|f^c_{m-1}\| = \epsilon(\delta_m) \le \epsilon(\frac{2}{m t_m \alpha})$$

and

$$\|f^c_{m-1}\| \le \frac{4}{t_m}\epsilon(\frac{2}{m t_m \alpha}) \to 0 \quad \text{as} \quad m \to \infty.$$

Theorem 2.1 is proved now.

Proof of Theorem 2.2. By Lemma 2.3 we have for $f \in \mathcal{A}_1(\mathcal{D})$ that

$$\|f^{c,\tau}_m\| \le \|f^{c,\tau}_{m-1}\| \inf_\lambda (1 - \lambda t_m + 2\gamma(\frac{\lambda}{\|f^{c,\tau}_{m-1}\|})^q). \tag{2.8}$$

Choose λ_m from the equation

$$\frac{1}{2}\lambda t_m = 2\gamma(\frac{\lambda}{\|f^{c,\tau}_{m-1}\|})^q$$

what implies that

$$\lambda_m = \|f^{c,\tau}_{m-1}\|^{\frac{q}{q-1}} (4\gamma)^{-\frac{1}{q-1}} t_m^{\frac{1}{q-1}}.$$

Denote

$$A_q := 2(4\gamma)^{\frac{1}{q-1}}.$$

Using the notation $p := \frac{q}{q-1}$ we get from (2.8)

$$\|f^c_m\| \le \|f^c_{m-1}\|(1 - \frac{1}{2}\lambda_m t_m) = \|f^c_{m-1}\|(1 - t^p_m \|f^c_{m-1}\|^p / A_q).$$

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

$$\|f^c_m\|^p \le \|f^c_{m-1}\|^p (1 - t^p_m \|f^c_{m-1}\|^p / A_q).$$

By Lemma 3.1 from [T] using the estimate $\|f\|^p \le 1 < A_q$ we get

$$\|f^c_m\|^p \le A_q (1 + \sum_{n=1}^m t^p_n)^{-1}$$

what implies

$$\|f^c_m\| \le C(q)(1 + \sum_{n=1}^m t^p_n)^{-1/p}.$$

Theorem 2.2 is proved now.

3. Convergence and rate of approximation of WRGA

We prove here the analogs of Theorems 2.1 and 2.2 for the Weak Relaxed Greedy Algorithm.

Theorem 3.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Assume that a sequence $\tau := \{t_k\}_{k=1}^\infty$ satisfies the conditions $1 \ge t_1 \ge t_2 \ge \dots$ and*

$$\lim_{n\to\infty} n\rho(\frac{A}{nt_n}) = 0,$$

for any positive A. Then for any $f \in \mathcal{A}_1(\mathcal{D})$ we have

$$\lim_{m\to\infty} \|f_m^{r,\tau}\| = 0.$$

Theorem 3.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q < \infty$. Then for a sequence $\tau := \{t_k\}_{k=1}^\infty$, $t_k \le 1$, $k = 1, 2, \dots,$ we have for any $f \in \mathcal{A}_1(\mathcal{D})$ that*

$$\|f_m^{r,\tau}\| \le C_1(q,\gamma)(1 + \sum_{k=1}^m t_k^p)^{-1/p}, \quad p := \frac{q}{q-1},$$

with a constant $C_1(q,\gamma)$ which may depend on q and γ.

Proof of Theorems 3.1 and 3.2. This proof is similar to the proof of Theorems 2.1 and 2.2. Instead of Lemma 2.3 we use the following lemma here.

Lemma 3.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Then for any $f \in \mathcal{A}_1(\mathcal{D})$ we have*

$$\|f_m^{r,\tau}\| \le \|f_{m-1}^{r,\tau}\| \inf_{0\le\lambda\le1}(1 - \lambda t_m + 2\rho(\frac{2\lambda}{\|f_{m-1}^{r,\tau}\|})), \quad m = 1, 2, \dots.$$

Proof. We have

$$f_m^r := f - ((1-\lambda_m)G_{m-1}^r + \lambda_m\varphi_m^r) = f_{m-1}^r - \lambda_m(\varphi_m^r - G_{m-1}^r)$$

and

$$\|f_m^r\| = \inf_{0\le\lambda\le1} \|f_{m-1}^r - \lambda(\varphi_m^r - G_{m-1}^r)\|.$$

Similarly to (2.4) we have for any λ

$$\|f_{m-1}^r - \lambda(\varphi_m^r - G_{m-1}^r)\| + \|f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)\| \le \tag{3.1}$$

$$2\|f_{m-1}^r\|(1 + \rho(\frac{\lambda\|\varphi_m^r - G_{m-1}^r\|}{\|f_{m-1}^r\|})).$$

Next we get for $\lambda \ge 0$

$$\|f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)\| \ge F_{f_{m-1}^r}(f_{m-1}^r + \lambda(\varphi_m^r - G_{m-1}^r)) =$$

$$\|f^r_{m-1}\| + \lambda F_{f^r_{m-1}}(\varphi^r_m - G^r_{m-1}) \geq \|f^r_{m-1}\| + \lambda t_m \sup_{g\in\mathcal{D}} F_{f^r_{m-1}}(g - G^r_{m-1}).$$

Using Lemma 2.2 we continue

$$= \|f^r_{m-1}\| + \lambda t_m \sup_{\phi\in\mathcal{A}_1(\mathcal{D})} F_{f^r_{m-1}}(\phi - G^r_{m-1}) \geq \|f^r_{m-1}\| + \lambda t_m \|f^r_{m-1}\|.$$

Using the trivial estimate $\|\varphi^r_m - G^r_{m-1}\| \leq 2$ we obtain from (3.1)

$$\|f^r_{m-1} - \lambda(\varphi^r_m - G^r_{m-1})\| \leq \|f^r_{m-1}\|(1 - \lambda t_m + 2\rho(\frac{2\lambda}{\|f^r_{m-1}\|})), \tag{3.2}$$

what proves Lemma 3.1.

The remaining part of the proof uses the inequality (3.2) in the same way as the relation (2.6) has been used in the proof of Theorems 2.1 and 2.2. The only additional difficulty here is that we are optimizing over $0 \leq \lambda \leq 1$. However, it is easy to check that the corresponding λ_m chosen in the same way as above always satisfies the restriction $0 \leq \lambda_m \leq 1$. In the proof of Theorem 3.1 we choose λ_m satisfying the equation

$$2\rho(\frac{2\lambda_m}{\|f^r_{m-1}\|}) = \frac{1}{2}\lambda_m t_m,$$

and in the proof of Theorem 3.2 from the equation

$$\frac{1}{2}\lambda t_m = 2\gamma(2\lambda)^q \|f^r_{m-1}\|^{-q}.$$

4. Rate of convergence of a modified WRGA

We consider a modification of WRGA in this section. This modification is motivated by results from [DDGS] (see also [DT]) and from [T]. It was observed in [DDGS] that one can replace λ_m in the definition of Relaxed Greedy Algorithm by $1/m$ without loss of order of approximation for $\mathcal{A}_1(\mathcal{D})$. In [T] this idea was used in the case of Weak Relaxed Greedy Algorithm in a Hilbert space. We consider here the following modification of WRGA which we call WRGA of type 2 and denote WRGA(2). The WRGA defined in Section 1 will be also called WRGA of type 1.

For a given $\tau = \{t_k\}_{k=1}^\infty$ and $1 < q \leq 2$ let N be such that

$$1 + t_1^p + \cdots + t_{N-1}^p < 2 \leq 1 + t_1^p + \cdots + t_N^p, \quad p := \frac{q}{q-1}.$$

Then for $m \leq N$ we define WRGA(2) as WRGA :

$$G^{rq}_m := G^r_m, \quad f^{rq} := f^r_m.$$

All notations for WRGA(2) are those for WRGA with r replaced by rq. For $m > N$ the definition of WRGA(2) differs from the definition of WRGA only at the step 2). We define

$$G^{rq}_m := G^{rq,\tau}_m := (1 - \beta_m)G^{rq}_{m-1} + \beta_m \varphi^{rq}_m$$

with

$$\beta_m := \frac{2t_m^{\frac{1}{q-1}}}{1 + \sum_{k=1}^{m-1} t_k^p}, \quad p := \frac{q}{q-1}.$$

The assumption on N guarantees that $0 \leq \beta_m \leq 1$, $m > N$. We prove the following theorem.

Theorem 4.1. *Let X be a uniformly smooth Banach space with the 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 $f \in \mathcal{A}_1(\mathcal{D})$ that*

$$\|f_m^{rq,\tau}\| \le C_2(q,\gamma)(1 + \sum_{k=1}^m t_k^p)^{-1/p}, \quad p := \frac{q}{q-1},$$

with a constant $C_2(q,\gamma)$ which may depend on q and γ.

Proof. Denote

$$T_0 := 1, \quad T_n := (1 + \sum_{k=1}^n t_k^p)^{1/p}.$$

Then obviously $T_n \le 2T_{n-1}$ for all n. We want to prove that

$$\|f_m^{rq,\tau}\| \le C T_m^{-1}, \quad m = 1, \dots,$$

with a constant C independent of m. The proof goes by induction. For convenience let us drop rq,τ in the notations. Theorem 3.2 gives the required estimate with $C \ge C_1(q,\gamma)$ for $m = 1, \dots, N$. We prove this estimate for $m > N$. Assume

$$\|f_{m-1}\| \le C_1 T_{m-1}^{-1}$$

with some constant $C_1 < C$ which will be specified later. Then

$$\|f_{m-1} - \beta_m(\varphi_m - G_{m-1})\| \le C_1 T_{m-1}^{-1} + 2\beta_m = \tag{4.1}$$

$$T_{m-1}^{-1}(C_1 + 2t_m^{\frac{1}{q-1}} T_{m-1}^{1-p}) \le (C_1 + 2)T_{m-1}^{-1} \le 2(C_1+2)T_m^{-1}.$$

Thus if C_1 satisfies the inequality

$$2(C_1 + 2) \le C \tag{4.2}$$

then we get the required estimate for $\|f_m\|$ in this case. Assume now that

$$C_1 T_{m-1}^{-1} \le \|f_{m-1}\| \le C T_{m-1}^{-1}. \tag{4.3}$$

By (3.2) with $\rho(u) \le \gamma u^q$ we get

$$\|f_m\| \le \|f_{m-1}\|(1 - \beta_m t_m + C_2 \beta_m^q \|f_{m-1}\|^{-q}) \tag{4.4}$$

with $C_2 := 2^{1+q}\gamma$. Assume that C_1 satisfies the inequality

$$2^{q-1} \le C_2^{-1} C_1^q. \tag{4.5}$$

Then using (4.3) we see that

$$\beta_m t_m \ge C_2 \beta_m^q \|f_{m-1}\|^{-q}.$$

This implies that

$$\mu_m := 1 - \beta_m t_m + C_2 \beta_m^q \|f_{m-1}\|^{-q} \le 1$$

and

$$\mu_m^q \le \mu_m. \tag{4.6}$$

Raising (4.4) into the power q and using (4.6) we obtain

$$\|f_m\|^q \le \|f_{m-1}\|^q(1-\beta_m t_m+C_2\beta_m^q\|f_{m-1}\|^{-q}) = \|f_{m-1}\|^q(1-\beta_m t_m)+C_2\beta_m^q. \tag{4.7}$$

Using (4.3) we get from here that

$$\|f_m\|^q T_m^q \le C^q(T_m/T_{m-1})^q(1-2t_m^pT_{m-1}^{-p})+C_2 2^q t_m^p T_{m-1}^{-pq}.$$

Next, for $1<q\le 2$ we have

$$(T_m/T_{m-1})^q = (1+t_m^pT_{m-1}^{-p})^{q/p} \le 1+t_m^pT_{m-1}^{-p}$$

and

$$(T_m/T_{m-1})^q(1-2t_m^pT_{m-1}^{-p}) \le 1-t_m^pT_{m-1}^{-p}.$$

Therefore,

$$\|f_m\|^qT_m^q \le C^q(1-t_m^pT_{m-1}^{-p}(1-C_2 2^qC^{-q})) \le C^q \tag{4.8}$$

if

$$C^q \ge C_2 2^q. \tag{4.9}$$

Now we specify the constants C and C_1 as

$$C_1 := 4\gamma^{1/q}$$

and

$$C := \max(4+8\gamma^{1/q}, C_1(q,\gamma)). \tag{4.10}$$

It is easy to check that these C_1 and C satisfy (4.2), (4.5), (4.9) and with the C from (4.10) the inequality (4.8) gives

$$\|f_m\|^qT_m^q \le C^q$$

what comletes the proof of Theorem 4.1.

Remark. Theorem 4.1 covers the case $1<q\le 2$. We can modify the definition of WRGA(2) to extend Theorem 4.1 to the case $2\le q<\infty$. Denote in this case $s:=q/p$ and choose N such that

$$1+t_1^p+\dots+t_{N-1}^p < 2^s \le 1+t_1^p+\dots+t_N^p, \quad p:=\frac{q}{q-1},$$

and

$$\beta_m := \frac{2^s t_m^{\frac{1}{q-1}}}{1+\sum_{k=1}^{m-1}t_k^p}, \quad p:=\frac{q}{q-1}.$$

Then using the inequalties

$$(1+x)^s \le 1+(2^s-1)x, \quad 0\le x\le 1, \quad s\ge 1,$$
$$(1+x)^s(1-2^sx) \le 1-x, \quad 0\le x\le 2^{-s},$$

in the estimating the right hand side of (4.7) we get the required estimate with some C.

References

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

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

[T] V.N. Temlyakov, *Weak Greedy Algorithms*, Advances in computational Mathematics (to appear).