Industrial Mathematics Institute

2004:27

On greedy algorithms with restricted depth search

V.N. Temlyakov

IMI

Preprint Series

Department of Mathematics
University of South Carolina

On Greedy Algorithms with restricted depth search[1]

V.N.Temlyakov

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

Abstract. We continue to study efficiency of approximation and convergence of greedy type algorithms in uniformly smooth Banach spaces. This paper is a development of two recent papers [T1] and [T5] in the direction of making practical algorithms out of theoretical approximation methods. The Weak Chebyshev Greedy Algorithm (WCGA) has been introduced and studied in [T1]. The WCGA is a general approximation method that works well in an arbitrary uniformly smooth Banach space X for any dictionary $\mathcal{D}$. It is an inductive procedure with each step of implementation consisting of several substeps. We describe the first substep of a particular case of the WCGA. Let $t \in (0,1]$. Then at the first substep of the mth step we are looking for an element φ_m from a given symmetric dictionary $\mathcal{D}$ satisfying

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

where f_{m-1} is a residual after $(m-1)$th step and $F_{f_{m-1}}$ is a norming functional of f_{m-1}. It is a greedy step of the WCGA. It is clear that in the case of infinite dictionary $\mathcal{D}$ there is no direct computationally feasible way of evaluating $\sup_{g\in\mathcal{D}} F_{f_{m-1}}(g)$. This is the main issue that we address in the paper. We consider countable dictionaries $\mathcal{D} = \{\pm\psi_j\}_{j=1}^{\infty}$ and replace (1) by

$$F_{f_{m-1}}(\varphi_m) \geq t \sup_{1\leq j\leq N_m} |F_{f_{m-1}}(\psi_j)|, \quad \varphi_m \in \{\pm\psi_j\}_{j=1}^{N_m}.$$

The retriction $j \leq N_m$ is known in the literature ([Do]) as the depth search condition. We prove convergence and rate of convergence results for such a modification of the WCGA.

1. Introduction

In this paper we discuss approximation by linear combinations of elements that are taken from a redundant (overcomplete) system of elements. We begin with a brief discussion of the question: why do we need redundant systems? Answering this question we first of all mention three classical redundant systems that are used in different areas of mathematics. Perhaps the first example of m-term approximation with regard to redundant dictionary was considered by E. Schmidt in 1907 [S] who considered the approximation of functions $f(x,y)$ of two variables by bilinear forms

$$\sum_{i=1}^{m} u_i(x)v_i(y)$$

[1] This research was supported by the National Science Foundation Grant DMS 0200187

in $L_2([0,1]^2)$. This problem is closely connected with properties of the integral operator

$$J_f(g) := \int_0^1 f(x,y)g(y)dy$$

with kernel $f(x,y)$. E. Schmidt [S] gave an expansion (known as the Schmidt expansion)

$$f(x,y) = \sum_{j=1}^{\infty} s_j(J_f)\phi_j(x)\psi_j(y)$$

where $\{s_j(J_f)\}$ is a nonincreasing sequence of singular numbers of J_f, i.e. $s_j(J_f) := \lambda_j(J_f^* J_f)^{1/2}$, $\{\lambda_j(A)\}$ is a sequence of eigenvalues of an operator A, J_f^* is the adjoint operator to J_f. The two sequences $\{\phi_j(x)\}$ and $\{\psi_j(y)\}$ form orthonormal sequences of eigenfunctions of the operators $J_f J_f^*$ and $J_f^* J_f$ respectively. He also proved that

$$\|f(x,y) - \sum_{j=1}^{m} s_j(J_f)\phi_j(x)\psi_j(y)\|_{L_2} = \inf_{u_j,v_j \in L_2, \quad j=1,\dots,m} \|f(x,y) - \sum_{j=1}^{m} u_j(x)v_j(y)\|_{L_2}.$$

Another example which is well known in statistics is the projection pursuit regression problem. We formulate the related setting in the function theory language. The problem is to approximate in $L_2(\Omega)$, $\Omega \subset \mathbb{R}^d$ is a bounded domain, a given function $f \in L_2(\Omega)$ by a sum of ridge functions, i.e. by

$$\sum_{j=1}^{m} r_j(\omega_j \cdot x), \quad x, \omega_j \in \mathbb{R}^d, \quad j = 1, \dots, m,$$

where r_j, $j = 1, \dots, m$, are univariate functions.

The third example is from signal processing. In signal processing the most popular means of approximation are wavelets and the system of Gabor functions $\{g_{a,b}(x-c), g_{a,b}(x) := e^{iax}e^{-bx^2}, \quad a, c \in \mathbb{R}, b \in \mathbb{R}_+\}$. The Gabor system gives more flexibility in constructing an approximant but it is a redundant (not minimal) system. It also seems natural (see discussion in [Do]) to use redundant systems in modeling analyzing elements for the visual system.

Thus, in order to address the contemporary needs of approximation theory and computational mathematics a very general model of approximation with regard to a redundant system (dictionary) has been considered in many recent papers. We refer the reader for a servey of some of these results to [D], [T4]. As such a model we choose a Banach space X with elements as target functions and an arbitrary system $\mathcal{D}$ of elements of this space such that $\overline{\text{span}}\mathcal{D} = X$ as an approximating system. We would like to have an algorithm of constructing m-term approximants that adds at each step only one new element from $\mathcal{D}$ and keeps elements of $\mathcal{D}$ obtained at the previous steps. This requirement is an analog of *on-line* computation property that is very desirable in practical algorithms. Clearly, we are looking for good algorithms which at a minimum converge for each target function. It is not obvious that such an algorithm exists in a setting at the above level of generality (X, $\mathcal{D}$ are arbitrary). It turned out that there is one fundamental principal that allows us to build good algorithms both for arbitrary redundant systems and for very simple well structured bases like the Haar basis. This principal is the use of a greedy step in searching for a new element to be added to a given m-term approximant. The common feature of all algorithms of m-term approximation discussed in this paper is the presence of a greedy step. By a greedy step in choosing an mth element $g_m(f) \in \mathcal{D}$ to be used in an m-term approximant, we mean one which maximizes a certain functional determined by information from the previous steps of the algorithm. We obtain different types of greedy algorithms by varying the above mentioned functional and also by using different ways of constructing (choosing coefficients of the linear combination) the m-term approximant from already found m elements of the dictionary.

We begin presentation of new results by a discussion of previous known results closely related to the results of this paper. The following general tendency in the development of these results will be seen in the discussion. We will be going step by step from theorecical approximation schemes to practically implementable algorithms.

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

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

and $\overline{\operatorname{span}}\mathcal{D} = X$. We note that in [T1] we required in the definition of a dictionary normalization of its elements ($\|g\| = 1$). However, it is easy to check that the arguments from [T1] work under assumption $\|g\| \le 1$ instead of $\|g\| = 1$. It will be more convenient for us to have an assumption $\|g\| \le 1$ than normalization of a dictionary.

For an 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 the 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 (see [T1]) the Weak Chebyshev Greedy Algorithm (WCGA) 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

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

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

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

In the case $t_k = 1$, $k = 1, 2, \dots$ we call the WCGA the Chebyshev Greedy Algorithm (CGA). Both the WCGA and the CGA are theoretical greedy approximation methods. The term *weak* in the above definition 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 do not 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. However, it is clear that in the case of infinite dictionary $\mathcal{D}$ there is no direct computationally feasible way to evaluate $\sup_{g\in\mathcal{D}} F_{f_{m-1}^c}(g)$.

At the second step we are looking for the best approximant of f from Φ_m. We know that such an approximant does exist. However, in practice we connot find it exactly. We can only find it approximately with some error.

The above observations motivated us to consider a variant of the WCGA toward practically implementable algorithm.

We studied in [T5] the following modification of the WCGA. Let three sequences $\tau = \{t_k\}_{k=1}^\infty$, $\delta = \{\delta_k\}_{k=0}^\infty$, $\eta = \{\eta_k\}_{k=1}^\infty$ of numbers from $[0,1]$ be given.

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

1). F_{m-1} is a functional with properties

$$\|F_{m-1}\| \le 1, \qquad F_{m-1}(f_{m-1}) \ge \|f_{m-1}\|(1-\delta_{m-1});$$

and $\varphi_m := \varphi_m^{\tau,\delta,\eta} \in \mathcal{D}$ is any satisfying

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

2). Define

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

and denote

$$E_m(f) := \inf_{\varphi\in\Phi_m} \|f - \varphi\|.$$

Let $G_m \in \Phi_m$ be such that

$$\|f - G_m\| \le E_m(f)(1 + \eta_m).$$

3). Denote

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

The term *approximate* in this definition means that we use a functional F_{m-1} that is an approximation to the norming (peak) functional $F_{f_{m-1}}$ and also we use an approximant $G_m \in \Phi_m$ which satisfies a weaker assumption than being a best approximant of f from Φ_m. Thus, in the *approximate* version of the WCGA we have addressed the issue of nonexact evaluation of the norming functional and the best approximant. We did not address the issue of finding the $\sup_{g\in\mathcal{D}} F_{f^c_{m-1}}(g)$. In this paper we address this issue. We will do it in two steps. First, we will consider the corresponding modification of the WCGA and then the modification of the AWCGA. These modifications are done in a style of the concept of *depth search* from [Do].

We now consider a countable dictionary $\mathcal{D} = \{\pm\psi_j\}_{j=1}^\infty$. We denote $\mathcal{D}(N) := \{\pm\psi_j\}_{j=1}^N$. Let $\mathcal{N} := \{N_j\}_{j=1}^\infty$ be a sequence of natural numbers.

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

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

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

2). Define

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

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

3). Denote

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

We present results on the behavior of the RWCGA in Section 2. In Section 3 we give a definition of the Restricted Approximate Weak Chebyshev Greedy Algorithm (RAWCGA) and give some convergence results for the RAWCGA.

2. Convergence and rate of approximation of the RWCGA

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 an even convex function satisfying the inequalities

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

We begin this section with a theorem from [T1] on convergence of the WCGA. In the formulation of this theorem we need a special sequence which is defined for a given modulus of smoothness $\rho(u)$ and a given $\tau = \{t_k\}_{k=1}^\infty$.

Definition 2.1. *Let $\rho(u)$ be an even convex function on $(-\infty,\infty)$ with the property: $\rho(2)\ge 1$ and*

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

For any $\tau=\{t_k\}_{k=1}^\infty$, $0<t_k\le 1$, and $0<\theta\le 1/2$ we define $\xi_m := \xi_m(\rho,\tau,\theta)$ as a number u satisfying the equation

$$\rho(u)=\theta t_m u. \tag{2.1}$$

Remark 2.1. *Assumptions on $\rho(u)$ imply that the function*

$$\epsilon(u) := \rho(u)/u, \quad u\neq 0, \quad \epsilon(0)=0,$$

is a continuous increasing on $[0,\infty)$ function with $\epsilon(2)\ge 1/2$. Thus (2.1) has a unique solution $0<\xi_m\le 2$.

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 condition: for any $\theta>0$ we have*

$$\sum_{m=1}^\infty t_m\xi_m(\rho,\tau,\theta)=\infty.$$

Then for any $f\in X$ we have

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

Corollary 2.1. *Let a Banach space X have modulus of smoothness $\rho(u)$ of power type $1<q\le 2$; ($\rho(u)\le\gamma u^q$). Assume that*

$$\sum_{m=1}^\infty t_m^p=\infty, \quad p=\frac{q}{q-1}. \tag{2.2}$$

Then the WCGA converges for any $f\in X$.

We prove here the following convergence result for the RWCGA.

Theorem 2.2. *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 condition: for any $\theta>0$ we have*

$$\sum_{m=1}^\infty t_m\xi_m(\rho,\tau,\theta)=\infty.$$

Suppose also that $\lim_{m\to\infty}N_m=\infty$. Then for any $f\in X$ we have

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

Corollary 2.2. *Let a Banach space X have modulus of smoothness $\rho(u)$ of power type $1 < q \le 2$; ($\rho(u) \le \gamma u^q$). Assume that $\lim_{m\to\infty} N_m = \infty$ and*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1}. \tag{2.3}$$

Then the RWCGA converges for any $f \in X$.

We now proceed to study the rate of convergence of the RWCGA. The following theorem has been proved in [T1] for the WCGA. We denote the closure of the convex hull of $\mathcal{D}$ by $\mathcal{A}_1(\mathcal{D})$.

Theorem 2.3. *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*

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

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

For $b > 0$, $K > 0$ we define the class

$$\mathcal{A}_1^b(K,\mathcal{D}) := \{f : d(f, \mathcal{A}_1(\mathcal{D}(n)) \le Kn^{-b}, \quad n = 1, 2, \dots\}.$$

Here, $\mathcal{A}_1(\mathcal{D}(n))$ is a convex hull of $\{\pm\psi_j\}_{j=1}^n$ and for a compact set F

$$d(f, F) := \inf_{\phi\in F} \|f - \phi\|.$$

Theorem 2.4. *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 $t \in (0,1]$ there exist $C_1(t,\gamma,q,K)$, $C_2(t,\gamma,q,K)$ such that for $\mathcal{N}$ with $N_m \ge C_1(t,\gamma,q,K)m^{r/b}$, $m = 1, 2, \dots$ we have for any $f \in \mathcal{A}_1^b(K,\mathcal{D})$*

$$\|f_m^{c,\tau,\mathcal{N}}\| \le C_2(t,\gamma,q,K)m^{-r}, \quad \tau = \{t\}, \quad r := 1 - 1/q.$$

We note that we can choose an algorithm from Theorem 2.4 that satisfies the *polynomial depth search* condition $N_m \le Cm^a$ from [Do].

We will use the following two simple well known lemmas (see, for instance, [T1]).

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 $\phi \in L$.

Lemma 2.2. *For any bounded linear functional F and any set S of elements we have*

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

where $\mathcal{A}_1(S)$ is the closure of the convex hull of S.

We will prove the following analog of Lemma 2.3 from [T1].

Lemma 2.3. *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*

$$\|f - f^\epsilon\| \leq \epsilon$$

and

$$f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D}(N(\epsilon))),$$

with some numbers $A(\epsilon)$, $N(\epsilon)$. Suppose $\lim_{m\to\infty} N_m = \infty$. Then there ixists $m(\epsilon)$ such that we have for $m \geq m(\epsilon)$

$$\|f_m^{c,\tau,\mathcal{N}}\| \leq \|f_{m-1}^{c,\tau,\mathcal{N}}\| \inf_\lambda (1 - \lambda t_m A(\epsilon)^{-1}(1 - \frac{\epsilon}{\|f_{m-1}^{c,\tau,\mathcal{N}}\|}) + 2\rho(\frac{\lambda}{\|f_{m-1}^{c,\tau,\mathcal{N}}\|})).$$

Proof. Let $m(\epsilon)$ be such that $N_m \geq N(\epsilon)$ for $m \geq m(\epsilon)$. Consider $m \geq m(\epsilon)$. We have for any λ

$$\|f_{m-1} - \lambda\varphi_m\| + \|f_{m-1} + \lambda\varphi_m\| \leq 2\|f_{m-1}\|(1 + \rho(\frac{\lambda}{\|f_{m-1}\|})) \tag{2.6}$$

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

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

$$t_m \sup_{\phi\in\mathcal{A}_1(\mathcal{D}(N_m))} F_{f_{m-1}}(\phi) \geq t_m A(\epsilon)^{-1} F_{f_{m-1}}(f^\epsilon).$$

By Lemma 2.1 we obtain

$$F_{f_{m-1}}(f^\epsilon) = F_{f_{m-1}}(f + f^\epsilon - f) \geq F_{f_{m-1}}(f) - \epsilon =$$

$$F_{f_{m-1}}(f_{m-1}) - \epsilon = \|f_{m-1}\| - \epsilon.$$

Using the inequality

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

we get from (2.6) and the above estimates

$$\|f_m\| \leq \inf_\lambda \|f_{m-1} - \lambda\varphi_m\| \leq \tag{2.7}$$

$$\|f_{m-1}\| \inf_\lambda (1 - \lambda t_m A(\epsilon)^{-1}(1 - \frac{\epsilon}{\|f_{m-1}\|}) + 2\rho(\frac{\lambda}{\|f_{m-1}\|}))$$

what proves the lemma.

Proof of Theorem 2.2. The definition of RWCGA implies that $\{\|f_m\|\}$ is a nonincreasing sequence. Therefore we have

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

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

$$\|f_m\| \geq \alpha.$$

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

$$\|f - f^\epsilon\| \leq \epsilon \quad \text{and} \quad f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D}(N(\epsilon)))$$

with some $A(\epsilon)$, $N(\epsilon)$. Then by Lemma 2.3 we get for $m \geq m(\epsilon)$

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

Let us specify $\theta := \frac{\alpha}{8A(\epsilon)}$ and take $\lambda = \alpha\xi_m(\rho, \tau, \theta)$. Then we obtain

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

The assumption

$$\sum_{m=1}^\infty t_m \xi_m = \infty$$

implies that

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

We got a contradiction which proves the theorem.

Proof of Theorem 2.4. By Lemma 2.3 with $\epsilon = KN_m^{-b}$ and $A(\epsilon) = 1$ we have for $f \in \mathcal{A}_1^b(K, \mathcal{D})$ that

$$\|f_m\| \leq \|f_{m-1}\| \inf_\lambda (1 - \lambda t(1 - \frac{KN_m^{-b}}{\|f_{m-1}\|}) + 2\gamma(\frac{\lambda}{\|f_{m-1}\|})^q). \tag{2.8}$$

First, we explain the idea of the proof and then we will give technical details. We will specify $C_1(t, \gamma, q, K)$, A_q, and λ in such a way that the assumption

$$\|f_{m-1}\| \geq A_q^{1/p} t^{-1}(m-1)^{-r}$$

will imply by (2.8) the estimate

$$\|f_m\| \leq \|f_{m-1}\|(1 - 1/(m-1)).$$

Then we will use the following lemma from [T3].

Lemma 2.4. *Let three positive numbers $\alpha < \beta \le 1$, $A > 1$ be given and let a sequence of positive numbers $1 \ge a_1 \ge a_2 \ge \dots$ satisfy the condition: if for some $\nu \in \mathbb{N}$ we have*

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

then

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

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

$$a_n \le Bn^{-\alpha}.$$

We choose λ from the equation

$$\frac{1}{4}\lambda t = 2\gamma(\frac{\lambda}{\|f_{m-1}\|})^q.$$

This implies that

$$\lambda = \|f_{m-1}\|^{\frac{q}{q-1}}(8\gamma)^{-\frac{1}{q-1}}t^{\frac{1}{q-1}}.$$

We set

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

Now, if $\|f_{m-1}\| \ge A_q^{1/p}/(t(m-1)^r)$ then we have for $N_m \ge (\frac{2Ktm^r}{A_q^{1/p}})^{1/b}$ that $\frac{KN_m^{-b}}{\|f_{m-1}\|} \le \frac{1}{2}$ and we get from (2.8)

$$\|f_m\| \le \|f_{m-1}\|(1-\frac{1}{4}\lambda t) = \|f_{m-1}\|(1-t^p\|f_{m-1}\|^p/A_q) \le \|f_{m-1}\|(1-1/(m-1)).$$

We use Lemma 2.4 and complete the proof.

We give an example of performance of the RWCGA. The problem concerns the trigonometric m-term approximation in the L_p-norm. Let $\mathcal{T}(N)$ be the subspace of real trigonometric polynomials of order N and let $\mathcal{T}$ be the real trigonometric system

$$\frac{1}{2}, \sin x, \cos x, \sin 2x, \cos 2x, \dots \quad .$$

Denote for $f \in L_p(\mathbb{T})$

$$\sigma_m(f,\mathcal{T})_p := \inf_{c_1,\dots,c_m;\phi_1,\dots,\phi_m\in\mathcal{T}} \|f - \sum_{j=1}^m c_j\phi_j\|_p$$

the best m-term trigonometric approximation of f in the L_p-norm. It is clear that one can get an upper estimate for $\sigma_{2m+1}(f,\mathcal{T})_p$ by approximating f by trigonometric polynomials of order m. Denote

$$E_m(f,\mathcal{T})_p := \inf_{u\in\mathcal{T}(m)} \|f-u\|_p.$$

Let

$$\mathcal{A}_1 := \mathcal{A}_1(\mathcal{T}) := \{f : \sum_{k=0}^{\infty}(|a_k(f)| + |b_k(f)|) \le 1\}$$

where $a_k(f)$, $b_k(f)$ are the corresponding Fourier coefficients. From the general results on convergence rate of the WCGA (see Theorem 2.3 above) it follows that for $f \in \mathcal{A}_1$, $t_k = t \in (0,1)$, $k = 1, 2, \dots,$

$$\|f_m^{c,\tau}\|_p \le C(p,t)m^{-1/2}, \quad 2 \le p < \infty.$$

Let us apply Theorem 2.4 in the same situation. Now, in addition to $f \in \mathcal{A}_1$ we require

$$E_n(f,\mathcal{T})_p \le Dn^{-b}, \quad n = 1, 2, \dots, \tag{2.9}$$

with some $b > 0$. Then it is easy to derive from Theorem 2.4 that there exist two constants $C_1(p,t,D)$, $C_2(p,t,D)$ such that for $\tau = \{t\}$ and $\mathcal{N}$ with $N_m \ge C_1(p,t,D)m^{-1/(2b)}$, $m = 1, 2, \dots$ we have for any $f \in \mathcal{A}_1$ satisfying (2.9) that

$$\|f_m^{c,\tau,\mathcal{N}}\|_p \le C_2(p,t,D)m^{-1/2}. \tag{2.10}$$

We note that for the above class one cannot obtain an esimate better than (2.10) (clearly, for $b \le 1/2$). Indeed, let m be given. Consider

$$f(x) := (2m)^{-1}R(x), \quad R(x) = \sum_{k=1}^{2m} \pm \cos kx,$$

where $R(x)$ is the Rudin-Shapiro polynomial such that

$$\|R\|_\infty \le Cm^{1/2}.$$

Then $f \in \mathcal{A}_1$ and

$$E_n(f,\mathcal{T})_\infty \le Dn^{-1/2}, \quad n = 1, 2, \dots.$$

Also,

$$\sigma_m(f,\mathcal{T})_2 \ge m^{-1/2}/2.$$

We now make some general remarks on m-term approximation with the depth search constraint. The depth search constraint means that for a given m we restrict ourselves to systems of elements (subdictionaries) containing at most $N := N(m)$ elements. Let X be a linear metric space and for a set $\mathcal{D} \subset X$, let $\mathcal{L}_m(\mathcal{D})$ denote the collection of all linear spaces spanned by m elements of $\mathcal{D}$. For a linear space $L \subset X$, the ϵ-neighborhood $U_\epsilon(L)$ of L is the set of all $x \in X$ which are at a distance not exceeding ϵ from L (i.e. those $x \in X$ which can be approximated to an error not exceeding ϵ by the elements of L). For any compact set $F \subset X$ and any integers $N, m \ge 1$, we define the (N,m)-entropy numbers (see [T4, p.94])

$$\epsilon_{N,m}(F,X) := \inf_{\#\mathcal{D}=N} \inf\{\epsilon : F \subset \cup_{L\in\mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

We can express $\sigma_m(F,\mathcal{D})$ as

$$\sigma_m(F,\mathcal{D}) = \inf\{\epsilon : F \subset \cup_{L\in\mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

It follows therefore that

$$\inf_{\#\mathcal{D}=N} \sigma_m(F,\mathcal{D}) = \epsilon_{N,m}(F,X).$$

In other words, finding best dictionaries consisting of N elements for m-term approximation of F is the same as finding sets $\mathcal{D}$ which attain the (N,m)-entropy numbers $\epsilon_{N,m}(F,X)$. It is easy to see that $\epsilon_{m,m}(F,X) = d_m(F,X)$ where $d_m(F,X)$ is the Kolmogorov width of F in X. This establishes a connection between (N,m)-entropy numbers and the Kolmogorov widths. One can find further discussion on the nonlinear Kolmogorov (N,m)-widths and the entropy numbers in [T4].

3. Convergence and rate of convergence of the RAWCGA

We study here the following modification of the AWCGA. Let three sequences $\tau = \{t_k\}_{k=1}^\infty$, $\delta = \{\delta_k\}_{k=0}^\infty$, $\eta = \{\eta_k\}_{k=1}^\infty$ of numbers from $[0,1]$ be given. Let $\mathcal{N} := \{N_j\}_{j=1}^\infty$ be a sequence of natural numbers.

Restricted Approximate Weak Chebyshev Greedy Algorithm (RAWCGA).

We define $f_0 := f_0^{\tau,\delta,\eta,\mathcal{N}} := f$. Then for each $m \ge 1$ we inductively define

1). F_{m-1} is a functional with properties

$$\|F_{m-1}\| \le 1, \qquad F_{m-1}(f_{m-1}) \ge \|f_{m-1}\|(1-\delta_{m-1});$$

and $\varphi_m := \varphi_m^{\tau,\delta,\eta,\mathcal{N}} \in \mathcal{D}(N_m)$ is any satisfying

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

2). Define

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

and denote

$$E_m(f) := \inf_{\varphi\in\Phi_m} \|f-\varphi\|.$$

Let $G_m \in \Phi_m$ be such that

$$\|f - G_m\| \le E_m(f)(1+\eta_m).$$

3). Denote

$$f_m := f_m^{\tau,\delta,\eta,\mathcal{N}} := f - G_m.$$

We begin with the convergence theorem. The following convergence theorem for the AWCGA and its corollaries have been proved in [T5].

Theorem 3.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Assume that sequences τ, δ, η satisfy the conditions: for any $\theta > 0$ we have*

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho,\tau,\theta) = \infty$$

and

$$\delta_m = o(t_m \xi_m(\rho,\tau,\theta)), \qquad \eta_m = o(t_m \xi_m(\rho,\tau,\theta)).$$

Then for any $f \in X$ we have

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

Corollary 3.1. *Let a Banach space X have modulus of smoothness $\rho(u)$ of power type $1 < q \le 2$; $(\rho(u) \le \gamma u^q)$. Assume that*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1};$$

and

$$\delta_m = o(t_m^p), \qquad \eta_m = o(t_m^p).$$

Then the AWCGA converges for any $f \in X$.

Corollary 3.2. *Let X be a uniformly smooth Banach space. Assume that $\tau = \{t\}$, $t \in (0,1]$. Then for any two sequences $\delta, \eta \in c_0$ the corresponding AWCGA converges for any $f \in X$.*

We prove here the following convergence result for the RAWCGA.

Theorem 3.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Assume that sequences τ, δ, η satisfy the conditions: for any $\theta > 0$ we have*

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho,\tau,\theta) = \infty$$

and

$$\delta_m = o(t_m \xi_m(\rho,\tau,\theta)), \qquad \eta_m = o(t_m \xi_m(\rho,\tau,\theta)).$$

Suppose also that $\lim_{m\to\infty} N_m = \infty$. Then for any $f \in X$ we have

$$\lim_{m\to\infty} \|f_m^{\tau,\delta,\eta,\mathcal{N}}\| = 0.$$

Corollary 3.3. *Let a Banach space X have modulus of smoothness $\rho(u)$ of power type $1 < q \le 2$; $(\rho(u) \le \gamma u^q)$. Assume that $\lim_{m\to\infty} N_m = \infty$,*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1},$$

and

$$\delta_m = o(t_m^p), \qquad \eta_m = o(t_m^p).$$

Then the RAWCGA converges for any $f \in X$.

Corollary 3.4. *Let X be a uniformly smooth Banach space. Assume that $\tau = \{t\}$, $t \in (0,1]$. Then for any two sequences $\delta, \eta \in c_0$ the corresponding RAWCGA converges for any $f \in X$ provided $\lim_{m\to\infty} N_m = \infty$.*

The proof of Theorem 3.2 is similar to that of Theorem 2.2. Instead of Lemma 2.1 we use the following one from [T5,Lemma 2.1].

Lemma 3.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. For a finite-dimensional subspace L of X and an element $f \in X$ denote*

$$E_L(f) := \inf_{l\in L} \|f - l\|.$$

Assume that an element $g \in L$ and a functional F satisfy the following conditions

$$0 < \|f^L\| \le E_L(f)(1+a), \quad f^L := f - g, \quad a \in [0,1];$$
$$F(f^L) \ge \|f^L\|(1-b), \quad \|F\| \le 1, \quad b \in [0,1].$$

Then

$$|F(g)| \le \inf_{v\ge 0}(a + b + 2\rho(3v\|f\|))/v.$$

We also replace Lemma 2.3 by the following lemma that is an analog of Lemma 2.3 from [T5] modified in a style of Lemma 2.3 from Section 2.

Lemma 3.2. *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*

$$\|f - f^\epsilon\| \le \epsilon$$

and

$$f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D}(N(\epsilon))),$$

with some numbers $A(\epsilon)$, $N(\epsilon)$. Then for the RAWCGA with τ, δ, η and $\mathcal{N}$ such that $\lim_{m\to\infty} N_m = \infty$ there exists $m(\epsilon)$ such that we have for $m \ge m(\epsilon)$

$$E_m(f) \le \|f_{m-1}\| \inf_\lambda (1 + \delta_{m-1} - \lambda t_m A(\epsilon)^{-1}(1 - \delta_{m-1} - \frac{\beta_{m-1}+\epsilon}{\|f_{m-1}\|}) + 2\rho(\frac{\lambda}{\|f_{m-1}\|})),$$

provided $\|f_{m-1}\| > 0$, where

$$\beta_{m-1} := \inf_{v\ge 0}(\delta_{m-1} + \eta_{m-1} + 2\rho(3v\|f\|))/v.$$

We obtain here the rate of convergence for an *adaptive* RAWCGA where *adaptive* means that sequences δ and η are determined by the RAWCGA applied to a given element $f \in \mathcal{A}_1^b(K,\mathcal{D})$.

Theorem 3.3. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let $t \in (0,1]$. There exist $C_1(t,\gamma,q,K)$, $C_2(t,\gamma,q,K)$ such that if for a given $f \in \mathcal{A}_1^b(K,\mathcal{D})$ we apply the RAWCGA with $\tau = \{t\}$, $N_m \ge C_1(t,\gamma,q,K)m^{r/b}$, $r := 1 - 1/q$,*

$$\delta_{m-1} := t^p \|f_{m-1}\|^p 3^{-p}(16A_q)^{-1}, \qquad m = 1,2,\dots,$$

$$\eta_{m-1} := t^p E_{m-1}(f)^p 3^{-p}(16A_q)^{-1}, \qquad m = 2,\dots, \quad p := \frac{q}{q-1},$$

where

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

then we obtain

$$\|f_m^{\tau,\delta,\eta,\mathcal{N}}\| \le C_2(t,\gamma,q,K)m^{-r}.$$

Proof. By Lemma 3.2 with $\epsilon = KN_m^{-b}$ and $A(\epsilon) = 1$ we have for $f \in \mathcal{A}_1^b(K,\mathcal{D})$ that

$$E_m(f) \le \tag{3.1}$$

$$\|f_{m-1}\| \inf_\lambda \big(1 + \delta_{m-1} - \lambda t(1 - \delta_{m-1} - (\beta_{m-1} + \epsilon)/\|f_{m-1}\|) + 2\gamma\big(\frac{\lambda}{\|f_{m-1}\|}\big)^q\big).$$

We estimate β_{m-1} by choosing

$$v = \|f_{m-1}\|^{\frac{1}{q-1}} 3^{-p}/A_q.$$

We have

$$\beta_{m-1} \le (\delta_{m-1} + \eta_{m-1})/v + 2\gamma 3^q v^{q-1} \le (1/16 + 1/16 + 1/4)\|f_{m-1}\| = \frac{3}{8}\|f_{m-1}\|.$$

Assume $E_{m-1}(f) \ge A_q^{1/p}/(t(m-1)^r)$. Using $\delta_{m-1} \le 1/16$ we get from (3.1) for $N_m \ge (32Ktm^r A_q^{-1/p})^{1/b}$

$$E_m(f) \le \|f_{m-1}\| \inf_\lambda \big(1 + \delta_{m-1} - \frac{17}{32}\lambda t + 2\gamma\big(\frac{\lambda}{\|f_{m-1}\|}\big)^q\big). \tag{3.2}$$

We choose λ from the equation

$$\frac{1}{4}\lambda t = 2\gamma\big(\frac{\lambda}{\|f_{m-1}\|}\big)^q$$

what implies that

$$\lambda = \|f_{m-1}\|^{\frac{q}{q-1}}(8\gamma)^{-\frac{1}{q-1}} t^{\frac{1}{q-1}} = 4t^{\frac{1}{q-1}}\|f_{m-1}\|^p/A_q.$$

With this λ using the notation $p := \frac{q}{q-1}$ we get from (3.2)

$$E_m(f) \le \|f_{m-1}\|(1 + \delta_{m-1} - \frac{9}{32}\lambda t) \le \|f_{m-1}\|(1 - t^p\|f_{m-1}\|^p/A_q) \le$$

$$E_{m-1}(f)(1 + t^p E_{m-1}(f)^p/(4A_q))(1 - t^p\|f_{m-1}\|^p/A_q) \le$$

$$E_{m-1}(f)(1 - 3t^p E_{m-1}(f)^p/(4A_q)) \le E_{m-1}(f)(1 - \frac{3}{4}(m-1)^{-1}).$$

By Lemma 2.4 we get

$$E_m(f) \le C_2' m^{-r}$$

what implies

$$\|f_m\| \le C_2 m^{-r}.$$

Theorem 3.3 is proved now.

References

[D] R.A. DeVore, *Nonlinear approximation*, Acta Numerica (1998), 51–150.

[Do] D.L. Donoho, *Sparse Components of Images and Optimal Atomic Decompositions*, Constructive Approx. **17** (2001), 353–382.

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

[S] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I*, Math. Annalen **63** (1906-1907), 433–476.

[T1] V.N. Temlyakov, *Greedy algorithms in Banach spaces*, Advances in Comput. Math. **14** (2001), 277–292.

[T2] V.N. Temlyakov, *Weak Greedy Algorithms*, Advances in Comput. Math. **12** (2000), 213–227.

[T3] V.N. Temlyakov, *Greedy Algorithms and m-term Approximation With Regard to Redundant Dictionaries*, J. Approx. Theory **98** (1999), 117–145.

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

[T5] V.N. Temlyakov, *Greedy Type Algorithms in Banach Spaces and Applications*, IMI Preprints **18** (2002), 1–36.