Constr. Approx. (2005) 21: 257–292
DOI: 10.1007/s00365-004-0565-6

CONSTRUCTIVE APPROXIMATION

Greedy-Type Approximation in Banach Spaces and Applications

V. N. Temlyakov

Abstract. We continue to study the efficiency of approximation and convergence of greedy-type algorithms in uniformly smooth Banach spaces. Two greedy-type approximation methods, the Weak Chebyshev Greedy Algorithm (WCGA) and the Weak Relaxed Greedy Algorithm (WRGA), have been introduced and studied in [24]. These methods (WCGA and WRGA) are very general approximation methods that work well in an arbitrary uniformly smooth Banach space X for any dictionary $\mathcal{D}$. It turns out that these general approximation methods are also very good for specific dictionaries. It has been observed in [7] that the WCGA and WRGA provide constructive methods in m-term trigonometric approximation in L_p, $p \in [2, \infty)$, which realize an optimal rate of m-term approximation for different function classes. In [25] the WCGA and WRGA have been used in constructing deterministic cubature formulas for a wide variety of function classes with error estimates similar to those for the Monte Carlo Method. The WCGA and WRGA can be considered as a constructive deterministic alternative to (or substitute for) some powerful probabilistic methods. This observation encourages us to continue a thorough study of the WCGA and WRGA.

In this paper we study modifications of the WCGA and WRGA that are motivated by numerical applications. In these modifications we are able to perform steps of the WCGA (or WRGA) approximately with some controlled errors. We prove that the modified versions of the *WCGA and WRGA perform as well as the WCGA and WRGA*.

We give two applications of greedy-type algorithms. First, we use them to provide a constructive proof of optimal estimates for best m-term trigonometric approximation in the uniform norm. Second, we use them to construct deterministic sets of points $\{\xi^1, \dots, \xi^m\} \subset [0, 1]^d$ with the L_p discrepancy less than $Cp^{1/2}m^{-1/2}$, C is an effective absolute constant.

1. Introduction

The purpose of this paper is to continue the 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 paper [24] which in turn is a combination of ideas and methods developed for Banach spaces in a fundamental paper [8], with the approach used in [23] in the case of Hilbert spaces. Papers [8] and

Date received: February 19, 2003. Date revised: December 11, 2003. Date accepted: December 16, 2003. Communicated by Peter Oswald. Online publication: June 15, 2004.
AMS classification: 41A65, 42A10, 46B20.
Key words and phrases: Greedy approximation, Greedy algorithms, Uniformly smooth Banach spaces, Trigonometric approximation, Discrepancy, Constructive proof.

[23] contain detailed historical remarks and we refer the reader to those papers. Two greedy-type approximation methods, the Weak Chebyshev Greedy Algorithm (WCGA) and the Weak Relaxed Greedy Algorithm (WRGA), have been introduced and studied in [24]. These methods (WCGA and WRGA) are very general approximation methods that work well in an arbitrary uniformly smooth Banach space X for any dictionary $\mathcal{D}$ (see below). Surprisingly, it turns out that these general approximation methods are also very good for specific dictionaries. It has been observed in [7] that the WCGA and WRGA provide constructive methods in m-term trigonometric approximation in L_p, $p \in [2, \infty)$, which realize an optimal rate of m-term approximation for different function classes. In [25] the WCGA and WRGA have been used in constructing deterministic cubature formulas for a wide variety of function classes with error estimates similar to those for the Monte Carlo Method. It appears that the WCGA and WRGA can be considered as a constructive deterministic alternative to (or substitute for) some powerful probabilistic methods. This observation encourages us to continue a thorough study of the WCGA and WRGA.

In Sections 2 and 3 we study modifications of the WCGA and WRGA that are motivated by numerical applications. In these modifications we are able to perform steps of the WCGA (or WRGA) approximately with some controlled errors. We prove that the modified versions of the WCGA and WRGA perform as well as the WCGA and WRGA. In Sections 2 and 3 we develop the theory of the Approximate Weak Chebyshev Greedy Algorithm (AWCGA) and the Approximate Weak Relaxed Greedy Algorithm (AWRGA) in a general setting: X is an arbitrary uniformly smooth Banach space and $\mathcal{D}$ is any dictionary. We keep the term *greedy algorithm* in the names of these two approximation methods for two reasons. First, this term has been used in previous papers and has become a standard name for procedures like WCGA and WRGA. For more discussion of the terminology see [26, Remark 1.1, p. 38]. Second, clearly, in the above general setting the term *algorithm* cannot be confused with the same term used in a more restricted sense, say, in computer science. We note that in the case of finite-dimensional X and finite $\mathcal{D}$ the above methods are algorithms in a strict sense.

In Section 4 we use the WCGA and WRGA to build a constructive method for m-term trigonometric approximation in the uniform norm. It is known that the case of approximating by m-term trigonometric polynomials in the uniform norm is the most difficult. We note that in the case of L_p-norms with $p < \infty$ the corresponding constructive method has been provided in [7].

In Section 5 we study a slight modification of the incremental-type algorithm from [8]. We apply that algorithm to construct deterministic sets of points with small L_p discrepancy and also with small symmetrized L_p discrepancy.

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 less than or equal to one ($\|g\| \leq 1$),

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

and $\overline{\operatorname{span}}\,\mathcal{D} = X$. We note that in [24] we required in the definition of a dictionary normalization of its elements ($\|g\| = 1$). However, it is easy to check that the arguments from [24] work under assumption $\|g\| \leq 1$ instead of $\|g\| = 1$. In the applications in Section 5 it will be more convenient for us to have an assumption $\|g\| \leq 1$ than normalization of a dictionary.

In this paper we will study two types of greedy algorithms with regard to $\mathcal{D}$. 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 first define (see [24]) the WCGA that is a generalization for Banach spaces of the Weak Orthogonal Greedy Algorithm defined and studied in [23] (see also [6] for the 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 study here 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. We note that there exists (see [10]) a version of the Approximate Weak Greedy Algorithm in Hilbert spaces.

The following WRGA has been studied in [24]:

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 $m \geq 1$ we inductively define:

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

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

(2) Find $0 \leq \lambda_m \leq 1$ such that

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

and define

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

(3) Denote

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

We will study here the following *approximate* version of the WRGA:

Approximate Weak Relaxed Greedy Algorithm (AWRGA). We define $f_0^{ar} := f_0^{ar,\tau,\delta,\eta} := f$ and $G_0^{ar} := G_0^{ar,\tau,\delta,\eta} := 0$. Then for each $m \geq 1$ we inductively define:

(1) F_{m-1}^{ar} is a functional with properties

$$\|F_{m-1}^{ar}\| \leq 1, \qquad F_{m-1}^{ar}(f_{m-1}^{ar}) \geq \|f_{m-1}^{ar}\|(1-\delta_{m-1});$$

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

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

(2) Find $0 \leq \lambda_m \leq 1$ such that

$$\|f - ((1-\lambda_m)G_{m-1}^{ar} + \lambda_m\varphi_m^{ar})\|$$
$$\leq \min\left(\|f_{m-1}^{ar}\|, \inf_{0\leq\lambda\leq 1} \|f - ((1-\lambda)G_{m-1}^{ar} + \lambda\varphi_m^{ar})\|(1+\eta_m)\right)$$

and define

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

(3) Denote

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

In Sections 2 and 3 we study the questions of convergence and the rate of convergence for the two methods of approximation, AWCGA and AWRGA. It is clear that in the case of AWRGA the assumption that f belongs to the closure of the convex hull of $\mathcal{D}$ is natural. We denote the closure of the convex hull of $\mathcal{D}$ by $\mathcal{A}_1(\mathcal{D})$. It has been proven in [23] 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

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

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

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

A uniformly smooth Banach space is 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

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

It has been established in [8] 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 the modulus of smoothness of a Banach space. Namely, if the 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 proved in [24] 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 $\rho(u) \le \gamma u^q$, $1 < q \le 2$, of order

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

It also has been proved in [24] that the WCGA converges for any $f \in X$ and the WRGA converges for any $f \in \mathcal{A}_1(\mathcal{D})$ if τ satisfies the condition

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

The sequences $\{\xi_m(\rho,\tau,\theta)\}$ are defined in Definition 2.1 of Section 2. In a particular case of $\rho(u) \asymp u^q$, $1 < q \le 2$, relation (1.3) is equivalent to

(1.4) $$\sum_{k=1}^{m} t_k^p = \infty, \qquad p := \frac{q}{q-1}.$$

In [24] we gave an example which showed that (1.4) is sharp for Banach spaces with modulus of smoothness of power type q.

It is well-known (see, for instance, [8, Lemma B.1]) that in the case $X = L_p$, $1 \le p < \infty$, we have

$$\rho(u) \le \begin{cases} u^p/p & \text{if } 1 \le p \le 2, \\ (p-1)u^2/2 & \text{if } 2 \le p < \infty. \end{cases} \tag{1.5}$$

It is also known (see [15, 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, $\dim X \ge 2$,

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

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

In Sections 2 and 3 we prove that under some reasonable assumptions on sequences δ and η the AWCGA and AWRGA are as good as the corresponding WCGA and WRGA. As an example we formulate here only one result (see Corollary 2.3 in Section 2 below).

Theorem 1.1. *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 recall that c_0 is the space of all convergent-to-zero sequences.

In Sections 4 and 5 we demonstrate the power of the WCGA and WRGA in classical areas of harmonic analysis and numerical integration. The first problem concerns the trigonometric m-term approximation in the uniform norm. Let $\mathcal{T}(N)$ be the subspace of real trigonometric polynomials of order N and let $\mathcal{T}$ be the real trigonometric system

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

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

$$\sigma_m(f, \mathcal{T})_p := \inf_{c_1,\dots,c_m;\varphi_1,\dots,\varphi_m \in \mathcal{T}} \left\| f - \sum_{j=1}^m c_j \varphi_j \right\|_p$$

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

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

The first result that indicated an advantage of m-term approximation over approximation by trigonometric polynomials of order m is due to R. S. Ismagilov [12]

$$\sigma_m(|\sin x|, \mathcal{T})_\infty \le C_\varepsilon m^{-6/5+\varepsilon} \qquad \text{for any} \quad \varepsilon > 0. \tag{1.6}$$

Let us compare it to the well-known result due to de La Vallée Poussin and S. N. Bernstein

$$E_m(|\sin x|, \mathcal{T})_\infty \asymp m^{-1}. \tag{1.7}$$

V. E. Maiorov [16] improved estimate (1.6):

$$\sigma_m(|\sin x|, \mathcal{T})_\infty \asymp m^{-3/2}. \tag{1.8}$$

Both R. S. Ismagilov [12] and V. E. Maiorov [16] used constructive methods to get their estimates (1.6) and (1.8). V. E. Maiorov [16] applied a number-theoretical method based on Gaussian sums. The key point of that technique can be formulated in terms of best m-term approximation of trigonometric polynomials. Using the Gaussian sums one can prove (constructively) the estimate

$$\sigma_m(t, \mathcal{T})_\infty \leq CN^{3/2}m^{-1}\|t\|_1, \qquad t \in \mathcal{T}(N). \tag{1.9}$$

Denote

$$\left\| a_0/2 + \sum_{k=1}^{N}(a_k \cos kx + b_k \sin kx) \right\|_A := |a_0| + \sum_{k=1}^{N}(|a_k| + |b_k|).$$

We note that, by simple inequality,

$$\|t\|_A \leq (2N+1)\|t\|_1, \qquad t \in \mathcal{T}(N),$$

the estimate (1.9) follows from the estimate

$$\sigma_m(t, \mathcal{T})_\infty \leq C(N^{1/2}/m)\|t\|_A. \tag{1.10}$$

Thus (1.10) is stronger than (1.9). The following estimate is known (see [5]):

$$\sigma_m(t, \mathcal{T})_\infty \leq Cm^{-1/2}(\ln(1 + N/m))^{1/2}\|t\|_A. \tag{1.11}$$

In a way, (1.11) is much stronger than (1.10) and (1.9). However, the existing proof of (1.11) (see [5]) is not constructive. Estimate (1.11) has been proved in [5] with the help of a nonconstructive theorem of Gluskin [9]. In Section 4 we give a constructive proof of (1.11). The key ingredient of that proof is the WCGA (or WRGA). In paper [7] we already pointed out that the WCGA provides a constructive proof of the estimate

$$\sigma_m(t, \mathcal{T})_p \leq C(p)m^{-1/2}\|t\|_A, \qquad p \in [2, \infty). \tag{1.12}$$

The known proofs (before [7]) of (1.12) were nonconstructive (see the discussion in [7, Section 5]).

We formulate here a general result from Section 4 (see Theorem 4.6).

Theorem 1.2. *Let $\Phi := \{\varphi_j\}_{j=1}^\infty$ be a uniformly bounded orthonormal system defined on a bounded domain. Assume Φ has the* (VP) *property. Then there exists a constructive algorithm $A(\Phi, N, m)$ such that for any $\varphi \in \Phi(N)$ it provides an m-term Φ-polynomial $A(\Phi, N, m)(\varphi)$ with the following approximation property*:

$$\|\varphi - A(\Phi, N, m)(\varphi)\|_\infty \leq Cm^{-1/2}(\ln(1 + N/m))^{1/2}\|\varphi\|_A$$

with a constant C which may depend on Φ.

The (VP) property is a property that guarantees the existence of a sequence of the de La Vallée Poussin operators. See Section 4 for precise definition.

In Section 5 we apply greedy-type algorithms to construct points with small discrepancy and small symmetrized discrepancy. Let $1 \le p < \infty$. We will first define the L_p discrepancy (the L_p-star discrepancy) of points $\{\xi^1, \dots, \xi^m\} \subset \Omega_d := [0,1]^d$. Let $\chi_{[a,b]}(\cdot)$ be a characteristic function of the interval $[a,b]$. Denote, for $x, y \in \Omega_d$,

$$B(x, y) := \prod_{j=1}^{d} \chi_{[0,x_j]}(y_j).$$

Then the L_p discrepancy of $\xi := \{\xi^1, \dots, \xi^m\} \subset \Omega_d$ is defined by

$$D(\xi, m, d)_p := \left\| \int_{\Omega_d} B(x, y)\, dy - \frac{1}{m} \sum_{\mu=1}^{m} B(x, \xi^\mu) \right\|_{L_p(\Omega_d)}.$$

We are interested in ξ with small discrepancy. Consider

$$D(m, d)_p := \inf_{\xi} D(\xi, m, d)_p.$$

The concept of discrepancy is a fundamental concept in numerical integration. There are many books and survey papers on discrepancy and related topics. We will mention some of them as a reference for the history of the subject: [14], [1], [18], [2], [20], [25]. For $1 < p < \infty$ the following relation is known (see [1, p. 5]):

$$D(m, d)_p \asymp m^{-1} (\ln m)^{(d-1)/2} \tag{1.13}$$

with constants in $\asymp$ depending on p and d. The right order of $D(m,d)_p$, $p = 1, \infty$, for $d \ge 3$ is unknown. Recently, driven by possible applications (see [20]) in numerical integration, the tendancy to control dependence of $D(m,d)_p$ on both variables m and d has appeared. Very interesting results in this direction have been obtained in [11]. They proved the estimate

$$D(m, d)_\infty \le C d^{1/2} m^{-1/2} \tag{1.14}$$

with C an absolute constant. It is pointed out in [11] that (1.14) is only an existence theorem and even a constant C in (1.14) is unknown. Their proof is a probabilistic one. There are also some other estimates in [11] with explicit constants. We mention one of them,

$$D(m, d)_\infty \le C (d \ln d)^{1/2} ((\ln n)/n)^{1/2} \tag{1.15}$$

with an explicit constant C. The proof of (1.15) is also probabilistic.

In Section 5 we give constructive proofs of the following two upper estimates:

$$\begin{aligned} D(m, d)_p &\le C_1 p^{1/2} m^{-1/2}, \qquad p \in [2, \infty), \\ D(m, d)_\infty &\le C_2 d^{3/2} (\max(\ln d, \ln m))^{1/2} m^{-1/2}, \qquad d, m \ge 2, \end{aligned}$$

with effective absolute constants C_1 and C_2. The term *constructive proof* goes back to Kronecker who outlined the program of giving constructive proofs of theorems that were established as existence theorems. Following the traditions of approximation theory we understand a constructive proof as a proof that provides a construction of an object and this construction has a potential of being implemented numerically. For instance, a proof by contradiction, or a probabilistic proof establishing the existence of an object, is not a constructive proof for us. In Section 5 we provide a method which consists of maximizing (approximately) certain functions of d variables at each step. For a given $p \in [2, \infty)$, after m steps of this method we obtain a set $\xi = \{\xi^1, \dots, \xi^m\} \subset \Omega_d$ of points with small L_p discrepancy

$$D(\xi, m, d)_p \le C_1 p^{1/2} m^{-1/2}$$

with effective absolute constant C_1. The above method is a greedy-type algorithm which is a slight modification of the corresponding procedure from [8]. Here we do not assume that a dictionary $\mathcal{D}$ is symmetric: $g \in \mathcal{D}$ implies $-g \in \mathcal{D}$. To indicate this we will use the notation $\mathcal{D}^+$ for such a dictionary. We do not assume that elements of a dictionary $\mathcal{D}^+$ are normalized ($\|g\| = 1$ if $g \in \mathcal{D}^+$), we only assume that $\|g\| \le 1$ if $g \in \mathcal{D}^+$. By $\mathcal{A}_1(\mathcal{D}^+)$ we denote the closure of the convex hull of $\mathcal{D}^+$. Let $\varepsilon = \{\varepsilon_n\}_{n=1}^\infty$, $\varepsilon_n > 0$, $n = 1, 2, \dots$.

Incremental Algorithm with Schedule ε (IA(ε)). Let $f \in \mathcal{A}_1(\mathcal{D}^+)$. Denote $f_0^{i,\varepsilon} := f$ and $G_0^{i,\varepsilon} := 0$. Then for each $m \ge 1$ we inductively define:

(1) $\varphi_m^{i,\varepsilon} \in \mathcal{D}^+$ is any satisfying

$$F_{f_{m-1}^{i,\varepsilon}}(\varphi_m^{i,\varepsilon} - f) \ge -\varepsilon_m.$$

(2) Define

$$G_m^{i,\varepsilon} := (1 - 1/m)G_{m-1}^{i,\varepsilon} + \varphi_m^{i,\varepsilon}/m.$$

(3) Denote

$$f_m^{i,\varepsilon} := f - G_m^{i,\varepsilon}.$$

Let us make a brief comparison of the above three types of greedy algorithms. The AWCGA contains a step of finding an approximant $G_m \in \Phi_m$ that provides approximation close to the best approximation. The corresponding steps of the AWRGA and IA(ε) are simpler: Optimization over $\lambda \in [0, 1]$ in the AWRGA and simple convex combination in the IA(ε). Next, the AWCGA can be applied to any $f \in X$. The AWRGA can be applied only to $f \in \mathcal{A}_1(\mathcal{D})$ (in other words, to f such that $\|f\|_{\mathcal{A}_1(\mathcal{D})} \le 1$). The IA($\varepsilon$) can be applied only to $f \in \mathcal{A}_1(\mathcal{D}^+)$ ($\|f\|_{\mathcal{A}_1(\mathcal{D}^+)} = 1$). In some cases (like in Section 5) a problem itself implies $\|f\|_{\mathcal{A}_1(\mathcal{D}^+)} = 1$. However, if the condition $\|f\|_{\mathcal{A}_1(\mathcal{D}^+)} = 1$ (or $\|f\|_{\mathcal{A}_1(\mathcal{D})} \le 1$) is not satisfied automatically, then it could be a difficult problem to find $\|f\|_{\mathcal{A}_1(\mathcal{D}^+)}$ and even estimate $\|f\|_{\mathcal{A}_1(\mathcal{D})}$. In such a case we would recommend using the AWCGA. Clearly, the AWCGA is the only option if $\|f\|_{\mathcal{A}_1(\mathcal{D})} = \infty$.

2. Convergence and Rate of Approximation of AWCGA

We begin this section with a known theorem on the convergence of WCGA [24]. 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: $\rho(2) \geq 1$ and

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

For any $\tau = \{t_k\}_{k=1}^{\infty}$, $0 < t_k \leq 1$, and $0 < \theta \leq \frac{1}{2}$ 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

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

is a continuous increasing on $[0, \infty)$ function with $\varepsilon(2) \geq \frac{1}{2}$. Thus (2.1) has a unique solution $0 < \xi_m \leq 2$.

The following theorem and a corollary have been proved in [24]:

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*

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) = \infty. \tag{2.2}$$

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 $1 < q \leq 2$; $(\rho(u) \leq \gamma u^q)$. Assume that*

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

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

We will prove the following theorem for convergence of the AWCGA:

Theorem 2.2. *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*

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) = \infty \tag{2.4}$$

and

$$\delta_m = o(t_m \xi_m(\rho, \tau, \theta)) \qquad \textit{and} \qquad \eta_m = o(t_m \xi_m(\rho, \tau, \theta)). \tag{2.5}$$

Then for any $f \in X$ *we have*

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

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*

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

and

$$\delta_m = o(t_m^p) \qquad \textit{and} \qquad \eta_m = o(t_m^p).$$

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

Corollary 2.3. *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* $f \in X$.

Lemma 2.1. *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 elemnt* $f \in X$ *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*:

$$0 < \|f^L\| \le E_L(f)(1+a), \qquad f^L := f - g, \quad a \in [0,1], \tag{2.6}$$

$$F(f^L) \ge \|f^L\|(1-b), \qquad \|F\| \le 1, \quad b \in [0,1]. \tag{2.7}$$

Then

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

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

$$\|f^L - \lambda g\| + \|f^L + \lambda g\| \le 2\|f^L\| \left(1 + \rho\left(\frac{\lambda\|g\|}{\|f^L\|}\right)\right). \tag{2.8}$$

Next, assume $|F(g)| = \beta > 0$. Then either $F(g) = \beta$ or $F(-g) = \beta$. We will carry out the proof under assumption $F(g) = \beta$ and note that the case $F(-g) = \beta$ is similar. We have

$$\|f^L + \lambda g\| \ge F(f^L + \lambda g) \ge \|f^L\|(1-b) + \lambda\beta \tag{2.9}$$

and, by (2.8),

$$\|f^L - \lambda g\| \le \|f^L\| \left(1 + b + 2\rho\left(\frac{3\lambda\|f\|}{\|f^L\|}\right)\right) - \lambda\beta. \tag{2.10}$$

On the other hand, for any λ,

$$\|f^L - \lambda g\| \geq E_L(f) \geq \|f^L\|(1+a)^{-1} \geq \|f^L\|(1-a).$$

Therefore, for any λ,

$$\frac{\lambda\beta}{\|f^L\|} \leq a + b + 2\rho\left(\frac{3\lambda\|f\|}{\|f^L\|}\right).$$

This proves the lemma. ■

We will need the following simple lemma (see [24]):

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

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

Lemma 2.3. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$. Take a number $\varepsilon \geq 0$ and two elements f, f^ε from X such that*

$$\|f - f^\varepsilon\| \leq \varepsilon$$

and

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

with some number $A(\varepsilon)$. Then for the AWCGA with τ, δ, η we have, for $m = 1, 2, \ldots,$

$$E_m(f) \;\leq\; \|f_{m-1}\| \inf_\lambda \left(1 + \delta_{m-1} - \lambda t_m A(\varepsilon)^{-1}\left(1 - \delta_{m-1} - \frac{\beta_{m-1}+\varepsilon}{\|f_{m-1}\|}\right)\right.$$
$$\left. +\; 2\rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right),$$

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

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

Proof. We have, for any λ,

$$\|f_{m-1} - \lambda\varphi_m\| + \|f_{m-1} + \lambda\varphi_m\| \leq 2\|f_{m-1}\|\left(1 + \rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right) \tag{2.11}$$

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

$$F_{m-1}(\varphi_m) \geq t_m \sup_{g\in\mathcal{D}} F_{m-1}(g) = t_m \sup_{\varphi\in\mathcal{A}_1(\mathcal{D})} F_{m-1}(\varphi) \geq t_m A(\varepsilon)^{-1} F_{m-1}(f^\varepsilon).$$

By Lemma 2.1 we obtain

$$\begin{aligned} F_{m-1}(f^\varepsilon) \;&=\; F_{m-1}(f + f^\varepsilon - f) \geq F_{m-1}(f) - \varepsilon \\ &=\; F_{m-1}(f_{m-1} + G_{m-1}) - \varepsilon \geq F_{m-1}(f_{m-1}) - |F_{m-1}(G_{m-1})| - \varepsilon \\ &\geq\; \|f_{m-1}\|(1-\delta_{m-1}) - \beta_{m-1} - \varepsilon. \end{aligned}$$

Thus similarly to (2.9) and (2.10) we get, from (2.11),

$$(2.12)\quad E_m(f) \le \inf_\lambda \|f_{m-1} - \lambda\varphi_m\|$$
$$\le \|f_{m-1}\| \inf_\lambda \left(1 + \delta_{m-1} - \lambda t_m A(\varepsilon)^{-1}\left(1 - \delta_{m-1} - \frac{\beta_{m-1}+\varepsilon}{\|f_{m-1}\|}\right) + 2\rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right),$$

which proves the lemma. ■

Proof of Theorem 2.2. The definition of $\{E_m(f)\}$ implies that it is a nonincreasing sequence. Therefore, we have

$$\lim_{m\to\infty} E_m(f) = \alpha.$$

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

$$\|f_m\| \ge E_m(f) \ge \alpha.$$

We set $\varepsilon = \alpha/4$ and find f^ε such that

$$\|f - f^\varepsilon\| \le \varepsilon \qquad \text{and} \qquad f^\varepsilon/A(\varepsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some $A(\varepsilon)$. It is clear that $\lim_{m\to\infty} \beta_m = 0$. We choose M such that for all $m \ge M$ we have

$$\delta_{m-1} + (\beta_{m-1} + \varepsilon)/\alpha \le \tfrac{1}{2}.$$

Then by Lemma 2.3 we get

$$E_m(f) \le \|f_{m-1}\| \inf_\lambda (1 + \delta_{m-1} - \lambda t_m A(\varepsilon)^{-1}/2 + 2\rho(\lambda/\alpha)).$$

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

$$E_m(f) \le \|f_{m-1}\|(1 + \delta_{m-1} - 2\theta t_m \xi_m)$$

and

$$\|f_m\| \le \|f_{m-1}\|(1 + \delta_{m-1} - 2\theta t_m \xi_m)(1 + \eta_m).$$

Using assumption (2.5) we get for big enough m that

$$(1 + \delta_{m-1} - 2\theta t_m \xi_m)(1 + \eta_m) \le 1 - \theta t_m \xi_m.$$

The assumption

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

implies that

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

We get a contradiction which proves the theorem. ■

We now proceed to study the rate of convergence of the AWCGA. The following theorem has been proved in [24] for the WCGA:

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 $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k \le 1$, $k = 1, 2, \ldots,$ we have for any $f \in \mathcal{A}_1(\mathcal{D})$ that*

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

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

Remark 2.2. It follows from the proof of Theorem 2.3 in [24] that

$$C(q,\gamma) = (2(4\gamma)^{1/(q-1)})^{1/p} \le C\gamma^{1/q}$$

with absolute constant C.

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

Theorem 2.4. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let a weakness sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k \le 1$, $k = 1, 2, \ldots,$ be such that*

$$\sum_{k=1}^{\infty} t_k^p = \infty, \qquad p = \frac{q}{q-1}.$$

For a given $f \in \mathcal{A}_1(\mathcal{D})$ apply the AWCGA with

$$\begin{aligned} \delta_{m-1} &:= t_m^p \|f_{m-1}\|^p 3^{-p}(16A_q)^{-1}, \qquad && m = 1, 2, \ldots, \\ \eta_{m-1} &:= t_m^p E_{m-1}(f)^p 3^{-p}(16A_q)^{-1}, \qquad && m = 2, \ldots, \end{aligned}$$

where

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

Then we have

$$\|f_m^{\tau,\delta,\eta}\| \le C\gamma^{1/q}\left(1 + \sum_{k=1}^{m} t_k^p\right)^{-1/p}, \qquad p := \frac{q}{q-1},$$

with absolute constant C.

Proof. By Lemma 2.3 with $\varepsilon = 0$ and $A(\varepsilon) = 1$ we have for $f \in \mathcal{A}_1(\mathcal{D})$ that

$$E_m(f) \le \|f_{m-1}\| \inf_\lambda \left(1 + \delta_{m-1} - \lambda t_m (1 - \delta_{m-1} - \beta_{m-1}/\|f_{m-1}\|) + 2\gamma \left(\frac{\lambda}{\|f_{m-1}\|}\right)^q\right). \tag{2.13}$$

We estimate β_{m-1} by choosing

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

We have

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

Using $\delta_{m-1} \le \frac{1}{16}$ we get, from (2.13),

$$E_m(f) \le \|f_{m-1}\| \inf_\lambda \left(1 + \delta_{m-1} - \tfrac{9}{16}\lambda t_m + 2\gamma \left(\frac{\lambda}{\|f_{m-1}\|}\right)^q\right). \tag{2.14}$$

We choose λ from the equation

$$\tfrac{1}{4}\lambda t_m = 2\gamma \left(\frac{\lambda}{\|f_{m-1}\|}\right)^q$$

which implies that

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

With this λ using the notation $p := q/(q-1)$ we get, from (2.14),

$$\begin{aligned} E_m(f) &\le \|f_{m-1}\|(1 + \delta_{m-1} - \tfrac{5}{16}\lambda t_m) \le \|f_{m-1}\|(1 - t_m^p \|f_{m-1}\|^p/A_q) \\ &\le E_{m-1}(f)(1 + t_m^p E_{m-1}(f)^p/(2A_q))(1 - t_m^p \|f_{m-1}\|^p/A_q) \\ &\le E_{m-1}(f)(1 - t_m^p E_{m-1}(f)^p/(2A_q)). \end{aligned}$$

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

$$E_m(f)^p \le E_{m-1}(f)^p (1 - t_m^p E_{m-1}(f)^p/(2A_q)).$$

By Lemma 3.1 from [23] using the estimate $\|f\|^p \le 1 < A_q$ we get

$$E_m(f)^p \le 2A_q \left(1 + \sum_{n=1}^m t_n^p\right)^{-1}$$

which implies

$$\|f_m\| \le C\gamma^{1/q} \left(1 + \sum_{n=1}^m t_n^p\right)^{-1/p}.$$

Theorem 2.3 is now proved. ■

We discussed above the performance of the AWCGA. The AWCGA is defined in a way to control relative errors of approximation of norming functional and best approximant (see the definition of the AWCGA). We now discuss a modification of the AWCGA with control of absolute errors of approximation. Let three sequences, $\tau = \{t_k\}_{k=1}^{\infty}$, $\varepsilon = \{\varepsilon_k\}_{k=0}^{\infty}$, $\alpha = \{\alpha_k\}_{k=1}^{\infty}$, of numbers from [0, 1] be given.

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

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

$$\|F_{m-1}\| \leq 1, \qquad F_{m-1}(f_{m-1}) \geq \|f_{m-1}\| - \varepsilon_{m-1},$$

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

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

(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

$$\|f - G_m\| \leq E_m(f) + \alpha_m.$$

(3) Denote

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

The following analog of Theorem 2.2 holds for the AWCGA(a):

Theorem 2.5. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Assume that sequences τ, ε, α satisfy the conditions: For any $\theta > 0$ we have*

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

and

$$\sum_{m=1}^{\infty} \varepsilon_m < \infty \qquad \textit{and} \qquad \sum_{m=1}^{\infty} \alpha_m < \infty.$$

Then for any $f \in X$ we have

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

The proof of this theorem is similar to the proof of Theorem 2.2. We will not present this proof here and remark that the only new ingredient of the proof of Theorem 2.5 is the following simple lemma:

Lemma 2.4. *Let*

$$\sum_{m=1}^{\infty} \gamma_m = \infty, \qquad \sum_{m=1}^{\infty} \alpha_m < \infty, \qquad \alpha_m, \gamma_m \in [0, 1].$$

Assume that a nonnegative sequence $\{x_k\}_{k=0}^{\infty}$ *satisfies the relation*

$$x_m \le x_{m-1}(1 - \gamma_m) + \alpha_m, \qquad m = 1, 2, \dots.$$

Then

$$\lim_{m\to\infty} x_m = 0.$$

3. Convergence and Rate of Approximation of the AWRGA

The following two theorems on the WRGA have been proved in [24]:

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

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

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* $\rho(u) \le \gamma u^q$, $1 < q \le 2$. *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*

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

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

Remark 3.1. It follows from the proof of Theorem 3.2 in [24] that

$$C_1(q, \gamma) \le C\gamma^{1/q}$$

with absolute constant C.

We prove here analogs of Theorems 2.2 and 2.4 for the AWRGA.

Theorem 3.3. *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*

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) = \infty \tag{3.1}$$

and

$$\delta_m = o(t_m \xi_m(\rho, \tau, \theta)) \qquad \text{and} \qquad \eta_m = o(t_m \xi_m(\rho, \tau, \theta)). \tag{3.2}$$

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

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

Corollary 3.1. *In the particular case of $\tau = \{t\}$, $t > 0$, assumption* (3.1) *is satisfied and assumption* (3.2) *takes a form $\delta_m = o(1)$ and $\eta_m = o(1)$. Thus in the case $\tau = \{t\}$, $t > 0$, the AWRGA converges for each $f \in \mathcal{A}_1(\mathcal{D})$ if $\delta, \eta \in c_0$.*

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, for $m = 1, 2, \dots,$*

$$\|f_m^{ar}\| \le \|f_{m-1}^{ar}\| \inf_{0\le\lambda\le 1}\left(1 + \delta_{m-1} - \lambda t_m(1-\delta_{m-1}) + 2\rho\left(\frac{2\lambda}{\|f_{m-1}^{ar}\|}\right)\right)(1+\eta_m).$$

Proof. We have

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

and

$$\|f_m^{ar}\| \le \inf_{0\le\lambda\le 1} \|f_{m-1}^{ar} - \lambda(\varphi_m^{ar} - G_{m-1}^{ar})\|(1+\eta_m).$$

Similarly to (2.11) we have, for any λ,

$$\begin{aligned} \|f_{m-1}^{ar} - \lambda(\varphi_m^{ar} - G_{m-1}^{ar})\| + \|f_{m-1}^{ar} + \lambda(\varphi_m^{ar} - G_{m-1}^{ar})\| \\ \le 2\|f_{m-1}^{ar}\|\left(1 + \rho\left(\frac{\lambda\|\varphi_m^{ar} - G_{m-1}^{ar}\|}{\|f_{m-1}^{ar}\|}\right)\right). \end{aligned} \tag{3.3}$$

Next we get, for $\lambda \ge 0$,

$$\begin{aligned} \|f_{m-1}^{ar} + \lambda(\varphi_m^{ar} - G_{m-1}^{ar})\| &\ge F_{m-1}^{ar}(f_{m-1}^{ar} + \lambda(\varphi_m^{ar} - G_{m-1}^{ar})) \\ &\ge \|f_{m-1}^{ar}\|(1-\delta_{m-1}) + \lambda F_{m-1}^{ar}(\varphi_m^{ar} - G_{m-1}^{ar}) \\ &\ge \|f_{m-1}^{ar}\|(1-\delta_{m-1}) + \lambda t_m \sup_{g\in\mathcal{D}} F_{m-1}^{ar}(g - G_{m-1}^{ar}). \end{aligned}$$

Using Lemma 2.2 we continue

$$\begin{aligned} &= \|f_{m-1}^{ar}\|(1-\delta_{m-1}) + \lambda t_m \sup_{\varphi\in\mathcal{A}_1(\mathcal{D})} F_{m-1}^{ar}(\varphi - G_{m-1}^{ar}) \\ &\ge \|f_{m-1}^{ar}\|(1-\delta_{m-1}) + \lambda t_m \|f_{m-1}^{ar}\|(1-\delta_{m-1}). \end{aligned}$$

Therefore, by a trivial estimate $\|\varphi_m^{ar} - G_{m-1}^{ar}\| \le 2$ we obtain, from (3.3),

$$\text{(3.4)} \quad \|f_{m-1}^{ar} - \lambda(\varphi_m^{ar} - G_{m-1}^{ar})\| \le \|f_{m-1}^{ar}\| \left(1 + \delta_{m-1} - \lambda t_m(1-\delta_{m-1}) + 2\rho\left(\frac{2\lambda}{\|f_{m-1}^{ar}\|}\right)\right),$$

which proves Lemma 3.1. ■

Proof of Theorem 3.3. By the definition the sequence $\{\|f_m^{ar}\|\}$ is nonincreasing. Let

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

Similarly to the proof of Theorem 2.2 we will use the contradiction argument. Assuming $\alpha > 0$ we get, from Lemma 3.1 for big enough m,

$$\text{(3.5)} \quad \|f_m^{ar}\| \le \|f_{m-1}^{ar}\| \inf_{0\le\lambda\le 1} \left(1 + \delta_{m-1} - \lambda t_m/2 + 2\rho\left(\frac{2\lambda}{\|f_{m-1}^{ar}\|}\right)\right)(1+\eta_m).$$

Specifying $\theta = \alpha/16$ and taking $\lambda = \alpha\xi_m(\rho, \tau, \theta)/2$ we obtain, from (3.5),

$$\text{(3.6)} \qquad \|f_m^{ar}\| \le \|f_{m-1}^{ar}\|(1 + \delta_{m-1} - 2\theta t_m \xi_m)(1+\eta_m).$$

The remaining part of the proof repeats the arguments from the proof of Theorem 2.2. ■

Theorem 3.4. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Let a weakness sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k \le 1$, $k = 1, 2, \dots,$ be such that*

$$\sum_{k=1}^{\infty} t_k^p = \infty, \qquad p = \frac{q}{q-1}.$$

For a given $f \in \mathcal{A}_1(\mathcal{D})$ apply the AWRGA with

$$\delta_{m-1} = \eta_m := t_m^p \|f_{m-1}^{ar}\|^p (2B_q)^{-1}, \qquad m = 1, 2, \dots,$$

where

$$B_q := 8(8\gamma)^{1/(q-1)} 2^p.$$

Then we have

$$\|f_m^{ar,\tau,\delta,\eta}\| \le C\gamma^{1/q}\left(1 + \sum_{k=1}^{m} t_k^p\right)^{-1/p}, \qquad p := \frac{q}{q-1},$$

with absolute constant C.

Proof. Using that $\delta_{m-1} \le \frac{1}{2}$ we get, by Lemma 3.1,

$$\|f_m^{ar}\| \le \|f_{m-1}^{ar}\| \inf_{0\le\lambda\le 1}\left(1+\delta_{m-1}-\lambda t_m/2+2\gamma\left(\frac{2\lambda}{\|f_{m-1}^{ar}\|}\right)^q\right)(1+\eta_m).$$

Choosing λ from the equation

$$\lambda t_m/4 = 2\gamma\left(\frac{2\lambda}{\|f_{m-1}^{ar}\|}\right)^q$$

we find

$$\lambda = (8\gamma)^{-1/(q-1)}2^{-p}t_m^{1/(q-1)}\|f_{m-1}^{ar}\|^p < 1$$

and

$$\|f_m^{ar}\| \le \|f_{m-1}^{ar}\|(1+\delta_{m-1}-2t_m^p\|f_{m-1}^{ar}\|^p/B_q)(1+\eta_m).$$

Using the definition of δ_{m-1}, η_m, and B_q we obtain

$$\|f_m^{ar}\| \le \|f_{m-1}^{ar}\|(1-t_m^p\|f_{m-1}^{ar}\|^p/B_q)$$

and complete the proof in the same way as in the proof of Theorem 2.4. ∎

Let us compare Theorem 3.3 with Theorem 3.4 from [8] (see Theorem 3.5 below). We recall some definitions from [8]. An incremental sequence is any sequence $a_1, a_2, \dots$ of X so that $a_1 \in \mathcal{D}$ and for each $n \ge 1$ there are some $g_n \in \mathcal{D}$ and $\lambda_n \in [0,1]$ so that

$$a_n = (1-\lambda_n)a_{n-1} + \lambda_n g_n \qquad (a_0 = 0).$$

We say that an incremental sequence $a_1, a_2, \dots$ is ε-greedy (with respect to f) if ($a_0 = 0$),

$$\|f-a_n\| < \inf_{\lambda\in[0,1];\,g\in\mathcal{D}}\|f-((1-\lambda)a_{n-1}+\lambda g)\| + \varepsilon_n, \qquad n = 1, 2, \dots. \tag{3.7}$$

Theorem 3.5 [8]. *Let X be a uniformly smooth Banach space, and let $\varepsilon = \{\varepsilon_n\}_{n=1}^\infty$ be such that*

$$\sum_{n=1}^{\infty}\varepsilon_n < \infty. \tag{3.8}$$

Then any ε-greedy (with respect to f) incremental sequence $\{a_n\}_{n=1}^\infty$ converges to f.

In order to find a sequence $\{a_n\}$ satisfying (3.7) one should solve a sequence of optimization problems:

$$\inf_{\lambda\in[0,1];\,g\in\mathcal{D}}\|f-((1-\lambda)a_{n-1}+\lambda g)\| + \varepsilon_n, \qquad n = 1, 2, \dots, \tag{3.9}$$

within accuracy ε_n satisfying (3.8). It is clear that the most difficult part of (3.9) is an optimization over $g \in \mathcal{D}$. The most important advantage of the WRGA and AWRGA is that they provide a way of obtaining a good element φ_m^r (or φ_m^{ar}) from the dictionary by checking much *weaker condition that being optimal* within accuracy ε_n. In the AWRGA the way of obtaining a φ_m^{ar} consists of two steps: First, we find an approximation of the norming (peak) functional of the residual f_{m-1}^{ar} with high accuracy ($F_{m-1}^{ar}(f_{m-1}^{ar}) \geq \|f_{m-1}^{ar}\|(1-\delta_{m-1})$); second, we look for φ_m^{ar} satisfying a very weak (compared to being optimal) condition

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

Another place in the AWRGA where we need high accuracy is the optimization over $\lambda \in [0, 1]$. Clearly, the above two tasks with high accuracy are much easier than the above selection of a dictionary element φ_m^{ar}.

4. Constructive Nonlinear Trigonometric m-Term Approximation

We describe the approximation method in detail in the univariate case. Consider the real $L_p(\mathbb{T})$ space with

$$\|f\|_p := \left(\frac{1}{\pi}\int_{\mathbb{T}} |f(x)|^p dx\right)^{1/p}, \qquad 1 \leq p < \infty,$$
$$\|f\|_\infty := \max_{x\in\mathbb{T}} |f(x)|, \qquad f\text{-continuous.}$$

Let $1 \leq p < \infty$. Denote $\mathcal{T}_p$ the real trigonometric system normalized in L_p,

$$2^{-1/p}, c_p \sin x, c_p \cos x, \dots,$$

where

$$c_p = \left(\frac{1}{\pi}\int_{\mathbb{T}} |\sin x|^p \, dx\right)^{-1/p}.$$

It is clear that $C^1 \leq c_p \leq C^2$ with two absolute constants C^1 and C^2. Let $\mathcal{T}(N)$ denote the set of trigonometric polynomials of order N.

We discuss first a simpler construction based on the particular case of $p = 4$ in order to illustrate the idea of the construction. For a trigonometric polynomial

$$t(x) = a_0/2 + \sum_{k=1}^{N} (a_k \cos kx + b_k \sin kx)$$

denote

$$\|t\|_A := |a_0| + \sum_{k=1}^{N} (|a_k| + |b_k|).$$

Then, by Theorems 2.3 (or 2.4) and 3.2 (or 3.4), each of the algorithms WCGA (or AWCGA), WRGA (or AWRGA) with $\tau = \{\frac{1}{2}\}$, $q = 2$ provide a constructive way of approximation in the L_4-norm: For any $t \in \mathcal{T}(N)$ we get an m-term trigonometric polynomial $G_m(t) \in \mathcal{T}(N)$ such that

$$\|t - G_m(t)\|_4 \leq C_1' m^{-1/2} \|t\|_A \tag{4.1}$$

with absolute constant C_1. By the Nikol'skii inequality this implies

$$\|t - G_m(t)\|_\infty \le C_2 N^{1/4} m^{-1/2} \|t\|_A \tag{4.2}$$

with absolute constant C_2.

We will build our constructive approximation operators $A^k(N, m)$ inductively from level $k = 1$ up to an arbitrary level k. We begin with the level $k = 1$. We set, for $t \in \mathcal{T}(N)$,

$$A^1(N, m)(t) := G_m(t).$$

Then (4.2) implies, for $m \le N$,

$$\|t - A^1(N, m)(t)\|_\infty \le C_2 N^{1/4} m^{-1/2} \|t\|_A \le A_1 N^{1/4} (N/m)^{1/2} m^{-1/2} \|t\|_A. \tag{4.3}$$

We continue the construction inductively. Suppose we have built operators $A^k(N, m)$, such that for any $t \in \mathcal{T}(N)$,

$$\|t - A^k(N, m)(t)\|_\infty \le A_k N^{2^{-k-1}} (N/m)^{1/2} m^{-1/2} \|t\|_A. \tag{4.4}$$

We will build operators $A^{k+1}(N, m)$ and will control the constant A_{k+1}. We will carry on the construction for even numbers m.

Step 1. Let $t \in \mathcal{T}(N)$. We approximate t using (4.1),

$$\|t - G_{m/2}(t)\|_4 \le C_1 m^{-1/2} \|t\|_A.$$

Denote

$$h := (t - G_{m/2}(t)) / \|t - G_{m/2}(t)\|_4.$$

Step 2. Take a positive number D and decompose

$$h = h^D + h_D, \qquad h_D(x) := \begin{cases} h(x) & \text{if } |h(x)| \le D, \\ 0 & \text{otherwise.} \end{cases}$$

We need the following simple well-known lemma:

Lemma 4.1. *Assume $p \in [2, \infty)$ and $\|f\|_p = 1$. Then*

$$\|f_D\|_\infty \le D \qquad \textit{and} \qquad \|f^D\|_2 \le D^{1-p/2}.$$

By Lemma 4.1 with $p = 4$ we get

$$\|h_D\|_\infty \le D \qquad \text{and} \qquad \|h^D\|_2 \le D^{-1}.$$

We would like to work with trigonometric polynomials instead of working with h_D and h^D. Let V_N be the de La Vallée Poussin operator. Consider $V_N(h_D)$ and $V_N(h^D)$. We have

$$h = V_N(h) = V_N(h_D) + V_N(h^D)$$

and

$$\|V_N(h_D)\|_\infty \le 3D, \qquad \|V_N(h^D)\|_2 \le D^{-1}, \qquad \|V_N(h^D)\|_A \le 2N^{1/2}D^{-1}.$$

Step 3. We approximate $V_N(h^D) \in \mathcal{T}(2N)$ using operators from level k. By (4.4) we have

$$\|V_N(h^D) - A^k(2N, m/2)(V_N(h^D))\|_\infty \le A_k(2N)^{2^{-k-1}} 3(N/m)^{1/2} m^{-1/2} \|V_N(h^D)\|_A.$$

For $t \in \mathcal{T}(N)$ define

$$A^{k+1}(N, m, D)(t) := G_{m/2}(t) + \|t - G_{m/2}(t)\|_4 A^k(2N, m/2)(V_N(h^D)).$$

Taking into account that $h \in \mathcal{T}(N)$ we get

$$\begin{aligned} t - A^{k+1}(N, m, D)(t) &= h\|t - G_{m/2}(t)\|_4 - \|t - G_{m/2}(t)\|_4 A^k(2N, m/2)(V_N(h^D)) \\ &= \|t - G_{m/2}(t)\|_4 (h - A^k(2N, m/2)(V_N(h^D))) \\ &= \|t - G_{m/2}(t)\|_4 (V_N(h_D) + V_N(h^D) \\ &\qquad - A^k(2N, m/2)(V_N(h^D))). \end{aligned}$$

Therefore,

$$\begin{aligned} (4.5)\quad & \|t - A^{k+1}(N, m, D)(t)\|_\infty \\ &\le \|t - G_{m/2}(t)\|_4 (3D + A_k(2N)^{2^{-k-1}} 6(N/m)^{1/2} m^{-1/2} N^{1/2} D^{-1}) \\ &\le (3D + A_k(2N)^{2^{-k-1}} 6(N/m) D^{-1}) C_1 m^{-1/2} \|t\|_A. \end{aligned}$$

Step 4. Choose

$$D = D(N, m, k) := (2A_k(2N)^{2^{-k-1}}(N/m))^{1/2}.$$

By (4.5) we obtain

$$\|t - A^{k+1}(N, m, D)(t)\|_\infty \le A'_{k+1} N^{2^{-k-2}} (N/m)^{1/2} m^{-1/2} \|t\|_A \tag{4.6}$$

with

$$A'_{k+1} := 6C_1 2^{1/2} 2^{2^{-k-2}} A_k^{1/2} \le C_3 A_k^{1/2}. \tag{4.7}$$

We recall that we have proved (4.6) with the constant A'_{k+1} from (4.7) under assumption that m is an even number. We complete the construction by setting

$$A^{k+1}(N, m) := A^{k+1}(N, 2[m/2], D(N, 2[m/2], k)), \qquad m \ge 2.$$

Clearly (4.6) implies

$$\|t - A^{k+1}(N, m, D)(t)\|_\infty \le A_{k+1} N^{2^{-k-2}} (N/m)^{1/2} m^{-1/2} \|t\|_A \tag{4.8}$$

for all m with $A_{k+1} = 2A'_{k+1}$. Relation (4.7) combined with $A_1 = C_2$ (see (4.3)) implies that $A_k \le C_4$ for all k.

Let N be given. Choose k satisfying $2^{k+1} \geq \ln N$. Then (4.4) gives, for any $t \in \mathcal{T}(N)$, the estimate

$$\|t - A^k(N,m)(t)\|_\infty \leq C_5(N^{1/2}/m)\|t\|_A \tag{4.9}$$

for any m.

We now proceed to a more elaborate construction that gives the following estimate:

Theorem 4.1. *There exists a constructive method $A(N,m)$ such that, for any $t \in \mathcal{T}(N)$, it provides an m-term trigonometric polynomial $A(N,m)(t)$ with the following approximation property*:

$$\|t - A(N,m)(t)\|_\infty \leq Cm^{-1/2}(\ln(1 + N/m))^{1/2}\|t\|_A \tag{4.10}$$

with an absolute constant C.

Proof. We will construct an analog of the sequence of operators $\{A^k(N,m)\}$ constructed above. A new ingredient of this construction is the following. We will now approximate t in the L_p-norm, $p \in [4,\infty)$, instead of the L_4-norm and will optimize over p.

Let N and m be given and let $t \in \mathcal{T}(N)$. We use either the WCGA (AWCGA) or the WRGA (AWRGA) with $\tau = \{\frac{1}{2}\}$, $q = 2$, $\mathcal{D}_N = \mathcal{T}_p \cap \mathcal{T}(N)$ to approximate t by an m-term trigonometric polynomial in the L_p-norm, $p \in [4,\infty)$. By Theorems 2.3 (2.4) or 3.2 (3.4), with $X = \mathcal{T}(N)_p$ where $\mathcal{T}(N)_p$ is the $\mathcal{T}(N)$ equipped with the L_p-norm, we get

$$\|t - G_m^p(t)\|_p \leq C_6 C(2,\gamma)m^{-1/2}\|t\|_A. \tag{4.11}$$

Let us estimate the constant $C(2,\gamma)$. By (1.5) we obtain $\gamma = (p-1)/2$. Thus by Remark 2.2 or Remark 3.1 we get

$$C_6 C(2,\gamma) \leq C_7 p^{1/2}. \tag{4.12}$$

We define the level $k = 1$ algorithms $A_p^1(N,m)$ by

$$A_p^1(N,m)(t) = G_m^p(t), \qquad t \in \mathcal{T}(N). \tag{4.13}$$

We note that by construction $A_p^1(N,m)(t) \in \mathcal{T}(N)$. By the Nikol'skii inequality we get, from (4.11)–(4.13),

$$\begin{aligned} \|t - A_p^1(N,m)(t)\|_\infty &\leq C_8 p^{1/2} N^{1/p} m^{-1/2}\|t\|_A \\ &\leq C_8 p^{1/2} N^{1/4}(N/m)^{1/(p-2)} m^{-1/2}\|t\|_A, \qquad m \leq N. \end{aligned} \tag{4.14}$$

We note here that taking $p_N := \ln N$ we get, from the first inequality in (4.14),

$$\|t - A_p^1(N,m)(t)\|_\infty \leq C(\ln N)^{1/2} m^{-1/2}\|t\|_A \tag{4.15}$$

with an absolute constant C. Thus the rest of the proof will be devoted to replacing $\ln N$ by $\ln(1 + N/m)$ in (4.15).

As in the case $p = 4$ we continue the construction by induction. Suppose we have built operators $A_p^k(N, m)$ such that, for any $t \in \mathcal{T}(N)$, $p \in [4, \infty)$,

$$\|t - A_p^k(N, m)\|_\infty \le A_k^p N^{2^{-k-1}} (N/m)^{1/(p-2)} m^{-1/2} \|t\|_A. \tag{4.16}$$

We will make steps similar to those from above.

Step 1. Let $t \in \mathcal{T}(N)$ and let m be an even number. We approximate t using (4.11), (4.12),

$$\|t - G_{m/2}^p(t)\|_p \le C_9 p^{1/2} m^{-1/2} \|t\|_A. \tag{4.17}$$

Denote

$$h[p] := (t - G_{m/2}^p(t))/\|t - G_{m/2}^p(t)\|_p.$$

Step 2. Take a positive number D and decompose

$$h[p] = h^D[p] + h_D[p].$$

By Lemma 4.1 we get

$$\|h_D[p]\|_\infty \le D \qquad \text{and} \qquad \|h^D[p]\|_2 \le D^{1-p/2}$$

and, therefore,

$$\begin{aligned} \|V_N(h_D[p])\|_\infty &\le 3D, \qquad \|V_N(h^D[p])\|_2 \le D^{1-p/2}, \\ \|V_N(h^D[p])\|_A &\le 2N^{1/2} D^{1-p/2}. \end{aligned}$$

Step 3. We approximate $V_N(h^D) \in \mathcal{T}(2N)$ using operators from level k. By (4.16) we have

$$\begin{aligned} &\|V_N(h^D[p]) - A_p^k(2N, m/2)(V_N(h^D[p]))\|_\infty \\ &\qquad \le A_k^p (2N)^{2^{-k-1}} (4N/m)^{1/(p-2)} 2^{1/2} m^{-1/2} \|V_N(h^D)\|_A. \end{aligned}$$

For $t \in \mathcal{T}(N)$ define

$$A_p^{k+1}(N, m, D)(t) := G_{m/2}^p(t) + \|t - G_{m/2}^p(t)\|_p A_p^k(2N, m/2)(V_N(h^D[p])).$$

Similarly to the case $p = 4$ (see (4.5)) we get

$$\begin{aligned} &\|t - A_p^{k+1}(N, m, D)(t)\|_\infty \\ &\qquad \le \|t - G_{m/2}^p(t)\|_p (3D + A_k^p (2N)^{2^{-k-1}} 6 (N/m)^{p/(2(p-2))} D^{1-p/2}). \end{aligned} \tag{4.18}$$

Step 4. Choose

$$D_p = D_p(N, m, k) := (2A_k^p (2N)^{2^{-k-1}} (N/m)^{p/(2(p-2))})^{2/p}.$$

By (4.17) we obtain from (4.18) for even m,

$$\|t - A_p^{k+1}(N, m, D_p)(t)\|_\infty \le A_{k+1}^{p,1} N^{2^{-k-2}} (N/m)^{1/(p-2)} m^{-1/2} \|t\|_A \tag{4.19}$$

with

$$A^{p,1}_{k+1} \le C_{10} p^{1/2} (A^p_k)^{2/p}, \qquad A^p_1 \le C_{11} p^{1/2}. \tag{4.20}$$

We note that (4.20) implies

$$A^p_k \le C_{12} p^{p/(2(p-2))}. \tag{4.21}$$

We set

$$A^{k+1}_p(N,m) := A^{k+1}_p(N, 2[m/2], D_p(N, 2[m/2], k))$$

and obtain (4.16) with k replaced by $k+1$ and a constant $A^p_{k+1} = 2A^{p,1}_{k+1}$.

Let N and m be given. First we choose k satisfying $2^{k+1} \ge \ln N$. Next we choose $p = 2 + \ln(1 + N/m)$. Then (4.16) and (4.21) give, for any $t \in \mathcal{T}(N)$, the estimate

$$\|t - A^k_p(N,m)(t)\|_\infty \le C_{13} m^{-1/2} (\ln(1 + N/m))^{1/2} \|t\|_A \tag{4.22}$$

for any m. This completes the proof of Theorem 4.1. ■

The same technique can also be used in the multivariate case. Let $L_p(\mathbb{T}^d)$ be the real Banach space with

$$\|f\|_p := \left(\frac{1}{\pi^d}\int_{\mathbb{T}^d} |f(x)|^p\,dx\right)^{1/p}, \qquad 1 \le p < \infty,$$
$$\|f\|_\infty := \max_{x\in\mathbb{T}^d} |f(x)|, \qquad f\text{-continuous}.$$

Denote $\mathcal{T}^d := \mathcal{T} \times \cdots \times \mathcal{T}$ (d times) the real multivariate trigonometric system. Let $\mathbf{N} = (N_1, \dots, N_d)$. Denote $\mathcal{T}(\mathbf{N})$ the space of trigonometric polynomials with degree N_j in the variable x_j, $j = 1, \dots, d$. Let $\upsilon(\mathbf{N})$ be the dimension of $\mathcal{T}(\mathbf{N})$. We formulate a generalization of Theorem 4.1 for the d-dimensional case and note that the proof repeats the proof of Theorem 4.1.

Theorem 4.2. *There exists a constructive method $A(\mathbf{N}, m)$ such that, for any $t \in \mathcal{T}(\mathbf{N})$, it provides an m-term trigonometric polynomial $A(\mathbf{N}, m)(t)$ with the following approximation property*:

$$\|t - A(\mathbf{N}, m)(t)\|_\infty \le C(d) m^{-1/2} (\ln(1 + \upsilon(\mathbf{N})/m))^{1/2} \|t\|_A \tag{4.23}$$

with a constant $C(d)$ which may depend on d.

This theorem can be applied to studying an m-term trigonometric approximation of function classes. We will consider here some examples. In paper [5] the following two types of function classes were studied from the point of view of best m-term trigonometric approximation. We begin with the first class. For $0 < \alpha < \infty$ and $0 < q \le \infty$, let $\mathcal{F}^\alpha_q$ denote the class of those functions in $L_1(\mathbb{T}^d)$ such that

$$|f|_{\mathcal{F}^\alpha_q} := \left(\sum_{k\in\mathbb{Z}^d} (\max(1, |k_1|, \dots, |k_d|)^{\alpha q} (|\hat{f}(k)|^q)\right)^{1/q} \le 1.$$

The following theorem has been proved in [5]:

Theorem 4.3. *If $\alpha > 0$ and $\lambda := \alpha/d + 1/q - \frac{1}{2}$, then for all $1 \le p \le \infty$ and all $0 < q \le \infty$,*

$$C_1 m^{-\lambda} \le \sigma_m(\mathcal{F}_q^\alpha, \mathcal{T}^d)_p \le C_2 m^{-\lambda}, \qquad \alpha > d(1-1/q)_+,$$

with $C_1, C_2 > 0$ constants depending only on d, α, q.

The second class is defined as follows. Let $\alpha > 0$, $0 < \tau, s \le \infty$, and $B_s^\alpha(L_\tau)$ denote the class of functions such that there exist trigonometric polynomials T_n of coordinate degree 2^n with the properties

$$f = \sum_{n=0}^{\infty} T_n, \qquad \|\{2^{n\alpha}\|T_n\|_\tau\}_{n=0}^\infty\|_{\ell_s(\mathbb{Z})} \le 1.$$

The following theorem has been proved in [5] for these classes:

Theorem 4.4. *Let $1 \le p \le \infty$, $0 < \tau, s \le \infty$, and define*

$$\alpha(p,\tau) := \begin{cases} d(1/\tau - 1/p)_+, & 0 < \tau \le p \le 2 \text{ and } 1 \le p \le \tau \le \infty, \\ \max(d/\tau, d/2), & \text{otherwise}. \end{cases}$$

Then for $\alpha > \alpha(p,\tau)$, we have

$$C_1 m^{-\mu} \le \sigma(B_s^\alpha(L_\tau), \mathcal{T}^d)_p \le C_2 m^{-\mu}, \qquad \mu := \alpha/d - (1/\tau - \max(1/p, \tfrac{1}{2}))_+,$$

with C_1, C_2 depending only on α, p, τ, and d.

It was proved in [21] that in the case $1 \le p \le 2$ the rate of best m-term approximation in Theorem 4.3 can be realized by the Thresholding Greedy Algorithm $G_m(\cdot, \mathcal{T}^d)$, that is, by a constructive method. It is well-known that for approximation by trigonometric polynomials of degree $m^{1/d}$ in each variable one has

$$E_m(B_s^\alpha(L_\tau), \mathcal{T}^d)_p := \sup_{f \in B_s^\alpha(L_\tau)} E_m(f, \mathcal{T}^d)_p \asymp m^{-\alpha/d + (1/\tau - 1/p)_+} \tag{4.24}$$

provided $\alpha/d - (1/\tau - 1/p)_+ > 0$. Comparing (4.24) with Theorem 4.4 we conclude that in the case $0 \le \tau \le p \le 2$ or $1 \le p \le \tau \le \infty$ the rate of $\sigma_m(B_s^\alpha(L_\tau), \mathcal{T}^d)_p$ can be realized by approximation by trigonometric polynomials of degree $m^{1/d}$ in each variable. Thus in the case $0 \le \tau \le p \le 2$ or $1 \le p \le \tau \le \infty$ there is a simple constructive method that realizes $\sigma_m(B_s^\alpha(L_\tau), \mathcal{T}^d)_p$. The remaining case is $1 \le \tau < p \le \infty$, $2 < p \le \infty$. In a subcase of the remaining case when $p < \infty$ it has been shown in [7] that the WCGA (or WRGA) can be used to build a constructive method of realizing $\sigma_m(B_s^\alpha(L_\tau), \mathcal{T}^d)_p$. This was done in the following way. In [5] the only nonconstructive step of the proof of Theorems 4.3 and 4.4 in the case $2 < p \le \infty$ was hidden in the following inequality (see [5, Corollary 5.1]):

$$\sigma_m(A_1(\mathcal{T}_n^d), \mathcal{T}^d)_\infty \le C m^{-1/2} \left(1 + \ln^+ \frac{n^d}{m}\right)^{1/2}, \tag{4.25}$$

where $\mathcal{T}_n^d$ denotes the subsystem of the trigonometric system $\mathcal{T}^d$ which forms a basis for the space of trigonometric polynomials of coordinate degree n. The inequality (4.25) was proved in [5] with the help of the following Gluskin theorem [9].

Theorem 4.5. *There exist absolute constants C_1 and $0 < \delta < 1$ such that for any finite collection V of M vectors from the unit Euclidean ball B_2^N of $\mathbb{R}^N$, there is a vector $z \in \mathbb{R}^N$ with $|z_i| = 0, 1$, $i = 1, \dots, N$, $\|z\|_{\ell_1^N} \geq \delta N$, and*

$$\max_{v \in V} |\langle v, z\rangle| \leq C_1 \left(1 + \ln^+ \frac{M}{N}\right)^{1/2}.$$

It was pointed out in [7] that in the case $2 < p < \infty$ the WCGA with the weakness sequence $\{t\}$, $t \in (0, 1]$, provides a constructive way to get an analog of (4.25). This follows immediately from Theorem 2.3: For $f \in A_1(\mathcal{T}_n^d)$ we have

$$\|f_m^{c,t}\|_p \leq C(p, t) m^{-1/2}, \qquad 2 \leq p < \infty. \tag{4.26}$$

Thus the only nonconstructive step in the proof of upper estimates in Theorems 4.3 and 4.4 was made constructive for $p < \infty$.

In the same way as in [7] one can use Theorem 4.2 instead of (4.26) to make the proofs of Theorems 4.3 and 4.4 [5] constructive in the case $p = \infty$. Therefore, we now have constructive proofs of Theorems 4.3 and 4.4 in all cases. It is interesting to compare this situation with the situation on finding a constructive proof for Kolmogorov's widths of the above function classes. We will make a comment only on classes $B_s^\alpha(L_\tau)$ in the case $\tau = 2$, $p = \infty$. We recall the definition of the Kolmogorov width

$$d_m(F, X) := \inf_{\varphi_1, \dots, \varphi_m} \sup_{f \in F} \inf_{c_1, \dots, c_m} \left\| f - \sum_{j=1}^m c_j \varphi_j \right\|.$$

By Kashin's [13] result

$$d_m(B_s^\alpha(L_2), L_\infty) \asymp m^{-\alpha/d}, \qquad \alpha > d/2. \tag{4.27}$$

Estimate (4.27) is only an existence theorem and it is an interesting open problem to find a constructive proof (constuct $\varphi_1, \dots, \varphi_m$) of (4.27).

One can check that the proof of Theorem 4.1 works in the following more general situation. Let $\Phi := \{\varphi_j\}_{j=1}^\infty$ be a uniformly bounded orthonormal system defined on a bounded domain. Denote

$$\Phi(N) := \operatorname{span}\{\varphi_1, \dots, \varphi_N\}$$

and assume that the system Φ admits a sequence of the de La Vallée Poussin operators:

(VP) There exist two positive constants K_1 and K_2 such that for any N there is an operator V_N^Φ with the properties

$$V_N^\Phi(\varphi_j) = \lambda_{N,j}\varphi_j,$$

$$\lambda_{N,j} = 1 \quad \text{for } j \in [1, N], \qquad \lambda_{N,j} = 0 \quad \text{for } j > K_1 N,$$

$$\|V_N^\Phi\|_{L_p \to L_p} \leq K_2 \qquad \text{for} \quad 1 \leq p \leq \infty \text{ and all } N. \tag{4.28}$$

For a system Φ having the (VP) property we can easily derive from (4.28) and the uniform boundedness of Φ that

$$\|V_N^{\Phi}\|_{L_2 \to L_\infty} \le CN^{1/2}.$$

By the interpolation theory of operators we get, from here and from (4.28) with $p = \infty$, that

$$\|V_N^{\Phi}\|_{L_p \to L_\infty} \le CN^{1/p}, \qquad p \in (2, \infty).$$

The last inequality implies the Nikol'skii inequality

$$\|\varphi\|_\infty \le CN^{1/p}\|\varphi\|_p, \qquad \varphi \in \Phi(N), \quad p \in (2, \infty).$$

Thus Φ has all the properties needed in the proof of Theorem 4.1. Therefore, we have the following generalization of Theorem 4.1. Denote

$$\left\| \sum_{j=1}^{N} c_j \varphi_j \right\|_A := \sum_{j=1}^{N} |c_j|.$$

Theorem 4.6. *Let $\Phi := \{\varphi_j\}_{j=1}^{\infty}$ be a uniformly bounded orthonormal system defined on a bounded domain. Assume Φ has the* (VP) *property. Then there exists a constructive method $A(\Phi, N, m)$ such that for any $\varphi \in \Phi(N)$ it provides an m-term Φ-polynomial $A(\Phi, N, m)(\varphi)$ with the following approximation property*:

$$\|\varphi - A(\Phi, N, m)(\varphi)\|_\infty \le Cm^{-1/2}(\ln(1 + N/m))^{1/2}\|\varphi\|_A$$

with a constant C which may depend on Φ.

We note that the decomposition technique used in the proof of Theorem 4.1 is a standard tool in the interpolation of operators. The idea of combining the decomposition technique with an inductive way of constructing approximations is also known in approximation theory. For instance, it has been used recently in [3].

5. The Discrepancy Estimates

Let $1 \le p < \infty$. We recall the definition of the L_p discrepancy (the L_p-star discrepancy) of points $\{\xi^1, \dots, \xi^m\} \subset \Omega_d := [0, 1]^d$. Let $\chi_{[a,b]}(\cdot)$ be a characteristic function of the interval $[a, b]$. Denote, for $x, y \in \Omega_d$,

$$B(x, y) := \prod_{j=1}^{d} \chi_{[0,x_j]}(y_j).$$

Then the L_p discrepancy of $\xi := \{\xi^1, \dots, \xi^m\} \subset \Omega_d$ is defined by

$$D(\xi, m, d)_p := \left\| \int_{\Omega_d} B(x, y)\, dy - \frac{1}{m} \sum_{\mu=1}^{m} B(x, \xi^\mu) \right\|_{L_p(\Omega_d)}.$$

It will be convenient for us to study a slight modification of $D(\xi, m, d)_p$. For $a, t \in [0, 1]$ denote

$$H(a, t) := \chi_{[0,a]}(t) - \chi_{[a,1]}(t),$$

and, for $x, y \in \Omega_d$,

$$H(x, y) := \prod_{j=1}^{d} H(x_j, y_j).$$

We define the symmetrized L_p discrepancy by

$$D^s(\xi, m, d)_p := \left\| \int_{\Omega_d} H(x, y)\, dy - \frac{1}{m} \sum_{\mu=1}^{m} H(x, \xi^\mu) \right\|_{L_p(\Omega_d)}.$$

The L_∞ discrepancies $D(\xi, m, d)_\infty$ and $D^s(\xi, m, d)_\infty$ are defined in the same way with the L_p-norm replaced by the L_∞-norm.

Using the identity

$$\chi_{[0,x_j]}(y_j) = \tfrac{1}{2}(H(1, y_j) + H(x_j, y_j))$$

we get a simple inequality

$$D(\xi, m, d)_\infty \le D^s(\xi, m, d)_\infty. \tag{5.1}$$

We are interested in ξ with small discrepancy. Consider

$$D(m, d)_p := \inf_\xi D(\xi, m, d)_p, \qquad D^s(m, d)_p := \inf_\xi D^s(\xi, m, d)_p.$$

For $1 < p < \infty$ the following relation is known (see [1, p. 5]):

$$D(m, d)_p \asymp m^{-1} (\ln m)^{(d-1)/2} \tag{5.2}$$

with constants in $\asymp$ depending on p and d. The right order of $D(m, d)_p$, $p = 1, \infty$, for $d \ge 3$ is unknown. As we mentioned in the Introduction the following estimate has been obtained in [11]:

$$D(m, d)_\infty \le C d^{1/2} m^{-1/2}. \tag{5.3}$$

It is pointed out in [11] that (5.3) is only an existence theorem and even a constant C in (5.3) is unknown. Their proof is a probabilistic one. There are also some other estimates in [11] with explicit constants. We mention one of them

$$D(m, d)_\infty \le C (d \ln d)^{1/2} ((\ln n)/n)^{1/2} \tag{5.4}$$

with an explicit constant C. The proof of (5.4) is also probabilistic.

In this section we apply greedy-type algorithms to obtain upper estimates of $D(m, d)_p$, $1 \le p \le \infty$, in a style of (5.3) and (5.4). The important feature of our proof is that it is deterministic and, moreover, it is constructive. Formally, the optimization problem

$$D(m, d)_p = \inf_\xi D(\xi, m, d)_p$$

is deterministic: One needs to minimize over $\{\xi^1, \dots, \xi^m\} \subset \Omega_d$. However, minimization by itself does not provide any upper estimate. It is known (see [4]) that simultaneous optimization over many parameters ($\{\xi^1, \dots, \xi^m\}$ in our case) is a very difficult problem. We note that

$$D(m,d)_p = \sigma_m^e(J,\mathcal{B})_p := \inf_{g_1,\dots,g_m\in\mathcal{B}} \left\| J(\cdot) - \frac{1}{m}\sum_{\mu=1}^m g_\mu \right\|_{L_p(\Omega_d)}$$

where

$$J(x) = \int_{\Omega_d} B(x,y)\,dy$$

and

$$\mathcal{B} = \{B(x,y),\ y \in \Omega_d\}.$$

It has been proved in [4] that if an algorithm finds best m-term approximation for each $f \in \mathbb{R}^N$ for every dictionary $\mathcal{D}$ with the number of elements of order N^k, $k \geq 1$, then this algorithm solves an NP-hard problem. Thus, in nonlinear m-term approximation we look for methods (algorithms) which provide approximation close to best m-term approximation and at each step solve an optimization problem over only one parameter (ξ^μ in our case). In this section we will provide such an algorithm for estimating $\sigma_m^e(J,\mathcal{B})_p$. We call this algorithm “constructive” because it provides an explicit construction with feasible one-parameter optimization steps.

We proceed to the construction. In this section we do not assume that a dictionary $\mathcal{D}$ is symmetric: $g \in \mathcal{D}$ implies $-g \in \mathcal{D}$. To indicate this we will use the notation $\mathcal{D}^+$ for such a dictionary. We do not assume that elements of a dictionary $\mathcal{D}^+$ are normalized ($\|g\| = 1$ if $g \in \mathcal{D}^+$) and assume only that $\|g\| \leq 1$ if $g \in \mathcal{D}^+$. By $\mathcal{A}_1(\mathcal{D}^+)$ we denote the closure of the convex hull of $\mathcal{D}^+$. We will use in our construction the IA(ε) which is a slight modification of the corresponding procedure from [8]. One can find results on the application of incremental algorithms in neural networks in [17]. For convenience we repeat here the definition of the IA(ε) from the Introduction. Let $\varepsilon = \{\varepsilon_n\}_{n=1}^\infty$, $\varepsilon_n > 0$, $n = 1, 2, \dots$.

Incremental Algorithm with Schedule ε (IA(ε)). Let $f \in \mathcal{A}_1(\mathcal{D}^+)$. Denote $f_0^{i,\varepsilon} := f$ and $G_0^{i,\varepsilon} := 0$. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_m^{i,\varepsilon} \in \mathcal{D}^+$ *is any satisfying*

$$F_{f_{m-1}^{i,\varepsilon}}(\varphi_m^{i,\varepsilon} - f) \geq -\varepsilon_m.$$

(2) Define

$$G_m^{i,\varepsilon} := (1 - 1/m)G_{m-1}^{i,\varepsilon} + \varphi_m^{i,\varepsilon}/m.$$

(3) Denote

$$f_m^{i,\varepsilon} := f - G_m^{i,\varepsilon}.$$

We note that similarly to Lemma 2.2 we have, for any bounded linear functional F and any $\mathcal{D}^+$,

$$\sup_{g\in\mathcal{D}^+} F(g) = \sup_{f\in\mathcal{A}_1(\mathcal{D}^+)} F(f). \tag{5.5}$$

Therefore, for any F and any $f \in \mathcal{A}_1(\mathcal{D}^+)$,

$$\sup_{g\in\mathcal{D}^+} F(g) \geq F(f).$$

This guarantees the existence of $\varphi_m^{i,\varepsilon}$.

Theorem 5.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Define*

$$\varepsilon_n := K_1\gamma^{1/q}n^{-1/w}, \qquad w = \frac{q}{q-1}, \quad n = 1, 2, \dots.$$

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

$$\|f_m^{i,\varepsilon}\| \leq C(K_1)\gamma^{1/q}m^{-1/w}, \qquad m = 1, 2 \dots.$$

Proof. We will use the abbreviated notation $f_m := f_m^{i,\varepsilon}$, $\varphi_m := \varphi_m^{i,\varepsilon}$, $G_m := G_m^{i,\varepsilon}$. Representing

$$f_m = f_{m-1} - (\varphi_m - G_{m-1})/m$$

we get immedietly the trivial estimate

$$\|f_m\| \leq \|f_{m-1}\| + 2/m. \tag{5.6}$$

Representing

$$\begin{aligned} f_m &= (1 - 1/m)f_{m-1} - (\varphi_m - f)/m \\ &= (1 - 1/m)(f_{m-1} - (\varphi_m - f)/(m-1)) \end{aligned} \tag{5.7}$$

we obtain, in a way similar to (2.10) or (3.4),

$$\begin{aligned} \|f_{m-1} - (\varphi_m - f)/(m-1)\| &\leq \|f_{m-1}\|(1 + 2\rho(2((m-1)\|f_{m-1}\|)^{-1})) \\ &\quad + \varepsilon_m(m-1)^{-1}. \end{aligned} \tag{5.8}$$

Using the definition of ε_m and the assumption $\rho(u) \leq \gamma u^q$ we make the following observation. There exists a constant $C(K_1)$ such that if

$$\|f_{m-1}\| \geq C(K_1)\gamma^{1/q}(m-1)^{-1/w}, \tag{5.9}$$

then

$$2\rho(2((m-1)\|f_{m-1}\|)^{-1}) + \varepsilon_m((m-1)\|f_{m-1}\|)^{-1} \leq 1/(4m) \tag{5.10}$$

and, therefore, by (5.7) and (5.8),

$$\|f_m\| \leq (1 - 3/(4m))\|f_{m-1}\|. \tag{5.11}$$

The following lemma is known ([22]):

Lemma 5.1. *Let three positive numbers $\alpha < \gamma \le 1$, $A > 1$ be given and let a sequence of positive numbers $1 \ge a_1 \ge a_2 \ge \cdots$ satisfy the condition: If for some $\nu \in \mathbb{N}$ we have*

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

then

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

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

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

Remark 5.1. It is easy to check that the proof of Lemma 5.1 from [22] works if we replace the assumption $a_m \le a_{m-1}$ by

$$a_m \le a_{m-1} + C(m-1)^{-\alpha}.$$

Taking into account (5.6) we apply Lemma 5.1 and Remark 5.1 to the sequence $a_n = \|f_n\|$, $n = 1, 2, \ldots$, with $\alpha = 1/w$, $\gamma = \frac{3}{4}$, and complete the proof of Theorem 5.1. ∎

Corollary 5.1. *We apply Theorem 5.1 for $X = L_p(\Omega_d)$, $p \in [2, \infty)$, $\mathcal{D}^+ = \{H(x, y), y \in \Omega_d\}$, $f = J^s(x)$, where*

$$J^s(x) = \int_{\Omega_d} H(x, y)\, dy \in \mathcal{A}_1(\mathcal{D}^+).$$

Using (1.5) we get by Theorem 5.1 a constructive set $\xi^1, \ldots, \xi^m$ such that

$$D^s(\xi, m, d)_p = \|(J^s)_m^{i,\varepsilon}\|_{L_p(\Omega_d)} \le Cp^{1/2}m^{-1/2}$$

with absolute constant C.

Corollary 5.2. *We apply Theorem 5.1 for $X = L_p(\Omega_d)$, $p \in [2, \infty)$, $\mathcal{D}^+ = \{B(x, y), y \in \Omega_d\}$, $f = J(x)$, where*

$$J(x) = \int_{\Omega_d} B(x, y)\, dy \in \mathcal{A}_1(\mathcal{D}^+).$$

Using (1.5) we get by Theorem 5.1 a constructive set $\xi^1, \ldots, \xi^m$ such that

$$D(\xi, m, d)_p = \|J_m^{i,\varepsilon}\|_{L_p(\Omega_d)} \le Cp^{1/2}m^{-1/2}$$

with absolute constant C.

Corollary 5.3. *We apply Theorem* 5.1 *for* $X = L_p(\Omega_d)$, $p \in [2,\infty)$, $\mathcal{D}^+ = \{B(x,y)/\|B(\cdot,y)\|_{L_p(\Omega_d)}, y \in \Omega_d\}$, $f = J(x)$. *Using* (1.5) *we get by Theorem* 5.1 *a constructive set* $\xi^1,\dots,\xi^m$ *such that*

$$\left\| \int_{\Omega_d} B(x,y)\,dy - \frac{1}{m}\sum_{\mu=1}^{m}\left(\frac{p}{p+1}\right)^d \left(\prod_{j=1}^{d}(1-\xi_j^{\mu})^{-1/p}\right) B(x,\xi^{\mu}) \right\|_{L_p(\Omega_d)}$$
$$\leq C\left(\frac{p}{p+1}\right)^d p^{1/2} m^{-1/2}$$

with absolute constant C.

We note that in the case $X = L_p(\Omega_d)$, $p \in [2,\infty)$, $\mathcal{D}^+ = \{H(x,y), x \in \Omega_d\}$, $f = J^s(y)$, the implementation of the IA(ε) is a sequence of maximization steps when we maximize functions of d variables. An important advantage of the L_p spaces is a simple and explicit form of the norming functional F_f of a function $f \in L_p(\Omega_d)$. The F_f acts as (for real L_p spaces)

$$F_f(g) = \int_{\Omega_d} \|f\|_p^{1-p}|f|^{p-2} f g\,dy.$$

Thus the IA(ε) should find at a step m an approximate solution to the following optimization problem (over $y \in \Omega_d$):

$$\int_{\Omega_d} |f_{m-1}^{i,\varepsilon}(x)|^{p-2} f_{m-1}^{i,\varepsilon}(x) H(x,y)\,dx \quad \to \quad \max.$$

Let us discuss a possible application of the AWRGA instead of the IA(ε). An obvious change is that instead of the cubature formula

$$\frac{1}{m}\sum_{\mu=1}^{m} H(x,\xi^{\mu})$$

in the case of IA(ε) we have a cubature formula

$$\sum_{\mu=1}^{m} w_{\mu}^{m} H(x,\xi^{\mu}), \qquad \sum_{\mu=1}^{m} |w_{\mu}^{m}| \leq 1,$$

in the case of the AWRGA. This is a disadvantage of the AWRGA. An advantage of the AWRGA is that we are more flexible in selecting an element φ_m^{ar}:

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

than an element $\varphi_m^{i,\varepsilon}$:

$$F_{f_{m-1}^{i,\varepsilon}}(\varphi_m^{i,\varepsilon} - f) \geq -\varepsilon_m.$$

We will now derive an estimate for $D(m,d)_\infty$ from Corollary 5.2.

Proposition 5.1. *For any m there exists a constructive set $\xi = \{\xi^1, \ldots, \xi^m\} \subset \Omega_d$ such that*

$$D(\xi, m, d)_\infty \le C d^{3/2} (\max(\ln d, \ln m))^{1/2} m^{-1/2}, \qquad d, m \ge 2, \tag{5.12}$$

with effective absolute constant C.

Proof. We use the inequality from [19],

$$D(\xi, m, d)_\infty \le c(d, p) d(3d + 4) D(\xi, m, d)_p^{p/(p+d)} \tag{5.13}$$

and the estimate for $c(d, p)$ from [11],

$$c(d, p) \le 3^{1/3} d^{-1+2/(1+p/d)}. \tag{5.14}$$

Specifying $p = d \max(\ln d, \ln m)$ and using Corollary 5.2 we get (5.12) from (5.13) and (5.14). ■

Acknowledgments. The author is grateful to E. Novak for an interesting discussion of results in Section 5. This research was supported by the National Science Foundation Grant DMS 0200187 and by ONR Grant N00014-91-J1343.

References

1. J. BECK, W. CHEN (1987): Irregularities of Distribution. Cambridge: Cambridge University Press.
2. B. CHAZELLE (2000): The Discrepancy Method. Cambridge: Cambridge University Press.
3. DAI FENG (2002): *Approximation of real smooth functions on the unit sphere S^{d-1}*. Ph.D. Thesis. Beijing Normal University, pp. 1–147.
4. G. DAVIS, S. MALLAT, M. AVELLANEDA (1997): *Adaptive greedy approximations*. Constr. Approx., **13**:57–98.
5. R. A. DEVORE, V. N. TEMLYAKOV (1995): *Nonlinear approximation by trigonometric sums*. J. Fourier Anal. Appl., **2**:29–48.
6. R. A. DEVORE, V. N. TEMLYAKOV (1996): *Some remarks on greedy algorithms*. Adv. in Comput. Math., **5**:173–187.
7. S. DILWORTH, D. KUTZAROVA, V. TEMLYAKOV (2002): *Convergence of some greedy algorithms in Banach spaces*. J. Fourier Anal. Appl., **8**:489–505.
8. M. DONAHUE, L. GURVITS, C. DARKEN, E. SONTAG (1997): *Rates of convex approximation in non-Hilbert spaces*. Constr. Approx., **13**:187–220.
9. E. D. GLUSKIN (1989): *Extremal properties of orthogonal parallelpipeds and their application to the geometry of Banach spaces*. Math USSR-Sb., **64**:85–96.
10. R. GRIBONVAL, M. NIELSEN (2001): *Approximate weak greedy algorithms*. Adv. in Comput. Math., **14**:361–368.
11. S. HEINRICH, E. NOVAK, G. WASILKOWSKI, H. WOZNIAKOWSKI (2001): *The inverse of the star-discrepancy depends linearly on the dimension*. Acta Arith., **96**:279–302.
12. R. S. ISMAGILOV (1974): *Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials*. Uspekhi Mat. Nauk, **29**:161–178; English transl. in Russian Math. Surveys, **29** (1974).
13. B. S. KASHIN (1977): *Widths of certain finite-dimensional sets and classes of smooth functions*. Izv. Akad. Nauk SSSR Ser. Mat., **41**:334–351; English transl. in Math. USSR-Izv., **11** (1977).
14. L. KUIPERS, H. NIEDERREITER (1974): Uniform Distribution of Sequences. New York: Wiley.
15. J. LINDENSTRAUSS, L. TZAFRIRI (1979): Classical Banach Spaces II. Berlin: Springer-Verlag.

16. V. E. MAIOROV (1986): *Trigonometric diameters of the Sobolev classes W_p^r in the space L_q*. Math. Notes, **40**:590–597.
17. V. E. MAIOROV, R. MEIR (2000): *On the optimality of networks approximation using incremental algorithms*. Trans. Neural Networks, **11**:323–337.
18. J. MATOUŠEK (1999): Geometric Discrepancy. Heidelberg: Springer-Verlag.
19. H. NIEDERREITER, R. F. TICHY, G. TURNWALD (1990): *An inequality for differences of distribution functions*. Arch. Math., **54**:166–172.
20. E. NOVAK, H. WOZNIAKOWSKI (2001): *When are integration and discrepancy tractable*? FoCM Proceedings. London Math. Soc. Lecture Notes Ser., **284**:211–266.
21. V. N. TEMLYAKOV (1998): *Greedy algorithm and m-term trigonometric approximation*. Constr. Approx., **14**:569–587.
22. V. N. TEMLYAKOV (1999): *Greedy algorithms and m-term approximation with regard to redundant dictionaries*. J. Approx. Theory, **98**:117–145.
23. V. N. TEMLYAKOV (2000): *Weak greedy algorithms*. Adv. in Comput. Math., **12**:213–227.
24. V. N. TEMLYAKOV (2001): *Greedy algorithms in Banach spaces*. Adv. in Comput. Math., **14**:277–292.
25. V. N. TEMLYAKOV (2003): *Cubature formulas and related questions*. J. Complexity, **19**:352–391.
26. V. N. TEMLYAKOV (2003): *Nonlinear methods of approximation*. Found. Comput. Math., **3**:33–107.

V. N. Temlyakov
Department of Mathematics
University of South Carolina
Columbia, SC 29208
USA