

# Greedy approximations with regard to bases

Vladimir N. Temlyakov

**Abstract.** This paper is a survey of recent results on greedy approximations with regard to bases. The theory of greedy approximations is a part of nonlinear approximations. The standard problem in this regard is the problem of  $m$ -term approximation where one fixes a basis and seeks to approximate a target function by a linear combination of  $m$  terms of the basis. When the basis is a wavelet basis or a basis of other waveforms, then this type of approximation is the starting point for compression algorithms. We are interested in the quantitative aspects of this type of approximation. Introducing the concept of best  $m$ -term approximation we obtain a lower bound for the accuracy of any method providing  $m$ -term approximation. It is known that a problem of simultaneous optimization over many parameters (like in best  $m$ -term approximation) is a very difficult problem. We would like to have an algorithm for constructing  $m$ -term approximants that adds at each step only one new element from the basis and keeps elements of the basis obtained at the previous steps. The primary object of our discussion is the Thresholding Greedy Algorithm (TGA) with regard to a given basis. The TGA, applied to a function  $f$ , picks at the  $m$ th step an element with the  $m$ th biggest coefficient (in absolute value) of the expansion of  $f$  in the series with respect to the basis. We show that this algorithm is very good for a wavelet basis and is not that good for the trigonometric system. We discuss in detail the behavior of the TGA with regard to the trigonometric system. We also discuss one example of an algorithm from a family of very general greedy algorithms that works in the case of a redundant system instead of a basis. It turns out that this general greedy algorithm is very good for the trigonometric system.

**Mathematics Subject Classification (2000).** Primary 41A25; Secondary 41A46.

**Keywords.** Nonlinear approximation, greedy algorithm, convergence, best  $m$ -term approximation, greedy basis.

## 1. Introduction. Historical remarks

In order to give the reader some ideas for comparing the quality of approximation methods we now discuss some classical results in approximation of periodic functions. In this section we briefly discuss various classical approaches, created in linear approximation, for the estimation of the quality of a method of approximation. We will use and refine these approaches in nonlinear approximation. We confine our discussion to the case of approximation of periodic functions of a single variable. The two main parameters of a method of approximation are accuracy and complexity. These concepts may be treated in various ways depending on the particular problems involved. Here we will start from the classical idea about approximation of functions by polynomials. After Fourier's article (1807) the representation of a  $2\pi$ -periodic

function by its Fourier series became natural. In other words, the function  $f(x)$  is approximately represented by a partial sum  $S_n(f, x)$  of its Fourier series.

We will be interested in the approximation of a function  $f$  by a polynomial  $S_n(f)$  in some  $L_p$ -norm,  $1 \leq p \leq \infty$ . In the case  $p = \infty$  we will assume that we deal with the uniform norm. As accuracy of the method of approximating a periodic function by its Fourier partial sum we will consider the quantity  $\|f - S(f)\|_p$ . The complexity of this method of approximation contains the two following characteristics. The order of the trigonometric polynomial  $S_n(f)$  is the quantitative characteristic. The following observation gives us the qualitative characteristic. The coefficients of this polynomial are found by the Fourier formulas which means that the operator  $S_n$  is the orthogonal projection onto the subspace of trigonometric polynomials of order  $n$ .

In 1854 Chebyshev suggested to represent a continuous function  $f$  by its polynomial of best approximation, namely by the polynomial  $t_n(f)$  such that

$$\|f - t_n(f)\|_\infty = E_n(f)_\infty := \inf_{\alpha_k, \beta_k} \left\| f(x) - \sum_{k=0}^n (\alpha_k \cos kx + \beta_k \sin kx) \right\|_\infty.$$

He proved the existence and uniqueness of such a polynomial. We will consider this method of approximation not only in the uniform norm, but in all  $L_p$ -norms,  $1 \leq p < \infty$ . The accuracy of the Chebyshev method can be easily compared with the accuracy of the Fourier method:

$$E_n(f)_p \leq \|f - S_n(f)\|_p.$$

The quantitative characteristics of complexity coincide for the two methods but the qualitative characteristics are different (for example, it is not difficult to understand that for  $p = \infty$  the mapping  $f \rightarrow t_n(f)$  is not a linear operator). The du Bois-Reymond example (1873) of a continuous function  $f$  such that  $\|f - S_n(f)\|_\infty \rightarrow \infty$  when  $n \rightarrow \infty$ , and the Weierstrass theorem which says that for each continuous function  $f$  we have  $E_n(f)_\infty \rightarrow 0$  as  $n \rightarrow \infty$ , showed the advantage of the Chebyshev method in comparison with the Fourier method from the point of view of accuracy. It is known that for each  $f \in L_2(\mathbb{T})$  the approximation with the error  $E_n(f)_2$  can be realized by the operator  $S_n$  of orthogonal projection onto the space of trigonometric polynomials of order  $n$ . The performance of the operator  $S_n$  was studied thoroughly in all  $L_p$  spaces,  $1 \leq p \leq \infty$ . It was proved that  $S_n$  provides almost optimal or close to optimal approximation for each  $f \in L_p(\mathbb{T})$ :

$$\begin{aligned} \|f - S_n(f)\|_p &\leq C(p)E_n(f)_p, & 1 < p < \infty, \\ \|f - S_n(f)\|_p &\leq C \ln(n+2)E_n(f)_p, & p = 1, \infty. \end{aligned}$$

The desire to construct methods of approximation which have the advantages of the Fourier and Chebyshev methods led to the study of various methods of summation of the Fourier series. The most important among them from the point of view of approximation are the de la Vallée Poussin, Fejér and Jackson methods which were

constructed early in the 20th century. All these methods are linear. For example, the de la Vallée Poussin method is the method of approximation of a function  $f$  by the polynomial

$$V_n(f) = \frac{1}{n} \sum_{l=n}^{2n-1} S_l(f)$$

of order  $2n - 1$ .

From the point of view of accuracy this method is close to the Chebyshev method; de la Vallée Poussin proved that

$$\|f - V_n(f)\|_p \leq 4E_n(f)_p, \quad 1 \leq p \leq \infty.$$

From the point of view of complexity it is close to the Fourier method, and the property of linearity essentially distinguishes it from the Chebyshev method.

We see that common to all these methods is the approximation by means of trigonometric polynomials; however, the ways of constructing these polynomials differ: orthogonal projections on the subspace of trigonometric polynomials of fixed order, the operator of best approximation, and linear operators.

In 1936 Kolmogorov introduced the concept of width  $d_n(F, X)$  of a class  $F$  in a Banach space  $X$ :

$$d_n(F, X) = \inf_{\{\phi_j\}_{j=1}^n} \sup_{f \in F} \inf_{\{c_j\}_{j=1}^n} \left\| f - \sum_{j=1}^n c_j \phi_j \right\|_X.$$

This concept is designed to find for a fixed  $n$  and for a class  $F$  a subspace of dimension  $n$ , optimal with respect to the construction of an approximating element as the element of best approximation. In other words, the Kolmogorov width gives the lower bound for accuracy of Chebyshev's methods, having the same quantitative characteristic of complexity (the dimension of the approximating subspace). In analogy to the concept of the Kolmogorov width, that is, to the problem concerning the best Chebyshev method, the problems concerning the best linear method and the best Fourier method were considered. Tikhomirov ([51]) introduced the concept of linear width

$$\lambda_n(F, X) = \inf_{A: \text{rank } A \leq n} \sup_{f \in F} \|f - Af\|_X.$$

The concept of orthonormal width (Fourier width) was introduced in [38]:

$$\varphi_n(F, X) := d_n^\perp(F, X) := \inf_{\substack{\text{orthonormal} \\ \text{system } \{u_i\}_{i=1}^n}} \sup_{f \in F} \left\| f - \sum_{i=1}^n \langle f, u_i \rangle u_i \right\|_X.$$

We discuss these widths in more detail later in this section. We present here some well-known results for the Sobolev classes

$$W_q^r := \{f : f^{(r-1)} \text{ is absolutely continuous, } \|f^{(r)}\|_q \leq 1\}.$$

The first result about widths, namely Kolmogorov's result (1936)

$$d_{2n+1}(W_2^r, L_2) = (n+1)^{-r},$$

showed that the best subspace of dimension  $2n+1$  for approximation of classes of periodic functions is the subspace of trigonometric polynomials of order  $n$ . This result confirmed that the approximation of functions in the class  $W_2^r$  by trigonometric polynomials is natural. Further estimates of the widths  $d_{2n+1}(W_q^r, L_p)$ ,  $1 \leq q, p \leq \infty$ , some of which are discussed here, showed that for some values of the parameters  $q, p$  the subspace of trigonometric polynomials of order  $n$  is optimal (in the sense of order) but for other values of  $q, p$  this subspace is not optimal.

The Ismagilov estimate [20] for the quantity  $d_n(W_1^r, L_\infty)$  gave the first example where the subspace of trigonometric polynomials of order  $n$  is not optimal. This phenomenon was thoroughly studied by Kashin [22]. We remark that from the point of view of orthowidth the Fourier operator  $S_n$  is optimal (in the sense of order of approximation in the  $L_p$ -norm) for all Sobolev classes  $W_q^r$  with  $1 \leq q, p \leq \infty$  with the exception of the two cases  $q = p = 1$  and  $q = p = \infty$ .

All the above defined widths have as a starting point a function class  $F$ . Thus in this setting we choose a priori a function class  $F$  and look for optimal subspaces for approximation of a given class. The following results are well known [39]. We present these results for  $r$  a positive integer. In the case  $q = 1, p = \infty$  we assume  $r > 1$ . For a number  $a$  we denote  $(a)_+ := \max(a, 0)$ .

**A.** In the case  $1 \leq p \leq q \leq \infty$  or  $1 \leq q \leq p \leq 2$  one has

$$\varphi_n(W_q^r, L_p) \asymp \lambda_n(W_q^r, L_p) \asymp d_n(W_q^r, L_p) \asymp n^{-r+(1/q-1/p)_+}. \quad (1.1)$$

**B.** In the case  $1 \leq q < p \leq \infty, p > 2$ , one has

$$\begin{aligned} d_n(W_q^r, L_p) &\asymp n^{-r+(1/q-1/2)_+}, \\ \lambda_n(W_q^r, L_p) &\asymp n^{-r+\max(1/q-1/2, 1/2-1/p)}, \\ \varphi_n(W_q^r, L_p) &\asymp n^{-r+1/q-1/p}. \end{aligned}$$

In case A the classical trigonometric system provides the optimal orders for all widths, except for  $\varphi_n$  for  $q = p = 1, \infty$ . Let us discuss the more interesting case B for the particular choice  $q = 2$  and  $p = \infty$ . We have

$$d_n(W_2^r, L_\infty) \asymp n^{-r}, \quad (1.2)$$

$$\lambda_n(W_2^r, L_\infty) \asymp \varphi_n(W_2^r, L_\infty) \asymp n^{-r+1/2}. \quad (1.3)$$

These relations show that if we drop the linearity requirement for the approximation method we gain in accuracy a factor  $n^{-1/2}$ . However, there is a big difficulty in realization of the estimate (1.2). We know by Kashin's result that there exists a subspace realizing (1.2) but we do not know a way to construct it. Thus it is only an existence theorem for now.

Let us discuss one more special case:  $q = 1$  and  $p = \infty$ . In this case we have

$$d_n(W_1^r, L_\infty) \asymp \lambda_n(W_1^r, L_\infty) \asymp n^{-r+1/2} \tag{1.4}$$

and

$$\varphi_n(W_1^r, L_\infty) \asymp n^{-r+1}. \tag{1.5}$$

Therefore, by (1.4) the best possible approximation (in the sense of order) can be realized by a linear method, say,  $A_n$ . However, by (1.5) this linear method  $A_n$  is certainly not an orthogonal projector. Moreover, by [39] it cannot satisfy even the following much weaker restriction  $\|A_n(e^{ikx})\|_2 \leq C, k \in \mathbb{Z}$ . This means that the optimal linear operator  $A_n$  is unstable. A small change in some of the Fourier coefficients of  $f$  may result in a big change of  $\|A_n(f)\|_2$ .

Let us make some conclusions now. In linear approximation of  $W_q^r$  in  $L_p$  the bottom line is given by  $\varphi_n(W_q^r, L_p)$  where the approximation method is the simplest, namely orthogonal projection. Partial sums with regard to classical systems provide an optimal error of approximation for this width. The trigonometric system works for all  $1 \leq q, p \leq \infty$  except for  $(q, p) = (1, 1), (\infty, \infty)$ . The wavelet systems (see [1]) work for all  $1 \leq q, p \leq \infty$ . In the example of the pair  $(W_1^r, L_\infty)$  we have seen that we need to sacrifice important and convenient properties of the approximating operator in order to achieve better accuracy. In the example of  $(W_2^r, L_\infty)$  we have seen that we need to pay even a bigger price for better accuracy in a form of proving only an existence theorem instead of providing a constructive method of approximation.

Our main interest in this paper is nonlinear approximation. We begin our discussion with the trigonometric system. Let  $\mathcal{T}$  be the complex trigonometric system  $\{e^{ikx}\}_{k \in \mathbb{Z}}$ . Denote for  $f \in L_p(\mathbb{T})$

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

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  $\mathcal{T}(m)$  the subspace of trigonometric polynomials of order  $m$  and define

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

The first result that indicated an advantage of  $m$ -term approximation over approximation by trigonometric polynomials of order  $m$  is due to Ismagilov [20]:

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

Let us compare it with the well-known result due to de la Vallée Poussin and Bernstein:

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

Maiorov [35] improved the estimate (1.6):

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

In [11] we proved the following rate of best  $m$ -term approximation of the Sobolev classes  $W_q^r$  in  $L_p$ ,  $1 \leq q, p \leq \infty$ :

$$\sigma_m(W_q^r, \mathcal{T})_p := \sup_{f \in W_q^r} \sigma_m(f, \mathcal{T})_p \asymp m^{-r+(1/q-\max(1/p, 1/2))_+}. \quad (1.9)$$

Comparing (1.9) with the above bounds for the Kolmogorov width we conclude that

$$\sigma_m(W_q^r, \mathcal{T})_p \asymp d_m(W_q^r, L_p).$$

In particular, this means, that in the case  $(W_2^r, L_\infty)$  the nonlinear  $m$ -term approximations provide much better accuracy than the trigonometric polynomials of order  $m$ . The best  $m$ -term approximations  $\sigma_m(f, \mathcal{T})_p$  may be considered as a nonlinear analogue (counterpart) of the best approximations  $E_m(f, \mathcal{T})_p$ . The main goal of this paper is to discuss a nonlinear analogue (counterpart) of the operator  $S_n(f)$ . We consider the greedy approximant to be a nonlinear analogue of the partial sum. In Sections 2 and 3 we discuss the general theory of greedy approximation with regard to bases. Our primary object of discussion is the Thresholding Greedy Algorithm (TGA). We return to a discussion of nonlinear approximations with regard to the trigonometric system in Sections 4 and 5. In Section 6 we deviate from the main stream of the paper of studying the TGA and discuss one example of a family of greedy algorithms that works in a very general situation. The most important feature of this algorithm is that it provides  $m$ -term approximation with regard to a very general system that may be redundant (overcomplete). It turns out (this will be seen from the discussion in Section 6) that this general approximation method is very good for the trigonometric system.

## 2. Greedy algorithms with regard to bases

Let  $X$  be a Banach space with a given basis  $\Psi = \{\psi_k\}_{k=1}^\infty$ . We assume that  $\|\psi_k\| \geq C > 0$ ,  $k = 1, 2, \dots$ , and consider the following theoretical greedy algorithm. For a given element  $f \in X$  we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f, \Psi) \psi_k. \quad (2.1)$$

For an element  $f \in X$  we call a permutation  $\rho$ ,  $\rho(j) = k_j$ ,  $j = 1, 2, \dots$ , of the positive integers decreasing and write  $\rho \in D(f)$  if

$$|c_{k_1}(f, \Psi)| \geq |c_{k_2}(f, \Psi)| \geq \dots \quad (2.2)$$

In the case of strict inequalities here  $D(f)$  consists of only one permutation. We define the  $m$ -th greedy approximant of  $f$  with regard to the basis  $\Psi$  corresponding to a permutation  $\rho \in D(f)$  by the formula

$$G_m(f) := G_m(f, \Psi) := G_m(f, \Psi, \rho) := \sum_{j=1}^m c_{k_j}(f, \Psi) \psi_{k_j}.$$

We note that there is another natural greedy type algorithm based on ordering  $\|c_k(f, \Psi) \psi_k\|$  instead of ordering absolute values of coefficients. In this case we do not need the restriction  $\|\psi_k\| \geq C > 0, k = 1, 2, \dots$ . Denote by  $\Lambda_m(f)$  a set of indices such that

$$\min_{k \in \Lambda_m(f)} \|c_k(f, \Psi) \psi_k\| \geq \max_{k \notin \Lambda_m(f)} \|c_k(f, \Psi) \psi_k\|.$$

We define  $G_m^X(f, \Psi)$  by the formula

$$G_m^X(f, \Psi) := S_{\Lambda_m(f)}(f, \Psi), \quad \text{where } S_E(f) := S_E(f, \Psi) := \sum_{k \in E} c_k(f, \Psi) \psi_k.$$

It is clear that in the case of the normalized basis ( $\|\psi_k\| = 1, k = 1, 2, \dots$ ) the above two greedy algorithms coincide.

In the case  $X = L_p$  we will write  $p$  instead of  $L_p$  in notations. It is a simple algorithm which describes the theoretical scheme for  $m$ -term approximation of an element  $f$ . We call this algorithm the Thresholding Greedy Algorithm (TGA). In order to understand the efficiency of this algorithm we compare its accuracy with the best possible when an approximant is a linear combination of  $m$  terms from  $\Psi$ . We define the best  $m$ -term approximation with regard to  $\Psi$  as follows:

$$\sigma_m(f) := \sigma_m(f, \Psi)_X := \inf_{c_k, \Lambda} \left\| f - \sum_{k \in \Lambda} c_k \psi_k \right\|_X,$$

where inf is taken over coefficients  $c_k$  and sets of indices  $\Lambda$  with cardinality  $|\Lambda| = m$ . The best we can achieve with the algorithm  $G_m$  is

$$\|f - G_m(f, \Psi, \rho)\|_X = \sigma_m(f, \Psi)_X,$$

or, a little weaker,

$$\|f - G_m(f, \Psi, \rho)\|_X \leq G \sigma_m(f, \Psi)_X \tag{2.3}$$

for all elements  $f \in X$  with a constant  $G = C(X, \Psi)$  independent of  $f$  and  $m$ . It is clear that in the case  $X = H$  is a Hilbert space and  $\Psi$  is an orthonormal basis we have

$$\|f - G_m(f, \Psi, \rho)\|_H = \sigma_m(f, \Psi)_H.$$

Let us begin our discussion with an important class of bases: wavelet type bases. Denote  $\mathcal{H} := \{H_k\}_{k=1}^\infty$  the Haar basis on  $[0, 1)$  normalized in  $L_2(0, 1)$ :  $H_1 = 1$  on  $[0, 1)$  and for  $k = 2^n + l, n = 0, 1, \dots, l = 1, 2, \dots, 2^n$ ,

$$H_k(x) = \begin{cases} 2^{n/2}, & x \in [(2l - 2)2^{-n-1}, (2l - 1)2^{-n-1}) \\ -2^{n/2}, & x \in [(2l - 1)2^{-n-1}, 2l2^{-n-1}) \\ 0, & \text{otherwise.} \end{cases}$$

We denote by  $\mathcal{H}_p := \{H_{k,p}\}_{k=1}^\infty$  the Haar basis  $\mathcal{H}$  renormalized in  $L_p(0, 1)$ . We will use the following definition of the  $L_p$ -equivalence of bases. We say that  $\Psi = \{\psi_k\}_{k=1}^\infty$  is  $L_p$ -equivalent to  $\Phi = \{\phi_k\}_{k=1}^\infty$  if for any finite set  $\Lambda$  and any coefficients  $c_k, k \in \Lambda$ , we have

$$C_1(p, \Psi, \Phi) \left\| \sum_{k \in \Lambda} c_k \phi_k \right\|_p \leq \left\| \sum_{k \in \Lambda} c_k \psi_k \right\|_p \leq C_2(p, \Psi, \Phi) \left\| \sum_{k \in \Lambda} c_k \phi_k \right\|_p$$

with two positive constants  $C_1(p, \Psi, \Phi), C_2(p, \Psi, \Phi)$  which may depend on  $p, \Psi$ , and  $\Phi$ . For sufficient conditions on  $\Psi$  to be  $L_p$ -equivalent to  $\mathcal{H}$  see [16] and [10]. In particular, it is known that all reasonable univariate wavelet type bases are  $L_p$ -equivalent to  $\mathcal{H}$  for  $1 < p < \infty$ . We proved the following theorem in [40].

**Theorem 2.1.** *Let  $1 < p < \infty$  and let a basis  $\Psi$  be  $L_p$ -equivalent to the Haar basis  $\mathcal{H}$ . Then for any  $f \in L_p(0, 1)$  we have*

$$\|f - G_m^p(f, \Psi)\|_p \leq C(p, \Psi)\sigma_m(f, \Psi)_p$$

with a constant  $C(p, \Psi)$  independent of  $f$  and  $m$ .

By a simple renormalization argument one obtains the following version of Theorem 2.1.

**Theorem 2.2.** *Let  $1 < p < \infty$  and let a basis  $\Psi$  be  $L_p$ -equivalent to the Haar basis  $\mathcal{H}_p$ . Then for any  $f \in L_p(0, 1)$  and any  $\rho \in D(f)$  we have*

$$\|f - G_m(f, \Psi, \rho)\|_p \leq C(p, \Psi)\sigma_m(f, \Psi)_p$$

with a constant  $C(p, \Psi)$  independent of  $f, \rho$ , and  $m$ .

We note that [40] also contains a generalization of Theorem 2.1 to the multivariate Haar basis obtained by the multiresolution analysis procedure. These theorems motivated us to consider the general setting of greedy approximation in Banach spaces. We concentrated on studying bases which satisfy (2.3) for all individual functions. The following Definitions 2.1, 2.2 and 2.3 are from [27].

**Definition 2.1.** We call a basis  $\Psi$  a *greedy basis* if for every  $f \in X$  there exists a permutation  $\rho \in D(f)$  such that

$$\|f - G_m(f, \Psi, \rho)\|_X \leq G\sigma_m(f, \Psi)_X \tag{2.4}$$

holds with a constant independent of  $f, m$ .

The following proposition has been proved in [27].

**Proposition 2.1.** *If  $\Psi$  is a greedy basis, then inequality (2.4) holds for any permutation  $\rho \in D(f)$ .*

Theorem 2.2 shows that each basis  $\Psi$  which is  $L_p$ -equivalent to the univariate Haar basis  $\mathcal{H}_p$  is a greedy basis for  $L_p(0, 1)$ ,  $1 < p < \infty$ . We note that in the case of Hilbert space each orthonormal basis is a greedy basis with a constant  $G = 1$  (see (2.4)).

We give now the definitions of unconditional and democratic bases.

**Definition 2.2.** A basis  $\Psi = \{\psi_k\}_{k=1}^{\infty}$  of a Banach space  $X$  is said to be *unconditional* if for every choice of signs  $\theta = \{\theta_k\}_{k=1}^{\infty}$ ,  $\theta_k = 1$  or  $-1$ ,  $k = 1, 2, \dots$ , the linear operator  $M_{\theta}$  defined by  $M_{\theta}(\sum_{k=1}^{\infty} a_k \psi_k) = \sum_{k=1}^{\infty} a_k \theta_k \psi_k$  is a bounded operator from  $X$  into  $X$ .

**Definition 2.3.** We say that a basis  $\Psi = \{\psi_k\}_{k=1}^{\infty}$  is a *democratic basis* for  $X$  if there exists a constant  $D := D(X, \Psi)$  such that for any two finite sets of indices  $P$  and  $Q$  with the same cardinality  $|P| = |Q|$  we have  $\|\sum_{k \in P} \psi_k\| \leq D \|\sum_{k \in Q} \psi_k\|$ .

We proved in [27] the following theorem.

**Theorem 2.3.** *A basis is greedy if and only if it is unconditional and democratic.*

This theorem gives a characterization of greedy bases. Further investigations ([41], [6], [25], [18], [21]) showed that the concept of greedy bases is very useful in direct and inverse theorems of nonlinear approximation and also in applications in statistics. The papers [27], [40] contain other results on greedy bases.

Let us discuss a question of weakening the property of a basis of being a greedy basis. We begin with the concept of quasi-greedy basis introduced in [27].

**Definition 2.4.** We call a basis  $\Psi$  a *quasi-greedy basis* if for every  $f \in X$  and every permutation  $\rho \in D(f)$  we have

$$\|G_m(f, \Psi, \rho)\|_X \leq C \|f\|_X \quad (2.5)$$

with a constant  $C$  independent of  $f$ ,  $m$ , and  $\rho$ .

It is clear that (2.5) is weaker than (2.4). P. Wojtaszczyk [53] proved the following theorem.

**Theorem 2.4.** *A basis  $\Psi$  is quasi-greedy if and only if for any  $f \in X$  and any  $\rho \in D(f)$  we have*

$$\|f - G_m(f, \Psi, \rho)\| \rightarrow 0 \quad \text{as } m \rightarrow \infty. \quad (2.6)$$

We proceed to an intermediate concept of almost greedy basis. This concept has been introduced and studied in [14]. Let

$$f = \sum_{k=1}^{\infty} c_k(f) \psi_k.$$

We define the following expansional best  $m$ -term approximation of  $f$ :

$$\tilde{\sigma}_m(f) := \tilde{\sigma}_m(f, \Psi) := \inf_{\Lambda, |\Lambda|=m} \left\| f - \sum_{k \in \Lambda} c_k(f) \psi_k \right\|.$$

It is clear that

$$\sigma_m(f, \Psi) \leq \tilde{\sigma}_m(f, \Psi).$$

It is also clear that for an unconditional basis  $\Psi$  we have

$$\tilde{\sigma}_m(f, \Psi) \leq C \sigma_m(f, \Psi).$$

**Definition 2.5.** We call a basis  $\Psi$  *almost greedy* if for every  $f \in X$  there exists a permutation  $\rho \in D(f)$  such that

$$\|f - G_m(f, \Psi, \rho)\|_X \leq C \tilde{\sigma}_m(f, \Psi)_X \quad (2.7)$$

holds with a constant independent of  $f, m$ .

The following proposition follows from the proof of Theorem 3.3 of [14] (see Theorem 2.5 below).

**Proposition 2.2.** *If  $\Psi$  is an almost greedy basis then (2.7) holds for any permutation  $\rho \in D(f)$ .*

The following characterization of almost greedy bases has been obtained in [14].

**Theorem 2.5.** *Suppose  $\Psi$  is a basis of a Banach space. The following are equivalent:*

- A.  $\Psi$  is almost greedy.
- B.  $\Psi$  is quasi-greedy and democratic.
- C. For any (respectively, every)  $\lambda > 1$  there is a constant  $C = C_\lambda$  such that

$$\|f - G_{[\lambda m]}(f, \Psi)\| \leq C_\lambda \sigma_m(f, \Psi).$$

We now proceed to a generalization of the concept of greedy bases from [26] that is useful in statistical applications. Let  $\Psi$  be a basis for  $X$ . If  $\inf_k \|\psi_k\| > 0$  then  $c_k(f) \rightarrow 0$  as  $k \rightarrow \infty$ , where

$$f = \sum_{k=1}^{\infty} c_k(f) \psi_k.$$

Then we can rearrange the coefficients  $\{c_k(f)\}$  in the decreasing way

$$|c_{k_1}(f)| \geq |c_{k_2}(f)| \geq \dots$$

and define the  $m$ th greedy approximant as

$$G_m(f, \Psi) := \sum_{j=1}^m c_{k_j}(f) \psi_{k_j}. \quad (2.8)$$

In the case  $\inf_k \|\psi_k\| = 0$  we define  $G_m(f, \Psi)$  by (2.8) for  $f$  of the form

$$f = \sum_{k \in Y} c_k(f) \psi_k, \quad |Y| < \infty. \quad (2.9)$$

Let a weight sequence  $w = \{w_k\}_{k=1}^{\infty}$ ,  $w_k > 0$ , be given. For  $\Lambda \subset \mathbb{N}$  denote  $w(\Lambda) := \sum_{k \in \Lambda} w_k$ . For a positive real number  $v > 0$  define

$$\sigma_v^w(f, \Psi) := \inf_{\{b_k\}, \Lambda: w(\Lambda) \leq v} \|f - \sum_{k \in \Lambda} b_k \psi_k\|,$$

where the sets  $\Lambda$  are finite.

**Definition 2.6.** We call a basis  $\Psi$  a *weight-greedy basis* (w-greedy basis) if for any  $f \in X$  in the case  $\inf_k \|\psi_k\| > 0$  or for any  $f \in X$  of the form (2.9) in the case  $\inf_k \|\psi_k\| = 0$  we have

$$\|f - G_m(f, \Psi)\| \leq C_G \sigma_{w(\Lambda_m)}^w(f, \Psi),$$

where

$$G_m(f, \Psi) = \sum_{k \in \Lambda_m} c_k(f) \psi_k, \quad |\Lambda_m| = m.$$

**Definition 2.7.** We call a basis  $\Psi$  *weight-democratic* (w-democratic basis) if for any finite  $A, B \subset \mathbb{N}$  such that  $w(A) \leq w(B)$  we have

$$\left\| \sum_{k \in A} \psi_k \right\| \leq C_D \left\| \sum_{k \in B} \psi_k \right\|.$$

Recently, we proved in [26] the following criterion for w-greedy bases.

**Theorem 2.6.** *A basis  $\Psi$  is a w-greedy basis if and only if it is unconditional and w-democratic.*

The reader can find a further discussion in the surveys [8], [28], [46], [47].

### 3. Optimal methods in nonlinear approximation

In the widths problem of linear approximation (see Section 1) we were looking for an optimal  $n$ -dimensional subspace for approximating a given function class. A nonlinear analogue of this setting is the following. Let a function class  $F$  and a Banach space  $X$  be given. Assume that on the basis of some additional information we know that our basis for  $m$ -term approximation should satisfy some structural properties, for instance, it has to be orthogonal. Then similarly to the setting for the widths  $d_n, \lambda_n, \varphi_n$  we get the optimization problems for  $m$ -term nonlinear approximation. Let  $\mathbb{B}$  be a collection of bases satisfying a given property.

**I.** Define an analogue of the Kolmogorov width

$$\sigma_m(F, \mathbb{B})_X := \inf_{\Psi \in \mathbb{B}} \sup_{f \in F} \sigma_m(f, \Psi)_X.$$

**II.** Define an analogue of the orthowidth

$$\gamma_m(F, \mathbb{B})_X := \inf_{\Psi \in \mathbb{B}} \sup_{f \in F} \|f - G_m(f, \Psi)\|_X.$$

We present here some results in the case  $\mathbb{B} = \mathbb{O}$ , the set of orthonormal bases,  $F = W_q^r$ ,  $X = L_p$ ,  $1 \leq q, p \leq \infty$ . First of all we formulate a result (see [24], [43]) that shows that in the case  $p < 2$  we need some more restrictions on  $\mathbb{B}$  in order to obtain meaningful results (lower bounds).

**Proposition 3.1.** *For any  $1 \leq p < 2$  there exists a complete in  $L_2(0, 1)$  orthonormal system  $\Phi$  such that for each  $f \in L_p(0, 1)$  we have  $\sigma_1(f, \Phi)_p = 0$ .*

Let us restrict our further discussion to the case  $p \geq 2$ . This case was also more interesting in the linear approximation discussion (see Section 1). Kashin [23] proved that

$$\sigma_m(W_\infty^r, \mathbb{O})_2 \gg m^{-r}. \quad (3.1)$$

We proved (see [11]) that

$$\sigma_m(W_2^r, \mathcal{T})_\infty \ll m^{-r}. \quad (3.2)$$

The estimates (3.1) and (3.2) imply that for  $2 \leq q, p \leq \infty$  we have

$$\sigma_m(W_q^r, \mathbb{O})_p \asymp \sigma_m(W_q^r, \mathcal{T})_p \asymp m^{-r}. \quad (3.3)$$

Let us compare this relation with (1.2). We see that the best  $m$ -term trigonometric approximation provides the same accuracy as the best approximation from an optimal  $m$ -dimensional subspace. An advantage of nonlinear approximation here is that we use a natural basis instead of an existing but nonconstructive subspace. However, we should note that the estimate (3.2) was proved in [11] as an existence theorem. We did not give an algorithm to get (3.2) in [11]. We gave such an algorithm in [49]

(see a further discussion in Section 6). The Thresholding Greedy Algorithm does not provide the estimate (3.2). We have (see [42])

$$\sup_{f \in W_2^r} \|f - G_m(f, \mathcal{T})\|_\infty \asymp m^{-r+1/2}.$$

It is known from different results (see [9], [8], [45]) that wavelets are well designed for nonlinear approximation.

In the multivariate periodic case the following basis  $U^d := U \times \dots \times U$  has approximation properties close to the corresponding properties of wavelets. We define the system  $U := \{U_I\}$  in the univariate case. Denote

$$\begin{aligned} U_n^+(x) &:= \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x} - 1}{e^{ix} - 1}, & n = 0, 1, 2, \dots; \\ U_{n,k}^+(x) &:= e^{i2^n x} U_n^+(x - 2\pi k 2^{-n}), & k = 0, 1, \dots, 2^n - 1; \\ U_{n,k}^-(x) &:= e^{-i2^n x} U_n^+(-x + 2\pi k 2^{-n}), & k = 0, 1, \dots, 2^n - 1. \end{aligned}$$

We normalize the system of functions  $\{U_{n,k}^+, U_{n,k}^-\}$  in  $L_2$  and enumerate it by dyadic intervals. We write

$$\begin{aligned} U_I(x) &:= 2^{-n/2} U_{n,k}^+(x) \quad \text{with } I = [(k + 1/2)2^{-n}, (k + 1)2^{-n}); \\ U_I(x) &:= 2^{-n/2} U_{n,k}^-(x) \quad \text{with } I = [k2^{-n}, (k + 1/2)2^{-n}); \end{aligned}$$

and

$$U_{[0,1)}(x) := 1.$$

P. Wojtaszczyk [52] proved that the system  $U$  is an unconditional basis for  $L_p(\mathbb{T})$ ,  $1 < p < \infty$ .

We define the anisotropic multivariate periodic Hölder–Nikol’skii classes  $NH_p^R$  in the following way. The class  $NH_p^R$ ,  $R = (R_1, \dots, R_d)$  and  $1 \leq p \leq \infty$ , is the set of periodic functions  $f \in L_p([0, 2\pi]^d)$  such that for each  $l_j = [R_j] + 1$ ,  $j = 1, \dots, d$ , the following relations hold:

$$\|f\|_p \leq 1, \quad \|\Delta_t^{l_j, j} f\|_p \leq |t|^{R_j}, \quad j = 1, \dots, d, \tag{3.4}$$

where  $\Delta_t^{l_j, j}$  is the  $l_j$ -th difference with step  $t$  in the variable  $x_j$ . In the case  $d = 1$   $NH_p^R$  coincides with the standard Hölder class  $H_p^R$ . For  $R = (R_1, \dots, R_d)$ ,  $R_j > 0$ ,  $j = 1, \dots, d$ , we define  $g(R) := (\sum_{j=1}^d R_j^{-1})^{-1}$ . The following result has been proved in [45].

**Theorem 3.1.** *Let  $1 < q, p < \infty$ . Then for  $R$  such that  $g(R) > (1/q - 1/p)_+$  we have*

$$\sup_{f \in NH_q^R} \|f - G_m^{L_p}(f, U^d)\|_p \ll m^{-g(R)}.$$

We also proved in [45] that the basis  $U^d$  is an optimal orthonormal basis for approximation of classes  $NH_q^R$  in  $L_p$ :

$$\sigma_m(NH_q^R, \mathbb{O})_p \asymp \sigma_m(NH_q^R, U^d)_p \asymp m^{-g(R)} \quad (3.5)$$

for  $1 < q < \infty$ ,  $2 \leq p < \infty$ ,  $g(R) > (1/q - 1/p)_+$ . It is important to remark that Theorem 3.1 guarantees that the estimate in (3.5) can be realized by the TGA with regard to  $U^d$ .

#### 4. The TGA with regard to the trigonometric system

Let us consider nonlinear approximation with regard to the trigonometric system  $\mathcal{T}^d := \mathcal{T} \times \dots \times \mathcal{T}$  ( $d$  times). The existence of best  $m$ -term trigonometric approximation was proved in [2] (see also [42]). The method  $G_m(f) := G_m(f, \mathcal{T}^d)$  has an advantage over the traditional approximation by trigonometric polynomials in the case of approximation of functions of several variables. In this case ( $d > 1$ ) there is no natural order of trigonometric system and the use of  $G_m$  allows us to avoid the problem of finding natural subspaces of trigonometric polynomials for approximation purposes. We proved in [42] the following results.

**Theorem 4.1.** *For each  $f \in L_p(\mathbb{T}^d)$  we have*

$$\|f - G_m(f)\|_p \leq (1 + 3m^{h(p)})\sigma_m(f)_p, \quad 1 \leq p \leq \infty,$$

where  $h(p) := |1/2 - 1/p|$ .

**Remark 4.1.** For all  $1 \leq p \leq \infty$

$$\|G_m(f)\|_p \leq m^{h(p)}\|f\|_p.$$

**Remark 4.2.** There is a positive absolute constant  $C$  such that for each  $m$  and  $1 \leq p \leq \infty$  there exists a function  $f \neq 0$  with the property

$$\|G_m(f)\|_p \geq Cm^{h(p)}\|f\|_p. \quad (4.1)$$

The above results show that the trigonometric system is not a quasi-greedy basis for  $L_p$ ,  $p \neq 2$ . This leads to a natural attempt to consider some other algorithms that may have some advantages over the TGA in the case of  $\mathcal{T}$ . We discuss here the performance of the WCGA (see Section 6) with regard to  $\mathcal{T}$ .

Let us compare the rate of approximation of the TGA and the WCGA. Let  $\mathcal{RT}$  denote the real trigonometric system  $1/2, \sin x, \cos x, \dots$ . We need to switch to this system from the complex trigonometric system because the algorithm WCGA is defined for the real Banach space. We note that the system  $\mathcal{RT}$  is not normalized in  $L_p$  but quasinormalized:  $C_1 \leq \|t\|_p \leq C_2$  for any  $t \in \mathcal{RT}$  with absolute constants

$C_1, C_2, 1 \leq p \leq \infty$ . It is sufficient for the application of the general methods developed in Section 6. For a function  $f$  with absolutely convergent Fourier series

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

denote

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

Define the class

$$A := A(\mathcal{RT}) := \{f : \|f\|_A \leq 1\}.$$

For a sequence  $\tau := \{t_k\}$  with  $t_k = t, k = 1, 2, \dots$ , we replace  $\tau$  by  $t$  in the notation. Theorem 6.1 and (6.2) imply the following result.

**Theorem 4.2.** *Let  $0 < t \leq 1$ . For  $f \in A$  we have*

$$\|f - G_m^{c,t}(f, \mathcal{RT})\|_p \leq C(p, t)m^{-1/2}, \quad 2 \leq p < \infty. \tag{4.2}$$

This estimate and Theorem 4.1 imply that for  $f \in A$  we have

$$\|f - G_m(f, \mathcal{RT})\|_p \leq C(p, t)m^{-1/p}, \quad 2 \leq p < \infty, \tag{4.3}$$

which is weaker than (4.2). It is proved in [15] that (4.3) can not be improved. Thus the WCGA works better than the TGA for the class  $A$ . We note that the restriction  $p < \infty$  in (4.2) is important. We gave a lower estimate for  $m$ -term approximation in  $L_\infty$  in [47].

**Proposition 4.1.** *For a given  $m$  define*

$$f := \sum_{k=0}^{2m} \cos 3^k x.$$

Then we have

$$\sigma_m(f, \mathcal{T})_\infty \geq m/8.$$

### 5. Convergence of the TGA with regard to the trigonometric system

We discuss in this section the following nonlinear method of summation of trigonometric Fourier series. Consider a periodic function  $f \in L_p(\mathbb{T}^d), 1 \leq p \leq \infty, (L_\infty(\mathbb{T}^d) = C(\mathbb{T}^d))$ , defined on the  $d$ -dimensional torus  $\mathbb{T}^d$ . Take a number  $t \in (0, 1]$ . Let a number  $m \in \mathbb{N}$  be given and  $\Lambda_m$  be a set of  $k \in \mathbb{Z}^d$  with the properties

$$\min_{k \in \Lambda_m} |\hat{f}(k)| \geq t \max_{k \notin \Lambda_m} |\hat{f}(k)|, \quad |\Lambda_m| = m, \tag{5.1}$$

where

$$\hat{f}(k) := (2\pi)^{-d} \int_{\mathbb{T}^d} f(x)e^{-i(k,x)} dx$$

is a Fourier coefficient of  $f$ . We define

$$G_m^t(f) := G_m^t(f, \mathcal{T}^d) := S_{\Lambda_m}(f) := \sum_{k \in \Lambda_m} \hat{f}(k)e^{i(k,x)}$$

and call it an  $m$ th weak greedy approximant of  $f$  with regard to the trigonometric system  $\mathcal{T}^d = \{e^{i(k,x)}\}_{k \in \mathbb{Z}^d}$ . We write  $G_m(f) = G_m^1(f)$  and call it an  $m$ th greedy approximant. Clearly, an  $m$ th weak greedy approximant and even an  $m$ th greedy approximant may not be unique. Here we do not impose any extra restrictions on  $\Lambda_m$  in addition to (5.1). Thus theorems formulated below hold for any choice of  $\Lambda_m$  satisfying (5.1) or, in other words, for any realization  $G_m^t(f)$  of the weak greedy approximation.

There has recently been (see the surveys [8], [47], [28]) much interest in approximation of functions by  $m$ -term approximants with regard to a basis (or minimal system). In this section we will discuss in detail only results concerning the trigonometric system. Answering a question raised by Carleson and Coifman, T. W. Körner constructed in [31] a function from  $L_2(\mathbb{T})$  and then in [32] a continuous function such that  $\{G_m(f, \mathcal{T})\}$  diverges almost everywhere. It has been proved in [42] for  $p \neq 2$  and in [7] for  $p < 2$  that there exists an  $f \in L_p(\mathbb{T})$  such that  $\{G_m(f, \mathcal{T})\}$  does not converge in  $L_p$ . It was remarked in [47] that the method from [42] gives a little bit more: 1) There exists a continuous function  $f$  such that  $\{G_m(f, \mathcal{T})\}$  does not converge in  $L_p(\mathbb{T})$  for any  $p > 2$ ; and 2) there exists a function  $f$  that belongs to any  $L_p(\mathbb{T})$ ,  $p < 2$ , such that  $\{G_m(f, \mathcal{T})\}$  does not converge in measure. Thus the above negative results show that the condition  $f \in L_p(\mathbb{T}^d)$ ,  $p \neq 2$ , does not guarantee convergence of  $\{G_m(f, \mathcal{T}^d)\}$  in the  $L_p$ -norm. The main goal of this section is to discuss an additional (to  $f \in L_p$ ) condition on  $f$  to guarantee that  $\|f - G_m(f, \mathcal{T}^d)\|_p \rightarrow 0$  as  $m \rightarrow \infty$ . Some results in this direction have been obtained in [29], [30]. In the case  $2 < p \leq \infty$  we found in [29] necessary and sufficient conditions on a decreasing sequence  $\{A_n\}_{n=1}^\infty$  to guarantee the  $L_p$ -convergence of  $\{G_m(f)\}$  for all  $f \in L_p$ , satisfying  $a_n(f) \leq A_n$ , where  $\{a_n(f)\}$  is a decreasing rearrangement of absolute values of the Fourier coefficients of  $f$ . We will formulate three theorems from [29].

For  $f \in L_1(\mathbb{T}^d)$  let  $\{\hat{f}(k(l))\}_{l=1}^\infty$  denote the decreasing rearrangement of  $\{\hat{f}(k)\}_{k \in \mathbb{Z}^d}$ , i.e.

$$|\hat{f}(k(1))| \geq |\hat{f}(k(2))| \geq \dots \tag{5.2}$$

Denote  $a_n(f) := |\hat{f}(k(n))|$ .

**Theorem 5.1.** *Let  $2 < p < \infty$  and let a decreasing sequence  $\{A_n\}_{n=1}^\infty$  satisfy the condition*

$$A_n = o(n^{1/p-1}) \text{ as } n \rightarrow \infty. \tag{5.3}$$

Then for any  $f \in L_p(\mathbb{T}^d)$  with the property  $a_n(f) \leq A_n, n = 1, 2, \dots$ , we have

$$\lim_{m \rightarrow \infty} \|f - G_m^t(f, \mathcal{T})\|_p = 0. \tag{5.4}$$

We also proved in [29] that for any decreasing sequence  $\{A_n\}$  satisfying

$$\limsup_{n \rightarrow \infty} A_n n^{1-1/p} > 0$$

there exists a function  $f \in L_p$  such that  $a_n(f) \leq A_n, n = 1, \dots$ , with divergent in the  $L_p$  norm sequence of greedy approximants  $\{G_m(f)\}$ .

**Theorem 5.2.** *Let a decreasing sequence  $\{A_n\}_{n=1}^\infty$  satisfy the following condition  $(\mathcal{A}_\infty)$ :*

$$\sum_{M < n \leq e^M} A_n = o(1) \text{ as } M \rightarrow \infty. \tag{5.5}$$

Then for any  $f \in C(\mathbb{T})$  with the property  $a_n(f) \leq A_n, n = 1, 2, \dots$ , we have

$$\lim_{m \rightarrow \infty} \|f - G_m^t(f, \mathcal{T})\|_\infty = 0. \tag{5.6}$$

The following theorem from [29] shows that the condition  $(\mathcal{A}_\infty)$  in Theorem 5.2 is sharp.

**Theorem 5.3.** *Assume that a decreasing sequence  $\{A_n\}_{n=1}^\infty$  does not satisfy the condition  $(\mathcal{A}_\infty)$ . Then there exists a function  $f \in C(\mathbb{T})$  with the property  $a_n(f) \leq A_n, n = 1, 2, \dots$ , and such that we have*

$$\limsup_{m \rightarrow \infty} \|f - G_m(f, \mathcal{T})\|_\infty > 0$$

for some realization  $G_m(f, \mathcal{T})$ .

In [30] we concentrated on imposing extra conditions in the following form. We assume that for some sequence  $\{M(m)\}, M(m) > m$ , we have

$$\|G_{M(m)}(f) - G_m(f)\|_p \rightarrow 0 \text{ as } m \rightarrow \infty. \tag{5.7}$$

In the case that  $p$  is an even number or  $p = \infty$  we found in [30] necessary and sufficient conditions on the growth of the sequence  $\{M(m)\}$  to provide convergence  $\|f - G_m(f)\|_p \rightarrow 0$  as  $m \rightarrow \infty$ . We proved the next theorem in [30].

**Theorem 5.4.** *Let  $p = 2q, q \in \mathbb{N}$ , be an even integer,  $\delta > 0$ . Assume that  $f \in L_p(\mathbb{T})$  and there exists a sequence of positive integers  $M(m) > m^{1+\delta}$  such that*

$$\|G_m(f) - G_{M(m)}(f)\|_p \rightarrow 0 \text{ as } m \rightarrow \infty.$$

Then we have

$$\|G_m(f) - f\|_p \rightarrow 0 \text{ as } m \rightarrow \infty.$$

In [30] we proved that the condition  $M(m) > m^{1+\delta}$  cannot be replaced by the condition  $M(m) > m^{1+o(1)}$ .

**Theorem 5.5.** *For any  $p \in (2, \infty)$  there exists a function  $f \in L_p(\mathbb{T})$  with divergent in the  $L_p(\mathbb{T})$  norm sequence  $\{G_m(f)\}$  of greedy approximations with the following property. For any sequence  $\{M(m)\}$  such that  $m \leq M(m) \leq m^{1+o(1)}$  we have*

$$\|G_{M(m)}(f) - G_m(f)\|_p \rightarrow 0 \quad (m \rightarrow \infty).$$

In [30] we also considered the case  $p = \infty$ . We proved there necessary and sufficient conditions for convergence of greedy approximations in the uniform norm. For a mapping  $\alpha : W \rightarrow W$  we denote by  $\alpha_k$  its  $k$ -fold iteration:  $\alpha_k := \alpha \circ \alpha_{k-1}$ .

**Theorem 5.6.** *Let  $\alpha : \mathbb{N} \rightarrow \mathbb{N}$  be strictly increasing. Then the following conditions are equivalent:*

- a) *For some  $k \in \mathbb{N}$  and for any sufficiently large  $m \in \mathbb{N}$  we have  $\alpha_k(m) > e^m$ .*
- b) *If  $f \in C(\mathbb{T})$  and*

$$\|G_{\alpha(m)}(f) - G_m(f)\|_\infty \rightarrow 0 \quad (m \rightarrow \infty)$$

*then*

$$\|f - G_m(f)\|_\infty \rightarrow 0 \quad (m \rightarrow \infty).$$

The proof of the necessary condition is based on the above Theorem 5.3 from [29]. In the proof of the sufficient condition we use the following special inequality (see [30]).

By  $\Sigma_m(\mathcal{T})$  we denote the set of all trigonometric polynomials with at most  $m$  nonzero coefficients.

**Theorem 5.7.** *For any  $h \in \Sigma_m(\mathcal{T})$  and any  $g \in L_\infty$  one has*

$$\|h + g\|_\infty \geq K^{-2} \|h\|_\infty - e^{C(K)m} \|\{\hat{g}(k)\}\|_{\ell_\infty}, \quad K > 1. \quad (5.8)$$

We note that in the proof of the above inequality we used a deep result on the uniform approximation property of the space  $C(X)$  (see [4]). The paper [30] contains some other inequalities in the style of (5.8).

## 6. General greedy algorithms

The purpose of this section is to discuss nonlinear  $m$ -term approximation and greedy algorithms with regard to a general system (dictionary). We concentrate here on a discussion of  $m$ -term approximation with regard to redundant dictionaries in Banach spaces. We will discuss only one example of an algorithm from the family of greedy

algorithms. The reader can find a further discussion of greedy approximation in Banach spaces in the survey [47]. This section is based on the paper [44] which in turn is a combination of ideas and methods developed for Banach spaces in a fundamental paper [13] with the approach used in [50] in the case of Hilbert spaces. The papers [13] and [50] 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 [44]. 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 turned out that these general approximation methods are also very good for specific dictionaries. It has been observed in [15] that the WCGA provides constructive methods in  $m$ -term trigonometric approximation in  $L_p$ ,  $p \in [2, \infty)$ , which realizes optimal rate of  $m$ -term approximation for different function classes. In [48] 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 looks like WCGA and WRGA can be considered as a constructive deterministic alternative to (substitute for) some powerful probabilistic methods. This observation encouraged us to continue thorough study of WCGA and WRGA.

In this section we discuss in detail only WCGA. In [44] we developed the theory of the Weak Chebyshev Greedy Algorithm 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 name of this approximation method for two reasons. First, this term has been used in previous papers and has become a standard name for procedures like WCGA. For more discussion of the terminology see [47, 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 this section we discuss the following two applications of general greedy algorithms from [49]. In [49] we used WCGA 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 [15]. In [49] we also studied a slight modification of incremental type algorithm from [13]. We applied that algorithm for constructing deterministic sets of points with small  $L_p$  discrepancy and also with small symmetrized  $L_p$  discrepancy.

We now proceed to a systematic presentation of the mentioned above results. 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  $\overline{\text{span}}\mathcal{D} = X$ . We note that in [44] we required in the definition of a dictionary normalization of its elements ( $\|g\| = 1$ ). However, it is pointed out in [49] that it is easy to check that the arguments from [44] work under the assumption  $\|g\| \leq 1$  instead of  $\|g\| = 1$ . In applications it is more convenient for us to have the assumption  $\|g\| \leq 1$  than normalization of a dictionary.

For an element  $f \in X$  we denote by  $F_f$  a norming (peak) functional for  $f$ :

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

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  $t_k \leq 1$ ,  $k = 1, \dots$ . We define (see [44]) the Weak Chebyshev Greedy Algorithm (WCGA) which is a generalization for Banach spaces of the Weak Orthogonal Greedy Algorithm defined and studied in [50] (see also [12] for the Orthogonal Greedy Algorithm).

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

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

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

- 2) Define

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

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

- 3) Denote

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

The term “weak” in this definition means that at step 1) we do not shoot for the optimal element of the dictionary which realizes the corresponding supremum but are satisfied with the weaker property rather 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.

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

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

The uniformly smooth Banach space is the one with the property

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

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

$$\max(0, u - 1) \leq \rho(u) \leq u, \quad u \in (0, \infty). \quad (6.1)$$

It is well known (see for instance [13], Lemma B.1) that in the case  $X = L_p$ ,  $1 \leq p < \infty$  we have

$$\rho(u) \leq \begin{cases} u^p/p & \text{if } 1 \leq p \leq 2, \\ (p-1)u^2/2 & \text{if } 2 \leq p < \infty. \end{cases} \quad (6.2)$$

It is also known (see [34], p. 63) that for any  $X$  with  $\dim X = \infty$  one has

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

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

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

This limits power type moduli of smoothness of nontrivial Banach spaces to the case  $1 \leq q \leq 2$ . Denote by  $A(\mathcal{D})$  the closure of the convex hull of  $\mathcal{D}$ . The following theorem from [44] gives the rate of convergence of the WCGA for  $f$  in  $A(\mathcal{D})$ .

**Theorem 6.1.** *Let  $X$  be a uniformly smooth Banach space with the modulus of smoothness  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Then for a sequence  $\tau := \{t_k\}_{k=1}^\infty$ ,  $t_k \leq 1$ ,  $k = 1, 2, \dots$ , we have for any  $f \in A(\mathcal{D})$  that*

$$\|f - G_m^{c,\tau}(f, \mathcal{D})\| \leq C(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p}, \quad p := \frac{q}{q-1},$$

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

In [49] we demonstrated the power of the WCGA in classical areas of harmonic analysis. The problem concerns the trigonometric  $m$ -term approximation in the uniform norm. Let  $\mathcal{RT}(N)$  be the subspace of real trigonometric polynomials of order  $N$ . Both R. S. Ismagilov [20] and V. E. Maiorov [35] used constructive methods to get their estimates (1.6) and (1.8). V. E. Maiorov [35] 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

$$\sigma_m(t, \mathcal{RT})_\infty \leq CN^{3/2}m^{-1}\|t\|_1, \quad t \in \mathcal{RT}(N). \quad (6.6)$$

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 the simple inequality

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

the estimate (6.6) follows from the estimate

$$\sigma_m(t, \mathcal{RT})_\infty \leq C(N^{1/2}/m)\|t\|_A, \quad t \in \mathcal{RT}(N). \quad (6.7)$$

Thus (6.7) is stronger than (6.6). The following estimate was proved in [11]:

$$\sigma_m(t, \mathcal{RT})_\infty \leq Cm^{-1/2}(\ln(1 + N/m))^{1/2}\|t\|_A, \quad t \in \mathcal{RT}(N). \quad (6.8)$$

In a way (6.8) is much stronger than (6.7) and (6.6). The proof of (6.8) from [11] is not constructive. The estimate (6.8) has been proved in [11] with the help of a nonconstructive theorem of Gluskin [17]. In [49] we gave a constructive proof of (6.8). The key ingredient of that proof is the WCGA. In the paper [15] we already pointed out that the WCGA provides a constructive proof of the estimate

$$\sigma_m(f, \mathcal{T})_p \leq C(p)m^{-1/2}\|f\|_A, \quad p \in [2, \infty). \quad (6.9)$$

The known proofs (before [15]) of (6.9) were nonconstructive (see discussion in [15, Section 5]).

We formulate here a result from [49] (see Theorem 4.1).

**Theorem 6.2.** *There exists a constructive method  $A(N, m)$  such that for any  $t \in \mathcal{RT}(N)$  it provides an  $m$ -term trigonometric polynomial  $A(N, m)(t)$  with the following approximation property:*

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

with an absolute constant  $C$ .

In [49] we applied greedy type algorithms for constructing points with small discrepancy and small symmetrized discrepancy. Let  $1 \leq p \leq \infty$ . We will define first 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  $\xi$  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 mention some of them as a reference for the history of the subject: [33], [3], [36], [5], [37], [48]. For  $1 < p < \infty$  the following relation is known (see [3, p. 5]):

$$D(m, d)_p \asymp m^{-1} (\ln m)^{(d-1)/2} \tag{6.10}$$

with constants in  $\asymp$  depending on  $p$  and  $d$ . The right order of  $D(m, d)_p$ ,  $p = 1, \infty$ , for  $d \geq 3$  is unknown. Recently, driven by possible applications (see [37]) in numerical integration the tendency 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 [19]. The authors established the estimate

$$D(m, d)_\infty \leq Cd^{1/2}m^{-1/2} \tag{6.11}$$

with  $C$  an absolute constant. It is pointed out in [19] that (6.11) is only an existence theorem and even a constant  $C$  in (6.11) is unknown. The proof is a probabilistic one. There are also some other estimates in [19] with explicit constants. We mention one of them:

$$D(m, d)_\infty \leq C(d \ln d)^{1/2}((\ln m)/m)^{1/2} \tag{6.12}$$

with an explicit constant  $C$ . The proof of (6.12) is also probabilistic.

In [49] we gave constructive proofs of the following two upper estimates:

$$D(m, d)_p \leq C_1 p^{1/2} m^{-1/2}, \quad p \in [2, \infty),$$

$$D(m, d)_\infty \leq C_2 d^{3/2} (\max(\ln d, \ln m))^{1/2} m^{-1/2}, \quad d, m \geq 2,$$

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 traditions of approximation theory we understand 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 existence of an object is not a constructive proof for us. In [49] we provided 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

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

with effective absolute constant  $C_1$ . The above method is a greedy type algorithm (see the IA( $\epsilon$ ) below) which is a slight modification of the corresponding procedure from [13]. 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  $\|g\| \leq 1$  if  $g \in \mathcal{D}^+$ . By  $\mathcal{A}_1(\mathcal{D}^+)$  we denote the closure of the convex hull of  $\mathcal{D}^+$ . Let  $\epsilon = \{\epsilon_n\}_{n=1}^\infty$ ,  $\epsilon_n > 0$ ,  $n = 1, 2, \dots$ .

**6.2. Incremental algorithm with schedule  $\epsilon$  ( $\mathbf{IA}(\epsilon)$ ).** Let  $f \in \mathcal{A}_1(\mathcal{D}^+)$ . Denote  $f_0^{i,\epsilon} := f$  and  $G_0^{i,\epsilon} := 0$ . Then for each  $m \geq 1$  we inductively define

1.  $\varphi_m^{i,\epsilon} \in \mathcal{D}^+$  is any element satisfying

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

2. Define

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

3. Denote

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

## References

- [1] Andrianov, A. V., Temlyakov, V. N., Best  $m$ -term approximation of functions from classes  $MW_q^r$ . In *Approximation Theory IX*, Innov. Appl. Math., Vanderbilt University Press, Nashville, TN, 1998, 7–14.
- [2] Baishanski, B. M., Approximation by polynomials of given length. *Illinois J. Math.* **27** (1983), 449–458.
- [3] Beck, J., Chen, W., *Irregularities of distribution*. Cambridge Tracts in Math. 89, Cambridge University Press, Cambridge 1987.
- [4] Bourgain, J., A remark on the behaviour of  $L^p$ -multipliers and the range of operators acting on  $L^p$ -spaces. *Israel J. Math.* **79** (1992), 193–206.
- [5] Chazelle, B., *The Discrepancy Method*. Cambridge University Press, Cambridge 2000.
- [6] Cohen, A., DeVore, R. A., Hochmuth, R., Restricted Nonlinear Approximation. *Constr. Approx.* **16** (2000), 85–113.
- [7] Cordoba, A., Fernandez, P., Convergence and divergence of decreasing rearranged Fourier series. *SIAM J. Math. Anal.* **29** (1998), 1129–1139.
- [8] DeVore, R. A., Nonlinear approximation. *Acta Numerica* (1998), 51–150.
- [9] DeVore, R. A., Jawerth, B., Popov, V., Compression of wavelet decompositions. *Amer. J. Math.* **114** (1992), 737–785.
- [10] DeVore, R. A., Konyagin, S. V., Temlyakov, V. N., Hyperbolic wavelet approximation. *Constr. Approx.* **14** (1998), 1–26.
- [11] DeVore, R. A., Temlyakov, V. N., Nonlinear approximation by trigonometric sums. *J. Fourier Anal. Appl.* **2** (1995), 29–48.
- [12] DeVore, R. A., Temlyakov, V. N., Some remarks on Greedy Algorithms. *Adv. Comput. Math.* **5** (1996), 173–187.
- [13] Donahue, M., Gurvits, L., Darken, C., Sontag, E., Rate of convex approximation in non-Hilbert spaces. *Constr. Approx.* **13** (1997), 187–220.
- [14] Dilworth, S. J., Kalton, N. J., Kutzarova, D., Temlyakov, V. N., The Tresholding Greedy Algorithm, Greedy Bases, and Duality. *Constr. Approx.* **19** (2003), 575–597.

- [15] Dilworth, S. J., Kutzarova, D., Temlyakov, V. N., Convergence of some Greedy Algorithms in Banach spaces. *J. Fourier Anal. Appl.* **8** (2002), 489–505.
- [16] Frazier, M., Jawerth, B., A discrete transform and decomposition of distribution spaces. *J. Funct. Anal.* **93** (1990), 34–170.
- [17] Gluskin, E. D., Extremal properties of orthogonal parallelepipeds and their application to the geometry of Banach spaces. *Math. USSR-Sb.* **64** (1989), 85–96.
- [18] Gribonval, R., Nielsen, M., Some remarks on non-linear approximation with Schauder bases. *East J. Approx.* **7** (2001), 267–285.
- [19] Heinrich, S., Novak, E., Wasilkowski, G., Wozniakowski, H., The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.* **96** (2001), 279–302.
- [20] Ismagilov, R. S., Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials. *Uspekhi Mat. Nauk* **29** (1974), 161–178.
- [21] Kamont, A., Temlyakov, V. N., Greedy Approximation and the Multivariate Haar system. *Studia Math.* **161** (3) (2004), 199–223.
- [22] Kashin, B. S., Widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.* **41** (1977), 334–351.
- [23] Kashin, B. S., On approximation properties of complete orthonormal systems. *Trudy Mat. Inst. Steklov* **172** (1985), 187–191.
- [24] Kashin, B. S., Temlyakov, V. N., On best  $m$ -term approximations and the entropy of sets in the space  $L^1$ . *Math. Notes* **56** (1994), 57–86.
- [25] Kerkyacharian, G., Picard, D., Entropy, universal coding, approximation and bases properties. *Constr. Approx.* **20** (2004), 1–37.
- [26] Kerkyacharian, G., Picard, D., Temlyakov, V.N., Some inequalities for the tensor product of greedy bases and weight-greedy bases. Manuscript, 2005, 1–14.
- [27] Konyagin, S. V., Temlyakov, V. N., A remark on greedy approximation in Banach spaces. *East J. Approx.* **5** (1999), 1–15.
- [28] Konyagin, S. V., Temlyakov, V. N., Greedy approximation with regard to bases and general minimal systems. *Serdica Math. J.* **28** (2002), 305–328.
- [29] Konyagin, S. V., Temlyakov, V. N., Convergence of greedy approximation II. The trigonometric system. *Studia Math.* **159(2)** (2003), 161–184.
- [30] Konyagin, S. V., Temlyakov, V. N., Convergence of Greedy Approximation for the trigonometric system. *Anal. Math.* **31** (2005), 85–115
- [31] Körner, T. W., Divergence of decreasing rearranged Fourier series. *Ann. of Math.* **144** (1996), 167–180.
- [32] Körner, T. W., Decreasing rearranged Fourier series. *J. Fourier Anal. Appl.* **5** (1999), 1–19.
- [33] Kuipers, L., Niederreiter, H., *Uniform distribution of sequences*. Pure and Applied Mathematics, Wiley, New York, London, Sydney 1974.
- [34] Lindenstrauss, J., Tzafriri, L., *Classical Banach Spaces I*. Ergebnisse Math Grenzgeb. 92, Springer-Verlag, Berlin 1977.
- [35] Maiorov, V. E., Trigonometric diameters of the Sobolev classes  $W_p^r$  in the space  $L_q$ . *Math. Notes* **40** (1986), 590–597.
- [36] Matoušek, J., *Geometric Discrepancy*. Algorithms Combin. 18, Springer-Verlag, Berlin 1999.

- [37] Novak, E., Wozniakowski, H., When are Integration and Discrepancy tractable? In *Foundations of computational mathematics* (Oxford, 1999), London Math. Soc. Lecture Notes Ser. 284, Cambridge University Press, Cambridge 2001, 211–266.
- [38] Temlyakov, V. N., Widths of Some Classes of Functions of Several Variables. *Soviet Math. Dokl.* **26** (1982), 619–622.
- [39] Temlyakov, V. N., *Approximation of periodic functions*. Comput. Math. Anal. Ser., Nova Science Publishers, Inc., Commack, NY, 1993.
- [40] Temlyakov, V. N., The best  $m$ -term approximation and Greedy Algorithms. *Adv. Comput. Math.* **8** (1998), 249–265.
- [41] Temlyakov, V. N., Nonlinear  $m$ -term approximation with regard to the multivariate Haar system. *East J. Approx.* **4** (1998), 87–106.
- [42] Temlyakov, V. N., Greedy Algorithm and  $m$ -term Trigonometric Approximation. *Constr. Approx.* **14** (1998), 569–587.
- [43] Temlyakov, V. N., Greedy algorithms with regard to multivariate systems with special structure. *Constr. Approx.* **16** (2000), 399–425.
- [44] Temlyakov, V. N., Greedy algorithms in Banach spaces. *Adv. Comput. Math.* **14** (2001), 277–292.
- [45] Temlyakov, V. N., Universal bases and greedy algorithms for anisotropic function classes. *Constr. Approx.* **18** (2002), 529–550.
- [46] Temlyakov, V. N., Nonlinear approximation with regard to bases. In *Approximation theory X* (St. Louis, MO, 2001), Innov. Appl. Math., Vanderbilt University Press, Nashville, TN, 2002, 373–402.
- [47] Temlyakov, V. N., Nonlinear Methods of Approximation. *Found. Comput. Math.* **3** (2003), 33–107.
- [48] Temlyakov, V. N., Cubature formulas and related questions. *J. Complexity* **19** (2003), 352–391.
- [49] Temlyakov, V. N., Greedy Type Algorithms in Banach Spaces and Applications. *Constr. Approx.* **21** (2005), 257–292.
- [50] Temlyakov, V. N., Weak Greedy Algorithms. *Adv. Comput. Math.* **12** (2000), 213–227.
- [51] Tikhomirov, V. M., Widths of sets in function spaces and the theory of best approximations. *Uspekhi Mat. Nauk* **15** (1960), 81–120.
- [52] Wojtaszczyk, P., On unconditional polynomial bases in  $L_p$  and Bergman spaces. *Constr. Approx.* **13** (1997), 1–15.
- [53] Wojtaszczyk, P., Greedy algorithms for general systems, *J. Approx. Theory* **107** (2000), 293–314.

Department of Mathematics, University of South Carolina, Columbia, SC 29208, U.S.A.

E-mail: temlyak@math.sc.edu