



INDUSTRIAL
MATHEMATICS
INSTITUTE

1997:10

The best m -term approximation
and greedy algorithms

V. Temlyakov

IMI

Preprint Series

Department of Mathematics
University of South Carolina

THE BEST m -TERM APPROXIMATION AND GREEDY ALGORITHMS¹

V.N. TEMLYAKOV

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

ABSTRACT. Two theorems on nonlinear m -term approximation in L_p , $1 < p < \infty$, are proved in this paper. The first one (Theorem 2.1) says that if a basis $\Psi := \{\psi_I\}_I$ is L_p -equivalent to the Haar basis then near best m -term approximation to any $f \in L_p$ can be realized by the following simple greedy type algorithm. Take the expansion $f = \sum_I c_I \psi_I$ and form a sum of m terms with the biggest $\|c_I \psi_I\|_p$ out of this expansion.

The second one (Theorem 3.3) states that nonlinear m -term approximations with regard to two dictionaries: the Haar basis and the set of all characteristic functions of intervals are equivalent in a very strong sense.

1. INTRODUCTION

This paper deals with nonlinear approximation in Banach spaces. Let B be a separable Banach space and \mathcal{D} be a system of elements in B such that $\overline{\text{span}}\mathcal{D} = B$. Consider the best m -term approximation of an element $f \in B$ with regard to the given system (dictionary) \mathcal{D}

$$\sigma_m(f, \mathcal{D})_B := \inf \|f - \sum_{j=1}^m c_j g_j\|_B \quad , m = 1, 2, \dots, \quad \sigma_0(f, \mathcal{D})_B := \|f\|_B \quad ,$$

where \inf is taken over elements $g_j \in \mathcal{D}$ and coefficients $c_j, j = 1, \dots, m$. The quantity $\sigma_m(f, \mathcal{D})_B$ gives the best possible error of approximation of f by a linear combination of m elements from the given dictionary \mathcal{D} . The fundamental question in this study is how to construct an algorithm which provides an error of approximation of f comparable with $\sigma_m(f, \mathcal{D})_B$. The answer to this question in some particular cases is simple. For instance if $B = H$ is a Hilbert space, and \mathcal{D} is an orthonormal basis then the Pure Greedy Algorithm which picks the m biggest in absolute value Fourier coefficients of f with regard to \mathcal{D} realizes the best m -term approximation. In the paper [T1] we studied the performance of Pure Greedy Algorithm with regard to the trigonometric system $\mathcal{T} := \{e^{i(k,x)}\}_{k \in \mathbb{Z}^d}$ in the Banach spaces $L_p(\mathbb{T}^d), 1 \leq p \leq \infty$. We proved there the inequality for approximation of individual function

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

where $h(p) := |1/2 - 1/p|$. This inequality is sharp (in the sense of order) and can be extended to other orthonormal uniformly bounded bases. We note here that the

¹This research was supported by the National Science Foundation Grant DMS 9622925 and by ONR Grant N00014-96-1-1003

use of Pure Greedy Algorithm G_m for approximation in L_p seems reasonable in the case of uniformly bounded basis $\{\phi_j\}$ because in this case we have

$$1/M \leq \|\phi_j\|_1 \leq \|\phi_j\|_p \leq \|\phi_j\|_\infty \leq M.$$

This means that L_p -norm of each summand $\langle f, \phi_j \rangle \phi_j$ is of order of $|\langle f, \phi_j \rangle|$.

In this paper we consider another important class of bases. A typical representative of this class is the Haar basis $\mathcal{H} := \{H_I\}_I$, where I are dyadic intervals of the form $I = [(j-1)2^{-n}, j2^{-n})$, $j = 1, \dots, 2^n$; $n = 0, 1, \dots$ and $I = [0, 1]$ with

$$H_{[0,1]}(x) = 1 \quad \text{for } x \in [0, 1) \quad ,$$

$$H_{[(j-1)2^{-n}, j2^{-n})} = \begin{cases} 2^{n/2}, & x \in [(j-1)2^{-n}, (j-1/2)2^{-n}) \\ -2^{n/2}, & x \in [(j-1/2)2^{-n}, j2^{-n}) \\ 0, & \text{otherwise.} \end{cases}$$

For the Haar basis \mathcal{H} we define for each $1 \leq p \leq \infty$ the Greedy Algorithm G^p which acts as follows. Denote

$$c_I(f) := \langle f, H_I \rangle = \int_0^1 f(x)H_I(x)dx \quad ,$$

and

$$c_I(f, p) := \|c_I(f)H_I\|_p \quad .$$

Let Λ be a set of m dyadic intervals I for which $c_I(f, p)$ take the biggest values. We set

$$G_m^p(f, \mathcal{H}) := \sum_{I \in \Lambda} c_I(f)H_I \quad .$$

Remark 1.1. *There is an algorithm which for any $f \in L_p$ gives an I with the biggest $\|c_I(f)H_I\|_p$ after finite number of steps provided we can calculate each $c_I(f)$ and the L_p -norm of a function in finite number of steps.*

We describe this algorithm now. Let $f \neq 0$. We find a nonzero coefficient $c_J(f)$ and denote $\epsilon := |c_J(f)|$. Next, we find n such that

$$\|f - \sum_{|I| \geq 2^{-n}} c_I(f)H_I\|_p < \epsilon.$$

This guarantees that for all $|I| < 2^{-n}$ we have $\|c_I(f)H_I\|_p < \epsilon$ and, therefore, we can restrict our search for the biggest $\|c_I(f)H_I\|_p$ to the finite number of I with $|I| \geq 2^{-n}$.

In Section 2 we prove that for any $1 < p < \infty$ we have for $f \in L_p$

$$\|f - G_m^p(f, \mathcal{H})\|_p \leq C(p)\sigma_m(f, \mathcal{H})_p \quad .$$

This means that the Greedy Algorithm G_m^p realizes near best m -term approximation. We also prove that the same inequality holds for bases equivalent to the Haar basis.

In Section 3 we study the following general problem in one special case. Assume we have two dictionaries \mathcal{D}_1 and \mathcal{D}_2 in B and want to compare their efficiency for m -term approximation. Let us introduce the following quasinorm in B for $0 < \alpha < \infty$ and $0 < \beta \leq \infty$

$$|f; \alpha, \beta, \mathcal{D}, B| := (\|f\|^\beta + \sum_{n=0}^{\infty} (2^{\alpha n} \sigma_{2^n}(f, \mathcal{D})_B)^\beta)^{1/\beta}.$$

We call two dictionaries \mathcal{D}_1 and \mathcal{D}_2 (α, β) -equivalent if for any $f \in B$ the quasinorms $|f; \alpha, \beta, \mathcal{D}_1, B|$ and $|f; \alpha, \beta, \mathcal{D}_2, B|$ are equivalent. In particular, (α, ∞) -equivalence means that

$$\sigma_m(f, \mathcal{D}_1)_B \ll m^{-\alpha}$$

if and only if

$$\sigma_m(f, \mathcal{D}_2)_B \ll m^{-\alpha}.$$

We discuss in Section 3 one concrete pair of dictionaries: $\mathcal{D}_1 = \chi := \{|J|^{-1/2} \chi_J, J \subset [0, 1]\}$ – the set of all characteristic functions of intervals (normalized in L_2); and $\mathcal{D}_2 = \mathcal{H}$ – the Haar basis. It is clear that

$$(1.1) \quad \sigma_m(f, \mathcal{H})_p \geq \sigma_{2m}(f, \chi)_p.$$

We prove that these two dictionaries are (α, β) -equivalent for any α and β .

We note that $\sigma_m(f, \chi)_p$ is closely related to approximation by splines with free knots. This is based on the following simple remark.

Remark 1.2. *For any set of intervals J_1, \dots, J_m in $[0, 1]$ there exists a set of disjoint intervals $J_1^d, \dots, J_{2m+1}^d, \cup_{j=1}^{2m+1} J_j^d = [0, 1]$, such that any function f of the form*

$$f = \sum_{i=1}^m c_i \chi_{J_i}$$

can be rewritten in the form

$$f = \sum_{j=1}^{2m+1} b_j \chi_{J_j^d}.$$

The proof of this remark can be carried over by induction on m .

Combining known results about approximation by splines with free knots (see [DL], Ch. 12, s. 8, p. 388) and known results on m -term Haar approximation and using Remark 1.2 we get that the dictionaries χ and \mathcal{H} are $(\alpha, (\alpha + 1/p)^{-1})$ -equivalent, $0 < \alpha < 1$.

2. GREEDY ALGORITHMS FOR BASES EQUIVALENT TO THE HAAR BASIS

Let $\Psi := \{\psi_I\}_I$ be a basis in $L_p[0, 1]$ indexed by dyadic intervals I . We say that Ψ is L_p -equivalent to \mathcal{H} if there exist two positive constants $C_1(p)$ and $C_2(p)$ such that for any finite set of coefficients c_I we have

$$(2.1) \quad C_1(p) \left\| \sum_I c_I H_I \right\|_p \leq \left\| \sum_I c_I \psi_I \right\|_p \leq C_2(p) \left\| \sum_I c_I H_I \right\|_p.$$

One can find a discussion and some results about bases L_p -equivalent to \mathcal{H} in the paper [DKT].

For a given basis Ψ we define the Greedy Algorithm $G^p(\cdot, \Psi)$ as follows. Let

$$f = \sum_I c_I(f, \Psi) \psi_I$$

and

$$c_I(f, p, \Psi) := \|c_I(f, \Psi) \psi_I\|_p.$$

Then $c_I(f, p, \Psi) \rightarrow 0$ as $|I| \rightarrow 0$. Denote Λ_m a set of m dyadic intervals I such that

$$(2.2) \quad \min_{I \in \Lambda_m} c_I(f, p, \Psi) \geq \max_{J \notin \Lambda_m} c_J(f, p, \Psi).$$

We define $G^p(\cdot, \Psi)$ by formula

$$G_m^p(f, \Psi) := \sum_{I \in \Lambda_m} c_I(f, \Psi) \psi_I.$$

Theorem 2.1. *Let $1 < p < \infty$ and a basis $\Psi := \{\psi_I\}_I$ be L_p -equivalent to \mathcal{H} . Then for any $f \in L_p$ we have*

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

Proof. Let us take a parameter $0 < t \leq 1$ and consider the following greedy type algorithm $G^{p,t}$ with regard to the Haar system. Denote $\Lambda_m(t)$ any set of m dyadic intervals such that

$$(2.3) \quad \min_{I \in \Lambda_m(t)} c_I(f, p) \geq t \max_{J \notin \Lambda_m(t)} c_J(f, p),$$

and define

$$G_m^{p,t}(f) := \sum_{I \in \Lambda_m(t)} c_I(f) H_I.$$

For a given function $f \in L_p$ we define

$$g(f) := \sum_I c_I(f, \Psi) H_I.$$

It is clear that $g(f) \in L_p$ and

$$(2.4) \quad \sigma_m(g(f), \mathcal{H})_p \leq C_1(p)^{-1} \sigma_m(f, \Psi)_p.$$

Next, for any two intervals $I \in \Lambda_m$, $J \notin \Lambda_m$ by the definition of Λ_m we have

$$c_I(f, p, \Psi) \geq c_J(f, p, \Psi).$$

Using (2.1) we get from here

$$(2.5) \quad \begin{aligned} \|c_I(g(f)) H_I\|_p &= \|c_I(f, \Psi) H_I\|_p \geq C_2(p)^{-1} \|c_I(f, \Psi) \psi_I\|_p = \\ &= C_2(p)^{-1} c_I(f, p, \Psi) \geq C_2(p)^{-1} c_J(f, p, \Psi) = \\ &= C_2(p)^{-1} \|c_J(f, \Psi) \psi_J\|_p \geq C_1(p) C_2(p)^{-1} \|c_J(g(f)) H_J\|_p. \end{aligned}$$

This inequality implies that for any m we can find a set $\Lambda_m(t)$, where $t = C_1(p) C_2(p)^{-1}$, such that $\Lambda_m(t) = \Lambda_m$ and, therefore,

$$(2.6) \quad \|f - G_m^p(f, \Psi)\|_p \leq C_2(p) \|g(f) - G_m^{p,t}(g(f))\|_p.$$

The relations (2.4) and (2.6) show that Theorem 2.1 will follow from Theorem 2.2.

Theorem 2.2. *Let $1 < p < \infty$ and $0 < t \leq 1$. Then for any $g \in L_p$ we have*

$$\|g - G_m^{p,t}(g)\|_p \leq C(p, t)\sigma_m(g, \mathcal{H})_p.$$

Proof. The Littlewood-Paley Theorem for the Haar system (see for instance [KS]) gives for $1 < p < \infty$

$$(2.7) \quad C_3(p)\left\|\left(\sum_I |c_I(g)H_I|^2\right)^{1/2}\right\|_p \leq \|g\|_p \leq C_4(p)\left\|\left(\sum_I |c_I(g)H_I|^2\right)^{1/2}\right\|_p.$$

We formulate first two simple corollaries from (2.7) :

$$(2.8) \quad \|g\|_p \leq C_5(p)\left(\sum_I \|c_I(g)H_I\|_p^p\right)^{1/p}, \quad 1 < p \leq 2,$$

$$(2.9) \quad \|g\|_p \leq C_6(p)\left(\sum_I \|c_I(g)H_I\|_p^2\right)^{1/2}, \quad 2 \leq p < \infty.$$

Analogous of these inequalities for the trigonometric system are known (see, for instance, [T2], p. 37). The same proof gives (2.8) and (2.9).

The dual inequalities to (2.8) and (2.9) are

$$(2.10) \quad \|g\|_p \geq C_7(p)\left(\sum_I \|c_I(g)H_I\|_p^2\right)^{1/2}, \quad 1 < p \leq 2,$$

$$(2.11) \quad \|g\|_p \geq C_8(p)\left(\sum_I \|c_I(g)H_I\|_p^p\right)^{1/p}, \quad 2 \leq p < \infty.$$

We proceed to the proof of Theorem 2.2. Let T_m be an m -term Haar polynomial of best m -term approximation to g in L_p (for existence see [D]):

$$T_m = \sum_{I \in \Lambda} a_I H_I, \quad |\Lambda| = m.$$

For any finite set Q of dyadic intervals we denote by S_Q the projector

$$S_Q(f) := \sum_{I \in Q} c_I(f)H_I.$$

From (2.7) we get

$$(2.12) \quad \|g - S_\Lambda(g)\|_p = \|g - T_m - S_\Lambda(g - T_m)\|_p \leq \|Id - S_\Lambda\|_{p \rightarrow p}\sigma_m(g, \mathcal{H})_p \leq C_4(p)C_3(p)^{-1}\sigma_m(g, \mathcal{H})_p,$$

where Id denotes the identical operator. Further, we have

$$G_m^{p,t}(g) = S_{\Lambda_m(t)}(g),$$

and

$$(2.13) \quad \|g - G_m^{p,t}(g)\|_p \leq \|g - S_\Lambda(g)\|_p + \|S_\Lambda(g) - S_{\Lambda_m(t)}(g)\|_p.$$

The first term in the right side of (2.13) has been estimated in (2.12). We estimate now the second term. We represent it in the form

$$S_\Lambda(g) - S_{\Lambda_m(t)}(g) = S_{\Lambda \setminus \Lambda_m(t)}(g) - S_{\Lambda_m(t) \setminus \Lambda}(g)$$

and remark that similarly to (2.12) we get

$$(2.14) \quad \|S_{\Lambda_m(t) \setminus \Lambda}(g)\|_p \leq C_9(p) \sigma_m(g, \mathcal{H})_p.$$

The key point of the proof of Theorem 2.2 is the estimate

$$(2.15) \quad \|S_{\Lambda \setminus \Lambda_m(t)}(g)\|_p \leq C(p, t) \|S_{\Lambda_m(t) \setminus \Lambda}(g)\|_p$$

which will be derived from the following two lemmas.

Lemma 2.1. *Consider*

$$f = \sum_{I \in Q} c_I H_I, \quad |Q| = N.$$

Let $1 \leq p < \infty$. Assume

$$(2.16) \quad \|c_I H_I\|_p \leq 1, \quad I \in Q.$$

Then

$$\|f\|_p \leq C_{10}(p) N^{1/p}.$$

Lemma 2.2. *Consider*

$$f = \sum_{I \in Q} c_I H_I, \quad |Q| = N.$$

Let $1 < p \leq \infty$. Assume

$$\|c_I H_I\|_p \geq 1, \quad I \in Q.$$

Then

$$\|f\|_p \geq C_{11}(p) N^{1/p}.$$

Proof of Lemma 2.1. We note that in the case $1 < p \leq 2$ the statement of Lemma 2.1 follows from (2.8). We will give a proof of this lemma for all $1 \leq p < \infty$. We have

$$\|c_I H_I\|_p = |c_I| |I|^{1/p-1/2}.$$

The assumption (2.16) implies

$$|c_I| \leq |I|^{1/2-1/p}.$$

Next, we have

$$(2.17) \quad \|f\|_p \leq \left\| \sum_{I \in Q} |c_I H_I| \right\|_p \leq \left\| \sum_{I \in Q} |I|^{-1/p} \chi_I(x) \right\|_p,$$

where $\chi_I(x)$ is a characteristic function of the interval I

$$\chi_I(x) = \begin{cases} 1, & x \in I \\ 0, & x \notin I. \end{cases}$$

In order to proceed further we need a lemma.

Lemma 2.3. *Let $n_1 < n_2 < \dots < n_s$ be integers and let $E_j \subset [0, 1]$ be measurable sets, $j = 1, \dots, s$. Then for any $0 < q < \infty$ we have*

$$\int_0^1 \left(\sum_{j=1}^s 2^{n_j/q} \chi_{E_j}(x) \right)^q dx \leq C_{12}(q) \sum_{j=1}^s 2^{n_j} |E_j|.$$

Proof. Denote

$$F(x) := \sum_{j=1}^s 2^{n_j/q} \chi_{E_j}(x)$$

and estimate it on the sets

$$E_l^- := E_l \setminus \cup_{k=l+1}^s E_k, \quad l = 1, \dots, s-1; \quad E_s^- := E_s.$$

We have for $x \in E_l^-$

$$F(x) \leq \sum_{j=1}^l 2^{n_j/q} \leq C(q) 2^{n_l/q}.$$

Therefore,

$$\int_0^1 F(x)^q dx \leq C(q)^q \sum_{l=1}^s 2^{n_l} |E_l^-| \leq C(q)^q \sum_{l=1}^s 2^{n_l} |E_l|,$$

what proves the lemma.

We return to the proof of Lemma 2.1. Denote by $n_1 < n_2 < \dots < n_s$ all integers such that there is $I \in Q$ with $|I| = 2^{-n_j}$. Introduce the sets

$$E_j := \cup_{I \in Q; |I|=2^{-n_j}} I.$$

Then the number N of elements in Q can be written in the form

$$(2.18) \quad N = \sum_{j=1}^s |E_j| 2^{n_j}.$$

Using these notations the right hand side of (2.17) can be rewritten as

$$Y := \left(\int_0^1 \left(\sum_{j=1}^s 2^{n_j/p} \chi_{E_j}(x) \right)^p dx \right)^{1/p}.$$

Applying Lemma 2.3 with $q = p$ we get

$$\|f\|_p \leq Y \leq C_{13}(p) \left(\sum_{j=1}^s |E_j| 2^{n_j} \right)^{1/p} = C_{13}(p) N^{1/p}.$$

On the last step we used (2.18). Lemma 2.1 is proved now.

Proof of Lemma 2.2. We derive Lemma 2.2 from Lemma 2.1. Define

$$u := \sum_{I \in Q} \bar{c}_I |c_I|^{-1} |I|^{1/p-1/2} H_I,$$

where the bar means complex conjugate number. Then for $p' = \frac{p}{p-1}$ we have

$$\|\bar{c}_I |c_I|^{-1} |I|^{1/p-1/2} H_I\|_{p'} = 1$$

and by Lemma 2.1

$$(2.19) \quad \|u\|_{p'} \leq C_{10}(p) N^{1/p'}.$$

Consider $\langle f, u \rangle$. We have on one hand

$$(2.20) \quad \langle f, u \rangle = \sum_{I \in Q} |c_I| |I|^{1/p-1/2} = \sum_{I \in Q} \|c_I H_I\|_p \geq N,$$

on the other hand

$$(2.21) \quad \langle f, u \rangle \leq \|f\|_p \|u\|_{p'}.$$

Combining (2.19) – (2.21) we get the statement of Lemma 2.2.

We complete now the proof of Theorem 2.2. It remained to prove the inequality (2.15). Denote

$$A := \max_{I \in \Lambda \setminus \Lambda_m(t)} \|c_I(g) H_I\|_p,$$

and

$$B := \min_{I \in \Lambda_m(t) \setminus \Lambda} \|c_I(g) H_I\|_p.$$

Then by the definition of $\Lambda_m(t)$ we have

$$(2.22) \quad B \geq tA.$$

Using Lemma 2.1 we get

$$(2.23) \quad \|S_{\Lambda \setminus \Lambda_m(t)}(g)\|_p \leq AC_{10}(p) |\Lambda \setminus \Lambda_m(t)|^{1/p} \leq t^{-1} BC_{10}(p) |\Lambda \setminus \Lambda_m(t)|^{1/p}.$$

Using Lemma 2.2 we get

$$(2.24) \quad \|S_{\Lambda_m(t) \setminus \Lambda}(g)\|_p \geq BC_{11}(p) |\Lambda_m(t) \setminus \Lambda|^{1/p}.$$

Taking into account that $|\Lambda_m(t) \setminus \Lambda| = |\Lambda \setminus \Lambda_m(t)|$ we get from (2.23) and (2.24) the relation (2.15).

The proof of Theorem 2.2 is complete now.

We discuss now the multivariate analog of Theorem 2.1. There are several natural generalizations of the Haar system to the d -dimensional case. We describe here that

one for which the statement of Theorem 2.1 and its proof coincide with the one-dimensional version. First of all we include in the system the constant function

$$H_{[0,1]^d}(x) = 1, \quad x \in [0, 1]^d.$$

Next we define $2^d - 1$ functions with support $[0, 1]^d$. Take any combination of intervals Q_1, \dots, Q_d where $Q_i = [0, 1]$ or $Q_i = [0, 1)$ with at least one $Q_j = [0, 1)$, and define for $Q = Q_1 \times \dots \times Q_d$, $x = (x_1, \dots, x_d)$,

$$H_Q(x) := \prod_{i=1}^d H_{Q_i}(x_i).$$

We shall also denote these functions by $H_{[0,1]^d}^k(x)$, $k = 1, \dots, 2^d - 1$. We define the basis of Haar functions with supports on dyadic cubes of the form

$$(2.25) \quad J = [(j_1 - 1)2^{-n}, j_1 2^{-n}) \times \dots \times [(j_d - 1)2^{-n}, j_d 2^{-n}),$$

$$j_i = 1, \dots, 2^n; \quad n = 0, 1, \dots$$

For each dyadic cube of the form (2.25) we define $2^d - 1$ basis functions

$$H_J^k(x) := 2^{nd/2} H_{[0,1]^d}^k(2^n(x - (j_1 - 1, \dots, j_d - 1)2^{-n})), \quad k = 1, \dots, 2^d - 1.$$

We can also use another enumeration of these functions. Let $H_{[0,1]^d}^k(x) = H_Q(x)$ with

$$Q = Q_1 \times \dots \times Q_d, \quad Q_i = [0, 1), \quad i \in E, \quad Q_i = [0, 1], \quad i \in \{1, d\} \setminus E, \quad E \neq \emptyset.$$

Consider a dyadic interval I of the form

$$(2.26) \quad I = I_1 \times \dots \times I_d, \quad I_i = [(j_i - 1)2^{-n}, j_i 2^{-n}), \quad i \in E,$$

$$I_i = [(j_i - 1)2^{-n}, j_i 2^{-n}], \quad i \in \{1, d\} \setminus E, \quad E \neq \emptyset$$

and define $H_I(x) := H_J^k(x)$. Denoting the set of dyadic intervals D as the set of all dyadic cubes of the form (2.26) amended by the cube $[0, 1]^d$ and denoting by \mathcal{H} the corresponding basis $\{H_I\}_{I \in D}$ we get the multivariate Haar system.

Remark 2.1. *Theorem 2.1 holds for the multivariate Haar system \mathcal{H} with the constant $C(p)$ allowed to depend also on d .*

We studied in this section approximation in $L_p([0, 1])$ and made a remark about approximation in $L_p([0, 1]^d)$. We can treat in the same way approximation in $L_p(\mathbb{R}^d)$.

Remark 2.2. *Theorem 2.1 holds for approximation in $L_p(\mathbb{R}^d)$.*

Results on approximation of function classes using multivariate greedy algorithm $G_m^p(\cdot, \Psi)$ can be found in [DJP].

3. COMPARISON OF TWO DICTIONARIES

We begin this section by one result about approximation of functions from $\mathcal{A}_\tau(\mathcal{H}, L_p, A)$ by m -term Haar polynomials. Let $0 < \tau < \infty$; we denote by $\mathcal{A}_\tau(\mathcal{H}, L_p, A)$ the set of functions $f \in L_p$ such that

$$(3.1) \quad |f; \mathcal{A}_\tau(\mathcal{H}, L_p)| := \left(\sum_I \|c_I(f)H_I\|_p^\tau \right)^{1/\tau} \leq A.$$

It seems that the set of classes $\mathcal{A}_\tau(\mathcal{H}, L_p, A)$ is a natural replacement for standard Hölder, Sobolev or Besov smoothness classes when we study nonlinear m -term approximation instead of linear approximation. This opinion is based on the following two arguments. 1). The terms $c_I H_I$ have the same weight in the definition of the class $\mathcal{A}_\tau(\mathcal{H}, L_p, A)$ for all I . 2). There are some results which describe the classes of functions with a given decay of their best m -term approximation in terms of \mathcal{A}_τ classes. We illustrate this statement by a particular case of one well-known result of Stechkin (see for instance [DT]). We have the equivalence

$$\sum_{m=0}^{\infty} \left(\frac{\sigma_m(f, \mathcal{H})_2}{(m+1)^{1/2}} \right)^\tau < \infty \Leftrightarrow |f; \mathcal{A}_\tau(\mathcal{H}, L_2)| < \infty.$$

Further results in this direction could be found in [DJP]. In particular, the following Theorem 3.1 can be derived from Theorem 2.1 in [DJP]. We give here another proof of Theorem 3.1.

Theorem 3.1. *Let $0 < \tau < p < \infty$ be given. There exists $C(\tau, p)$ such that for any $m \in \mathbb{N}$ we have*

$$\sigma_m(f, \mathcal{H})_p \leq C(\tau, p) m^{1/p-1/\tau} |f; \mathcal{A}_\tau(\mathcal{H}, L_p)|.$$

Proof. We start with the case $p \leq 1$ (under the assumption $\tau < 1$). In this case $\|\cdot\|_p$ is a quasinorm. We use the G_m^p algorithm defined in Section 1. Then

$$(3.2) \quad \|f - G_m^p(f, \mathcal{H})\|_p^p = \int_0^1 \left| \sum_{I \notin \Lambda} c_I(f)H_I \right|^p dx \leq \sum_{I \notin \Lambda} \|c_I(f)H_I\|_p^p.$$

We use now the following simple and well-known lemma (see for instance [T2], p.97).

Lemma 3.1. *Let $y_1 \geq y_2 \geq \dots \geq 0$ and for some $\tau > 0$*

$$\sum_{k=1}^{\infty} y_k^\tau \leq A^\tau.$$

Then for any $p > \tau$ we have

$$\left(\sum_{k=m}^{\infty} y_k^p \right)^{1/p} \leq A m^{1/p-1/\tau}.$$

Using this lemma with $y_j := \|c_{I_j}(f)H_{I_j}\|_p$, where

$$\|c_{I_1}(f)H_{I_1}\|_p \geq \|c_{I_2}(f)H_{I_2}\|_p \geq \dots$$

we get from (3.2) the required estimate.

In the case $1 < p \leq 2$ we use the above arguments with (3.2) replaced in accordance with (2.8) by

$$(3.3) \quad \|f - G_m^p(f, \mathcal{H})\|_p \leq C_5(p) \left(\sum_{I \notin \Lambda} \|c_I(f)H_I\|_p^p \right)^{1/p}.$$

Let us proceed to the remaining case $2 < p < \infty$. Our proof in this case will use Lemma 2.1. We keep the above notation $y_j, j = 1, 2, \dots$. Then we have

$$(3.4) \quad A := |f; \mathcal{A}_\tau(\mathcal{H}, L_p)| = \left(\sum_{j=1}^{\infty} y_j^\tau \right)^{1/\tau}.$$

In particular, this implies

$$(3.5) \quad y_m \leq Am^{-1/\tau}.$$

Denote $m_0 := m$ and denote by $m_l, l = 1, 2, \dots$, the index such that

$$y_{m_l} \geq y_m 2^{-l}, \quad y_{m_{l+1}} < y_m 2^{-l}.$$

Then for $m_{l-1} < k \leq m_l$ we have

$$(3.6) \quad y_m 2^{-l} \leq y_k \leq y_m 2^{-l+1}.$$

Define $N_l := m_l - m_{l-1}$. The relations (3.4) and (3.6) imply

$$(3.7) \quad \sum_{l=1}^{\infty} (y_m 2^{-l})^\tau N_l \leq A^\tau.$$

Further, we have

$$(3.8) \quad \delta := \|f - G_m^p(f, \mathcal{H})\|_p = \left\| \sum_{j=m+1}^{\infty} c_{I_j}(f)H_{I_j} \right\|_p \leq \sum_{l=1}^{\infty} \|f_l\|_p$$

where we have denoted

$$f_l := \sum_{j=m_{l-1}+1}^{m_l} c_{I_j}(f)H_{I_j}.$$

By Lemma 2.1 we get

$$\|f_l\|_p \leq y_m 2^{-l+1} C_{10}(p) N_l^{1/p}.$$

From here and (3.8) we obtain

$$(3.9) \quad \delta \leq 2C_{10}(p) \sum_{l=1}^{\infty} y_m 2^{-l} N_l^{1/p} = 2C_{10}(p) \sum_{l=1}^{\infty} ((y_m 2^{-l})^\tau N_l)^{1/p} (y_m 2^{-l})^{1-\tau/p} \leq \\ \leq C_1(\tau, p) \left(\sum_{l=1}^{\infty} (y_m 2^{-l})^\tau N_l \right)^{1/p} y_m^{1-\tau/p}.$$

Using (3.7), (3.5) and (3.4) we get from (3.9)

$$\delta \leq C_1(\tau, p) A m^{1/p-1/\tau} = C_1(\tau, p) m^{1/p-1/\tau} |f; \mathcal{A}_\tau(\mathcal{H}, L_p)|,$$

what completes the proof of Theorem 3.1.

We proceed now to studying approximation of linear combinations of characteristic functions by m -term Haar polynomials. We introduce some notations convenient for us. Denote for any interval J

$$U_J := |J|^{-1/2} \chi_J$$

and for $s \in \mathbb{N}$

$$F(s, \chi) := \{f : f = \sum_{i=1}^s b_i U_{J_i}, \quad J_i \subset [0, 1], i = 1, \dots, s\}.$$

Lemma 3.2. *For any $0 < \tau \leq 1 < p < \infty$ there exists $C(\tau, p)$ such that for $f \in F(s, \chi)$ we have*

$$|f; \mathcal{A}_\tau(\mathcal{H}, L_p)| \leq C(\tau, p) s^{1/\tau-1/p} \|f\|_p.$$

Proof. Consider first the Fourier-Haar expansion of U_J for some $J \subset [0, 1]$. For each level k of Haar functions $H_I, |I| = 2^{-k}$, at most two functions $H_{I^1(J)}$ and $H_{I^2(J)}$ will have nonzero inner product with U_J . For these $I^i(J), i = 1, 2$, we have

$$(3.10) \quad |\langle U_J, H_{I^i(J)} \rangle| \leq \min\left(\left(\frac{|J|}{|I^i(J)|}\right)^{1/2}, \left(\frac{|I^i(J)|}{|J|}\right)^{1/2}\right).$$

Thus we have for any $|\theta| < 1/2$ and $\tau > 0$

$$(3.11) \quad \sum_I (|c_I(U_J)| (|I|/|J|)^\theta)^\tau \leq C(\theta, \tau)^\tau.$$

Consider now

$$f = \sum_{i=1}^s b_i U_{J_i}.$$

Using Remark 1.2 we represent f in the form

$$f = \sum_{k=1}^{2s+1} a_k U_{J_k^d},$$

with disjoint J_1^d, \dots, J_{2s+1}^d . Then

$$(3.12) \quad \|f\|_p^p = \sum_{k=1}^{2s+1} \|a_k U_{J_k^d}\|_p^p = \sum_{k=1}^{2s+1} |a_k|^p |J_k^d|^{1-p/2}.$$

Take any $0 < \tau \leq 1$ and estimate

$$(3.13) \quad \sum_I \|c_I(f) H_I\|_p^\tau = \sum_I |c_I(f)|^\tau |I|^{\tau(1/p-1/2)}.$$

We have

$$|c_I(f)|^\tau = \left| \sum_{k=1}^{2s+1} a_k c_I(U_{J_k^d}) \right|^\tau \leq \sum_{k=1}^{2s+1} |a_k|^\tau |c_I(U_{J_k^d})|^\tau.$$

From here and (3.13) we get

$$\begin{aligned} \sum_I \|c_I(f) H_I\|_p^\tau &\leq \sum_{k=1}^{2s+1} |a_k|^\tau \sum_I |c_I(U_{J_k^d})|^\tau |I|^{\tau(1/p-1/2)} = \\ &\sum_{k=1}^{2s+1} (|a_k| |J_k^d|^{1/p-1/2})^\tau \sum_I (|c_I(U_{J_k^d})| (\frac{|I|}{|J_k^d|})^{1/p-1/2})^\tau \leq \end{aligned}$$

Using (3.11) with $\theta = 1/p - 1/2$ we continue

$$\leq C_1(\tau, p)^\tau \sum_{k=1}^{2s+1} (|a_k| |J_k^d|^{1/p-1/2})^\tau \leq$$

$$C_1(\tau, p)^\tau (2s+1)^{1-\tau/p} \left(\sum_{k=1}^{2s+1} (|a_k| |J_k^d|^{1/p-1/2})^p \right)^{\tau/p} = C_1(\tau, p)^\tau (2s+1)^{1-\tau/p} \|f\|_p^\tau,$$

what completes the proof of Lemma 3.2.

We use Lemma 3.2 and Theorem 3.1 to prove an upper estimate for $\sigma_m(f, \mathcal{H})_p$ in terms of $\sigma_n(f, \chi)_p$.

Theorem 3.2. *For any $1 < p < \infty$ and $r > 0$ there exist positive $C, C(p), C(p, r)$ and $\theta(p)$ such that for any $n = 1, 2, \dots$, we have*

$$\sigma_{C2^n}(f, \mathcal{H})_p \leq C(p, r) \sum_{k=0}^n \sigma_{2^k}(f, \chi)_p 2^{-r(n-k)} + C(p) \|f\|_p 2^{-\theta(p)2^{n/2}}.$$

Proof. Let $\epsilon > 0$ be an arbitrary number and let $t_k, k = 0, 1, \dots, n$, be a ϵ -best 2^k -term approximation of f with regard to χ in the L_p -norm:

$$\|f - t_k\|_p \leq \min(\sigma_{2^k}(f, \chi)_p + \epsilon, \|f\|_p).$$

We represent f in the form

$$(3.14) \quad f = f - t_n + t_n - t_{n-1} + \dots + t_1 - t_0 + t_0.$$

We have

$$(3.15) \quad \|f - t_n\|_p \leq \sigma_{2^n}(f, \chi)_p + \epsilon,$$

and

$$(3.16) \quad \|t_k - t_{k-1}\|_p \leq 2\sigma_{2^{k-1}}(f, \chi)_p + 2\epsilon, \quad k = 1, 2, \dots, n.$$

For t_0 we have

$$(3.17) \quad \|t_0\|_p \leq 2\|f\|_p.$$

For $m = m_0 + m_1 + \dots + m_n$ we get from the representation (3.14)

$$(3.18) \quad \sigma_m(f, \mathcal{H})_p \leq \|f - t_n\|_p + \sum_{k=1}^n \sigma_{m_k}(t_k - t_{k-1}, \mathcal{H})_p + \sigma_{m_0}(t_0, \mathcal{H})_p.$$

We choose now $m_k := \lceil 2^{(n+k)/2} \rceil$. Then

$$m = \sum_{k=0}^n m_k \leq C2^n.$$

Next,

$$t_k - t_{k-1} \in F(2^k + 2^{k-1}, \chi).$$

Using Lemma 3.2 and Theorem 3.1 with $\tau = (2r + 1/p)^{-1}$ we get

$$(3.19) \quad \sigma_{m_k}(t_k - t_{k-1}, \mathcal{H})_p \leq C(r, p) \left(\frac{2^k}{m_k}\right)^{2r} \|t_k - t_{k-1}\|_p \leq \\ C(r, p) 2^{-(n-k)} (\sigma_{2^{k-1}}(f, \chi)_p + \epsilon).$$

At the last step we used the definition of m_k and (3.16).

Let us consider now $\sigma_{m_0}(t_0, \mathcal{H})_p$. We estimate first $\sigma_{4l}(U_J, \mathcal{H})_p$. Using (3.10) we get

$$(3.20) \quad \sigma_{4l}(U_J, \mathcal{H})_p \leq \sum_{|I| \leq |J|2^{-l}} \|c_I(U_J)H_I\|_p + \sum_{|I| \geq |J|2^l} \|c_I(U_J)H_I\|_p \leq \\ \leq C(p) |J|^{1/p-1/2} 2^{-l(1/2-|1/p-1/2|)}.$$

Next, $t_0 = aU_J$ and by (3.17)

$$|a| \leq 2\|f\|_p |J|^{1/2-1/p}.$$

From here and (3.20) with $l := \lceil 2^{n/2-2} \rceil$ and $\theta(p) := (1/2 - |1/p - 1/2|)/4$ we get

$$(3.21) \quad \sigma_{m_0}(t_0, \mathcal{H})_p \leq C(p) 2^{-\theta(p)2^{n/2}}.$$

Combining (3.15), (3.19) and (3.21) we complete the proof of Theorem 3.2.

Theorem 3.3. For any $\alpha > 0$, $0 < \beta \leq \infty$ we have for $f \in L_p$

$$C_1(\alpha, \beta, p)|f; \alpha, \beta, \chi, L_p| \leq |f; \alpha, \beta, \mathcal{H}, L_p| \leq C_2(\alpha, \beta, p)|f; \alpha, \beta, \chi, L_p|.$$

Proof. The first inequality follows from Theorem 3.2 and Lemma 3.4 and the second inequality follows from (1.1).

Lemma 3.4. Let two sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_k\}_{k=1}^{\infty}$ of nonnegative numbers satisfy the inequalities

$$a_n \leq \sum_{k=1}^n b_k 2^{r(k-n)}, \quad n = 1, 2, \dots,$$

with some $r > 0$. Then for any $0 < \alpha < r$ and $0 < \beta \leq \infty$ we have

$$\left(\sum_{n=1}^{\infty} (2^{\alpha n} a_n)^{\beta}\right)^{1/\beta} \leq C(\alpha, \beta, r) \left(\sum_{k=1}^{\infty} (2^{\alpha k} b_k)^{\beta}\right)^{1/\beta}.$$

Proof. Consider first the case $0 < \beta \leq 1$. We have

$$a_n^{\beta} \leq \sum_{k=1}^n b_k^{\beta} 2^{\beta r(k-n)}$$

and

$$\begin{aligned} \sum_{n=1}^{\infty} (2^{\alpha n} a_n)^{\beta} &\leq \sum_{n=1}^{\infty} 2^{(\alpha-r)\beta n} \sum_{k=1}^n b_k^{\beta} 2^{\beta r k} = \\ &= \sum_{k=1}^{\infty} b_k^{\beta} 2^{\beta r k} \sum_{n=k}^{\infty} 2^{(\alpha-r)\beta n} \leq C(\alpha, r) \sum_{k=1}^{\infty} b_k^{\beta} 2^{\alpha \beta k}. \end{aligned}$$

Let us proceed now to the case $1 < \beta < \infty$. Take $\theta := (r - \alpha)/2$ and estimate using the Hölder inequality

$$a_n \leq C(\theta) \left(\sum_{k=1}^n (b_k 2^{(r-\theta)(k-n)})^{\beta}\right)^{1/\beta}$$

Similarly to the above we get the required inequality.

It remains to consider $\beta = \infty$. Let $b_k \leq 2^{-\alpha k}$, $k = 1, 2, \dots$, then

$$a_n \leq \sum_{k=1}^n 2^{-\alpha k + r(k-n)} \leq C(r - \alpha) 2^{-\alpha n}, \quad n = 1, 2, \dots$$

This completes the proof of Lemma 3.4 and Theorem 3.3.

Acknowledgement. The author thanks Professor Ronald DeVore for valuable discussions concerning the material in this paper.

REFERENCES

- [D] V.V. Dubinin, *Greedy Algorithms and Applications*, Dissertation USC, 1997.
- [DJP] R. DeVore, B. Jawerth, V. Popov, *Compression of wavelet decompositions*, American Journal of Math. **114** (1992), 737-785.
- [DKT] R.A. DeVore, S.V. Konyagin, and V.V. Temlyakov, *Hyperbolic Wavelet Approximation* (to appear).
- [DL] R. DeVore, G.Lorentz, *Constructive Approximation*, Springer Grundlehren, Heidelberg, 1993.
- [DT] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in Computational Mathematics **5** (1996), 173-187.
- [KS] B.S. Kashin and A.A. Saakyan, *Orthogonal Series*, American Math. Soc., Providence, R.I., 1989.
- [T1] V.N. Temlyakov, *Greedy Algorithm and m -term trigonometric approximation* (to appear).
- [T2] V. Temlyakov, *Approximation of functions with bounded mixed derivative*, Proc. Steklov Institute (1989, Issue 1).