GREEDY ALGORITHMS WITH REGARD TO MULTIVARIATE SYSTEMS WITH SPECIAL STRUCTURE[1]

V.N. Temlyakov

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

Abstract. The question of finding an optimal dictionary for nonlinear m-term approximation is studied in the paper. We consider this problem in the periodic multivariate (d variables) case for classes of functions with mixed smoothness. We prove that the well known dictionary U^d which consists of trigonometric polynomials (shifts of the Dirichlet kernels) is nearly optimal among orthonormal dictionaries. Next, it is established that for these classes near best m-term approximation with regard to U^d can be achieved by simple greedy type (thresholding type) algorithm.

The univariate dictionary U is used to construct a dictionary which is optimal among dictionaries with the tensor product structure.

1. Introduction

This paper is devoted to nonlinear approximation, namely, to m-term approximation. Nonlinear m-term approximation is important in applications in image and signal processing (see for instance the recent servey [D]). One of the major questions in approximation (theoretical and numerical) is: what is an optimal method? We discuss here this question in a theoretical setting with the only criterion of quality of approximating method its accuracy. One more important point in the setting of optimization problem is to specify a set of methods over which we are going to optimize. Most of the problems which approximation theory deals with are of this nature. Let us give some examples from classical approximation theory. These examples will help us to motivate the question we are studying in this paper.

Example 1. When we are searching for n-th best trigonometric approximation of a given function we are optimizing in the sense of accuracy over the subspace of trigonometric polynomials of degree n.

Example 2. When we are solving the problem on Kolmogorov's n-width for a given function class we are optimizing in the sense of accuracy for a given class over all subspaces of dimension n.

[1]This research was supported by the National Science Foundation Grant DMS 9622925 and by ONR Grant N0014-96-1-1003

Example 3. When we are finding best m-term approximation of a given function with regard to a given system of functions (dictionary) we are optimizing over all m-dimensional subspaces spanned by elements from a given dictionary.

Example 2 is a development of Example 1 in the sense that in Example 2 we are looking for an optimal n-dimensional subspace instead of being confined to a given one (trigonometric polynomials of degree n). Example 3 is a nonlinear analog of Example 1, where instead of trigonometric system we take a dictionary $\mathcal{D}$ and allow approximating elements from $\mathcal{D}$ to depend on a function. In this paper we make some steps in a direction of developing Example 3 to a setting which is a nonlinear analog of Example 2. In other words, we want to optimize over some sets of dictionaries. We discuss in this paper two classical structural properties of dictionaries:

1. Orthogonality;
2. Tensor product structure (multivariate case).

Denote by $\mathcal{D}$ a dictionary in a Banach space X and by

$$\sigma_m(f,\mathcal{D})_X := \inf_{g_i\in\mathcal{D},c_i,i=1,\dots,m} \|f - \sum_{i=1}^m c_i g_i\|_X$$

best m-term approximation of f with regard to $\mathcal{D}$. For a function class $F \subset X$ and a collection $\mathbb{D}$ of dictionaries we consider

$$\sigma_m(F,\mathcal{D})_X := \sup_{f\in F} \sigma_m(f,\mathcal{D})_X \quad ,$$

$$\sigma_m(F,\mathbb{D})_X := \inf_{\mathcal{D}\in\mathbb{D}} \sigma_m(F,\mathcal{D})_X \quad .$$

Thus the quantity $\sigma_m(F,\mathbb{D})_X$ gives the sharp lower bound for best m-term approximation of a given function class F with regard to any dictionary $\mathcal{D} \in \mathbb{D}$.

Denote by $\mathbb{O}$ the set of all orthonormal dictionaries defined on a given domain. B.S. Kashin [K] proved that for the class $H^{r,\alpha}$, $r = 0, 1, \dots, \quad \alpha \in [0, 1]$, of univariate functions such that

$$\|f\|_\infty + \|f^{(r)}\|_\infty \le 1 \quad \text{and} \quad |f^{(r)}(x) - f^{(r)}(y)| \le |x-y|^\alpha, \quad x, y \in [0,1]$$

we have

$$\sigma_m(H^{r,\alpha},\mathbb{O})_{L_2} \ge C(r,\alpha)m^{-r-\alpha}. \tag{1.1}$$

It is interesting to remark that we cannot prove anything like (1.1) with L_2 replaced by L_p, $p < 2$. We proved (see [KT]) that there exists $\Phi \in \mathbb{O}$ such that for any $f \in L_1(0,1)$ we have $\sigma_1(f,\Phi)_{L_1} = 0$. The proof from [KT] also works for L_p, $p < 2$, instead of L_1.

Remark 1.1. *For any $1 \le p < 2$ there exists a complete in $L_2(0,1)$ orthonormal system Φ such that for each $f \in L_p(0,1)$ we have $\sigma_1(f,\Phi)_{L_p} = 0$.*

This remark means that to obtain nontrivial lower bounds for $\sigma_m(f,\Phi)_{L_p}$, $p < 2$, we need to impose additional restrictions on $\Phi \in \mathbb{O}$. One way of imposing restrictions was discussed in [KT], and we present another way in Section 4.

We discuss in this paper approximation of multivariate functions. It is convenient for us to present results in the periodic case. In this paper we consider classes of functions with bounded mixed derivative MW^r_q (see the definition in Section 3) and classes with restriction of Lipschitz type on mixed difference MH^r_q (see the definition in Section 2). These classes are well known (see for instance [T2]) for their importance in numerical integration, in finding universal methods for approximation of functions of several variables, in the average case setting of approximation problems for the spaces equipped with the Wiener sheet measure (see [W]) and in other problems. In Section 4 we prove

$$\sigma_m(MH^r_q,\mathbb{O})_{L_2} \gg m^{-r}(\log m)^{(d-1)(r+1/2)}, \quad 1\le q<\infty, \tag{1.2}$$

$$\sigma_m(MW^r_q,\mathbb{O})_{L_2} \gg m^{-r}(\log m)^{(d-1)r}, \quad 1\le q<\infty. \tag{1.3}$$

In Sections 2 and 3 we prove that the orthogonal basis U^d which we construct at the end of this section provides optimal upper estimates (like (1.2) and (1.3)) in best m-term approximation of the classes MH^r_q and MW^r_q in the L_p-norm, $2\le p<\infty$. Moreover, we prove there that for all $1<q,p<\infty$ the order of best m-term approximation $\sigma_m(MH^r_q,U^d)_{L_p}$ and $\sigma_m(MW^r_q,U^d)_{L_p}$ can be achieved by a greedy type algorithm $G^p(\cdot,U^d)$. Assume a given system Ψ of functions ψ_I indexed by dyadic intervals can be enumerated in such a way that $\{\psi_{I^j}\}_{j=1}^\infty$ is a basis for L_p. Then we define the greedy algorithm $G^p(\cdot,\Psi)$ as follows. Let

$$f=\sum_{j=1}^{\infty} c_{I^j}(f,\Psi)\psi_{I^j}$$

and

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

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

$$\min_{I\in\Lambda_m} c_I(f,p,\Psi) \ge \max_{J\notin\Lambda_m} c_J(f,p,\Psi). \tag{1.4}$$

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

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

Remark 1.2. *Let $\Phi=\{\phi_k\}_{k=1}^\infty$ be a basis for a Banach space X and $\|\phi_k\|_X = 1,\quad k=1,2,\dots$. Assume that we can calculate the X-norm of a function $f\in X$ and each $c_k(f)$ from the expansion*

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

in a finite number of steps. Then there is an algorithm which for any $f \in X$ gives the biggest $|c_k(f)|$ after a finite number of steps.

Proof. We have for any $f \in X$

$$|c_k(f)| \le B\|f\|_X, \quad k = 1, 2, \dots,$$

with a constant B and

$$\lim_{n\to\infty} \|\sum_{k=n+1}^{\infty} c_k(f)\phi_k\|_X = 0.$$

Let $f \neq 0$. We find a nonzero coefficient $c_l(f)$ and denote $\epsilon := |c_l(f)|/B$. Next, we find n such that

$$\|\sum_{k=n+1}^{\infty} c_k(f)\phi_k\|_X < \epsilon.$$

This implies that for all $k > n$ we have $|c_k(f)| < |c_l(f)|$ and, therefore, we can restrict our search for the largest $|c_k(f)|$ to $1 \le k \le n$.

The question of constructing an algorithm which realizes (in the sense of order) the best possible accuracy is a very important one and we discuss it in detail in this paper. Let $A_m(\cdot, \mathcal{D})$ be a mapping which maps each $f \in X$ to a linear combination of m elements from a given dictionary $\mathcal{D}$. Then the best we can hope for with this mapping is to have for each $f \in X$

$$\|f - A_m(f, \mathcal{D})\|_X = \sigma_m(f, \mathcal{D})_X \tag{1.5}$$

or a little weaker

$$\|f - A_m(f, \mathcal{D})\|_X \le C(\mathcal{D}, X)\sigma_m(f, \mathcal{D})_X. \tag{1.6}$$

There are some known trivial and nontrivial examples when (1.5) holds in a Hilbert space X. We do not touch this kind of relations in this paper. Concerning (1.6) it is proved in [T3] that for any basis Ψ which is L_p-equivalent to the univariate Haar basis we have

$$\|f - G^p_m(f, \Psi)\|_{L_p} \le C(p)\sigma_m(f, \Psi)_p, \quad 1 < p < \infty. \tag{1.7}$$

However, as it is shown in [T4] and in Section 5 of this paper the inequality (1.7) does not hold for particular dictionaries with tensor product structure. We have for instance (see Section 5)

$$\sup_{f\in L_p} \|f - G^p_m(f, U^d)\|_{L_p}/\sigma_m(f, U^d)_{L_p} \gg (\log m)^{(d-1)|1/2-1/p|}. \tag{1.8}$$

The inequality (1.8) shows that using the algorithm $G^p(\cdot, U^d)$ we lose for sure for some functions $f \in L_p$, $p \neq 2$. In light of (1.8) the results of Sections 2 and 3 look encouraging for using $G^p(\cdot, U^d)$: we have for $1 < q, p < \infty$ and big enough r

$$\sup_{f\in MH^r_q} \|f - G^p_m(f, U^d)\|_p \asymp \sigma_m(MH^r_q, U^d)_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}, \tag{1.9}$$

$$\sup_{f\in MW^r_q} \|f-G^p_m(f,U^d)\|_p \asymp \sigma_m(MW^r_q,U^d)_p \asymp m^{-r}(\log m)^{(d-1)r}, \tag{1.10}$$

where we use the abreviated notation $\|\cdot\|_p := \|\cdot\|_{L_p}$.

Comparing (1.9) with (1.2) and (1.10) with (1.3), we conclude that the dictionary U^d is the best (in the sense of order) among all orthogonal dictionaries for m-term approximation of the classes MH^r_q and MW^r_q in L_p where $1 < q < \infty$ and $2 \le p < \infty$. The dictionary U^d has one more important feature. The near best m-term approximation of functions from MH^r_q and MW^r_q in the L_p-norm can be realized by the simple greedy type algorithm $G^p(\cdot,U^d)$ for all $1 < q,p < \infty$.

Let us now compare the performance of U^d with the performance of the best dictionary with tensor product structure. Denote by Π^d the set of all functions of the form $u_1(x_1)\dots u_d(x_d)$, where $u_j \in L_p$, $j = 1,\dots,d$. Then it is clear that for any dictionary $\mathcal{D}$, with tensor product structure, we have $\mathcal{D} \subset \Pi^d$ and

$$\sigma_m(f,\mathcal{D})_p \ge \sigma_m(f,\Pi^d)_p.$$

The problem of estimating $\sigma_m(f,\Pi^2)_2$ (best m-term bilinear approximation in L_2) is a classical one and was considered for the first time by E. Schmidt [S] in 1907. For many function classes F an asymptotic behavior of $\sigma_m(F,\Pi^2)_p$ is known. For instance, the relation

$$\sigma_m(MW^r_q,\Pi^2)_p \asymp \sigma_m(MH^r_q,\Pi^2)_p \asymp m^{-2r+(1/q-\max(1/2,1/p))_+} \tag{1.11}$$

for $r > 1$ and $1 \le q,p \le \infty$ follows from more general results in [T5]. In the case $d > 2$ almost nothing is known. There is (see [T6]) an upper estimate in the case $q = p = 2$

$$\sigma_m(MW^r_2,\Pi^d)_2 \ll m^{-dr/(d-1)}. \tag{1.12}$$

Comparing (1.9), (1.10) with (1.11),(1.12) we conclude that m-term approximation with regard to U^d does not provide an optimal rate of approximation among dictionaries with tensor product structure. This observation motivated us to study m-term approximation with regard to the following dictionary

$$Y := (U \times L_p) \cup (L_p \times U) = \{y(x_1,x_2)\}$$

with $y(x_1,x_2)$ of the form $y(x_1,x_2) = U_I(x_1)v(x_2),\quad U_I \in U, v \in L_p$, or $y(x_1,x_2) = v(x_1)U_I(x_2),\quad v \in L_p, U_I \in U$. We prove in Section 6 that we have for $r > (1/q-1/p)_+$

$$\sigma_m(MH^r_q,Y)_p \asymp \sigma_m(MW^r_q,Y)_p \asymp m^{-2r+(1/q-1/p)_+},\quad 1 < q,p < \infty. \tag{1.13}$$

Comparing (1.13) with (1.11) we realize that for $1 < q \le p \le 2$ and $1 < p \le q < \infty$ the dictionary Y which is much smaller than Π^2 provides optimal m-term bilinear approximation for the classes MH^r_q and MW^r_q. We also make the following important point. The error of approximation in (1.13) can be achieved by combination of a linear method and greedy algorithm $G^p(\cdot,U^2)$.

We define at the end of this section a system of orthogonal trigonometric polynomials which is optimal in a certain sense (see above) for m-term approximations. Variants of this system are well-known and very useful in interpolation of functions by trigonometric polynomials. We define first the system $U := \{U_I\}$ in the univariate case. Denote

$$U_n^+(x) := \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x}-1}{e^{ix}-1}, \quad n = 0, 1, 2, \dots;$$

$$U_{n,k}^+(x) := e^{i2^n x} U_n^+(x - 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1;$$

$$U_{n,k}^-(x) := e^{-i2^n x} U_n^+(-x + 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1.$$

It will be more convenient for us to normalize in L_2 the system of functions $\{U_{m,k}^+, U_{n,k}^-\}$ and enumerate it by dyadic intervals. We write

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

and

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

Denote

$$D_n^+ := \{I : I = [(k+1/2)2^{-n}, (k+1)2^{-n}), \quad k = 0, 1, \dots, 2^n - 1\}$$

and

$$D_n^- := \{I : I = [k2^{-n}, (k+1/2)2^{-n}), \quad k = 0, 1, \dots, 2^n - 1\}$$

$$D_0^+ = D_0^- = D_0 := [0, 1), \quad D := \cup_{n \ge 1}(D_n^+ \cup D_n^-) \cup D_0.$$

It is easy to check that for any $I, J \in D$, $I \neq J$ we have

$$\langle U_I, U_J \rangle = (2\pi)^{-1} \int_0^{2\pi} U_I(x) \bar{U}_J(x) dx = 0,$$

and

$$\|U_I\|_2^2 = 1.$$

We use the notations for $f \in L_1$

$$f_I := \langle f, U_I \rangle = (2\pi)^{-1} \int_0^{2\pi} f(x) \bar{U}_I(x) dx; \qquad \hat{f}(k) := (2\pi)^{-1} \int_0^{2\pi} f(x) e^{-ikx} dx$$

and

$$\delta_s^+(f) := \sum_{k=2^s}^{2^{s+1}-1} \hat{f}(k) e^{ikx}; \quad \delta_s^-(f) := \sum_{k=-2^{s+1}+1}^{-2^s} \hat{f}(k) e^{ikx}; \quad \delta_0(f) := \hat{f}(0).$$

Then for each s and $f \in L_1$ we have

$$\delta_s^+(f) = \sum_{I \in D_s^+} f_I U_I; \quad \delta_s^-(f) = \sum_{I \in D_s^-} f_I U_I; \quad \delta_0(f) = f_{[0,1)}.$$

Moreover, the following important for us analog of Marcinkiewicz theorem holds

$$\|\delta_s^+(f)\|_p^p \asymp \sum_{I \in D_s^+} \|f_I U_I\|_p^p; \quad \|\delta_s^-(f)\|_p^p \asymp \sum_{I \in D_s^-} \|f_I U_I\|_p^p \tag{1.14}$$

for $1 < p < \infty$ with constants depending only on p.

We remark that

$$\|U_I\|_p \asymp |I|^{1/p-1/2}, \quad 1 < p \leq \infty, \tag{1.15}$$

which implies for any $1 < q, p < \infty$

$$\|U_I\|_p \asymp \|U_I\|_q |I|^{1/p-1/q}. \tag{1.16}$$

In the multivariate case of $x = (x_1, \dots, x_d)$ we define the system U^d as the tensor product of the univariate systems U. Let $I = I_1 \times \cdots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

$$U_I(x) := \prod_{j=1}^{d} U_{I_j}(x_j).$$

For $s = (s_1, \dots, s_d)$ and $\epsilon = (\epsilon_1, \dots, \epsilon_d)$, $\epsilon_j = +$ or $-$, denote

$$D_s^\epsilon := \{I : I = I_1 \times \cdots \times I_d, \quad I_j \in D_{s_j}^{\epsilon_j}, \quad j = 1, \dots, d\}.$$

It is easy to see that (1.15) and (1.16) are also true in the multivariate case. It is not difficult to derive from (1.14) that for any ϵ we have

$$\|\delta_s^\epsilon(f)\|_p^p \asymp \sum_{I \in D_s^\epsilon} \|f_I U_I\|_p^p, \quad 1 < p < \infty, \tag{1.17}$$

with constants depending on p and d. Here we denote

$$\delta_s^\epsilon(f) := \sum_{k \in \rho(s,\epsilon)} \hat{f}(k) e^{i(k,x)}$$

where

$$\rho(s, \epsilon) := \epsilon_1[2^{s_1}, 2^{s_1+1} - 1) \times \cdots \times \epsilon_d[2^{s_d}, 2^{s_d+1} - 1).$$

We will often use the following inequalities

$$\Big(\sum_{s,\epsilon} \|\delta_s^\epsilon(f)\|_p^p\Big)^{1/p} \ll \|f\|_p \ll \Big(\sum_{s,\epsilon} \|\delta_s^\epsilon(f)\|_p^2\Big)^{1/2}, \quad 2 \leq p < \infty, \tag{1.18}$$

$$(\sum_{s,\epsilon}\|\delta_s^\epsilon(f)\|_p^2)^{1/2} \ll \|f\|_p \ll (\sum_{s,\epsilon}\|\delta_s^\epsilon(f)\|_p^p)^{1/p}, 1<p\le 2,$$

which are corollaries of the well-known Littlewood-Paley inequalities

$$\|f\|_p \asymp \|(\sum_s |\sum_\epsilon \delta_s^\epsilon(f)|^2)^{1/2}\|_p. \tag{1.20}$$

We note that the system U_I^d can be enumerated in such a way that $\{U_{I^l}\}_{l=1}^\infty$ forms a basis for each L_p, $1<p<\infty$. Indeed, let us first enumerate vectors $s=(s_1,\dots,s_d)$ with integer nonnegative components in such a way that for all $j=1,2,\dots$ we have $\|s^j\|_\infty \le \|s^{j+1}\|_\infty$. Then we enumerate all dyadic intervals in D following the rule: we proceed to enumerate the intervals from $D^\epsilon_{s^{j+1}}$ after enumerating all intervals from $D^\epsilon_{s^j}$ for all ϵ. Any partial sum with regard to $\{U_{I^l}\}_{l=1}^\infty$ can be represented in the form

$$\sum_{j=1}^{n-1}\sum_\epsilon \delta^\epsilon_{s^j}(f) + \sum_{I\in\Lambda_n} f_I U_I =: f^1+f^2,$$

where $\Lambda_n \subset \cup_\epsilon D^\epsilon_{s^n}$. Then we get from (1.20)

$$\|f^1\|_p \ll \|f\|_p. \tag{1.21}$$

In order to prove the estimate

$$\|f^2\|_p \ll \|f\|_p \tag{1.22}$$

we use the following inequalities

$$\|\delta_s^\epsilon(f)\|_p \ll \|\sum_\epsilon \delta_s^\epsilon(f)\|_p \ll \|f\|_p$$

and the relation (1.17). Thus, the norms of the operators of taking partial sums with regard to $\{U_{I^l}\}_{l=1}^\infty$ are uniformly bounded. This implies that $\{U_{I^l}\}_{l=1}^\infty$ is a basis for L_p, $1<p<\infty$. We remark that P. Wojtaszczyk [Wo] proved recently that the system U is equivalent to the Haar system in all L_p, $1<p<\infty$, and, therefore, is an unconditional basis for all L_p, $1<p<\infty$.

In Sections 2 and 3 we study efficiency of greedy algorithms with regard to U^d on the classes of functions with bounded mixed derivative or difference.

2. The upper estimates for the classes MH_q^r

In this section we study the classes MH_q^r. We define these classes as follows (see for instance [T2], p. 196). Let $\Delta_u^l(j)$ denote the operator of l-th difference with step u in the variable x_j. For a nonempty set e of natural numbers from $[1,d]$ and a vector $t=(t_1,\dots,t_d)$ we denote

$$\Delta_t^l(e) := \prod_{j\in e}\Delta^l_{t_j}(j).$$

We define the class MH_q^r as the set of $f\in L_q$ such that $\|f\|_q\le 1$ and for any nonempty set e we have

$$\|\Delta_t^l(e)f(x)\|_q \le \prod_{j\in e}|t_j|^r$$

with $l=[r]+1$, where $[a]$ denotes the integral part of a. We prove first two auxiliary results.

Lemma 2.1. *For a fixed real number a denote*

$$h_n(s) := 2^{-n(r+1/2)+a(\|s\|_1-n)}$$

and for $f \in MH_q^r$ consider the sets

$$A(f,n,a) := \{I : |f_I| \geq h_n(s), \quad \text{if} \quad I \in D_s^\epsilon\}, \quad n = 1, 2, \dots.$$

Then if $r > 1/q - 1/2 - a$ we have

$$\#A(f,n,a) \ll 2^n n^{d-1}$$

with a constant independent of n and f.

Proof. It is known (see [T1], p.33 and [T2], p.197) that for $f \in MH_q^r$ we have for all ϵ

$$\|\delta_s^\epsilon(f)\|_q \ll 2^{-r\|s\|_1}. \tag{2.1}$$

For convenience we will omit ϵ in the notations $\delta_s^\epsilon(f)$, D_s^ϵ, N_s^ϵ (see below) meaning that we are estimating a quantity $\delta_s^\epsilon(f)$ or N_s^ϵ for a fixed ϵ and all estimates we are going to do in this paper are the same for all ϵ.

Using the following two properties of the system $\{U_I\}$

$$\|\delta_s(f)\|_q^q \asymp \sum_{I \in D_s} \|f_I U_I\|_q^q, \tag{2.2}$$

$$\|U_I\|_q \asymp 2^{\|s\|_1(1/2-1/q)}, \quad I \in D_s, \tag{2.3}$$

we get from (2.1)

$$\sum_{I \in D_s} |f_I|^q \ll 2^{-\|s\|_1(rq+q/2-1)}. \tag{2.4}$$

Denote $N_s^\epsilon := \#(A(f,n,a) \cap D_s^\epsilon)$. Then (2.4) implies

$$N_s h_n(s)^q \ll 2^{-\|s\|_1(rq+q/2-1)}$$

and

$$N_s \ll 2^{n(r+1/2+a)q} 2^{-\|s\|_1(rq+q/2-1+aq)}.$$

Using the assumption $r > 1/q - 1/2 - a$ we get

$$\sum_{\|s\|_1 \geq n} N_s \ll 2^n n^{d-1}$$

and

$$\sum_\epsilon \sum_{\|s\|_1 \geq n} N_s^\epsilon \ll 2^n n^{d-1}. \tag{2.5}$$

It remains to remark that for $\|s\|_1 < n$ we have the following trivial estimates

$$\sum_\epsilon \sum_{\|s\|_1 < n} N_s^\epsilon \leq \sum_\epsilon \sum_{\|s\|_1 < n} \#D_s^\epsilon \ll 2^n n^{d-1}. \tag{2.6}$$

Combining (2.5) and (2.6) we complete the proof.

Lemma 2.2. *Let $h_n(s)$ and $A(f,n,a)$ be from Lemma 2.1 and $a > -1/2$. For each n denote*

$$g_n(f) := \sum_{I \in A(f,n,a)} f_I U_I, \qquad f^n := f - g_n(f).$$

Then for any $f \in MH^r_q$ and $p \geq 2$ satisfying $1 < q \leq p < \infty$ we have for $r > (a+1/2)(p/q-1)$

$$\|f^n\|_p \ll 2^{-rn} n^{(d-1)/2}$$

with a constant independent of n and f.

Proof. For $2 \leq p < \infty$ we have by a corollary to the Littlewood-Paley inequalities

$$\|f^n\|_p^2 \ll \sum_\epsilon (\sum_s \|\delta_s^\epsilon(f^n)\|_p^2) = \sum_\epsilon (\sum_{\|s\|_1 < n} \|\delta_s^\epsilon(f^n)\|_p^2 +$$

$$\sum_{\|s\|_1 \geq n} \|\delta_s^\epsilon(f^n)\|_p^2) =: \sum_\epsilon (\Sigma' + \Sigma'').$$

We estimate first Σ'. By the definition of $A(f,n,a)$ we have for all I

$$|f_I^n| < h_n(s), \quad I \in D_s. \tag{2.7}$$

Therefore,

$$\|\delta_s(f^n)\|_p^p \ll h_n(s)^p \sum_{I \in D_s} \|U_I\|_p^p \ll 2^{-n(r+1/2+a)p} 2^{\|s\|_1(a+1/2)p}$$

and

$$\sum_{\|s\|_1 < n} \|\delta_s(f^n)\|_p^2 \ll 2^{-2rn} n^{d-1}. \tag{2.8}$$

We proceed to estimating Σ'' now. We have

$$\|\delta_s(f^n)\|_p^p \ll \sum_{I \in D_s} \|f_I^n U_I\|_p^p \ll (h_n(s) 2^{\|s\|_1(1/2-1/p)})^{p-q} \sum_{I \in D_s} \|f_I^n U_I\|_p^q \ll \tag{2.9}$$

$$(h_n(s) 2^{\|s\|_1(1/2-1/p)})^{p-q} \sum_{I \in D_s} \|f_I^n U_I\|_q^q 2^{\|s\|_1(1/q-1/p)q}.$$

Using (2.1) we get

$$\sum_{I \in D_s} \|f_I^n U_I\|_q^q \leq \sum_{I \in D_s} \|f_I U_I\|_q^q \ll \|\delta_s(f)\|_q^q \ll 2^{-r\|s\|_1 q} \tag{2.10}$$

and from (2.9)

$$\|\delta_s(f^n)\|_p^p \ll 2^{-n(r+1/2+a)(p-q)} 2^{\|s\|_1(-rq+(a+1/2)(p-q))}. \tag{2.11}$$

Using the assumption $r > (a+1/2)(p/q-1)$ we get

$$\Sigma'' \ll 2^{-2rn} n^{d-1}. \tag{2.12}$$

Combining (2.8) and (2.12) we complete the proof of Lemma 2.2.

It is clear from the proof of Lemma 2.2 that the following statement holds.

Lemma 2.2′. *Let $h_n(s)$ be from Lemma 2.1 and $a > -1/2$. Assume that a function f satisfies the restrictions*

$$\|\delta_s^\epsilon(f)\|_q \ll 2^{-r\|s\|_1}, \quad 1 < q < \infty,$$

$$|f_I| \ll h_n(s), \quad I \in D_s^\epsilon,$$

with constants independent of f, n, and s. Then for $\max(2,q) \leq p < \infty$ and $r > (a+1/2)(p/q-1)$ we have

$$\|f\|_p \ll 2^{-rn} n^{(d-1)/2}$$

with a constant independent of n and f.

Consider the following greedy type algorithm $G^{c,a}$. Take a real number a and rearrange the sequence $|f_I||I|^a$ in the decreasing order

$$|f_{I^1}||I^1|^a \geq |f_{I^2}||I^2|^a \geq \dots \quad .$$

Define

$$G_m^{c,a}(f,U^d) := \sum_{k=1}^m f_{I^k} U_{I^k}.$$

Theorem 2.1. *Let $1 < q < \infty$ and $\max(2,q) \leq p < \infty$. Then for any $a > -1/2$ and $r > \max\{(a+1/2)(p/q-1), 1/q - 1/2 - a\}$ we have*

$$\sup_{f \in MH_q^r} \|f - G_m^{c,a}(f,U^d)\|_p \asymp \sigma_m(MH_q^r, U^d)_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

Proof. Let m be given. Denote by $n(m)$ the biggest n satisfying

$$\sup_{f \in MH_q^r} \#A(f,n,a) \leq m.$$

Lemma 2.1 implies

$$2^{n(m)} \gg m(\log m)^{1-d},$$

and for

$$g := f - G_m^{c,a}(f,U^d) \quad \text{we have} \quad |g_I| \leq h_{n(m)}(s), \quad I \in D_s.$$

Similarly to (2.10) it is easy to check that

$$\|\delta_s(g)\|_q \ll 2^{-r\|s\|_1}$$

with a constant independent of s and g. Applying Lemma 2.2′ to g we get

$$\|g\|_p \ll 2^{-rn(m)} n(m)^{(d-1)/2} \ll m^{-r}(\log m)^{(d-1)(r+1/2)},$$

what proves the upper estimate in Theorem 2.1

$$\sup_{f \in MH_q^r} \|f - G_m^{c,a}(f,U^d)\|_p \ll m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

The lower estimate

$$\sigma_m(MH^r_q, U^d)_p \gg m^{-r}(\log m)^{(d-1)(r+1/2)}$$

follows from Theorem 4.2 in Section 4.

The proof of Theorem 2.1 is complete.

Consider now the L_b-greedy algorithm $G^b(\cdot, U^d)$. Take a number $1 \le b \le \infty$ and rearrange the sequence $\{\|f_I U_I\|_b\}$ in the decreasing order

$$\|f_{I_1} U_{I_1}\|_b \ge \|f_{I_2} U_{I_2}\|_b \dots \quad .$$

Define

$$G^b_m(f, U^d) := \sum_{k=1}^{m} f_{I_k} U_{I_k}.$$

It is clear from the relation

$$\|f_I U_I\|_b \asymp |f_I||I|^{1/b-1/2}$$

that the algorithms G^b and $G^{c,a}$ with $a = 1/b - 1/2$ are closely connected. The following proposition can be proved similarly to Theorem 2.1.

Theorem 2.2. *Let $1 < q < \infty$ and $\max(2, q) \le p < \infty$. Then for any $1 < b < \infty$ and $r > \max\{(p/q-1)/b, 1/q - 1/b\}$ we have*

$$\sup_{f \in MH^r_q} \|f - G^b_m(f, U^d)\|_p \asymp \sigma_m(MH^r_q, U^d)_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

We formulate now the corollary of Theorem 2.2 in the most interesting case $b = p$.

Theorem 2.3. *Let $1 < q, p < \infty$. Then for all $r > r(q, p)$ we have*

$$\sup_{f \in MH^r_q} \|f - G^p_m(f, U^d)\|_p \asymp \sigma_m(MH^r_q, U^d)_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}$$

with

$$r(q, p) := \begin{cases} (1/q - 1/p)_+, & \text{for } p \ge 2 \\ (\max(2/q, 2/p) - 1)/p, & \text{otherwise.} \end{cases}$$

Proof. The lower estimates follow from Theorem 4.2 in Section 4. We prove the upper estimates. Consider first the case $2 \le p < \infty$. If $1 < q \le p$ we use Theorem 2.2 with $b = p$ and get a restriction $r > 1/q - 1/p$. If $p < q < \infty$ we use the inequality

$$\sup_{f \in MH^r_q} \|f - G^p_m(f, U^d)\|_p \le \sup_{f \in MH^r_p} \|f - G^p_m(f, U^d)\|_p \tag{2.13}$$

and reduce this case to the case $q = p$ which has already been considered above. It remains to consider the case $1 < p < 2$. If $1 < q \le p$ we use Theorem 2.2 with $p = 2$ and $b = p$ and get

$$\sup_{f \in MH^r_q} \|f - G^p_m(f, U^d)\|_p \le \sup_{f \in MH^r_q} \|f - G^p_m(f, U^d)\|_2 \ll m^{-r}(\log m)^{(d-1)(r+1/2)}$$

provided $r > (2/q - 1)/p$. If $p < q < \infty$ we use the inequality (2.13) to reduce this case to the case $q = p$. In this case we get a restriction $r > (2/p - 1)/p$.

Theorem 2.3 is proved now.

3. The upper estimates for the classes MW^r_q

In this section we study the classes MW^r_q which we define for positive r (not necessarily an integer). Let

$$F_r(u) := 1 + 2\sum_{k=1}^{\infty} k^{-r}\cos(ku - \pi r/2)$$

be the univariate Bernoulli kernel and let

$$F_r(x) := F_r(x_1,\dots,x_d) := \prod_{j=1}^{d} F_r(x_j)$$

be its multivariate analog. We define

$$MW^r_q := \{f : f = F_r * \phi, \quad \|\phi\|_q \le 1\},$$

where $*$ denotes the convolution.

Results and their proofs in this section are similar to those from the previous section. The technique in this section is a little more involved. We start with two lemmas.

Lemma 3.1. *For a fixed real number a denote*

$$w_n(s) := 2^{-n(r+1/2)+a(\|s\|_1-n)} n^{-(d-1)/2}$$

and for $f \in MW^r_q$ consider the sets

$$W(f,n,a) := \{I : |f_I| \ge w_n(s), \quad \text{if} \quad I \in D^\epsilon_s\}, \quad n = 1,2,\dots.$$

Then for $1 < q \le 2$ and $r > 1/q - 1/2 - a$ we have

$$\#W(f,n,a) \ll 2^n n^{d-1}$$

with a constant independent on n and f.

Proof. It is known ([T1], p.36 and [T2], p.242) that for $f \in MW^r_q$ we have

$$\|\sum_{\|s\|_1=l} \delta_s(f)\|_q \ll 2^{-rl}. \tag{3.1}$$

Further, for $1 < q \le 2$ we have as a corollary of the Littlewood-Paley inequalities

$$\|\sum_{\|s\|_1=l} \delta_s(f)\|_q \gg (\sum_{\|s\|_1=l} \|\delta_s(f)\|_q^2)^{1/2}. \tag{3.2}$$

Similarly to the proof of Lemma 2.1 we get for $N^\epsilon_s := \#(W(f,n,a) \cap D^\epsilon_s)$

$$N_s w_n(s)^q \ll \|\delta_s(f)\|_q^q 2^{-\|s\|_1(q/2-1)}. \tag{3.3}$$

Using (3.1) and (3.2) we obtain

$$\sum_{\|s\|_1=l} N_s \ll 2^{n(r+1/2+a)q} n^{(d-1)q/2} 2^{-l(q/2-1+aq)} \sum_{\|s\|_1=l} \|\delta_s(f)\|_q^q \ll$$

$$2^{n(r+1/2+a)q} n^{(d-1)q/2} 2^{-l(q/2-1+aq)} l^{(d-1)(1-q/2)} (\sum_{\|s\|_1=l} \|\delta_s(f)\|_q^2)^{q/2} \ll$$

$$2^{n(r+1/2+a)q} n^{(d-1)q/2} 2^{-l(q/2-1+aq+rq)} l^{(d-1)(1-q/2)}.$$

Using the asumption $r > 1/q - 1/2 - a$ we get from here

$$\sum_{l \geq n} \sum_{\|s\|_1=l} N_s \ll 2^n n^{d-1}. \tag{3.4}$$

For N_s with $\|s\|_1 \leq n$ we have

$$\sum_{\|s\|_1<n} N_s \leq \sum_{\|s\|_1<n} \#D_s \ll 2^n n^{d-1}. \tag{3.5}$$

Combining (3.4) and (3.5) and summating over ϵ we complete the proof of Lemma 3.1.

Lemma 3.2. *Let $w_n(s)$ be from Lemma 3.1 and $a > -1/2$. Assume that a function f satisfies the restrictions*

$$(\sum_{\|s\|_1=l} \|\delta_s^\epsilon(f)\|_q^2)^{1/2} \ll 2^{-rl}, \quad 1 < q < \infty,$$

$$|f_I| \ll w_n(s), \quad I \in D_s^\epsilon,$$

with constants independent of f, n, and s. Then for $\max(2,q) \leq p < \infty$ and $r > (a+1/2)(p/q-1)$ we have

$$\|f\|_p \ll 2^{-rn}$$

with a constant independent of n and f.

Proof. By a corollary to the Littlewood-Paley inequalities we have for $p \geq 2$

$$\|f\|_p^2 \ll \sum_\epsilon \sum_s \|\delta_s^\epsilon(f)\|_p^2 =$$

$$\sum_\epsilon (\sum_{\|s\|_1<n} \|\delta_s^\epsilon(f)\|_p^2 + \sum_{\|s\|_1 \geq n} \|\delta_s^\epsilon(f)\|_p^2) =: \sum_\epsilon (\Sigma' + \Sigma'').$$

Similarly to the corresponding part (see (2.8)) of the proof of Lemma 2.1 we obtain

$$\Sigma' \ll 2^{-2rn}. \tag{3.6}$$

Analogously to (2.9) and (2.11) we get

$$\|\delta_s(f)\|_p^p \ll \gamma_n^{p-q} 2^{\|s\|_1(a+1/2)(p-q)} \|\delta_s(f)\|_q^q, \tag{3.7}$$

where we use the notation

$$\gamma_n := 2^{-n(r+1/2+a)} n^{-(d-1)/2}.$$

Next,

$$\sum_{\|s\|_1=l} \|\delta_s(f)\|_p^2 \ll \gamma_n^{2(p-q)/p} 2^{2l(a+1/2)(p-q)/p} \sum_{\|s\|_1=l} \|\delta_s(f)\|_q^{2q/p} \le$$

$$\gamma_n^{2(p-q)/p} 2^{2l(a+1/2)(p-q)/p} l^{(d-1)(1-q/p)} \Big(\sum_{\|s\|_1=l} \|\delta_s(f)\|_q^2 \Big)^{q/p} \ll$$

$$\gamma_n^{2(p-q)/p} 2^{2l(-rq+(a+1/2)(p-q))/p} l^{(d-1)(1-q/p)}.$$

Using the assumption $r > (a+1/2)(p/q-1)$ we get from here

$$\Sigma'' \ll 2^{-2rn}. \tag{3.8}$$

Combining (3.6) and (3.8) we complete the proof.

Using Lemmas 3.1 and 3.2 instead of Lemmas 2.1 and 2.2$'$ we prove in the same way as in Section 2 the following analogs of Theorems 2.1 and 2.2. We note that the lower estimates follow from Theorem 4.1 in Section 4.

Theorem 3.1. *Let $1 < q \le 2 \le p < \infty$. Then for any $a > -1/2$ and $r > \max\{(a+1/2)(p/q-1), 1/q - 1/2 - a\}$ we have*

$$\sup_{f \in MW_q^r} \|f - G_m^{c,a}(f, U^d)\|_p \asymp \sigma_m(MW_q^r, U^d)_p \asymp m^{-r} (\log m)^{(d-1)r}.$$

Theorem 3.2. *Let $1 < q \le 2 \le p < \infty$. Then for any $1 < b < \infty$ and $r > \max\{(p/q-1)/b, 1/q - 1/b\}$ we have*

$$\sup_{f \in MW_q^r} \|f - G_m^{b}(f, U^d)\|_p \asymp \sigma_m(MW_q^r, U^d)_p \asymp m^{-r} (\log m)^{(d-1)r}.$$

We derive now one more theorem from Theorem 3.2.

Theorem 3.3. *Let $1 < q, p < \infty$. Then for all $r > r'(q,p)$ we have*

$$\sup_{f \in MW_q^r} \|f - G_m^{p}(f, U^d)\|_p \asymp \sigma_m(MW_q^r, U^d)_p \asymp m^{-r} (\log m)^{(d-1)r}$$

with

$$r'(q,p) := \begin{cases} \max(1/q, 1/2) - 1/p, & \text{for } p \ge 2; \\ (\max(2/q, 2/p) - 1)/p, & \text{for } p < 2. \end{cases}$$

Proof. The lower estimates follow from Theorem 4.1 in Section 4. Proving the upper estimates we consider first the case $2 \le p < \infty$. If $1 < q \le 2$ we use Theorem 3.2 with $b = p$. This will result in a restriction $r > 1/q - 1/p$. If $2 < q < \infty$ we use the inequality

$$\sup_{f \in MW^r_q} \|f - G^p_m(f, U^d)\|_p \le \sup_{f \in MW^r_2} \|f - G^p_m(f, U^d)\|_p \tag{3.9}$$

to reduce this case to the case that has already been treated. We get a restriction $r > 1/2 - 1/p$ in this case. We proceed to the case $1 < p < 2$ now. If $1 < q \le p$ we use Theorem 3.2 with $p = 2$ and $b = p$ and get

$$\sup_{f \in MW^r_q} \|f - G^p_m(f, U^d)\|_p \le \sup_{f \in MW^r_q} \|f - G^p_m(f, U^d)\|_2 \ll m^{-r}(\log m)^{(d-1)r}$$

provided $r > (2/q - 1)/p$. If $p < q < \infty$ we use an analog of inequality (3.9) to reduce this case to the case $q = p$. In this case we get a restriction $r > (2/p - 1)/p$.

Theorem 3.3 is proved now.

4. Lower bounds for best m-term approximation for the classes MH^r_q and MW^r_q

We begin this section by proving the following two lower estimates.

Theorem 4.1. *For any $1 < q, p < \infty$ and $r > (1/q - 1/p)_+$ we have*

$$\sigma_m(MW^r_q, U^d)_p \gg m^{-r}(\log m)^{(d-1)r}.$$

Theorem 4.2. *For any $1 < q, p < \infty$ and $r > (1/q - 1/p)_+$ we have*

$$\sigma_m(MH^r_q, U^d)_p \gg m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

For proving these theorems we use a method which is based on geometrical charateristics of the sets MW^r_q and MH^r_q. The first realizations (see [DT], [KT]) of this method used volume estimates of projections of the set under consideration onto appropriately chosen finite dimensional subspaces. We will use a variant of this method (see [T7]) expressed in terms of entropy numbers of the given set.

For a bounded set F in a Banach space X we denote for integer m

$$\epsilon_m(F, X) := \inf\{\epsilon : \exists f_1, \dots, f_{2^m} \in X : \quad F \subset \cup_{j=1}^{2^m}(f_j + \epsilon B(X))\},$$

where $B(X)$ is the unit ball of Banach space X and $f_j + \epsilon B(X)$ is the ball of radius ϵ with the center at f_j. The entropy numbers are closely connected with metric entropy. Both characteristics had been well studied for different function classes (see for instance [BS], [T8] and historical remarks there). In this section we will use the following two known estimates (see [T9], [T8]):

(W) for any $1 \le q < \infty$ and $r > 0$ we have

$$\epsilon_m(MW^r_q, L_1) \gg m^{-r}(\log m)^{(d-1)r}; \tag{4.1}$$

(H) for any $r > 0$ we have

$$(4.2) \qquad \epsilon_m(MH^r_\infty, L_1) \gg m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

These estimates will be used in the general method which roughly speaking states that m-term approximations with regard to any reasonable basis are bounded from below by the entropy numbers. We formulate now one result from [T7].

Assume a system $\Psi := \{\psi_j\}_{j=1}^\infty$ of elements in X satisfies the condition:

(VP) There exist three positive constants A_i, $i = 1, 2, 3$, and a sequence $\{n_k\}_{k=1}^\infty$, $n_{k+1} \le A_1 n_k$, $k = 1, 2, \dots$, such that there is a sequence of the de la Vallée-Poussin type operators V_k with the properties

(4.3)

$$V_k(\psi_j) = \lambda_{k,j}\psi_j, \qquad \lambda_{k,j} = 1 \quad \text{for} \quad j = 1, \dots, n_k; \qquad \lambda_{k,j} = 0 \quad \text{for} \quad j > A_2 n_k,$$

$$(4.4) \qquad \|V_k\|_{X\to X} \le A_3, \quad k = 1, 2, \dots \quad .$$

Theorem 4.3. *Assume that for some $a > 0$ and $b \in \mathbb{R}$ we have*

$$\epsilon_m(F, X) \ge C_1 m^{-a}(\log m)^b, \quad m = 2, 3 \dots \quad .$$

Then if a system Ψ satisfies the condition (VP) and also satisfies the following two conditions

$$(4.5) \quad E_n(F, \Psi) := \sup_{f\in F} \inf_{c_1,\dots,c_n} \|f - \sum_{j=1}^n c_j\psi_j\|_X \le C_2 n^{-a}(\log n)^b, \quad n = 1, 2, \dots;$$

$$(4.6) \qquad V_k(F) \subset C_3 F$$

we have

$$\sigma_m(F, \Psi)_X \gg m^{-a}(\log m)^b.$$

Proof of Theorem 4.1. Let $p > 1$ be fixed. We specify $\Psi = U^d$, $X = L_p$ and the sequence of operators $V_n = S_{Q_n}$, where S_{Q_n} is defined as follows

$$S_{Q_n}(f) := \sum_{k\in Q_n} \hat f(k) e^{i(k,x)}$$

with

$$Q_n := \cup_\epsilon \cup_{\|s\|_1\le n} \rho(s,\epsilon),$$

where $\rho(s,\epsilon)$ is defined at the end of Section 1. It is known ([T2], p. 20) that for any $1 < p < \infty$

$$(4.7) \qquad \|S_{Q_n}\|_{L_p\to L_p} \le C(p,d),$$

what implies in particular that

$$(4.8) \qquad S_{Q_n}(MW^r_q) \subset C_3(q,d) MW^r_q, \quad 1 < q < \infty.$$

It remains to check the relation (4.5). We use the known estimate ([T1], p. 36 and [T2], p. 242)

$$E_{Q_n}(MW^r_p)_p \ll 2^{-rn}, \quad 1 < p < \infty. \tag{4.9}$$

Let first $1 < p \le q < \infty$. Then

$$E_{Q_n}(MW^r_q)_p \le E_{Q_n}(MW^r_p)_p \ll 2^{-rn} \ll (\#Q_n)^{-r}(\log \#Q_n)^{(d-1)r}. \tag{4.10}$$

Using (4.1), (4.8) and (4.10) we get from Theorem 4.3 that

$$\sigma_m(MW^r_q, U^d)_p \gg m^{-r}(\log m)^{(d-1)r}. \tag{4.11}$$

It remains to prove (4.11) for $1 < q < p$. This follows right away from (4.11) with $q = p$ and the embedding

$$MW^r_p \subset MW^r_q, \quad q \le p.$$

Theorem 4.1 is proved now.

Proof of Theorem 4.2. This proof is similar to the previous one. We specify as above $\Psi = U^d$, $X = L_p$ and $V_n = S_{Q_n}$. The property (4.7) and the following characterization of the classes MH^r_q, $1 < q < \infty$, (see [T1], p. 33 and [T2], p. 197)

$$f \in MH^r_q \quad \Rightarrow \quad \|\delta_s(f)\|_q \ll 2^{-r\|s\|_1}, \tag{4.12}$$

$$\|\delta_s(f)\|_q \ll 2^{-r\|s\|_1} \quad \Rightarrow \quad C(q,d)f \in MH^r_q \tag{4.13}$$

imply that

$$S_{Q_n}(MH^r_q) \subset C'_3(q,d)MH^r_q, \quad 1 < q < \infty. \tag{4.14}$$

It is clear that it suffices to prove Theorem 4.2 for big q, say $2 \le q < \infty$, and small p, say $1 < p \le 2$. In this case we use the estimate ([T1], p. 37 and [T2], p. 244)

$$E_{Q_n}(MH^r_q)_p \le E_{Q_n}(MH^r_q)_q \ll 2^{-rn}n^{(d-1)/2} \ll \tag{4.15}$$

$$(\#Q_n)^{-r}(\log \#Q_n)^{(d-1)(r+1/2)}.$$

Using (4.2), (4.14) and (4.15) and applying Theorem 4.3 we get for $1 < p \le 2 \le q < \infty$, $r > 0$

$$\sigma_m(MH^r_q, U^d)_p \gg m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

The general case $1 < q, p < \infty$ follows from the case considered by embedding arguments. Theorem 4.2 is proved now.

We prove now the lower bounds (1.2) and (1.3).

Theorem 4.4. *For any orthonormal basis Φ we have for $r > (1/q - 1/2)_+$*

$$\sigma_m(MH^r_q, \Phi)_2 \geq C_1(r,q,d)m^{-r}(\log m)^{(d-1)(r+1/2)}, \quad 1 \leq q < \infty,$$

and

$$\sigma_m(MW^r_q, \Phi)_2 \geq C_2(r,q,d)m^{-r}(\log m)^{(d-1)r}, \quad 1 \leq q < \infty.$$

Proof. This proof is based on a proposition from [K] (see Corollary 2) which we formulate as a lemma.

Lemma 4.1. *There exists an absolute constant $c_0 > 0$ such that for any orthonormal basis Φ and any N-dimensional cube*

$$B_N(\Psi) := \{\sum_{j=1}^N a_j\psi_j, \quad |a_j| \leq 1, \quad j = 1,\dots,N; \quad \Psi := \{\psi_j\}_{j=1}^N \quad \text{an orthonormal system}\}$$

we have

$$\sigma_m(B_N, \Phi)_2 \geq \frac{3}{4}N^{1/2}$$

if $m \leq c_0N$.

Let $q < \infty$ be fixed and m be given. Denote

$$D(n) := \cup_{\|s\|_1=n} D_s^{(+,\dots,+)}$$

and find a minimal n such that

$$m \leq c_0 \# D(n).$$

then

$$m \asymp 2^n n^{d-1}. \tag{4.16}$$

We set $N := \#D(n)$ and choose in a place of $\{\psi_j\}_{j=1}^N$ the system $U(n) := \{U_I\}_{I\in D(n)}$. Then for any $f \in B_N(U(n))$ we have

$$\|\delta_s(f)\|_q^q \asymp \sum_{I\in D_s} \|f_I U_I\|_q^q \leq \sum_{I\in D_s} \|U_I\|_q^q \ll 2^{nq/2}. \tag{4.17}$$

This estimate and the relation (4.13) imply that for some positive $C(q,d)$ we have

$$C(q,d)2^{-n(r+1/2)}B_N(U(n)) \subset MH^r_q.$$

Therefore, Lemma 4.1 gives

$$\sigma_m(MH^r_q, \Phi)_2 \gg 2^{-rn}n^{(d-1)/2} \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}.$$

Next, for $2 \leq q < \infty$ for any $f \in B_N(U(n))$ we have

$$\|f\|_q \ll (\sum_s \|\delta_s(f)\|_q^2)^{1/2} \ll 2^{n/2}n^{(d-1)/2}. \tag{4.18}$$

By Bernstein inequality ([T1], p. 12 and [T2], p. 209) we get from (4.18)

$$\|f^{(r,\dots,r)}\|_q \ll 2^{rn}\|f\|_q \ll 2^{n(r+1/2)}n^{(d-1)/2}.$$

Consequently, for some positive $C(q,d)$ we have

$$C(q,d)2^{-n(r+1/2)}n^{-(d-1)/2}B_N(U(n)) \subset MW^r_q.$$

Therefore, by Lemma 4.1 we get

$$\sigma_m(MW^r_q, \Phi)_2 \gg m^{-r}(\log m)^{(d-1)r}.$$

It is clear that the general case $1 \leq q < \infty$ follows from the above considered case $2 \leq q < \infty$. Theorem 4.4 is proved now.

5. Efficiency of G^p for individual functions

We prove in this section that for each m and $1 < p < \infty$ there is a function $f_{m,p} \in L_p$ such that

$$\|f - G^p_m(f, U^d)\|_p / \sigma_m(f, U^d)_p \gg (\log m)^{(d-1)|1/2-1/p|}. \tag{5.1}$$

We prove this inequality for m of the form

$$m_n := \#D(n) = \sum_{\|s\|_1=n} \#D_s \asymp 2^n n^{d-1}, \quad D_s := D_s^{(+,\dots,+)}. \tag{5.2}$$

For a given n we construct two functions $f_1(n,x)$ and $f_2(n,x)$. The first function is defined as follows

$$f_1(n,x) := \sum_{\|s\|_1=n} e^{i(2^{s_1}x_1+\dots+2^{s_d}x_d)}.$$

Then for any $I \in D_\mu$, $\|\mu\|_1 \neq n$, we have $(f_1(n))_I = 0$ and for $I \in D_\mu$, $\|\mu\|_1 = n$, we have

$$f_1(n)_I = 2^{-n/2}, \quad \text{and} \quad \|f_1(n)_I U_I\|_p \asymp 2^{-n/p}.$$

Next, by the Littlewood-Paley inequalities we get

$$\|f_1(n)\|_p \asymp n^{(d-1)/2}, \quad 1 < p < \infty.$$

We proceed to define the second function. We set $l(n) = [(\log m_n)/d] + 1$ and define

$$f_2(n) := 2^{-n/2} \sum_{I \in \Lambda(n)} U_I$$

where $\Lambda(n) \subset D_{(l(n),\dots,l(n))}$ with $\#\Lambda(n) = m_n$. Then for each $I \in \Lambda(n)$ we have

$$f_2(n)_I = 2^{-n/2}$$

and

$$\|f_2(n)_I U_I\|_p \asymp 2^{-n/p}.$$

We also have

$$\|f_2(n)\|_p \asymp 2^{-n/2}(\#\Lambda(n) 2^{n(1/2-1/p)p})^{1/p} \asymp n^{(d-1)/p}.$$

Let $2 \le p < \infty$ and a constant $C_1(d,p)$ be such that

$$\min_{I \in \Lambda(n)} \|f_2(n)_I U_I\|_p > C_1(d,p) \max_I \|f_1(n)_I U_I\|_p. \tag{5.3}$$

Consider

$$f_{m_n,p} := C_1(d,p) f_1(n) + f_2(n).$$

Then by (5.3) we have

$$\|f_{m_n,p} - G^p_{m_n}(f_{m_n,p}, U^d)\|_p = C_1(d,p)\|f_1(n)\|_p \asymp n^{(d-1)/2}. \tag{5.4}$$

Next,

$$\sigma_{m_n}(f_{m_n,p}, U^d)_p \le \|f_2(n)\|_p \asymp n^{(d-1)/p}. \tag{5.5}$$

Combining (5.4) with (5.5) we get (5.1) for $m = m_n$.

Let now $1 < p \le 2$ and a constant $C_2(d,p)$ be such that

$$\min_{I\in D(n)} \|f_1(n)_I U_I\|_p > C_2(d,p) \max_{I\in\Lambda(n)} \|f_2(n)_I U_I\|_p.$$

Consider

$$f_{m_n,p} := f_1(n) + C_2(d,p) f_2(n).$$

Then we have on one hand

$$\|f_{m_n,p} - G^p_{m_n}(f_{m_n,p}, U^d)\|_p = C_2(d,p)\|f_2(n)\|_p \asymp n^{(d-1)/p} \tag{5.6}$$

and on the other hand

$$\sigma_{m_n}(f_{m_n,p}, U^d)_p \le \|f_1(n)\|_p \asymp n^{(d-1)/2}. \tag{5.7}$$

Combining (5.6) with (5.7) we get (5.1) for $m = m_n$. It is easy to see that the general case of m can be derived from the case $m = m_n$.

We note that using the result of P. Wojtaszczyk [Wo] on equivalence of U to the Haar system in all L_p, $1 < p < \infty$, we can derive some results on U^d from the corresponding results on the multivariate Haar system $\mathcal{H}^d$ (see [T4]). For instance, we get from [T4, Theorems 2.1, 2.2] that for any $f \in L_p$, $1 < p < \infty$, the inequality

$$\|f - G^p_m(f, U^d)\|_p \le C(p,d)(\log m)^d \sigma_m(f, U^d)_p \tag{5.8}$$

holds, and from [T4, Theorem 2.1, Section 3] that in the case $d = 2$, $4/3 \le p \le 4$, the factor $(\log m)^d$ in (5.8) can be replaced by $(\log m)^{(d-1)|1/2-1/p|}$. The last remark shows that the inequality (5.1) is sharp.

6. One special dictionary with tensor product structure

In this section we study m-term approximation with regard to the dictionary Y (see the Introduction for the definition) which is something intermediate between the dictionaries U^2 and Π^2. We prove here the following theorem.

Theorem 6.1. *For $d = 2$ and $1 < q, p < \infty$ we have*

$$\sigma_m(MH^r_q, Y)_p \asymp \sigma_m(MW^r_q, Y)_p \asymp m^{-2r+(1/q-1/p)_+},$$

provided $r > (1/q - 1/p)_+$.

Proof. The lower estimates in the case $1 < p \le q < \infty$ and in the case $1 < q \le p \le 2$ follow from the corresponding result for bilinear approximations (see (1.11)).

We remark only that the restriction $r > 1$ in (1.11) was used for proving upper estimates. For details see [T5]. We prove now the lower estimates in the case $1 < q \le p < \infty$. It is clear that it suffices to carry out the proof for m of the form $m = 2^{l-1}$. Consider a function

$$f(x_1, x_2) := \sum_{I \in D_l^+} U_I(x_1) U_I(x_2).$$

We have

$$\|f\|_q^q \asymp 2^{l(q-1)}, \quad 1 < q < \infty,$$

and by Bernstein inequality

$$\|f^{(r,r)}\|_q \ll 2^{l(2r+1-1/q)}. \tag{6.1}$$

Assume an m-term approximant with regard to Y has the form

$$g(x) = \sum_{I \in \Omega_1} U_I(x_1) v_I^1(x_2) + \sum_{I \in \Omega_2} v_I^2(x_1) U_I(x_2)$$

with $\#\Omega_1 + \#\Omega_2 = m$. Then for $1 < p < \infty$ we have

$$\|f - g\|_p^p \ge C(p) \|\delta_{(l,l)}^{(+,+)}(f - g)\|_p^p \gg \tag{6.2}$$

$$\|\sum_{I \in D_l^+ \setminus (\Omega_1 \cup \Omega_2)} U_I(x_1) U_I(x_2)\|_p^p \gg 2^{l(p-1)}.$$

The inequalities (6.1) and (6.2) imply

$$\sigma_m(MW_q^r, Y)_p \gg 2^{l(-2r+1/q-1/p)} \asymp m^{-2r+1/q-1/p}.$$

It remains to note that MW_q^r is embedded in MH_q^r.

We proceed with proving the upper estimates. It is sufficient to prove the upper estimates in the case $1 < q \le p < \infty$. We prove the upper estimates for the wider class MH_q^r. We use in the proof a combination of a linear method and the algorithm $G^p(\cdot, U^2)$. For a fixed n we define a linear operator S_n as follows

$$S_n'(f)(x) := \sum_{|I| \ge 2^{-n}} \langle f(\cdot, x_2), U_I(\cdot) \rangle U_I(x_1)$$

$$S_n(f)(x) := S_n'(f)(x) + \sum_{|I| \ge 2^{-n}} \langle f(x_1, \cdot) - S_n'(f)(x_1, \cdot), U_I(\cdot) \rangle U_I(x_2).$$

Then

$$f^n := f - S_n(f) = \sum_{|I_1| < 2^{-n}, |I_2| < 2^{-n}} f_I U_I.$$

We apply the greedy algorithm $G^p(\cdot, U^2)$ to f^n. The proof is similar to, but simpler than the corresponding proofs in Sections 2 and 3.

Lemma 6.1. *Let $1 < q \le p < \infty$. Denote*

$$h(n) := 2^{-n(2r-1/q+2/p)}.$$

Then for any function f of the form

$$f = \sum_{|I_1|<2^{-n},|I_2|<2^{-n}} f_I U_I, \quad \|\delta^\epsilon_s(f)\|_q \le C(r,d,q)2^{-r\|s\|_1}, \tag{6.3}$$

we have

$$\#H_n \ll 2^n$$

where

$$H_n := \{I : \|f_I U_I\|_p \ge h(n)\}.$$

Proof. Denote $N^\epsilon_s := \#(H_n \cap D^\epsilon_s)$. Then similarly to the proof of Lemma 2.1 we get

$$N_s \ll h(n)^{-q} 2^{\|s\|_1(-rq-q/p+1)},$$

and using $r > 1/q - 1/p$ we get

$$\sum_{s_1\ge n, s_2\ge n} N_s \ll h(n)^{-q} 2^{2n(-rq-q/p+1)} = 2^n.$$

Lemma 6.2. *Let $h(n)$ be from Lemma 6.1. Assume that a function f of the form (6.3) satisfies the restriction*

$$\|f_I U_I\|_p < h(n) \quad \textit{for all} \quad I.$$

Then we have

$$\|f\|_p \le 2^{-n(2r-1/q+1/p)}.$$

Proof. For each s we have

$$\|\delta_s(f)\|^p_p \ll \sum_{I\in D_s} \|f_I U_I\|^p_p \le h(n)^{p-q} \sum_{I\in D_s} \|f_I U_I\|^q_p \ll$$

$$h(n)^{p-q} \sum_{I\in D_s} \|f_I U_I\|^q_q 2^{\|s\|_1(1/q-1/p)q} \ll h(n)^{p-q} 2^{\|s\|_1(-rq+1-q/p)}$$

and

$$\|f\|_p \le \sum_{s_1\ge n,s_2\ge n} \|\delta_s(f)\|_p \ll h(n)^{1-q/p} \sum_{s_1\ge n,s_2\ge n} 2^{\|s\|_1(-rq+1-q/p)/p}.$$

Using $r > 1/q - 1/p$ we get from here

$$\|f\|_p \ll 2^{n(-2r+1/q-1/p)}.$$

Lemma 6.2 is proved now.

We continue the proof of Theorem 6.1. From Lemmas 6.1 and 6.2 we obtain for f^n

$$\|f^n - G^p_{2^n}(f^n, U^2)\|_p \ll 2^{-n(2r-1/q+1/p)}. \tag{6.4}$$

The estimate (6.4) implies the upper estimate in Theorem 6.1 for $m = 5(2^n)$. It is clear that this implies the general case of m.

7. Further remarks

The results we have developed in the previous sections in the periodic case can be extended to the nonperiodic case and to other systems Ψ instead of U^d. We discuss here in more detail a generalization of the results from Section 3. The key points of the proof of Theorems 3.1– 3.3 were the following.

1. The multivariate system U^d satisfies the relation (1.17)

$$\|\sum_{I\in D_s} f_I U_I\|_p^p \asymp \sum_{I\in D_s} \|f_I U_I\|_p^p, \quad 1<p<\infty,$$

which is a corollary of the corresponding relation (1.14) for the univariate system U. This system also satisfies (1.15)

$$\|U_I\|_p \asymp |I|^{1/p-1/2}, \quad 1<p<\infty.$$

2. The system U^d satisfies the Littlewood-Paley inequalities in a weak form

$$\|f\|_p \asymp \|(\sum_s |\sum_{I\in D_s} f_I U_I|^2)^{1/2}\|_p, \quad 1<p<\infty.$$

3. The function class MW^r_q has a certain approximation property (see (3.1)) which is equivalent to the Jackson inequality: for $f\in MW^r_q$, $1<q<\infty$, we have

$$\|f-\sum_{|I|\ge 2^{-n}} f_I U_I\|_q \ll 2^{-rn}; \tag{7.1}$$

and the embedding property $MW^r_{q_1}\subset MW^r_{q_2}$ if $q_1\ge q_2$.

Thus if some system Ψ and function classes F^r_q satisfy the conditions 1–3 above then Theorems 3.1–3.3 hold with U^d and MW^r_q replaced by Ψ and F^r_q.

In the paper [DKT] we gave some sufficient conditions on a system Ψ to be L_p-equivalent to the Haar system. We recall the definition of the Haar system and make some simple observations about systems L_p-equivalent to the Haar system. Denote the univariate Haar system by $\mathcal{H} := \{H_I\}_I$, where I are dyadic intervals of the form $I=[(j-1)2^{-n}, j2^{-n})$, $\quad j=1,\dots,2^n;\quad n=0,1,\dots$ and $I=[0,1]$ with

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

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

Consider the multivariate Haar basis $\mathcal{H}^d := \mathcal{H}\times\cdots\times\mathcal{H}$ which consists of functions

$$H_I(x)=\prod_{j=1}^d H_{I_j}(x_j), \qquad I=I_1\times\cdots\times I_d, \quad x=(x_1,\dots,x_d).$$

We say that a system $\Psi = \{\psi_I\}$ is L_p-equivalent to the Haar system $\mathcal{H}^d$ if for any finite set Λ and for any coefficients $\{c_I\}$ we have

$$C_1(\Psi,p,d)\|\sum_{I\in\Lambda} c_I H_I\|_p \le \|\sum_{I\in\Lambda} c_I\psi_I\|_p \le C_2(\Psi,p,d)\|\sum_{I\in\Lambda} c_I H_I\|_p. \tag{7.2}$$

It is well known (see for instance [KS]) that the Haar system satisfies the Littlewood-Paley inequalities in a strong form

$$\|\sum_I c_I H_I\|_p \asymp \|(\sum_I |c_I H_I|^2)^{1/2}\|_p, \quad 1<p<\infty. \tag{7.3}$$

It is clear from (7.2) and (7.3) and the corresponding properties of $\mathcal{H}^d$ that each system Ψ which is L_p-equivalent to $\mathcal{H}^d$ with $1<p<\infty$ satisfies property 1 and a stronger analog of property 2 from above. In the paper [DKT] we gave some sufficient conditions on a system Ψ to have the Jackson inequality (7.1). For particular examples of Ψ satisfying (7.1) see [Km].

Completing this section we conclude that the results on nice properties of the system U^d can be extended onto many other systems including wavelet type systems. For instance the following two theorems hold.

Theorem 7.1. *Assume a system Ψ is L_p-equivalent to the Haar system $\mathcal{H}^d$, $1<p<\infty$, and function classes F^r_q having the following property: for any $f\in F^r_q$ we have*

$$\|f - \sum_{|I|\ge 2^{-n}} c_I(f,\Psi)\psi_I\|_q \ll 2^{-rn}, \quad 1<q<\infty,$$

with a constant independent of f and n; and $F^r_{q_1}\subset F^r_{q_2}$ if $q_1\ge q_2$. Then we have

$$\sup_{f\in F^r_q}\|f - G^p_m(f,\Psi)\|_p \ll m^{-r}(\log m)^{(d-1)r}, \quad 1<q,p<\infty,$$

provided $r>r'(q,p)$ with $r'(q,p)$ from Theorem 3.3.

Theorem 7.2. *Assume a system Ψ is L_p-equivalent to the Haar system $\mathcal{H}^d$, $1<p<\infty$, and function classes F^r_q having the following property: for any $f\in F^r_q$ we have*

$$\|\sum_{I\in D_s} c_I(f,\Psi)\psi_I\|_q \ll 2^{-r\|s\|_1}, \quad 1<q<\infty,$$

with a constant independent of f and n; and $F^r_{q_1}\subset F^r_{q_2}$ if $q_1\ge q_2$. Then we have

$$\sup_{f\in F^r_q}\|f - G^p_m(f,\Psi)\|_p \ll m^{-r}(\log m)^{(d-1)(r+1/2)}, \quad 1<q,p<\infty,$$

provided $r>r(q,p)$ with $r(q,p)$ from Theorem 2.3.

References

[BS] M.Sh. Birman and M.Z. Solomyak, *Piecewise polynomial approximation of the classes W_p^α*, Mat. Sb. **73(115)** (1967), 331–355; English transl. in Math. USSR Sb. **2** (1967), 295–317.

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

[DKT] R.A. DeVore, S.V. Konyagin, and V.N. Temlyakov, *Hyperbolic wavelet approximation*, Constructive Approximation **14** (1998), 1–26.

[DT] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation by trigonometric sums*, J. Fourier Analysis and Applications **2** (1995), 29–48.

[K] B.S. Kashin, *On approximation properties of complete orthonormal systems*, Trudy Mat. Inst. Steklov **172** (1985), 187–191; English transl. in Proc. Steklov Inst. Math. **1987**, no. 3, 207–211.

[KS] B.S. Kashin and A.A. Saakyan, *Orthogonal Series*, Nauka, Moscow, 1984; English transl. in American Math. Soc., Providence, R.I., 1989.

[KT] B.S. Kashin and V.N. Temlyakov, *On best m-term approximation and the entropy of sets in the space L^1*, Mat. Zametki **56** (1994), 57–86; English transl. in Math. Notes **56** (1994), 1137–1157.

[Km] A. Kamont, *On hyperbolic summation and hyperbolic moduli of smoothness*, Constructive Approximation **12** (1996), 111–125.

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

[T1] V.N. Temlyakov, *Approximation of functions with bounded mixed derivative*, Trudy Mat. Inst. Steklov **178** (1986); English transl. in Proc. Steklov Institute **1989,**, no. 1.

[T2] V.N. Temlyakov, *Approximation of periodic functions*, Nova Science Publishers, Inc., New York, 1993.

[T3] V.N. Temlyakov, *The best m-term approximation and Greedy Algorithms*, Advances in Comp. Math. (to appear).

[T4] V.N. Temlyakov, *Nonlinear m-term approximation with regard to the multivariate Haar system*, East J. Approx. **4** (1998), 87–106.

[T5] V.N. Temlyakov, *Estimates of the best bilinear approximations of functions of two variables and some of their applications*, Mat. Sb. **134(176)** (1987), 93–107; English transl. in Math. USSR Sb. **62** (1989), 95–109.

[T6] V.N. Temlyakov, *Estimates of the best bilinear approximations of periodic functions*, Trudy Mat. Inst. Steklov **181** (1988), 250–267; English transl. in Proc. Steklov Inst. Math. **1989**, no. 4, 275–293.

[T7] V.N. Temlyakov, *Nonlinear Kolmogorov's widths*, Matem. zametki **63** (1998), 891–902.

[T8] V.N. Temlyakov, *Estimates of the asymptotic characteristics of classes of functions with bounded mixed derivative or difference*, Trudy Mat. Inst. Steklov **189** (1989), 138–168; English transl. in Proc. Steklov Inst. Math. **1990**, no. 4, 161–197.

[T9] V.N. Temlyakov, *On estimates of ϵ-entropy and widths of classes of functions with bounded mixed derivative or difference*, Dokl. Acad. of Science USSR **301** (1988), 288–291; English transl. in Soviet math. Dokl. **38** (1989), 84–87.

[Wo] P. Wojtaszczyk, *On unconditional polynomial bases in L_p and Bergman spaces*, Constructive Approximation **13** (1997), 1–15.

[W] H. Wozniakowski, *Average case complexity of linear multivariate problems*, J. Complexity **8** (1992), 337–372.