Nonlinear tensor product approximation of functions

D. Bazarkhanov † and V. Temlyakov ‡

Abstract

We are interested in approximation of a multivariate function \( f(x_1, \ldots, x_d) \) by linear combinations of products \( u^1(x_1) \cdots u^d(x_d) \) of univariate functions \( u^i(x_i), \ i = 1, \ldots, d \). In the case \( d = 2 \) it is the classical problem of bilinear approximation. In the case of approximation in the \( L_2 \) space the bilinear approximation problem is closely related to the problem of singular value decomposition (also called Schmidt expansion) of the corresponding integral operator with the kernel \( f(x_1, x_2) \). There are known results on the rate of decay of errors of best bilinear approximation in \( L_p \) under different smoothness assumptions on \( f \). The problem of multilinear approximation (nonlinear tensor product approximation) in the case \( d \geq 3 \) is more difficult and much less studied than the bilinear approximation problem. We will present results on best multilinear approximation in \( L_p \) under mixed smoothness assumption on \( f \).

1 Introduction

In this paper we study multilinear approximation (nonlinear tensor product approximation) of functions. For a function \( f(x_1, \ldots, x_d) \) denote

\[
\Theta_M(f)_X := \inf_{\{u^j\}_{j=1, \ldots, M}, i=1, \ldots, d} \| f(x_1, \ldots, x_d) - \sum_{j=1}^M \prod_{i=1}^d u^j_i(x_i) \|_X
\]
and for a function class $F$ define

$$\Theta_M(F)_X := \sup_{f \in F} \Theta_M(f)_X.$$ 

In the case $X = L_p$ we write $p$ instead of $L_p$ in the notation. In other words we are interested in studying $M$-term approximations of functions with respect to the dictionary

$$\Pi^d := \{ g(x_1, \ldots, x_d) : g(x_1, \ldots, x_d) = \prod_{i=1}^{d} u^i(x_i) \}$$

where $u^i(x_i)$ are arbitrary univariate functions. We discuss the case of $2\pi$-periodic functions of $d$ variables and approximate them in the $L_p$ spaces. Denote by $\Pi^d_p$ the normalized in $L_p$ dictionary $\Pi^d$ of $2\pi$-periodic functions. We say that a dictionary $\mathcal{D}$ has a tensor product structure if all its elements have a form of products $u^1(x_1) \cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i = 1, \ldots, d$. Then any dictionary with tensor product structure is a subset of $\Pi^d$. The classical example of a dictionary with tensor product structure is the $d$-variate trigonometric system $\{ e^{i(k,x)} \}$. Other examples include the hyperbolic wavelets and the hyperbolic wavelet type system $\mathcal{U}^d$ defined in Section 3.

Modern problems in approximation, driven by applications in biology, medicine, and engineering, are being formulated in very high dimensions, which brings to the fore new phenomena. For instance, partial differential equations in a phase space of large spacial dimensions (e.g. Schrödinger and Fokker-Plank equations) are very important in applications. It is known (see, for instance, [1]) that such equations involving large number of spacial variables pose a serious computational challenge because of the so-called curse of dimensionality, which is caused by the use of classical notions of smoothness as the regularity characteristics of the solution. The authors of [1] show that replacing the classical smoothness assumptions by structural assumptions in terms of sparsity with respect to the dictionary $\Pi^d$, they overcome the above computational challenge. They prove that the solutions of certain high-dimensional equations inherit sparsity, based on tensor product decompositions, from given data. Thus, our algorithms, which provide good sparse approximation with respect to $\Pi^d$ for individual functions might be useful in applications to PDEs of the above type. The nonlinear tensor product approximation is very important in numerical applications. We refer the reader to the monograph [4] which presents the state of the art on the topic. Also, the reader can find a very recent discussion of related results in [7].
In the case $d = 2$ the multilinear approximation problem is the classical problem of bilinear approximation. In the case of approximation in the $L_2$ space the bilinear approximation problem is closely related to the problem of singular value decomposition (also called Schmidt expansion) of the corresponding integral operator with the kernel $f(x_1, x_2)$. There are known results on the rate of decay of errors of best bilinear approximation in $L_p$ under different smoothness assumptions on $f$. We only mention some known results for classes of functions which are studied in this paper. We study the classes $W^{r}_q$ of functions with bounded mixed derivative which we define for positive $r$ (not necessarily an integer). Let

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

be the univariate Bernoulli kernel and let

$$F_r(x) := F_r(x_1, \ldots, x_d) := \prod_{i=1}^{d} F_r(x_i)$$

be its multivariate analog. We define

$$W^{r}_q := \{ f : f = F_r \ast \varphi, \| \varphi \|_q \leq 1 \},$$

where $\ast$ denotes the convolution.

The problem of estimating $\Theta_M(f)_2$ in case $d = 2$ (best $M$-term bilinear approximation in $L_2$) is a classical one and was considered for the first time by E. Schmidt [6] in 1907. For many function classes $F$ an asymptotic behavior of $\Theta_M(F)_p$ is known. For instance, the relation

$$\Theta_M(W^{r}_q)_p \asymp M^{-2r + (1/q - \max(1/2,1/p))}, \tag{1.1}$$

for $r > 1$ and $1 \leq q \leq p \leq \infty$ follows from more general results in [10]. In the case $d > 2$ almost nothing is known. There is (see [11]) an upper estimate in the case $q = p = 2$

$$\Theta_M(W^{r}_2)_2 \ll M^{-rd/(d-1)}. \tag{1.2}$$

Results of this paper are around the bound (1.2). First of all we discuss the lower bound matching the upper bound (1.2). In the case $d = 2$ the lower bound

$$\Theta_M(W^{r}_p)_p \gg M^{-2r}, \quad 1 \leq p \leq \infty, \tag{1.3}$$

3
follows from more general results in [10] (see (1.1) above). A stronger result

\[ \Theta_M(W_\infty^r) \gg M^{-2r} \quad (1.4) \]

follows from Theorem 1.1 in [12].

We could not prove the lower bound matching the upper bound (1.2) for \( d > 2 \). Instead, we prove a weaker lower bound. For a function \( f(x_1, \ldots, x_d) \) denote

\[ \Theta^b_M(f)_X := \inf_{\{u^j_i, \|u^j_i\|_x \leq b\}} \|f(x_1, \ldots, x_d) - \sum_{j=1}^{M} \prod_{i=1}^{d} u^j_i(x_i)\|_X \]

and for a function class \( F \) define

\[ \Theta^b_M(F)_X := \sup_{f \in F} \Theta^b_M(f)_X. \]

In Section 2 we prove the following lower bound (see Corollary 2.2)

\[ \Theta^b_M(W_\infty^r) \gg (M \ln M)^{-\frac{rd}{d-1}}. \]

This lower bound indicates that probably the exponent \( \frac{rd}{d-1} \) is the right one in the power decay of the \( \Theta_M(W_p^r)_p \).

Secondly, we discuss some upper bounds which extend the bound (1.2). The relation (1.1) shows that for \( 2 \leq p \leq \infty \) in the case \( d = 2 \) one has

\[ \Theta_M(W_2^r)_p \ll M^{-2r}. \quad (1.5) \]

In Section 3 we extend (1.5) for \( d > 2 \).

**Theorem 1.1.** Let \( 2 \leq p < \infty \) and \( r > (d - 1)/d \). Then

\[ \Theta_M(W_2^r)_p \ll \left( \frac{M}{(\log M)^{d-1}} \right)^{-\frac{rd}{d-1}}. \]

The proof of Theorem 1.1 in Section 3 is not constructive. It goes by induction and uses a nonconstructive bound in the case \( d = 2 \), which is obtained in [11]. The corresponding proof from [11] uses the bounds for the Kolmogorov width \( d_n(W_2^r, L_\infty) \), proved by Kashin [5]. Kashin’s proof is a probabilistic one, which provides existence of a good linear subspace.
for approximation, but there is no known explicit constructions of such subspaces. This problem is related to a problem from compressed sensing on construction of good matrices with Restricted Isometry Property. It is an outstanding difficult open problem. In Section 4 we discuss constructive ways of building good multilinear approximations. The simplest way would be to use known results about $M$-term approximation with respect to special systems with tensor product structure. We illustrate this idea on the example of the system $U^d$ defined and discussed in Section 3. We define a well-known Thresholding Greedy Algorithm with respect to a basis. It is convenient for us to enumerate the basis functions by dyadic intervals. Assume a given system $\Psi$ of functions $\psi_I$ indexed by dyadic intervals can be enumerated in such a way that $\{\psi_I\}_{j=1}^{\infty}$ is a basis for $L_p$. Then we define the greedy algorithm $G^p(\cdot, \Psi)$ as follows. Let

$$f = \sum_{j=1}^{\infty} c_I(f, \Psi) \psi_I,$$

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 $\Lambda_m$ a set of $m$ dyadic intervals $I$ such that

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

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

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

For a system (dictionary) of elements $D$ define the best $M$-term approximation in $X$ as follows

$$\sigma_M(f, D)_X := \inf_{g_j \in D, c_j, j=1,\ldots,M} \|f - \sum_{j=1}^{M} c_j g_j\|_X.$$

With this standard notation we have

$$\Theta_M(f)_p = \sigma_M(f, \Pi^d)_L_p.$$

It is proved in [14] that for $1 < q, p < \infty$ and big enough $r$

$$\sup_{f \in W^r_q} \|f - G^p_M(f, U^d)\|_p \approx \sigma_M(W^r_q, U^d)_p \approx M^{-r} (\log M)^{(d-1)r}. \quad (1.6)$$
The above relation (1.6) illustrates two phenomena: (I) for the class $W^r_q$ the simple Thresholding Greedy Algorithm provides near best $M$-term approximation; (II) the rate $M^{-r}(\log M)^{(d-1)r}$ of best $M$-term approximation with respect to the basis $\mathcal{U}^d$, which has a tensor product structure, is not as good as best $M$-term approximation with respect to $\Pi^d$ (we have exponent $r$ for $\mathcal{U}^d$ instead of $\frac{rd}{d+1}$ for $\Pi^d$).

As we pointed out in a discussion of Theorem 1.1 the upper bound in Theorem 1.1 is proved with a help of probabilistic results. There is no known deterministic constructive methods (theoretical algorithms), which provide the corresponding upper bounds. In Section 4 we apply greedy type algorithms to obtain upper estimates of $\Theta_M(W^r_2)_p$ (see Theorems 4.3 and 4.4). The important feature of our proof is that it is deterministic and moreover it is constructive. Formally, the optimization problem

$$\Theta_M(f)_X := \inf_{\{u^j\}_{j=1,\ldots,M}} \|f(x_1, \ldots, x_d) - \sum_{j=1}^{M} \prod_{i=1}^{d} u^j_i(x_i)\|_X$$

is deterministic: one needs to minimize over $u^j_i$. However, minimization by itself does not provide any upper estimate. It is known (see [2]) that simultaneous optimization over many parameters is a very difficult problem. Thus, in nonlinear $M$-term approximation we look for methods (algorithms), which provide approximation close to best $M$-term approximation and at each step solve an optimization problem over only one parameter ($\prod_{i=1}^{d} u^j_i(x_i)$ in our case). In Section 4 we provide such an algorithm for estimating $\Theta_M(f)_p$. We call this algorithm constructive because it provides an explicit construction with feasible one parameter optimization steps. We stress that in the setting of approximation in an infinite dimensional Banach space, which is considered in our paper, the use of term algorithm requires some explanation. In this paper we discuss only theoretical aspects of the efficiency (accuracy) of $M$-term approximation and possible ways to realize this efficiency. The greedy algorithms used in Section 4 give a procedure to construct an approximant, which turns out to be a good approximant. The procedure of constructing a greedy approximant is not a numerical algorithm ready for computational implementation. Therefore, it would be more precise to call this procedure a theoretical greedy algorithm or stepwise optimizing process. Keeping this remark in mind we, however, use the term greedy algorithm in this paper because it has been used in previous papers and has become a standard name for procedures used in Section 4 and for more general proce-
dures of this type (see for instance [3], [17]). Also, the theoretical algorithms, which we use in Section 4, become algorithms in a strict sense if instead of an infinite dimensional setting we consider a finite dimensional setting, replacing, for instance, the $L_p$ space by its restriction on the set of trigonometric polynomials. We note that the greedy-type algorithms are known to be very efficient in numerical applications (see, for instance, [20] and [8]).

In Section 4 we use two very different greedy-type algorithms to provide a constructive multilinear approximant. The first greedy-type algorithm, which gives Lemma 4.1, is based on a very simple dictionary similar to $U^d$ consisting of shifts of the de la Vallée Poussin kernels. The algorithm uses function (dyadic blocks of a function) evaluations and picks the largest of them. The second greedy-type algorithm, which gives Lemma 4.4, is more complex. It is based on the dictionary $\Pi^d$ and uses the Weak Chebyshev Greedy Algorithm with respect to $\Pi^d$ to update the approximant. Surprisingly, these two algorithms give the same error bound. For instance, Theorems 4.3 and 4.4 give for big enough $r$ the following constructive upper bound for $2 \leq p < \infty$

$$
\Theta_M(W_2^r)_p \ll \left( \frac{M}{(\ln M)^{d-1}} \right)^{-\frac{rd}{2(d-1)} + \frac{\beta}{d-1}} \cdot \beta := \frac{1}{2} - \frac{1}{p}.
$$

This constructive upper bound has an extra term $\frac{\beta}{d-1}$ in the exponent compared to the best $M$-term approximation. It would be interesting to find a constructive proof of Theorem 1.1.

Clearly, the problem of estimating $\Theta_M(f)_p$ is a nonlinear problem. We already made a comment that a step of replacing a very redundant dictionary $\Pi^d$ by an orthonormal dictionary $U^d$ with tensor product structure results in loosing in accuracy from, roughly, $M^{-\frac{d}{2d-1}}$ to $M^{-r}$. We now make a comment on a step of replacing the nonlinear $M$-term approximation problem with respect to $\Pi^d$ by a linear approximation problem. We begin with the Kolmogorov width, where we are optimizing over linear subspaces of fixed dimension:

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

It is known (see [13], Ch.3, Theorem 4.4) that for $2 \leq p < \infty$

$$
d_M(W_2^r, L_p) \asymp M^{-r} (\log M)^{r(d-1)}. \quad (1.7)
$$
Comparing (1.7) with (1.6) and Theorem 1.1 we see that there is a linear subspace of dimension $M$, which gives the same (in the sense of order) error bound as nonlinear $M$-term approximation with respect to $U^d$, which is worse than $M$-term approximation with respect to $\Pi^d$. However, there are two problems with such a linear subspace. First of all, the proof of the upper bound in (1.7) is based on Kashin’s result [5] and, therefore, we only know existence of such a subspace, but do not have a construction of it. Second, we do not know if this subspace has a tensor product structure.

Next, let us look at the linear widths introduced by Tikhomirov [18]:

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

It is known (see [13], Ch.1, Theorem 4.2) that for $2 \leq p < \infty$

$$\lambda_M(W^r_2, L_p) \gg M^{-r+1/2-1/p}. \quad (1.8)$$

It is also known (see [13], Ch.3, Theorem 3.2) that approximation by the hyperbolic cross polynomials provides accuracy similar to the one in (1.8) (up to the factor $(\log M)^{(d-1)(r-1/2+1/p)}$). Thus, we have the following conclusion. By linear methods we can achieve accuracy $M^{-r+1/2-1/p}$ in the power scale and cannot do better than that. Accuracy $M^{-r+1/2-1/p}$ is substantially worse than the one provided by Theorem 1.1 and Theorems 4.3, 4.4.

We conclude the Introduction by a brief remark on the technique used in this paper. In the proof of (1.2) (see [11]) the following simple observation was used:

$$t(x) := \sum_{|k_d| \leq N_d} \cdots \sum_{|k_1| \leq N_1} c_{k_1,\ldots,k_d} e^{i(k,x)}$$

$$= \sum_{|k_d| \leq N_d} \cdots \sum_{|k_2| \leq N_2} e^{i(k_2x_2+\cdots+k_dx_d)} u_{k_2,\ldots,k_d}(x_1), \quad (1.9)$$

where

$$u_{k_2,\ldots,k_d}(x_1) := \sum_{|k_1| \leq N_1} c_{k_1,\ldots,k_d} e^{i k_1 x_1}.$$ 

Representation (1.9) means that any element $t$ from the $\prod_{j=1}^d (2N_j + 1)$-dimensional space of trigonometric polynomials of order $N = (N_1, \ldots, N_d)$ is $\prod_{j=2}^d (2N_j + 1)$ sparse with respect to the dictionary $\Pi^d$. It is a simple observation, which is useful in general theory of representation of tensors (see [4],
We use this observation in the proof of upper bounds (see, for instance, Lemma 3.1). For constructive proofs in Section 4 we use greedy-type algorithms. The one used in the proof of Lemma 4.1 is a development of the corresponding algorithm from [11]. The greedy algorithm used in the proof of Lemma 4.4 is from the general theory of greedy algorithms in Banach spaces developed recently (see [17]). Probably, this paper is the first one, which contains an application of that algorithm and general theory behind it in tensor product approximation.

The technique used in Section 2 for proving the lower bounds is a development of techniques from [10] and [12]. It is based on known results about the entropy of unit balls in spaces of trigonometric polynomials. We could not fully extend the technique, which works well in the case of bilinear approximation \((d = 2)\), to the case of tensor approximation with \(d \geq 3\). We need an extra assumption on boundedness of each term of the tensor approximant. Similar assumptions appear in other problems of tensor approximations, for instance, in the existence of best approximations and stability of approximations (see [4], s.9.4.3).

## 2 The lower bound

Let \(X\) be a Banach space and let \(B_X\) denote the unit ball of \(X\) with the center at 0. Denote by \(B_X(y, r)\) a ball with center \(y\) and radius \(r\): \(\{x \in X : \|x - y\| \leq r\}\). For a compact set \(K\) and a positive number \(\epsilon\) we define the covering number \(N_\epsilon(K)\) as follows

\[
N_\epsilon(K) := N_\epsilon(K, X) := \min\{n : \exists y^1, \ldots, y^n : K \subseteq \bigcup_{j=1}^n B_X(y^j, \epsilon)\}.
\]

The following bound is well known (see, for instance, [17], Ch. 3).

**Lemma 2.1.** For any \(n\)-dimensional Banach space \(X\) we have

\[
\epsilon^{-n} \leq N_\epsilon(B_X, X) \leq (1 + 2/\epsilon)^n.
\]

For \(N = (N_1, \ldots, N_d)\) let \(T(N)\) be the set of trigonometric polynomials of order \(N_j\) in the \(j\)th variable. Denote

\[
T(N)_p := \{t \in T(N) : \|t\|_p \leq 1\}
\]
and

\[ \Pi^d(N, n, b) := \{ f \in T(N)_2, f(x) = \sum_{j=1}^{n} u_j^1(x_1) \cdots u_j^d(x_d), \|u_j^i\|_2 \leq b \}. \]

**Lemma 2.2.** We have

\[ N_\epsilon(\Pi^d(N, n, b), L_2) \leq \left(\frac{Cb}{\delta}\right) \sqrt[2n]{\sum_{i=1}^{d} (2N_i+1)}, \quad 0 < \epsilon \leq 1. \]

**Proof.** First of all it is clear that we can assume that \( u_j^i \in T(N_i), j = 1, \ldots, n, \) \( i = 1, \ldots, d. \) Second, in the \( b \)-ball of the \( T(N_i)_2 \) we build a \( \delta \)-net. It is known (see Lemma 2.1) that we can build a net with cardinality \( S_i \) satisfying

\[ S_i \leq \left(\frac{Cb}{\delta}\right) 2^{N_i+1}. \]

Third, for each \( u_j^i(x_i) \) choose an \( v_{s(i,j)}(x_i) \), \( s(i,j) \in [1, S_i] \) from the corresponding \( \delta \)-net such that

\[ \|u_j^i(x_i) - v_{s(i,j)}(x_i)\|_2 \leq \delta. \]

Then

\[ \| \prod_{i=1}^{d} u_j^i(x_i) - \prod_{i=1}^{d} v_{s(i,j)}(x_i)\|_2 \leq db^{d-1}\delta \]

and

\[ \| \sum_{j=1}^{n} \prod_{i=1}^{d} u_j^i(x_i) - \sum_{j=1}^{n} \prod_{i=1}^{d} v_{s(i,j)}(x_i)\|_2 \leq ndb^{d-1}\delta. \]

The total number of functions \( \sum_{j=1}^{n} \prod_{i=1}^{d} v_{s(i,j)}(x_i) \) when \( v_{s(i,j)}(x_i) \) are taken from sets of cardinalities \( S_i, i = 1, \ldots, d \), does not exceed

\[ \left( \prod_{i=1}^{d} S_i \right)^n \leq \left(\frac{Cb}{\delta}\right) \sqrt[2n]{\sum_{i=1}^{d} (2N_i+1)}. \]

Specifying \( \delta = \frac{\epsilon}{ndb^{d-1}} \) we obtain

\[ \left(\frac{Cb}{\delta}\right)^{n \sum_{i=1}^{d} (2N_i+1)} \leq n^n \sqrt[2n]{\sum_{i=1}^{d} (2N_i+1)} \left(\frac{C(b, d)}{\epsilon}\right)^{n \sum_{i=1}^{d} (2N_i+1)} \]

which completes the proof. \( \square \)
We are interested in the lower bounds for the quantities $\Theta^b_M(f)_X$ and $\Theta^b_M(F)_X$ defined in the Introduction.

**Theorem 2.1.** Let $N_1 = \cdots = N_d = N$. There is $c(b, d) > 0$ such that for any $M$ satisfying $M \ln M \leq c(b, d)N^{d-1}$ there exists an $f \in T(N)_\infty$ with the property: for any $u^j_i(x_i)$, $\|u^j_i\|_1 \leq b$, we have

$$\|f(x) - \sum_{j=1}^M \prod_{i=1}^d u^j_i(x_i)\|_1 \geq C(b, d) > 0.$$  

**Proof.** The proof repeats the proof of Theorem 1.1 from [12]. We use notations from [12]. Denoting

$$\epsilon := \Theta^b_M(T(N)_\infty)_1$$

we prove, using Lemma 2.2, in the same way as in [12] the following bound

$$N_{2\epsilon}(K_N(T(N)_\infty)_1 \leq C_1(b, d)N^{d}(C_2(b, d)/\epsilon)^C_3(d)N^{M\ln M}, N > 0. \quad (2.1)$$

Lemma 1.2 from [12] gives the lower bound

$$N_{2\epsilon}(K_N(T(N)_\infty)_1 \geq (C(d)/\epsilon)^{N^d}, N > 0. \quad (2.2)$$

Comparing (2.1) and (2.2) we complete the proof of Theorem 2.1.  

**Corollary 2.1.** Let $N_1 = \cdots = N_d = N$. There is $c(b, d) > 0$ such that for any $M$ satisfying $M \ln M \leq c(b, d)N^{d-1}$ we have

$$\Theta^b_M(T(N)_\infty)_1 \geq C(b, d) > 0.$$ 

**Proof.** By the definition of $\Theta^b_M(f)_1$ for all $f \in T(N)_\infty$ we can only use $u^j_i$ satisfying the condition

$$\|u^j_i\|_1 \leq b\|f\|_1^{1/d} \leq b\|f\|_\infty^{1/d} \leq b.$$ 

Therefore, Theorem 2.1 implies Corollary 2.1.  

**Corollary 2.2.** One has

$$\Theta^b_M(W^r_\infty)_{L_1} \gg (M \ln M)^{-\frac{rd}{d-1}}.$$ 

**Proof.** By the Bernstein inequality

$$CN^{-rd}T(N)_\infty \subset W^r_\infty.$$ 

By Theorem 2.1 with $N \simeq (M \ln M)^{\frac{1}{d-1}}$ we obtain the required bound.  

11
3 Upper bounds. Proof of Theorem 1.1

We define the system $\mathcal{U} := \{U_I\}$ in the univariate case. Denote

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

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

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

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

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

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

and

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

Denote

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

and

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

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

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

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

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

$$U_I(x) := \prod_{j=1}^d U_{I_j}(x_j).$$
It is known (see [19]) that $U^d$ is an unconditional basis for $L_p$, $1 < p < \infty$.

We use the notations for $f \in L_1$

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

and for $s = (s_1, \ldots, s_d) \in \mathbb{N}_0^d$

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

where

$$\rho(s) := \{k = (k_1, \ldots, k_d) \in \mathbb{Z}^d : [2^{s_j}-1] \leq |k_j| < 2^{s_j}, j = 1, \ldots, d\}.$$

The convergence

$$\lim_{\min_j \mu_j \to \infty} \|f - \sum_{s_j \leq \mu_j, j=1, \ldots, d} \delta_s(f)\|_p = 0, \quad 1 < p < \infty, \quad (3.1)$$

and the Littlewood-Paley inequalities

$$\|f\|_p \lesssim \|(\sum_s |\delta_s(f)|^2)^{1/2}\|_p, \quad 1 < p < \infty, \quad (3.2)$$

are well-known.

We now proceed to the key lemma of this section. For any nonnegative integer number $n$ we denote $\bar{n} := \max(n, 1)$

**Lemma 3.1.** Let $f \in T(\mathbb{N})$. Denote $v(N) := \prod_{j=1}^d \tilde{N}_j$. Then for $2 \leq p < \infty$ one has

$$\Theta_M(f)_p \ll v(N)^{1-\frac{1}{2}}(\bar{M})\|f\|_2.$$  

**Proof.** The proof is by induction. In the case $d = 2$ it follows from Lemma 2.2 of [11]. Let $d > 2$. Assume $N_j = \min_i N_i$. Represent

$$f = \frac{1}{2N_j+1} \sum_{k=0}^{2N_j} D_{N_j}(x_j - x_j^k)\psi_k(x^j),$$

where $D_N(t)$ is the univariate Dirichlet kernel, $x_j^k = \frac{2\pi k}{2N_j+1}$, and $\psi_k(x^j) = f(x_1, \ldots, x_{j-1}, x_j^k, x_{j+1}, \ldots, x_d)$. Then it is well known that

$$\|f\|^2 = \frac{1}{2N_j+1} \sum_{k=0}^{2N_j} \|\psi_k(x^j)\|^2_2.$$  

13
By the induction assumption we obtain for \( m = \sum_k m_k \)

\[
\Theta_m(f)_p \ll \frac{1}{2N_j + 1} \sum_{k=0}^{2N_j} \Theta_{m_k}(\psi_k)_p
\]

\[
\ll \left( \prod_{i \neq j} (2N_i + 1) \right)^{(1 - \frac{1}{d+1})p} (2N_j + 1)^{-1} \sum_{k=0}^{2N_j} ((\bar{m}_k)^{-1}\|\psi_k\|_2)_p.
\]

Define

\[
m_k := \left[ \frac{\|\psi_k\|_2 M}{\|f\|_2 2N_j + 1} \right].
\]

Then

\[
\sum_{k=0}^{2N_j} m_k \leq \frac{M}{(2N_j + 1)\|f\|_2} (2N_j + 1)^{1/2} \left( \sum_{k=0}^{2N_j} \|\psi_k\|_2^2 \right)^{1/2} = M.
\]

We continue

\[
\Theta_M(f)_p \ll \left( \prod_{i \neq j} (2N_i + 1) \right)^{(1 - \frac{1}{d+1})} (2N_j + 1)^{-1/p} \left( \sum_{k=0}^{2N_j} ((2N_j + 1)\|f\|_2 M^{-1})^p \right)^{1/p}
\]

\[
= \left( \prod_{i \neq j} (2N_i + 1) \right)^{(1 - \frac{1}{d+1})} (2N_j + 1) M^{-1}\|f\|_2.
\]

By our choice of \( N_j \) we have

\[
\left( \prod_{i \neq j} (2N_i + 1) \right)^{\frac{d-2}{d+1}} (2N_j + 1) \leq \left( \prod_{i=1}^{d} (2N_i + 1) \right)^{\frac{d-1}{d+1}},
\]

which follows from

\[
(2N_j + 1)^{1/d} \leq \left( \prod_{i \neq j} (2N_i + 1) \right)^{\frac{1}{d(d-1)}}.
\]

This completes the proof of Lemma 3.1.
Remark 3.1. It is clear that the approximating functions $u^i_j(x_i)$ in Lemma 3.1 can be chosen from $T(N_i)$.

Proof of Theorem 1.1. We consider the following class of functions which is equivalent to the class of functions with bounded mixed derivatives in $L^2$:

$$W^r_A := \{ f : \sum_s 2^{2r\|s\|_1} \|\delta_s(f)\|_2^2 \leq A^2 \}.$$  

For $\|s\|_1 \leq n$ set $m_s := 2^{|s|_1(d-1)/d}$ such that for any $t \in T(\rho(s))$, $\Theta_{m_s}(t) = 0$. For $\|s\|_1 > n$ set

$$m_s := \left[2^{(n-r(\|s\|_1-n))(d-1)/d}\right]$$

with $\kappa > 0$ small enough to satisfy $r > \frac{d-1}{d} + \kappa$. Then

$$M_1 := \sum_{\|s\|_1 \leq n} m_s \asymp 2^{n(d-1)/d} n^{d-1}$$

and

$$M_2 := \sum_{\|s\|_1 > n} m_s \asymp 2^{n(d-1)/d} n^{d-1}.$$

By Lemma 3.1 and Remark 3.1 we obtain for $M := M_1 + M_2$

$$\Theta_M(f)_p \ll \left( \sum_{\|s\|_1 > n} \Theta_{m_s}(\delta_s(f))_p^2 \right)^{1/2}$$

$$\ll \left( \sum_{\|s\|_1 > n} (2-r\|s\|_1) 2^{\|s\|_1(d-1)/d} (\bar{m}_s)^{-1} \|\delta_s(f)\|_2 2^{r\|s\|_1})^2 \right)^{1/2}$$

$$\ll 2^{n(-r+(d-1)/d-(d-1)/d)} A \ll 2^{-rn} \ll \left( \frac{M}{(\log M)^{d-1}} \right)^{-\frac{rd}{d-1}}.$$

4 Constructive upper bounds

In this section we discuss two algorithms for construction of good multilinear approximations. As in Section 3 we concentrate on the case $2 \leq p < \infty$. Our constructive upper bounds are not as good as the corresponding upper bounds for best approximations from Section 3. We begin with two main lemmas.
Lemma 4.1. Suppose that $f \in T(N)$. Denote $v(N) := \prod_{j=1}^{d} \tilde{N}_j$. Then for $1 \leq q \leq p \leq \infty$

$$\Theta_m(f)_p \ll v(N)^{\beta}(\tilde{m})^{-\beta}\|f\|_q, \quad \beta := \frac{1}{q} - \frac{1}{p}, \quad \tilde{m} := \max(1, m). \quad (4.1)$$

The bound (4.1) is realized by a simple greedy-type algorithm.

Proof. In the case $1 \leq q \leq p \leq 2$, $d = 2$, this lemma follows from Lemma 1.1 of [11]. That proof from [11] works in the general case $1 \leq q \leq p \leq \infty$, $d \geq 2$. We will give a sketch of this proof to illustrate the algorithm used in the construction of the approximant. Let $P(N)$ denote the set of points $z^h = (z_1^{h_1}, \ldots, z_d^{h_d})$, $h = (h_1, \ldots, h_d)$ such that $z_j^{h_j} := \frac{\pi h_j}{4N_j}$, $h_j = 0, 1, \ldots, 8\tilde{N}_j - 1$, $j = 1, \ldots, d$. Denote by $V_n(t)$ the univariate de la Vallée Poussin kernel of order $2n - 1$ for $n \geq 1$ and $V_0(t) = 1$. Define the multivariate de la Vallée Poussin kernel as follows

$$V_N(z) := \prod_{j=1}^{d} V_{N_j}(z_j), \quad N = (N_1, \ldots, N_d).$$

Then it is well known that any $f \in T(N)$ has the representation

$$f(z) = \left(\prod_{j=1}^{d}(8\tilde{N}_j)\right)^{-1} \sum_{z^h \in P(N)} f(z^h)V(z - z^h). \quad (4.2)$$

We have the following equivalence relation.

Theorem 4.1. For all $1 \leq q \leq \infty$ and for $f \in T(N)$

$$\|f\|_q \asymp v(N)^{-1/q} \left(\sum_{z^h \in P(N)} |f(z^h)|^q\right)^{1/q}.$$

Theorem 4.1 was used in [11] (see Theorem 1). Theorem 4.1 is the Marcinkiewicz-Zygmund theorem in the case $d = 1$ (see [21], Vol II, pp. 28–33). It is pointed out in [11] that in the general case ($d > 1$) Theorem 4.1
is an immediate corollary of the one-dimensional theorem. We note that Theorem 4.1 and Lemma 4.2 (see below) hold with $P(N)$ replaced by a smaller net of points

$$P'(N) := \{ z^h : z^h_j := \frac{\pi h_j}{2N_j}, \quad h_j = 0, 1, \ldots, 4\tilde{N}_j - 1, \quad j = 1, \ldots, d \}.$$ 

Also, Theorem 4.1 and Lemma 4.2 hold in the case of vector $L_p$. Theorem 4.1 in the case of vector $L_p$ is not as straightforward corollary of the univariate case as in the case of scalar $p$. The reader can find detailed proofs of the corresponding results in [13] Chapter 2, Theorem 2.4 and Lemma 2.6.

We have the following inequality (see [11], Lemma 2).

**Lemma 4.2.** For arbitrary numbers $A(z^h)$

$$\left\| \sum_{z^h \in P(N)} A(z^h) V_N(z, z^h) \right\|_p \ll v(N)^{1-1/p} \left( \sum_h |A(z^h)|^p \right)^{1/p}.$$ 

**Proof.** For completeness we give a proof of this known lemma. Denote $\nu(N) := |P(N)|$. Let $V$ be an operator on $\ell_p(\nu)$ defined as follows:

$$V(A) = \sum_{z^h \in P(N)} A(z^h) V_N(z, z^h).$$

It is obvious from the known properties of the univariate de la Vallée Poussin kernels that

$$\|V\|_{\ell_p(\nu) \rightarrow L_1} \leq 3^d. \quad (4.3)$$

Using the known estimate

$$|V_n(x)| \leq C \min(n, (nx^2)^{-1})$$

it is not hard to prove that

$$\|V\|_{\ell_p(\nu) \rightarrow L_\infty} \leq C(d) \nu(N). \quad (4.4)$$

From the relations (4.3) and (4.4), using the Riesz-Torin theorem we find

$$\|V\|_{\ell_p(\nu) \rightarrow L_p} \leq C \nu(N)^{1-1/p},$$

which implies the conclusion of the lemma. □
We now complete the proof of Lemma 4.1. Using representation (4.2) we choose a set \( G(m) \) of \( m \) points \( z^h \) with the largest \( |f(z^h)| \). Then we use Theorem 4.1, Lemma 4.2 and the following known lemma (see, for instance, [9]).

**Lemma 4.3.** Let \( b_1 \geq b_2 \geq \ldots b_n \geq 0 \), \( 1 \leq q \leq p \leq \infty \) and

\[
\sum_{j=1}^{n} b_j^q \leq A^q.
\]

Then for any \( m \leq n \) we have (with natural modification for \( p = \infty \))

\[
\left( \sum_{j=m}^{n} b_j^p \right)^{1/p} \leq m^{1/p-1/q} A.
\]

It gives us

\[
\|f(z) - \sum_{z_h \in G(m)} f(z^h) \mathcal{V}_N(z - z^h)\|_p \ll v(N)^\beta m^{-\beta} \|f\|_q.
\]

\[\square\]

The algorithm used above in the proof of Lemma 4.1 is a simple greedy-type algorithm which uses a special dictionary \( \{\mathcal{V}_N(z - z^h)\}_{z^h \in P(N)} \). We now proceed to a discussion of general greedy-type algorithms which will use the dictionary \( \Pi^d \). We begin with a brief description of greedy approximation methods in Banach spaces. The reader can find a detailed discussion of greedy approximation in the book [17]. Let \( X \) be a Banach space with norm \( \| \cdot \| \). We say that a set of elements (functions) \( \mathcal{D} \) from \( X \) is a symmetric dictionary, if each \( g \in \mathcal{D} \) has norm bounded by one (\( \|g\| \leq 1 \)),

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

and the closure of span \( \mathcal{D} \) is \( X \). We denote the closure (in \( X \)) of the convex hull of \( \mathcal{D} \) by \( A_1(\mathcal{D}) \). In other words \( A_1(\mathcal{D}) \) is the closure of \( \text{conv}(\mathcal{D}) \). We use this notation because it has become a standard notation in relevant greedy approximation literature. For a nonzero element \( f \in X \) we let \( F_f \) denote a norming (peak) functional for \( f \) that is a functional with the following properties

\[ \|F_f\| = 1, \quad F_f(f) = \|f\|. \]
The existence of such a functional is guaranteed by the Hahn-Banach theorem. The norming functional $F_f$ is a linear functional (in other words is an element of the dual to $X$ space $X^*$) which can be explicitly written in some cases. In a Hilbert space $F_f$ can be identified with $f \| f \|^{-1}$. In the real $L_p$, $1 < p < \infty$, it can be identified with $f \| f \|^{p-2} \| f \|^{-p}$. We describe a typical greedy algorithm which uses a norming functional. We call this family of algorithms dual greedy algorithms. Let $\tau := \{ t_k \}_{k=1}^\infty$ be a given sequence of nonnegative numbers $t_k \leq 1$, $k = 1, \ldots$. We call this sequence weakness sequence. We first define the Weak Chebyshev Greedy Algorithm (WCGA) (see [15]) that is a generalization for Banach spaces of the Weak Orthogonal Greedy Algorithm.

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

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

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

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 a best approximant to $f$ from $\Phi_m$.

3. Let

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

The index $c$ in the notation refers to Chebyshev. We use the name Chebyshev in this algorithm because at step (2) of the algorithm we use best approximation operator which bears the name of the Chebyshev projection or the Chebyshev operator. In the case of Hilbert space the Chebyshev projection is the orthogonal projection and it is reflected in the name of the algorithm.

For a Banach space $X$ we define the modulus of smoothness

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

A uniformly smooth Banach space is one with the property

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

The following proposition is well-known (see, [17], p.336).
Proposition 4.1. Let $X$ be a uniformly smooth Banach space. Then, for any $x \neq 0$ and $y$ we have

$$F_x(y) = \left( \frac{d}{du} \|x + uy\| \right)(0) = \lim_{u \to 0} (\|x + uy\| - \|x\|)/u.$$ 

Proposition 4.1 shows that in the WCGA we are looking for an element $\varphi_m \in D$ that provides a big derivative of the quantity $\|f_{m-1} + u\varphi_m\|$. Here is one more important greedy algorithm.

Weak Greedy Algorithm with Free Relaxation (WGAFR). Let $\tau := \{t_m\}_{m=1}^\infty$, $t_m \in [0, 1]$, be a weakness sequence. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we have the following inductive definition.

1. $\varphi_m \in D$ is any element satisfying

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

2. Find $w_m$ and $\lambda_m$ such that

$$\|f - ((1 - w_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{\lambda, w} \|f - ((1 - w)G_{m-1} + \lambda\varphi_m)\|$$

and define

$$G_m := (1 - w_m)G_{m-1} + \lambda_m\varphi_m.$$ 

3. Let

$$f_m := f - G_m.$$ 

It is known that both algorithms WCGA and WGAFR converge in any uniformly smooth Banach space under mild conditions on the weakness sequence $\{t_k\}$, for instance, $t_k = t$, $k = 1, 2, \ldots, t > 0$, guarantees such convergence. The following theorem provides rate of convergence (see [17], pp. 347, 353).

Theorem 4.2. Let $X$ be a uniformly smooth Banach space with modulus of smoothness $\varrho(u) \leq \gamma u^q$, $1 < q \leq 2$. Take a number $\epsilon \geq 0$ and two elements $f, f^\epsilon$ from $X$ such that

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/B \in A_1(D),$$

with some number $B = C(f, \epsilon, D, X) > 0$. Then, for both algorithms WCGA and WGAFR we have ($p := q/(q - 1)$)

$$\|f_m\| \leq \max \left(2\epsilon, C(q, \gamma)(B + \epsilon)(1 + \sum_{k=1}^m t_k^p)^{-1/p}\right).$$
Theorem 4.2 completes our brief introduction to the theory of greedy algorithms with respect to dictionaries in Banach spaces. We now show how the above general greedy algorithms can be applied for estimating $\Theta_M(f)_p$. We begin with a lemma.

**Lemma 4.4.** Let $f \in T(N)$. Then for $2 \leq p < \infty$

$$\Theta_m(f)_p \ll v(N)^{1/2 - \frac{1}{2p}}(\bar{m})^{-1/2}\|f\|_2.$$  

The above bound is realized by the WCGA and the WGAFR with $\tau = \{t\}$.

**Proof.** Assume $N_j = \max_i N_i$. Represent

$$f(x) = \sum_{k \in Q(N)} \hat{f}(k)e^{i(k,x)} = \sum_{k^j \in Q(N^j)} u_{k^j}(x^j)e^{i(k^j,x^j)},$$

where $k^j := (k_1, \ldots, k_{j-1}, k_{j+1}, \ldots, k_d)$ and $x^j := (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)$,

$$u_{k^j}(x^j) := \sum_{|k_j| \leq N_j} \hat{f}(k)e^{ik_j x_j},$$

$$Q(N) := \{k = (k_1, \ldots, k_d) : |k_i| \leq N_i, i = 1, \ldots, d\}.$$ 

Denote

$$\psi_{k^j}(x) := u_{k^j}(x^j)e^{i(k^j,x^j)}.$$ 

It is clear that $\psi_{k^j} \in \Pi^d$. We now bound

$$\sum_{k^j \in Q(N^j)} \|\psi_{k^j}\|_p = \sum_{k^j \in Q(N^j)} \|u_{k^j}\|_p \leq CN_j^{1/2 - 1/p} \sum_{k^j \in Q(N^j)} \|u_{k^j}\|_2 \leq CN_j^{1/2 - 1/p}v(N^j)^{1/2} \left( \sum_{k^j \in Q(N^j)} \|u_{k^j}\|^2_2 \right)^{1/2} = CN_j^{-1/p}v(N)^{1/2}\|f\|_2 \leq Cv(N)^{1/2 - \frac{1}{2p}}\|f\|_2.$$ 

Therefore,

$$f/B \in A_1(\Pi^d), \quad B = Cv(N)^{1/2 - \frac{1}{2p}}\|f\|_2. \quad (4.5)$$

We proved (4.5) for complex trigonometric polynomials. Clearly, the same proof works for real trigonometric polynomials from $T(N)$. We switch to
real polynomials because the theory of greedy approximation, in particular the theory for the WCGA and WGAFR, is developed in real Banach spaces. We apply Theorem 4.2. It is known that the \(L_p\) space with \(2 \leq p < \infty\) is a uniformly smooth Banach space with modulus of smoothness \(\rho(u) \leq \gamma u^2\). Applying Theorem 4.2 with \(\epsilon = 0\) and \(\tau = \{t\}\) we obtain the required bound.

Remark 4.1. It is clear that the approximant in Lemma 4.4 and the approximant in Lemma 4.1 in case \(1 < p < \infty\) can be taken from \(T(N)\).

Theorem 4.3. Let \(2 \leq p < \infty\). Denote \(\beta := 1/2 - 1/p\). Then there is a constructive way provided by Lemma 4.4 to obtain the bound

\[
\Theta_M(W_p^r) \ll \left( \frac{M}{(\ln M)^{d-1}} \right)^{-\frac{p}{d} + \frac{\beta}{d-1}}, \quad r > \frac{1}{2} - \frac{1}{pd}.
\]

Proof. For \(\|s\|_1 \leq n\) set \(m_s \asymp 2^{|s|_1(d-1)/d}\) such that for any \(t \in T(\rho(s))\), \(\Theta_{m_s}(t)_p = 0\). For \(\|s\|_1 > n\) set

\[
m_s := \left[2^{(n-\kappa(\|s\|_1-n))(d-1)/d}\right]
\]

with \(\kappa > 0\) small enough to satisfy \(r > \frac{1}{2} - \frac{1}{pd} + \kappa\). Then

\[
M_1 := \sum_{\|s\|_1 \leq n} m_s \asymp 2^{n(d-1)/dn^{d-1}}
\]

and

\[
M_2 := \sum_{\|s\|_1 > n} m_s \asymp 2^{n(d-1)/dn^{d-1}}.
\]

By Lemma 4.4 we obtain for \(M := M_1 + M_2\)

\[
\Theta_M(f)_p \ll \left( \sum_{\|s\|_1 > n} \Theta_{m_s}(\delta_s(f))_p^2 \right)^{1/2}
\]

\[
\ll \left( \sum_{\|s\|_1 > n} \left( 2^{-r\|s\|_1}2^{(1/2 - \frac{1}{pd})\|s\|_1} (\tilde{m}_s)^{-1/2}\|\tilde{\delta}_s(f)\|_2 2^{r\|s\|_1} \right)^2 \right)^{1/2}
\]

\[
\ll 2^{n(-r + \left(\frac{1}{2} - \frac{1}{pd}\right) - \frac{1}{2}(d-1)/d)} = 2^{n(-r + \beta/d)}.
\]

\[\square\]
Theorem 4.4. Let $2 \leq p < \infty$. Denote $\beta := 1/2 - 1/p$. Then there is a constructive way provided by Lemma 4.1 to obtain the bound

$$\Theta_M(W_2^r)^p \ll \left( \frac{M}{(\ln M)^{d-1}} \right)^{-\frac{rd}{d-1} + \frac{\beta}{d-1}}, \quad r > \beta.$$ 

Proof. The proof of this theorem is similar to the proof of Theorem 4.3. We use the same notations as above. Then by Lemma 4.1 we obtain for $M := M_1 + M_2$

$$\Theta_M(f)^p \ll \left( \sum_{\|s\|_1 > n} \Theta_m(\delta_s(f))^2 \right)^{1/2}$$

$$\ll \left( \sum_{\|s\|_1 > n} (2^{-r\|s\|_1} 2^{\beta\|s\|_1} (\bar{m}_s)^{-\beta} \|\delta_s(f)\|_2 2^{r\|s\|_1})^2 \right)^{1/2}$$

$$\ll 2^{n(-r+\beta-\beta(d-1)/d)} = 2^{n(-r+\beta/d)}.$$

Some improvements. In this subsection we explain how Theorems 1.1, 4.3, and 4.4 can be slightly improved by changing the subdivision of the set of $s$. The rest of the proofs including the use of Lemmas 3.1, 4.1, and 4.4 is the same. For a nonnegative $(d - 1)$-dimensional integer vector $w$ define

$$S(w,j) := \{ s : s_j = \max_i s_i, \ s^j = w \}, \quad j = 1, \ldots, d.$$

For the set $\cup_{s \in S(w,j)} \rho(s)$ we set $m_w \asymp 2^{\|w\|_1}$ in such a way that for any $t \in T(\cup_{s \in S(w,j)} \rho(s))$ we have $\Theta_{m_w}(t)^p = 0$. Then

$$M'_1 := \sum_{j=1}^d \sum_{\|w\|_1 \leq n(d-1)/d} m_w \leq 2^{n(d-1)/d} n^{d-2}.$$ 

For the remaining set $S^c$ of $s$ we have

$$S^c := \{ s : s \notin \cup_{j=1}^d \cup_{\|w\|_1 \leq n(d-1)/d} S(w,j) \} = \{ s : \min_j \|s^j\|_1 > n(d-1)/d \}.$$

For $s \in S^c$ we define as above

$$m_s := \left[ 2^{(n-\kappa(\|s\|_1-n))(d-1)/d} \right]$$
with $\kappa > 0$ small enough. Then

$$M_2' := \sum_{s \in S^c} m_s \leq \sum_{\|s^1\|_1 \geq n(d-1)/d} \sum_{s_1 \geq \|s^1\|_1/(d-1)} m_s \ll \sum_{\|s^1\|_1 \geq n(d-1)/d} 2^{(n-\kappa(\|s^1\|_1 + \|s^1\|_1/(d-1)-n))(d-1)/d} \ll 2^{n(d-1)/d} n^{d-2}.$$ 

The rest of the proofs is the same as in Theorems 1.1, 4.3, and 4.4. We only need to notice that for $s \in S^c$ we have

$$\|s\|_1 = \sum_{j=1}^d \|s^j\|_1 > n.$$ 

The above argument gives us the following slightly stronger versions of Theorems 1.1, 4.3, and 4.4.

**Theorem 4.5.** Let $2 \leq p < \infty$ and $r > (d-1)/d$. Then

$$\Theta_M(W^r_{2^p}) \ll \left( \frac{M}{(\log M)^{d-2}} \right)^{-\frac{\alpha d}{d-1}}.$$

**Theorem 4.6.** Let $2 \leq p < \infty$. Denote $\beta := 1/2 - 1/p$. Then there is a constructive way provided by Lemma 4.4 to obtain the bound

$$\Theta_M(W^r_{2^p}) \ll \left( \frac{M}{(\ln M)^{d-2}} \right)^{-\frac{\alpha d + \beta d-1}{d-1}}, \quad r > \frac{1}{2} - \frac{1}{pd}.$$

**Theorem 4.7.** Let $2 \leq p < \infty$. Denote $\beta := 1/2 - 1/p$. Then there is a constructive way provided by Lemma 4.1 to obtain the bound

$$\Theta_M(W^r_{2^p}) \ll \left( \frac{M}{(\ln M)^{d-2}} \right)^{-\frac{\alpha d + \beta d-1}{d-1}}, \quad r > \beta.$$
References

[1] W. Dahmen, R. DeVore, L. Grasedyck, and E. Süli, Tensor-sparsity of solutions to high-dimensional elliptic Partial Differential Equations, Manuscript, July 23, 2014.

[2] G. Davis, S. Mallat, and M. Avellaneda, Adaptive greedy approximations, Constructive Approximation 13 (1997), 57–98.

[3] R.A. DeVore, Nonlinear approximation, Acta Numerica, 7 (1998), 51–150.

[4] W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus, Springer-Verlag, Heidelberg, 2012.

[5] B.S. Kashin, Widths of certain finite-dimensional sets and classes of smooth functions, Izv. AN SSSR, 41 (1977), 334–351; English transl. in Math. Izv. 11 (1977).

[6] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Math. Ann., 63 (1907), 433-476.

[7] R. Schneider and A. Uschmajew, Approximation rates for the hierarchical tensor format in periodic Sobolev spaces, J. Complexity 30 (2014), 56–71.

[8] S. Shalev-Shwartz, N. Srebro, and T. Zhang, Trading accuracy for sparsity in optimization problems with sparsity constrains, SIAM Journal on Optimization, 20(6) (2010), 2807–2832.

[9] V.N. Temlyakov, Approximation of periodic functions of several variables by combinations of functions depending on fewer variables, Trudy Mat. Inst. Steklov, 173 (1986), 243–252; English transl. in Proc. Steklov Inst. Math. 4 (1987), 275–293.

[10] V.N. Temlyakov, Estimates of the best bilinear approximations of functions of two variables and some of their applications, Mat. Sb., 134 (1987), 93–107; English transl. in Math. USSR-Sb 62 (1989), 95–109.
[11] V.N. Temlyakov, Estimates of best bilinear approximations of periodic functions, Trudy Mat. Inst. Steklov, 181 (1988), 250–267; English transl. in Proc. Steklov Inst. of Math. 4 (1989), 275–293.

[12] V.N. Temlyakov, Estimates of Best Bilinear Approximations of Functions and Approximation Numbers of Integral Operators, Matem. Zametki, 51 (1992), 125–134; English translation in Math. Notes, 51 (1992), 510–517.

[13] V.N. Temlyakov, Approximation of Periodic Functions, Nova Science Publishers, Inc., 1993.

[14] V.N. Temlyakov, Greedy Algorithms with Regard to Multivariate Systems with Special Structure, Constr. Approx., 16 (2000), 399–425.

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

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

[17] V.N. Temlyakov, Greedy approximation, Cambridge University Press, 2011.

[18] V. Tikhomirov, Widths of sets in function spaces and the theory of best approximations, Uspekhi Mat. Nauk, 15(3), (93), (1960), 81-120. English translation in Russian Math. Survey, 15, 1960.

[19] P. Wojtaszczyk, On unconditional polynomial bases in $L_p$ and Bergman spaces, Constr. Approx., 13 (1997), 1–15.

[20] T. Zhang, Sequential greedy approximation for certain convex optimization problems, IEEE Transactions on Information Theory, 49(3) (2003), 682–691.

[21] A. Zygmund, Trigonometric series, Cambridge Univ. Press, 1959.