Constructive sparse trigonometric approximation and other problems for functions with mixed smoothness

V. N. Temlyakov

Abstract. Our main interest in this paper is to study some approximation problems for classes of functions with mixed smoothness. We use a technique based on a combination of results from hyperbolic cross approximation, which were obtained in 1980s–1990s, and recent results on greedy approximation to obtain sharp estimates for best \(m\)-term approximation with respect to the trigonometric system. We give some observations on the numerical integration and approximate recovery of functions with mixed smoothness. We prove lower bounds, which show that one cannot improve the accuracy of sparse grids methods with \(\asymp 2^n n^{d-1}\) points in the grid by adding \(2^n\) arbitrary points. In the case of numerical integration these lower bounds provide the best available lower bounds for optimal cubature formulae and for sparse grids based cubature formulae.

Bibliography: 31 titles.

Keywords: nonlinear approximation, sparse approximation, trigonometric system, constructive methods.

§1. Introduction

Sparse approximation with respect to dictionaries is a very important topic in high-dimensional approximation. The main motivation for the study of sparse approximation is that it gives a good approximation to many real world signals. Sparse approximation automatically implies a need for nonlinear approximation, in particular, for greedy approximation.

We give a brief description of a sparse approximation problem and present a discussion of the results we obtain and their relation to previous work. In a general setting we are working in a Banach space \(X\) with a redundant system of elements (dictionary) \(\mathcal{D}\). There is a solid justification of the importance of a Banach space setting in numerical analysis in general and in sparse approximation in particular (see the preface in [1], for instance). Let \(X\) be a real Banach space with norm \(\|\cdot\| := \|\cdot\|_X\). We say that a set of elements (functions) \(\mathcal{D}\) from \(X\) is a dictionary if each \(g \in \mathcal{D}\) has norm one (\(\|g\| = 1\)), and the closure of span \(\mathcal{D}\) is \(X\). A symmetrized dictionary is \(\mathcal{D}^\pm := \{\pm g: g \in \mathcal{D}\}\). For a nonzero element \(g \in X\) we let

This research was supported by the NSF (grant no. DMS-1160841).

AMS 2010 Mathematics Subject Classification. Primary 41A60, 42A10, 46E35; Secondary 41A65.

© 2015 Russian Academy of Sciences (DoM), London Mathematical Society, Turpion Ltd.
$F_g$ denote a norming (peak) functional for $g$:

$$\|F_g\|_{X^*} = 1, \quad F_g(g) = \|g\|_X.$$  

The existence of this functional is guaranteed by the Hahn-Banach theorem.

An element (function, signal) $s \in X$ is said to be $m$-sparse with respect to $\mathcal{D}$ if it has a representation $s = \sum_{i=1}^{m} c_i g_i$, $g_i \in \mathcal{D}$, $i = 1, \ldots, m$. The set of all $m$-sparse elements is denoted by $\Sigma_m(\mathcal{D})$. For a given element $f$ we introduce the error of the best $m$-term approximation

$$\sigma_m(f, \mathcal{D})_X := \inf_{s \in \Sigma_m(\mathcal{D})} \|f - s\|_X.$$  

For a function class $\mathcal{W}$ we set

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

Let $t \in (0, 1]$ be a given nonnegative number. We define (see [2]) the Weak Chebyshev Greedy Algorithm (WCGA), which is a generalization to Banach spaces of the Weak Orthogonal Greedy Algorithm defined and was studied in [3] (see also [1]).

**Weak Chebyshev Greedy Algorithm (WCGA).** Set $f_0 := f_0^{c,t} := f$. Then for each $m \geq 1$ we define by induction:

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

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|;$$

2) $\Phi_m := \Phi_m^t := \text{span}\{\varphi_j\}_{j=1}^m$

and we define $G_m := G_m^{c,t}$ to be the best approximant to $f$ in $\Phi_m$;

3) set

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

We proved in [4] that the Weak Chebyshev Greedy Algorithm (WCGA) is very good for $m$-term approximation with respect to a special class of dictionaries, in particular, for the trigonometric system. The trigonometric system is a classical system that is known to be difficult to study. In [4] amongst other problems we studied the problem of nonlinear sparse approximation with respect to it. Let $\mathcal{RT}$ denote the real trigonometric system $1, \sin 2\pi x, \cos 2\pi x, \ldots$ on $[0, 1]$ and let $\mathcal{RT}_p$ be the version of it normalized in $L_p([0, 1])$. Set $\mathcal{RT}_p^d := \mathcal{RT}_p \times \cdots \times \mathcal{RT}_p$ to be the $d$-variate trigonometric system. We need to consider the real trigonometric system because the algorithm WCGA has been well studied in a real Banach space. We proved the following Lebesgue-type inequality for the WCGA in [4].

**Theorem 1.1.** Let $\mathcal{D}$ be the real $d$-variate trigonometric system normalized in $L_p$, $2 \leq p < \infty$. Then for any $f \in L_p$ the WCGA with weakness parameter $t$ gives

$$\|f_{C(t,p,d)\ln(m+1)}\|_p \leq C \sigma_m(f, \mathcal{D})_p.$$  

(1.1)
The above Lebesgue-type inequality guarantees that the WCGA works very well for each individual function \( f \). To complement this inequality, we would like to obtain results which relate the rate of decay of \( \sigma_m(f, T^d)_p \) to some smoothness type properties of \( f \). It is the main goal of this paper. We measure smoothness in terms of mixed derivative and mixed differences. We note that the function classes with a bounded mixed derivative are not only interesting and challenging objects for approximation theory, they are important in numerical computations. Griebel and his group use approximation methods designed for these classes in elliptic variational problems. Yserentant’s recent work on new regularity models for the Schrödinger equation shows that the eigenfunctions of the electronic Schrödinger operator have a certain mixed smoothness similar to having a bounded mixed derivative. This makes approximation techniques developed for classes of functions with a bounded mixed derivative an appropriate choice for the numerical treatment of the Schrödinger equation.

Sparse trigonometric approximation of periodic functions began with Stechkin’s paper [5], where it was used in the criterion for absolute convergence of trigonometric series. Ismagilov [6] found nontrivial estimates for \( m \)-term approximation of functions with singularities of the type \(|x|\) and gave interesting and important applications to the widths of Sobolev classes. He used a deterministic method based on number theoretical constructions. His method was developed by Maiorov [7], who used a method based on Gaussian sums. Further powerful results were obtained in [8] with the help of a nonconstructive result from the theory of finite dimensional Banach spaces due to Gluskin [9]. Another powerful nonconstructive method, based on a probabilistic argument, was used by Makovoz [10] and Belinskii [11], [12]. Different methods were developed in [13]–[16] to prove lower bounds for function classes. It was discovered in [17] and [18] that greedy algorithms can be used for constructive \( m \)-term approximation with respect to the trigonometric system. In this paper we show how greedy algorithms can be used to prove optimal or the best available upper bounds for \( m \)-term approximation of classes of functions with mixed smoothness. It is a simple and powerful method for proving upper bounds. However, we do not know how to use it for small smoothness. The reader can find a detailed study of \( m \)-term approximation of classes of functions with mixed smoothness, including small smoothness, in A.S. Romanyuk’s paper [19]. We note that in the case \( 2 < p < \infty \) the upper bounds in [19] are not constructive.

We begin with some notation. Let \( s = (s_1, \ldots, s_d) \) be a vector whose coordinates are nonnegative integers. Then

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

\[
Q_n := \bigcup_{\|s\|_1 \leq n} \rho(s) \quad \text{is a step hyperbolic cross},
\]

\[
\Gamma(N) := \left\{ \mathbf{k} \in \mathbb{Z}^d : \prod_{j=1}^d \max(|k_j|, 1) \leq N \right\} \quad \text{is a hyperbolic cross}.
\]

For \( f \in L_1(\mathbb{T}^d) \)

\[
\delta_s(f, x) := \sum_{\mathbf{k} \in \rho(s)} \hat{f}(\mathbf{k}) e^{i<\mathbf{k}, \mathbf{x}>}. \]
Constructive sparse approximation 1631

If \( G \) is a finite set of points in \( \mathbb{Z}^d \), then we set

\[
\mathcal{F}(G) := \left\{ t : t(x) = \sum_{k \in G} c_k e^{i(k,x)} \right\}.
\]

For the sake of simplicity we write \( \mathcal{F}(\Gamma(N)) = \mathcal{F}(N) \).

We study some approximation problems for classes of functions with mixed smoothness. We define these classes now. We begin with the case of univariate periodic functions. For \( r > 0 \) let

\[
F_r(x) := 1 + 2 \sum_{k=1}^{\infty} k^{-r} \cos \left( kx - \frac{r\pi}{2} \right),
\]

(1.2)

\[
W^r_p := \{ f : f = \varphi \ast F_r, \|\varphi\|_p \leq 1 \}.
\]

(1.3)

It is well known that for \( r > 1/p \) the class \( W^r_p \) is embedded into the space of continuous functions \( C(\mathbb{T}) \). In the particular case of \( W^1_1 \) we also have embedding into \( C(\mathbb{T}) \).

In the multivariate case for \( x = (x_1, \ldots, x_d) \) denote

\[
F_r(x) := \prod_{j=1}^{d} F_r(x_j), \quad W^r_p := \{ f : f = \varphi \ast F_r, \|\varphi\|_p \leq 1 \}.
\]

For \( f \in W^r_p \) we set \( f^{(r)} := \varphi \) where \( \varphi \) is such that \( f = \varphi \ast F_r \).

The main results in §2 are the following two theorems. We use the notation

\[
\beta := \beta(q,p) := \frac{1}{q} - \frac{1}{p} \quad \text{and} \quad \eta := \eta(q) := \frac{1}{q} - \frac{1}{2).
\]

In the case of the trigonometric system \( \mathcal{T}^d \) we drop it from the notation:

\[
\sigma_m(W)_p := \sigma_m(W, \mathcal{T}^d)_p.
\]

**Theorem 1.2.** The following relations hold:

\[
\sigma_m(W^r_p)_{q,p} \lesssim \begin{cases} 
m^{-r+\beta} (\log m)^{(d-1)(r-2\beta)}, & 1 < q \leq p \leq 2, \quad r > 2\beta, \\
m^{-r+\eta} (\log m)^{(d-1)(r-2\eta)}, & 1 < q \leq 2 \leq p < \infty, \quad r > \frac{1}{q}, \\
m^{-r} (\log m)^{r(d-1)}, & 2 \leq q < \infty, \quad r > \frac{1}{2}, \end{cases}
\]

**Theorem 1.3.** The following relations hold:

\[
\sigma_m(W^r_p)_{\infty} \ll \begin{cases} 
m^{-r+\eta} (\log m)^{(d-1)(r-2\eta)+1/2}, & 1 < q \leq 2, \quad r > \frac{1}{q}, \\
m^{-r} (\log m)^{r(d-1)+1/2}, & 2 \leq q < \infty, \quad r > \frac{1}{2}. \end{cases}
\]
Theorem 1.4. For $p \in (1, \infty)$ and $\mu > 0$ there exist constructive methods $A_m(f, p, \mu)$, which provide an $m$-term approximation for $f \in \mathbf{W}_q^r$ such that

$$\|f - A_m(f, p, \mu)\|_p \leq \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-2\beta)}, & 1 < q \leq p \leq 2, \quad r > 2\beta + \mu, \\ m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)}, & 1 < q \leq 2 < p < \infty, \quad r > \frac{1}{q} + \mu, \\ m^{-r}(\log m)^r(d-1), & 2 < q \leq p < \infty, \quad r > \frac{1}{2} + \mu. \end{cases}$$

An analogue of Theorem 1.3 holds for $p = \infty$. We do not have matching lower bounds for the upper bounds in Theorem 1.3 in the case of approximation in the uniform norm $L_\infty$. In §3 we use known results on the entropy numbers to prove one lower bound in the case of functions of two variables. We note that it is of interest for small smoothness, when $r < 1/2$.

As a direct corollary of Theorems 1.1 and 1.2 we obtain the following result.

Theorem 1.5. Let $p \in [2, \infty)$. Apply the WCGA with weakness parameter $t \in (0, 1]$ to $f \in L_p$ with respect to the real trigonometric system $\mathbb{R} T_p^d$. If $f \in \mathbf{W}_q^r$, then

$$\|f_m\|_p \leq \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+r-\eta}, & 1 < q \leq 2, \quad r > \frac{1}{q}, \\ m^{-r}(\log m)^rd, & 2 < q < \infty, \quad r > \frac{1}{2}. \end{cases}$$

In §4 and §5 we give some observations on numerical integration and approximate recovery of functions with mixed smoothness. We prove some lower bounds which show that we cannot improve the accuracy of sparse grids methods with $\asymp 2^n n^{d-1}$ points in the grid by adding $2^n$ arbitrary points. In the case of numerical integration these lower bounds provide the best available lower bounds for optimal cubature formulae and sparse-grids based cubature formulae.

Our technique is based on a combination of results in hyperbolic cross approximation, which were obtained in the 1980s–1990s, and recent results on greedy approximation. We give some known results from hyperbolic cross approximation theory, which will be used in our analysis. We begin with the problem of estimating $\|f\|_p$ in terms of the array $\{\|\delta_s(f)\|_q\}$. Here and below $p$ and $q$ are scalars such that $1 \leq q, p \leq \infty$. Let an array $\varepsilon = \{\varepsilon_s\}$ be given, where $\varepsilon_s \geq 0$, $s = (s_1, \ldots, s_d)$, and the $s_j$ are nonnegative integers, $j = 1, \ldots, d$. We denote the following sets of functions by $G(\varepsilon, q)$ and $F(\varepsilon, q)$ ($1 \leq q \leq \infty$):

$$G(\varepsilon, q) := \{f \in L_q; \|\delta_s(f)\|_q \leq \varepsilon_s \quad \text{for all } s\},$$

$$F(\varepsilon, q) := \{f \in L_q; \|\delta_s(f)\|_q \geq \varepsilon_s \quad \text{for all } s\}.$$
The next theorem is Theorem 3.3 in [20], Ch. 1. For the special case $q = 2$, which will be used in this paper, see [13] and [20], Ch. 4, Theorem 1.1.

**Theorem 1.6.** The following relations hold:

\[
\begin{align*}
\sup_{f \in G(\varepsilon, q)} \|f\|_p & \asymp \left( \sum_s \varepsilon^{p/q} \|s\|_1 (p/q - 1) \right)^{1/p}, \quad 1 \leq q < p < \infty, \quad (1.4) \\
\inf_{f \in F(\varepsilon, q)} \|f\|_p & \asymp \left( \sum_s \varepsilon^{p/q} \|s\|_1 (p/q - 1) \right)^{1/p}, \quad 1 < p < q \leq \infty, \quad (1.5)
\end{align*}
\]

with constants independent of $\varepsilon$.

We need a corollary of Theorem 1.6 (see [20], Ch. 1, Theorem 2.2), which we formulate as a theorem.

**Theorem 1.7.** Let $1 < q \leq 2$. Then for any $t \in \mathcal{T}(N)$

\[
\|t\|_A := \sum_k |\hat{t}(k)| \leq C(q, d) N^{1/q} (\log N)^{(d-1)(1-1/q)} \|t\|_q.
\]

Let

\[
\Pi(N, d) := \{(a_1, \ldots, a_d) \in \mathbb{R}^d : |a_j| \leq N_j, \ j = 1, \ldots, d\},
\]

where $N_j$ are nonnegative integers and $N := (N_1, \ldots, N_d)$. We set

\[
\mathcal{T}(N, d) := \left\{ t : t = \sum_{k \in \Pi(N,d)} c_k e^{i(k,x)} \right\}.
\]

Then

\[
\dim \mathcal{T}(N, d) = \prod_{j=1}^d (2N_j + 1) =: \vartheta(N).
\]

The following theorem is Theorem 1.1 in [21], Ch. 2 (see, also, [22]).

**Theorem 1.8.** Let $\varepsilon > 0$ and let the subspace $\Psi \subset \mathcal{T}(N, d)$ be such that $\dim \Psi \geq \varepsilon \dim \mathcal{T}(N, d)$. Then there exists $t \in \Psi$ such that

\[
\|t\|_\infty = 1, \quad \|t\|_2 \geq C(\varepsilon, d) > 0.
\]

§ 2. Sparse approximation

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

\[
\lim_{u \to 0} \frac{\rho(u)}{u} = 0.
\]
It is well known (see, for instance, [23], Lemma B.1) that when $X = L^p$, $1 \leq p < \infty$, we have

$$\rho(u) \leq \begin{cases} \frac{u^p}{p} & \text{if } 1 \leq p \leq 2, \\ \frac{(p-1)u^2}{2} & \text{if } 2 \leq p < \infty. \end{cases} \quad (2.1)$$

Denote the closure in $X$ of the convex hull of $\mathcal{D}$ by $A_1(\mathcal{D}) := A_1(\mathcal{D}, X)$. The following theorem from [2] gives the rate of convergence of the WCGA for $f$ in $A_1(\mathcal{D}^\pm)$.

**Theorem 2.1.** Let $X$ be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q < 2$. Then for $t \in (0,1]$ and any $f \in A_1(\mathcal{D}^\pm)$

$$\|f - G^{c,t}_m(f, \mathcal{D})\| \leq C(q, \gamma)(1 + mt^p)^{-1/p}, \quad p := \frac{q}{q-1},$$

where the constant $C(q, \gamma)$ can only depend on $q$ and $\gamma$.

**Remark 2.1.** It follows from the proof of Theorem 2.1 that $C(q, \gamma) \leq C(q)^{1/q}$.

We proceed to the Incremental Greedy Algorithm (see [18] and [1], Ch.6). Let $\epsilon = \{\epsilon_n\}_{n=1}^\infty$, $\epsilon_n > 0$, $n = 1, 2, \ldots$. We note that the Incremental Greedy Algorithm belongs to the family of relaxed greedy algorithms (see [1], Ch.6).

**Incremental Algorithm with schedule $\epsilon$ (IA($\epsilon$)).** Set $f^{i, \epsilon}_0 := f$ and $G^{i, \epsilon}_0 := 0$. Then, for each $m \geq 1$ we have the following inductive definition:

1) the element $\varphi^{i, \epsilon}_m \in \mathcal{D}$ is any element satisfying

$$F_{f^{i, \epsilon}_{m-1}}(\varphi^{i, \epsilon}_m - f) \geq -\epsilon_m;$$

2) set

$$G^{i, \epsilon}_m := \left(1 - \frac{1}{m}\right)G^{i, \epsilon}_{m-1} + \frac{\varphi^{i, \epsilon}_m}{m};$$

3) set

$$f^{i, \epsilon}_m := f - G^{i, \epsilon}_m.$$

**Theorem 2.2.** Let $X$ be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q < 2$. Define

$$\epsilon_n := v\gamma^{1/q}n^{-1/p}, \quad p = \frac{q}{q-1}, \quad n = 1, 2, \ldots.$$  

Then for any $f \in A_1(\mathcal{D})$,

$$\|f^{i, \epsilon}_m\| \leq C(v)\gamma^{1/q}m^{-1/p}, \quad m = 1, 2, \ldots.$$  

In [18] we demonstrated the power of the WCGA in classical areas of harmonic analysis. The problem concerns the trigonometric $m$-term approximation in the uniform norm. The first result to indicate the advantage of $m$-term approximation
with respect to the real trigonometric system $\mathcal{RT}$ over approximation by trigonometric polynomials of order $m$ is due to Ismagilov (see [6]).

$$\sigma_m(|\sin 2\pi x|, \mathcal{RT})_\infty \leq C \epsilon m^{-6/5+\epsilon} \quad \text{for any} \quad \epsilon > 0. \quad (2.2)$$

Maiorov [7] improved the estimate (2.2):

$$\sigma_m(|\sin 2\pi x|, \mathcal{RT})_\infty \precsim m^{-3/2}. \quad (2.3)$$

Both Ismagilov [6] and Maiorov [7] used constructive methods to get their estimates (2.2) and (2.3). Maiorov [7] applied a number theoretical method based on Gaussian sums. The key point of that technique can be formulated in terms of the best $m$-term approximation of trigonometric polynomials. Let $\mathcal{RT}(N)$ be the subspace of real trigonometric polynomials of order $N$. Using Gaussian sums one can (constructively) prove the estimate

$$\sigma_m(t, \mathcal{RT})_\infty \leq C N^{3/2} m^{-1} \|t\|_1, \quad t \in \mathcal{RT}(N). \quad (2.4)$$

Set

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

Note that using the simple inequality

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

estimate (2.4) follows from the estimate

$$\sigma_m(t, \mathcal{RT})_\infty \leq C \left(\frac{N^{1/2}}{m}\right) \|t\|_A, \quad t \in \mathcal{RT}(N). \quad (2.5)$$

Thus, (2.5) is stronger than (2.4). The following estimate was proved in [8]:

$$\sigma_m(t, \mathcal{RT})_\infty \leq C m^{-1/2} \left(\ln \left(1 + \frac{N}{m}\right)\right)^{1/2} \|t\|_A, \quad t \in \mathcal{RT}(N). \quad (2.6)$$

In a way, (2.6) is much stronger than both (2.5) and (2.4). The proof of (2.6) in [8] is not constructive. The estimate (2.6) was proved in [8] with the help of a nonconstructive theorem of Gluskin’s [9]. Belinskii [12] used a probabilistic method to prove the following inequality:

$$\sigma_m(t, \mathcal{RT})_\infty \leq C \left(\frac{N}{m}\right)^{1/p} \left(\ln \left(1 + \frac{N}{m}\right)\right)^{1/p} \|t\|_p, \quad t \in \mathcal{RT}(N),$$

for $2 \leq p < \infty$. His proof is also nonconstructive. In [18] we gave a constructive proof of (2.6). The key ingredient of that proof is the WCGA. In [17] we had already pointed out that the WCGA provides a constructive proof of the estimate

$$\sigma_m(f, \mathcal{RT})_p \leq C(p) m^{-1/2} \|f\|_A, \quad p \in [2, \infty). \quad (2.7)$$

The proofs of (2.7) which were known before [17] were nonconstructive (see the discussion in [17], §5). Thus, the WCGA provides a way of building a good $m$-term approximant. We now state a result from [18].
Theorem 2.3. There exists a constructive method $A(N, m)$ that provides an $m$-term trigonometric polynomial $A(N, m)(t)$ for any $t \in \mathcal{R}(N)$ with the following approximation property:

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

with an absolute constant $C$.

However, step 2) of the WCGA makes it difficult to control the coefficients of the approximant—they are obtained through the Chebyshev projection of $f$ onto $\Phi_m$. This motivates us to consider the algorithm $IA(\epsilon)$ which gives the coefficients of the approximant explicitly. An advantage of the $IA(\epsilon)$ over other greedy-type algorithms is that the $IA(\epsilon)$ gives precise control of the coefficients of the approximant. For all approximants $G_{i,\epsilon}^m$ we have the property $\|G_{i,\epsilon}^m\|_A = 1$. Moreover, we know that all nonzero coefficients of the approximant have the form $a/m$ where $a$ is a natural number. We prove the following result.

Theorem 2.4. For any $t \in \mathcal{R}(N, d)$ after $m$ iterations the $IA(\epsilon)$ applied to $f := t/\|t\|_A$ provides an $m$-term trigonometric polynomial $G_m(t) := G_{i,\epsilon}^m(f)\|t\|_A$ with the following approximation property

$$
\|t - G_m(t)\|_\infty \leq C(d)(m)^{-1/2}(\ln \vartheta(N))^{1/2}\|t\|_A,
$$

where the constant $C(d)$ can only depend on $d$.

Proof. It is clear that it is sufficient to prove Theorem 2.4 for $t \in \mathcal{R}(N, d)$ with $\|t\|_A = 1$. Then $t \in (A_1(\mathcal{R}(N, d) \cap \mathcal{D})^\pm, L_p)$ for all $p \in [2, \infty)$. Now, applying Theorem 2.2 with $X = L_p$ and $\mathcal{D}$, where $\mathcal{D} := \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$, $n = \vartheta(N)$, is the real trigonometric system

$$
\varphi_l := \prod_{j \in E} \cos k_j x_j \prod_{j \in [1, d] \setminus E} \sin k_j x_j,
$$

we obtain that

$$
\left\|t - \sum_{j \in \Lambda} a_j \frac{\varphi_j}{m} \right\|_p \leq C\gamma^{1/2}m^{-1/2}, \quad \sum_{j \in \Lambda} |a_j| = m,
$$

where $\sum_{j \in \Lambda} a_j \frac{\varphi_j}{m}$ is $G_{i,\epsilon}^m(t)$. From (2.1) we see that $\gamma \leq p/2$. Next, using the Nikol’skii inequality, from (2.8) we obtain

$$
\left\|t - \sum_{j \in \Lambda} a_j \frac{\varphi_j}{m} \right\|_\infty \leq C(d)N^{1/p}\left\|t - \sum_{j \in \Lambda} a_j \frac{\varphi_j}{m} \right\|_p \leq C(d)p^{1/2}N^{1/p}m^{-1/2}.
$$

Choosing $p \asymp \ln N$ we obtain the required bound in Theorem 2.4.

Note that the above proof of Theorem 2.4 yields the following statement.
Theorem 2.5. Let $2 \leq p < \infty$. For any $t \in \mathcal{R}(N,d)$ after $m$ iterations the IA($\epsilon$) applied to $f := t/\|t\|_A$ provides an $m$-term trigonometric polynomial $G_m(t) := G^m_\epsilon(f)\|t\|_A$ with the following approximation property

$$\|t - G_m(t)\|_p \leq C(d)(\overline{m})^{-1/2}p^{1/2}\|t\|_A, \quad \overline{m} := \max(1,m), \quad \|G_m(t)\|_A = \|t\|_A,$$

where the constant $C(d)$ can only depend on $d$.

We note that the implementation of the IA($\epsilon$) depends on the dictionary and the ambient space $X$. For example, for $d = 1$ the IA($\epsilon$) from Theorem 2.4 acts with respect to the real trigonometric system $1, \cos 2\pi x, \sin 2\pi x, \ldots, \cos N2\pi x, \sin N2\pi x$ in the space $X = L_p$ with $p > \ln N$.

The above Theorems 2.4 and 2.5 are formulated for $m$-term approximation with respect to the real trigonometric system because the general Theorem 2.2 is proved for real Banach spaces. Clearly, as a corollary of Theorems 2.4 and 2.5 we obtain the corresponding results for the complex trigonometric system $\mathcal{T}^d := \{e^{i(k,x)}\}_{k \in \mathbb{Z}^d}$. As above, set $\overline{m} := \max(1,m)$.

Theorem 2.6. There exist constructive greedy-type approximation methods $G^p_m(\cdot)$ that provide $m$-term polynomials with respect to $\mathcal{T}^d$ with the following properties: for $2 \leq p < \infty$

$$\|f - G^p_m(f)\|_p \leq C_1(d)(\overline{m})^{-1/2}p^{1/2}\|f\|_A, \quad \|G^p_m(f)\|_A \leq C_2(d)\|f\|_A,$$

and for $p = \infty$, $f \in \mathcal{T}(N,d)$

$$\|f - G^\infty_m(f)\|_\infty \leq C_3(d)(\overline{m})^{-1/2}(\ln \vartheta(N))^{1/2}\|f\|_A, \quad \|G^\infty_m(f)\|_A \leq C_4(d)\|f\|_A.$$

We now apply Theorem 2.6 to $m$-term approximation of functions with mixed smoothness. The following theorem was proved in [13] (see also [20], Ch. 4). The proofs in [13] and [20] are constructive. We use the following notation

$$\beta := \beta(q,p) := \frac{1}{q} - \frac{1}{p}, \quad \eta := \eta(q) := \frac{1}{q} - \frac{1}{2}.$$

Theorem 2.7. Let $1 < q \leq p \leq 2$ and let $r > 2\beta$. Then

$$\sigma_m(W^r_q) \precsim m^{-r+\beta(\log m)^{(d-1)(r-2\beta)}}.$$

First, we extend Theorem 2.7 to the case $1 < q \leq p < \infty$.

Theorem 2.8. The following relations hold:

$$\sigma_m(W^r_q) \precsim \begin{cases} m^{-r+\beta(\log m)^{(d-1)(r-2\beta)}}, & 1 \leq q \leq 2, \quad r > 2\beta, \\ m^{-r+\eta(\log m)^{(d-1)(r-2\eta)}}, & 1 \leq q \leq 2 \leq p < \infty, \quad r > \frac{1}{q}, \\ m^{-r}(\log m)^{r(d-1)}, & 2 \leq q \leq p < \infty, \quad r > \frac{1}{2}. \end{cases}$$
Proof. The case \( p \leq 2 \), which corresponds to the first line, follows from Theorem 2.7. We note that in the case \( p > 2 \) Theorem 2.8 was proved in [15]. However, the proof given there is not constructive: it uses a nonconstructive result from [8]. We provide a constructive proof, which is based on greedy algorithms. In addition, this proof works under weaker conditions on \( r: r > 1/q \) instead of \( r > 1/q + \eta \) for \( 1 < q \leq 2 \). The following lemma plays the key role in the proof.

**Lemma 2.1.** For \( f \in L_1 \) set

\[
f_l := \sum_{\|s\|_1 = l} \delta_s(f), \quad l \in \mathbb{N}_0, \quad \mathbb{N}_0 := \mathbb{N} \cup \{0\}.
\]

Consider the class

\[
W_{ab}^A := \{ f: \| f \|_A \leq 2^{-al}(d-1)b \}.
\]

Then for \( 2 \leq p \leq \infty \) and \( 0 < \mu < a \) there is a constructive method \( A_m(\cdot, p, \mu) \) based on greedy algorithms, which gives the following bound for \( f \in W_{ab}^A \)

\[
\| f - A_m(f, p, \mu) \|_p \ll m^{-a-1/2}(log m)(d-1)(a+b), \quad 2 \leq p < \infty,
\]

\[
\| f - A_m(f, \infty, \mu) \|_\infty \ll m^{-a-1/2}(log m)(d-1)(a+b)+1/2.
\]

**Proof.** We prove the lemma for \( m \asymp 2^n n^{d-1}, \quad n \in \mathbb{N} \). Let \( f \in W_{ab}^A \). We approximate \( f_l \) in \( L_p \). By Theorem 2.6, for \( p \in [2, \infty) \) we obtain

\[
\| f_l - G_{m_l}^p (f_l) \|_p \ll (m_l)^{-1/2} \| f_l \|_A \ll (m_l)^{-1/2} 2^{-al}(d-1)b.
\]

We take \( \mu \in (0, a) \) and take

\[
m_l := [2^{n-\mu(l-n)}(d-1)], \quad l = n, n+1, \ldots.
\]

In addition we include

\[
S_n(f) := \sum_{\|s\|_1 \leq n} \delta_s(f)
\]

in the approximant. Set

\[
A_m(f, p, \mu) := S_n(f) + \sum_{l > n} G_{m_l}^p (f_l).
\]

Then we have built an \( m \)-term approximant to \( f \) with

\[
m \ll 2^n n^{d-1} + \sum_{l \geq n} m_l \ll 2^n n^{d-1}.
\]

The error of this approximation in \( L_p \) is bounded above by

\[
\| f - A_m(f, p, \mu) \|_p \leq \sum_{l \geq n} \| f_l - G_{m_l}^p (f_l) \|_p \ll (m_l)^{-1/2} 2^{-al}(d-1)b
\]

\[
\ll \sum_{l \geq n} 2^{-1/2(n-\mu(l-n))} l^{-(d-1)/2} 2^{-al}(d-1)b \ll 2^{-n(a+1/2)} n^{(d-1)(b-1/2)}.
\]

This completes the proof of lemma in the case \( 2 \leq p < \infty \).
We now discuss the case $p = \infty$. The proof repeats the above proof for $p < \infty$, with the following change. Instead of using (2.9) to estimate an $m_l$-term approximation of $f_l$ in $L^p$, we use (2.10) to estimate an $m_l$-term approximation of $f_l$ in $L^\infty$.

Then (2.13) is replaced by

$$
\|f_l - G_{m_l}(f_l)\|_\infty \ll (m_l)^{-1/2}(\ln 2)^{1/2}\|f_l\|_A \ll (m_l)^{-1/2}l^{1/2}2^{-a_l l^{(d-1)b}}. \tag{2.14}
$$

The extra factor $l^{1/2}$ in (2.14) gives an extra factor $(\log m)^{1/2}$ in (2.12). The proof is complete.

We now complete the proof of Theorem 2.8. First, consider the case $1 < q \leq 2 \leq p < \infty$. It is well known that for $f \in W^r_q$

$$
\|f_l\|_q \ll 2^{-lr} \tag{2.15}
$$

(see, for instance, [20], Ch. 2, Theorem 2.1). Theorem 1.7 implies that

$$
\|f_l\|_A \ll 2^{-(r-1)/q)l^{(d-1)(1-1/q)}. \tag{2.16}
$$

Therefore, it is sufficient to use Lemma 2.1 with $a = r - 1/q$ and $b = 1 - 1/q$ to obtain the upper bounds.

Second, in the case $2 \leq q < p < \infty$ the upper bounds follow from the above case $1 < q \leq 2 \leq p < \infty$ with $q = 2$. The lower bounds follow from Theorem 2.7 with $p = 2$. The lower bounds in the case $2 \leq q < p < \infty$ follow from known results for the case $1 < p \leq q < \infty$ in [14] (see Theorem 2.17 below). Theorem 2.8 is proved.

Next, we discuss the case $p = \infty$. In the same way as Theorem 2.8 was derived from (2.11) of Lemma 2.1 the following upper bounds in case $p = \infty$ are derived from (2.12) of Lemma 2.1.

**Theorem 2.9.** The following relations hold:

$$
\sigma_m(W^r_q) \ll \begin{cases} 
  m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+1/2}, & 1 < q \leq 2, \quad r > \frac{1}{q}, \\
  m^{-r}(\log m)^{r(d-1)+1/2}, & 2 \leq q < \infty, \quad r > \frac{1}{2}.
\end{cases}
$$

The upper bounds are provided by a constructive method $A_m(\cdot, \infty, \mu)$ based on greedy algorithms.

Consider the case $\sigma_m(W^r_{1,\alpha})$, which is not covered by Theorems 2.7 and 2.8. The function $F_r(x)$ belongs to the closure of $W^r_{1,\alpha}$ in $L^p$, $r > 1 - 1/p$, and, therefore, on the one hand

$$
\sigma_m(W^r_{1,\alpha}) \geq \sigma_m(F_r(x)).
$$

On the other hand, it follows from the definition of $W^r_{1,\alpha}$ that for any $f \in W^r_{1,\alpha}$

$$
\sigma_m(f) \leq \sigma_m(F_r(x)).
$$

Thus,

$$
\sigma_m(W^r_{1,\alpha}) = \sigma_m(F_r(x)) \tag{2.16}
$$

We now prove some results concerning $\sigma_m(F_r(x))$. 
Theorem 2.10. The following relations hold:

\[ \sigma_m(F_r(x))_p = \begin{cases} 
  m^{-r+1-1/p}(\log m)^{(d-1)(r-1+2/p)}, & 1 < p \leq 2, \quad r > 1 - \frac{1}{p}, \\
  m^{-r+1/2}(\log m)^{r(d-1)}, & 2 \leq p < \infty, \quad r > 1.
\] 

The upper bounds are provided by a constructive method \( A_m(\cdot, p, \mu) \) based on greedy algorithms.

Proof. We begin with the case \( 1 < p \leq 2 \). The following error bound for approximation by the hyperbolic cross polynomials is known:

\[ E_{Q_n}(F_r)_p := \inf_{t \in \mathcal{T}(Q_n)} \| F_r - t \|_p \ll 2^{-n(r-1+1/p)}n^{(d-1)/p} \quad (2.17) \]

(see, for instance, [20], Ch. 2, Theorem 3.1). Now, \( |Q_n| \asymp 2^n n^{d-1} \) and so from (2.17) we obtain the required upper bound in the case \( 1 < p \leq 2 \). Thus, it remains to prove the matching lower bound in the case \( 1 < p \leq 2 \). Set

\[ \theta_n := \{ s \in \mathbb{N}^d : \|s\|_1 = n \}, \quad \Delta Q_n := \bigcup_{s \in \theta_n} \rho(s). \]

Let \( K_m := \{ k^j \}_{j=1}^m \) be given. Choose \( n \) such that it is the smallest that satisfies

\[ |\Delta Q_n| \geq 4m. \]

Clearly

\[ 2^n n^{d-1} \asymp m. \]

Set

\[ \theta'_n := \left\{ s \in \theta_n : |K_m \cap \rho(s)| \leq \frac{|\rho(s)|}{2} \right\}. \]

Note that for \( s \in \mathbb{N}^d \) we have \( |\rho(s)| = 2^n. \) Then

\[ \frac{(|\theta_n| - |\theta'_n|)2^n}{2} \leq m \leq \frac{|\theta_n|2^n}{4}, \]

which implies

\[ |\theta'_n| \geq \frac{|\theta_n|}{2}. \]

Applying relation (1.5) in Theorem 1.6, with \( q = 2 \) and \( 1 < p < 2 \), for any

\[ t = \sum_{j=1}^m c_j e^{i(k^j \cdot x)} \]

we obtain

\[ \| F_r - t \|_p \gg \left( \sum_{s \in \theta'_n} \| \delta_s(F_r - t) \|_2^{p/2} 2^{n(p/2-1)} \right)^{1/p} \]

\[ \gg \left( \sum_{s \in \theta'_n} 2^{pn(-r+1/2)} 2^{n(p/2-1)} \right)^{1/p} \gg 2^{-n(r-1+1/p)}n^{(d-1)/p}. \]

This gives the required lower bound for \( 1 < p < 2 \).
The above argument gives the lower bound in the case $p = 2$ without the use of Theorem 1.6: it is sufficient to use the Parseval identity.

We now proceed to the case $2 \leq p < \infty$. The analysis here is similar to that in the proof of Theorem 2.8. For

$$F_r^l := \sum_{\|s\|_1 = l} \delta_s(F_r)$$

we get

$$\|F_r^l\|_A \ll 2^{-lr}2^l1^{d-1}.$$ 

The required upper bound follows from Lemma 2.1 with $a = r - 1$ and $b = 1$.

The lower bound follows from the case $p = 2$. The proof is complete.

In the same way as a modification of the proof of Theorem 2.8 gave Theorem 2.9 the corresponding modification of the argument in the proof of Theorem 2.10 gives the following result.

**Theorem 2.11.** The estimate

$$\sigma_m(F_r(x))_\infty \ll m^{-r+1/2}(\log m)^{r(d-1)+1/2}, \quad r > 1,$$

holds. The bounds are provided by a constructive method $A_m(\cdot, \infty, \mu)$ based on greedy algorithms.

We now turn to the classes $H_q^r$ and $B_{q, \theta}^r$. Set

$$\|f\|_{H_q^r} := \sup_s \|\delta_s(f)\|_q 2^r\|s\|_1$$

and for $1 \leq \theta < \infty$ let

$$\|f\|_{B_{q, \theta}^r} := \left( \sum_s (\|\delta_s(f)\|_q 2^r\|s\|_1)^{\theta} \right)^{1/\theta}.$$ 

We write $B_{q, \infty}^r := H_q^r$. With a little abuse of notation, we can denote the corresponding unit ball by

$$B_{q, \theta}^r := \{ f : \|f\|_{B_{q, \theta}^r} \leq 1 \}.$$

It will be convenient for us to use the following slight modification of the classes $B_{q, \theta}^r$. Define

$$\|f\|_{H_{q, \theta}^r} := \sup_n \left( \sum_{s : \|s\|_1 = n} (\|\delta_s(f)\|_q 2^r\|s\|_1)^{\theta} \right)^{1/\theta}$$

and

$$H_{q, \theta}^r := \{ f : \|f\|_{H_{q, \theta}^r} \leq 1 \}.$$

The best $m$-term approximations of classes $B_{q, \theta}^r$ were studied in detail by Romanyuk [19]. The following theorem was proved in [13] (see also [20], Ch. 4). The proofs in [13] and [20] are constructive.
Theorem 2.12. Let \( 1 < q \leq p \leq 2 \) and let \( r > \beta \). Then
\[
\sigma_m(H^r_q)_p \asymp m^{-r+\beta} (\log m)^{(d-1)(r-\beta+1/p)}.
\]

The following analogue of Theorem 2.8 for classes \( H^r_q \) was proved in [19]. The proof in [19] in the case \( p > 2 \) is not constructive.

Theorem 2.13. The following relations hold:
\[
\sigma_m(H^r_q)_p \asymp \begin{cases} 
m^{-r+\beta} (\log m)^{(d-1)(r-\beta+1/p)}, & 1 < q \leq p \leq 2, \quad r > \beta, 
m^{-r+\eta} (\log m)^{(d-1)(r-1/q+1)}, & 1 < q \leq 2 < p < \infty, \quad r > \frac{1}{q}, 
m^{-r} (\log m)^{(d-1)(r+1/2)}, & 2 \leq q \leq p < \infty, \quad r > \frac{1}{2}. 
\end{cases}
\]

Proposition 2.1. The upper bounds in Theorem 2.13 are provided by a constructive method \( A_m(\cdot, p, \mu) \) based on greedy algorithms.

Proof. The case \( p \leq 2 \), which corresponds to the first line, follows from Theorem 2.12. We now consider \( p > 2 \). From the definition of the classes \( H^r_q \) for \( 1 < q < \infty \) we have
\[
f \in H^r_q \iff \|\delta_s(f)\|_q \leq 2^{-r\|s\|_1}.
\]
Next,
\[
\|\delta_s(f)\|_A \ll 2^{\|s\|_1/q} \|\delta_s(f)\|_q.
\]
Therefore, for \( f \in H^r_q \) we obtain
\[
\|f_1\|_A \ll 2^{-(r-1/q)(d-1)}.
\]
Applying Lemma 2.1 with \( a = r-1/q \) and \( b = 1 \) in the case \( 1 < q \leq 2 \leq p < \infty \), we obtain the required upper bounds. The upper bounds in the case \( 2 \leq q \leq p < \infty \) follow from the above case with \( q = 2 \). The lower bounds in the case \( 1 < q \leq 2 \leq p < \infty \) follow from Theorem 2.12.

The lower bounds in the case \( 2 \leq q \leq p < \infty \) follow from the known results in [14] (see Theorem 2.18 below).

Proposition 2.1 and Theorem 2.13 are proved.

In the case \( p = \infty \) we have the following.

Theorem 2.14. The following relations hold:
\[
\sigma_m(H^r_q)_\infty \ll \begin{cases} 
m^{-r+\eta} (\log m)^{(d-1)(r-1/q+1)+1/2}, & 1 < q \leq 2, \quad r > \frac{1}{q}, 
m^{-r} (\log m)^{(r+1/2)(d-1)+1/2}, & 2 \leq q < \infty, \quad r > \frac{1}{2}. 
\end{cases}
\]
The upper bounds are provided by a constructive method \( A_m(\cdot, \infty, \mu) \) based on greedy algorithms.

For a nonconstructive proof of the bound in Theorem 2.14 in the case \( 2 \leq q < \infty \), see [24].

We now proceed to the classes \( B^r_{q, \theta} \). The following extension of Theorem 2.13 is valid (see [19]).
Theorem 2.15. The following relations hold:

\[
s_m(\mathcal{B}_{q, \theta}^r)_p \begin{cases} 
  m^{-r+\beta} (\log m)^{(d-1)(r-\beta+1/p-1/\theta)} +, & 1 < q \leq p \leq 2, \ r > \beta, \\
  m^{-r+\eta} (\log m)^{(d-1)(r-1/q+1-1/\theta)} +, & 1 < q \leq 2 < p < \infty, \ r > \frac{1}{q}, \\
  m^{-r} (\log m)^{(d-1)(r+1/2-1/\theta)}, & 2 < q \leq p < \infty, \ r > \frac{1}{2}.
\end{cases}
\]

Proposition 2.2. The upper bounds in Theorem 2.15 are provided by a constructive method based on greedy algorithms.

Proof. We can prove the proposition for the first relation in Theorem 2.15 (see [19], Theorem 3.1) in the same way as Theorems 2.7 and 2.12 were proved constructively in [20]. We now consider \( p \geq 2 \) and establish an error bound for both the classes \( \mathcal{H}_{q, \theta}^r \) and \( \mathcal{B}_{q, \theta}^r \). From the definition of the classes \( \mathcal{H}_{q, \theta}^r \) for \( 1 < q < \infty \) we get

\[
f \in \mathcal{H}_{q, \theta}^r \iff \left( \sum_{\|s\|_1 = l} \|\delta_s(f)\|_q \right)^{1/\theta} \ll 2^{-r_l}.
\]

Next,

\[
\|f_l\|_A \ll \sum_{\|s\|_1 = l} \|\delta_s(f)\|_A \ll 2^{l/q} \sum_{\|s\|_1 = l} \|\delta_s(f)\|_q
\ll 2^{l/q} l^{(d-1)(1-1/\theta)} \left( \sum_{\|s\|_1 = l} \|\delta_s(f)\|_q \right)^{1/\theta} \ll 2^{-l((r-1)/q)l^{(d-1)(1-1/\theta)}}.
\]

Therefore, for \( f \in \mathcal{H}_{q, \theta}^r \) we obtain

\[
\|f_l\|_A \ll 2^{-(r-1)/q)l^{(d-1)(1-1/\theta)}).
\]

Applying Lemma 2.1 with \( a = r - 1/q \) and \( b = 1 - 1/\theta \) in the case \( 1 < q \leq 2 \leq p < \infty \), we obtain the required upper bounds. The upper bounds in the case \( 2 \leq q \leq p < \infty \) follow from the above case with \( q = 2 \). The lower bounds in the case \( 1 < q \leq 2 \leq p < \infty \) follow from the case \( 1 < q \leq 2, \ p = 2 \).

It was proved in [14] that

\[
\sigma_m(\mathcal{H}_{q, \theta}^r \cap \mathcal{F}(\Delta Q_n))_p \gg m^{-r} (\log m)^{(d-1)(r+1/2)} \quad (2.18)
\]

with some \( n \) such that \( m \approx 2^n n^{d-1} \). It is easy to see that for any \( f \in \mathcal{H}_{q, \theta}^r \cap \mathcal{F}(\Delta Q_n) \), we have

\[
\|f\|_{\mathcal{B}_{q, \theta}^r} \ll n^{(d-1)/\theta}. \quad (2.19)
\]

Relations (2.18) and (2.19) imply the lower bound in the case \( 2 \leq q \leq p < \infty \).

Proposition 2.2 and Theorem 2.15 are proved.

In the case \( p = \infty \) we have the following.
Theorem 2.16. The following relations hold:

\[
\sigma_m(\mathcal{H}^r_{q,\theta})_\infty \ll \begin{cases} 
  m^{-r+\eta(\log m)^{(d-1)(r-1/q+1-1/\theta)+1/2}}, & 1 < q \leq 2, \quad r > \frac{1}{q}, \\
  m^{-r}(\log m)^{(r+1/2-1/\theta)(d-1)+1/2}, & 2 \leq q < \infty, \quad r > \frac{1}{2}.
\end{cases}
\]

The upper bounds are provided by a constructive method \( A_m(\cdot, \infty, \mu) \) based on greedy algorithms.

We formulate some known results in the case \( 1 < p \leq q < \infty \).

Theorem 2.17. Let \( 1 < p \leq q < \infty \) and \( r > 0 \). Then

\[
\sigma_m(\mathcal{W}^r_q)_p \asymp m^{-r}(\log m)^{(d-1)r}.
\]

The upper bound in Theorem 2.17 follows from error bounds for approximation by the hyperbolic cross polynomials (see [20], Ch.2, §2)

\[
E_{Q_n}(\mathcal{W}^r_q, L_q) \ll 2^{-rn}, \quad 1 < q < \infty.
\]

The lower bound in Theorem 2.17 was proved in [14].

The following result for \( \mathcal{H}^r_q \) classes is known.

Theorem 2.18. Let \( p \leq q, 2 \leq q \leq \infty, 1 < p < \infty \) and \( r > 0 \). Then

\[
\sigma_m(\mathcal{H}^r_q)_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2)}.
\]

The lower bound for all \( p > 1 \),

\[
\sigma_m(\mathcal{H}^r_\infty)_p \gg m^{-r}(\log m)^{(d-1)(r+1/2)},
\]

was obtained in [14]. The matching upper bound follows from approximation by the hyperbolic cross polynomials (see [20], Ch.2, Theorem 2.2)

\[
E_{Q_n}(\mathcal{H}^r_q)_q := \sup_{f \in \mathcal{H}^r_q} E_{Q_n}(f)_q \asymp n^{(d-1)/2}2^{-rn}, \quad 2 \leq q < \infty.
\]

The following result for \( \mathcal{B}^r_q \) classes was proved in [19].

Theorem 2.19. Let \( 1 < p \leq q < \infty, 2 \leq q < \infty, 1 < p < \infty \) and \( r > 0 \). Then

\[
\sigma_m(\mathcal{B}^r_{q,\theta})_p \asymp m^{-r}(\log m)^{(d-1)(r+1/2-1/\theta)}.\]

§ 3. Application of the entropy numbers

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

\[
N_\epsilon(A) := N_\epsilon(A, X) := \min \left\{ n : \exists y^1, \ldots, y^n : A \subseteq \bigcup_{j=1}^n B_X(y^j, \epsilon) \right\}.
\]
Given a compact $A$ we define an $\epsilon$-distinguishable set $\{x^1, \ldots, x^m\} \subseteq A$ to be a set with the property
\[
\|x^i - x^j\| > \epsilon \quad \text{for all } i, j : i \neq j.
\]
Denote the maximum cardinality of $\epsilon$-distinguishable sets of a compact set $A$ by $M_\epsilon(A) := M_\epsilon(A, X)$. The following simple theorem is well known.

**Theorem 3.1.** For any compact set $A$
\[
M_{2\epsilon}(A) \leq N(\epsilon)(A) \leq M_\epsilon(A).
\]

Consider the entropy numbers
\[
\epsilon_k(A, X) := \inf \left\{ \epsilon : \exists y^1, \ldots, y^{2^k} \in X : A \subseteq \bigcup_{j=1}^{2^k} B_X(y^j, \epsilon) \right\}.
\]

The following theorem can be found in [16].

**Theorem 3.2.** Let the compact set $F \subset X$ be such that there exists a normalized system $\mathcal{D}$, $|\mathcal{D}| = N$, and a number $r > 0$ such that
\[
\sigma_k(F, \mathcal{D})X \leq k^{-r}, \quad k \leq N.
\]
Then for $k \leq N$
\[
\epsilon_k(F, X) \leq C(r) \left( \frac{\log(2N/k)}{k} \right)^r.
\]

We use the above theorem to prove the following lower bound for best $m$-term approximations.

**Theorem 3.3.** When $d = 2$ the following lower bound holds for any $q < \infty$ and $r > 1/q$:
\[
\sigma_m(W^r_q)_\infty \gg m^{-r}(\log m)^{1/2}.
\]

**Proof.** We will use a special inequality from [25], which is called the Small Ball Inequality. For an even number $n$ define
\[
Y_n := \left\{ s = (2n_1, 2n_2), n_1 + n_2 = \frac{n}{2} \right\}.
\]
Then for any coefficients $\{c_k\}$
\[
\left\| \sum_{s \in Y_n} \sum_{k \in \rho(s)} c_k e^{i(k,x)} \right\|_\infty \geq C \left\| \sum_{s \in Y_n} \sum_{k \in \rho(s)} c_k e^{i(k,x)} \right\|_1,
\]
where $\rho(s)$ was defined in §1 and $C$ is a positive number. Inequality (3.4) plays a key role in the proof of lower bounds for the entropy numbers.

Take any even $n \in \mathbb{N}$, which will be chosen later, depending on $m$. Consider the following compact set:
\[
F(Y_n)_\infty := \left\{ t = \sum_{s \in Y_n} t_s : t_s \in \mathcal{T}(\rho(s)), \quad \|t_s\|_\infty \leq 1 \right\}.
\]
The known results on sets of Fourier coefficients of trigonometric polynomials imply the following lemma (see [22] and [25]).
Lemma 3.1. There exist $2^{n2^{n-1}}$ functions
\[ f_j \in F(Y_n)_\infty, \quad j = 1, \ldots, 2^{n2^{n-1}}, \]
such that for $i \neq j$
\[ \|f_i - f_j\|_2 \gg n^{1/2}. \]

We now show that for $f_j$ from Lemma 3.1 we have
\[ \|f_i - f_j\|_\infty \gg n. \]

Indeed, for any $f \in F(Y_n)_\infty$ we have
\[ \|f\|_2^2 = \sum_{s \in Y_n} \|t_s\|_2^2 \leq \sum_{s \in Y_n} \|t_s\|_1 \|t_s\|_\infty \leq \sum_{s \in Y_n} \|t_s\|_1. \]

It remains to apply the Small Ball Inequality (3.4). Therefore, for $k = n2^{n-1}$, by Theorem 3.1 we obtain
\[ \epsilon_k(F(Y_n)_\infty) \gg \log k. \]

We now use Theorem 3.2. We set $F := F(Y_n)_\infty$, $\mathcal{D} := \{e^{i(k,x)} : \|k\|_\infty \leq 2^{n+1}\}$, $X := L_\infty$. It is clear that for $l \geq \dim T(Y_n) \asymp n2^n \times k$
\[ \sigma_l(F, \mathcal{D}) = 0. \]

Also, for any $f \in F$ we have
\[ \|f\|_\infty \ll \frac{n}{2} \ll \log k. \]

Put
\[ B := \max_l l' \sigma_l(F, \mathcal{D})_\infty. \]

By Theorem 3.2 we obtain
\[ \log k \ll Bn^r k^{-r}, \quad B \gg n^{-r} k^r \log k. \]

This implies that there exists $l \asymp k$ such that
\[ \sigma_l(F, \mathcal{D})_\infty \gg n^{-r} \log k \asymp (\log k)^{1-r}. \]

Next, it is clear that for any $m$
\[ \sigma_m(F, \mathcal{D})_\infty \ll \sigma_m(F, \mathcal{D})_\infty. \]

Further, by the Littlewood-Paley theorem there is $c_1(q) > 0$ such that
\[ c_1(q)n^{-1/2}2^{-rn} F \subset W_q^r, \quad q < \infty. \]

This completes the proof.
§ 4. Numerical integration

4.1. Notation. The setting for the problem. Numerical integration looks for ways to obtain a good approximation of the integral

$$\int_{\Omega} f(x) \, d\mu$$

using an expression of the form

$$\Lambda(f, X_m) := \sum_{j=1}^{m} \lambda_j f(\xi_j), \quad (4.1)$$

where $X_m := \{\xi^1, \ldots, \xi^m\}$, $\xi^j \in \Omega$, $j = 1, \ldots, m$. For a function class $\mathcal{W}$ set

$$\Lambda(\mathcal{W}, X_m) := \sup_{f \in \mathcal{W}} \left| \int_{\Omega} f(x) \, d\mu - \Lambda(f, X_m) \right|.$$ 

We are interested in the way the best $m$-knot error of numerical integration

$$\delta_m(\mathcal{W}) := \inf_{\lambda_1, \ldots, \lambda_m; \xi^1, \ldots, \xi^m} \Lambda(\mathcal{W}, X_m)$$

depends on $m$, for some function classes $\mathcal{W}$.

4.2. Known lower bounds. The reader can find results and historical comments on numerical integration of classes of functions with mixed smoothness in Temlyakov [21], Ch. 4 and in [26]. The following theorem was proved in [27].

**Theorem 4.1.** The following lower estimate is valid for any cubature formula $(\Lambda, X_m)$ with $m$ knots

$$\Lambda(\mathcal{W}^r_p, X_m) \geq C(r, d, p)m^{-r}(\log m)^{(d-1)/2}, \quad 1 \leq p < \infty, \quad r > \frac{1}{p}.$$ 

The proof of this theorem is based on Theorem 1.8 from the introduction. Theorem 1.8 is used to prove the following assertion.

**Lemma 4.1.** Let the coordinates of the vector $s$ be natural numbers and suppose that $\|s\|_1 = n$. Then for any $N \leq 2^n - 1$ and an arbitrary cubature formula $(\Lambda, X_N)$ with $N$ knots there is a $t_s \in \mathcal{T}(2^n, d)$ such that $\|t_s\|_{\infty} \leq 1$ and

$$\hat{t}_s(0) - \Lambda(t_s, X_N) \geq C(d) > 0. \quad (4.2)$$

For a given $m$ choose $n$ such that

$$m \leq 2^n - 1 < 2m.$$ 

Consider the polynomial

$$t(x) = \sum_{\|s\|_1 = n} t_s(x),$$
where the $t_s$ are polynomials from Lemma 4.1 with $N = m$. Then
\[ \hat{t}(0) - \Lambda(t, X_m) \geq C(d)n^{d-1}. \] (4.3)

The proof of Theorem 4.1 was completed by establishing that
\[ \|t\|_{W_p^r} \ll \|t\|_{B^r_{p,2}} \ll 2^{-n}n^{(d-1)/2}. \] (4.4)

Theorem 4.1 gives the same lower bound for different parameters $1 \leq p < \infty$. It is clear that the bigger the value of $p$ the stronger the statement.

### 4.3. New lower bounds

We obtain lower bounds for numerical integration with respect to a special class of knots. Let $s = (s_1, \ldots, s_d)$, $s_j \in \mathbb{N}_0$, $j = 1, \ldots, d$. We associate with $s$ a web $W(s)$ as follows: set
\[ w(s, x) := \prod_{j=1}^d \sin(2^{s_j}x_j) \]
and define
\[ W(s) := \{x : w(s, x) = 0\}. \]

**Definition 4.1.** A set of knots $X_m := \{\xi^i\}_{i=1}^m$ is an $(n, l)$-net if $|X_m \setminus W(s)| \leq 2^l$ for all $s$ such that $\|s\|_1 = n$.

**Theorem 4.2.** Any cubature formula $(\Lambda, X_m)$ with respect to an $(n, n-1)$-net $X_m$ satisfies
\[ \Lambda(W_{p}^r, X_m) \gg 2^{-n}n^{(d-1)/2}, \quad 1 \leq p < \infty. \]

**Proof.** This proof is similar to the proof of Theorem 4.1 in [27]. Take $s$ such that $\|s\|_1 = n$ and consider $\mathcal{T}(N, d)$ with $N_j := 2^{s_j-1}$, $j = 1, \ldots, d$. Then
\[ \dim \mathcal{T}(N, d) \geq 2^{\|s\|_1} = 2^n. \]

Let $I(s)$ be a set of indices such that
\[ X_m \setminus W(s) = \{\xi^i\}_{i \in I(s)}. \]
Then by assumption $|I(s)| \leq 2^{n-1}$. Consider
\[ \Psi(s) := \{t \in \mathcal{T}(N, d) : t(\xi^i) = 0, i \in I(s)\}. \]
Then $\dim \Psi(s) \geq 2^{n-1}$. By Theorem 1.8 we can find $t_1^s \in \Psi(s)$ such that $\|t_1^s\|_\infty = 1$ and $\|t_1^s\|_2 \geq c(d) > 0$. Consider
\[ t(x) := \sum_{\|s\|_1 = n} t_s(x), \quad t_s(x) := |t_1^s(x)|^2w(s, x)^2. \]

We have
\[ t_1^s(x)w(s_1, x_1) = (2i)^{-1}\left( \sum_{k : |k_j| \leq 2^{s_j-1}} \hat{t}_s^1(k)e^{i(k_j)x_1 + i2^{s_j-1}x_1} + \sum_{k : |k_j| \leq 2^{s_j-1}} \hat{t}_s^1(k)e^{i(k_j)x_1 - i2^{s_j-1}x_1} \right) \]
and 
\[ \|t_s^1(x)w(s_1, x_1)\|_2^2 = 2^{-1}\|t_s^1\|_2^2. \]

Therefore, 
\[ \|t_s^1(x)w(s, x)\|_2^2 = 2^{-d}\|t_s^1\|_2^2 \geq c_1(d) > 0. \]

Then (4.3) is obviously satisfied for our \( t \). Relation (4.4) is proved in the same way as it was proved in [27].

The example that was constructed in the proof of Theorem 4.1 (see above) provides the lower bound for Besov-type classes. Another proof of Theorem 4.3 is given in [28].

**Theorem 4.3.** The following lower estimate holds for any cubature formula \((\Lambda, X_m)\) with \( m \) knots

\[ \Lambda(B_{p, \theta}^r, X_m) \geq C(r, d, p) m^{-r} (\log m)^{(d-1)(1-1/\theta)}, \]

\[ 1 \leq p \leq \infty, \quad 1 \leq \theta \leq \infty, \quad r > \frac{1}{p}. \]

Indeed, the proof of (4.4) in [27] implies

\[ \|t\|_{B_{p, \theta}^r} \ll 2^{rn}n^{(d-1)/\theta}. \] (4.5)

In the same way the proof of Theorem 4.2 gives the following result.

**Theorem 4.4.** Any cubature formula \((\Lambda, X_m)\) with respect to an \((n, n-1)\)-net \( X_m \) satisfies

\[ \Lambda(B_{p, \theta}^r, X_m) \gg 2^{-rn}n^{(d-1)(1-1/\theta)}, \quad 1 \leq p \leq \infty. \]

We note that Theorems 4.2 and 4.4 provide lower bounds for numerical integration with respect to sparse grids and their modifications. For \( n \in \mathbb{N} \) we define the sparse grid \( SG(n) \) as follows

\[ SG(n) := \{ \xi(n, k) = (\pi k_1 2^{-n_1}, \ldots, \pi k_d 2^{-n_d}), \]

\[ 0 \leq k_j < 2^{n_j}, j = 1, \ldots, d, \|n\|_1 = n \}. \]

Then it is easy to check that \( SG(n) \subset W(s) \) with any \( s \) such that \( \|s\|_1 = n \). Let \( \xi(n, k) \in SG(n) \). Take any \( s \) with \( \|s\|_1 = n \). Then \( \|s\|_1 = \|n\|_1 \) and there exists \( j \) such that \( s_j \geq n_j \). For this \( j \) we have

\[ \sin 2^{s_j} \xi(n, k)_j = \sin 2^{s_j} \pi k_j 2^{-n_j} = 0, \quad w(s, \xi(n, k)) = 0. \]

This means that \( SG(n) \) is an \((n, l)\)-net for any \( l \). We note that \( |SG(n)| \asymp 2^n n^{d-1} \). It is known (see [30]) that there exists a cubature formula \((\Lambda, SG(n))\) such that

\[ \Lambda(H_p^r, SG(n)) \ll 2^{-rn}n^{d-1}, \quad 1 \leq p \leq \infty, \quad r > \frac{1}{p}. \] (4.6)

Theorem 4.4 with \( \theta = \infty \) shows that the bound (4.6) is sharp. Moreover, Theorem 4.4 shows that even the addition of an extra \( 2^n \) arbitrary knots to \( SG(n) \) will not improve the bound in (4.6). When \( X_m = SG(n) \) another proof of Theorem 4.4 is given in [28].
§ 5. Approximate recovery

Consider the following recovery operator. For fixed $m$, $X_m := \{\xi^j\}^m_{j=1}$ and $\psi_1(x), \ldots, \psi_m(x)$ define the linear operator

$$\Psi(f, X_m) := \sum_{j=1}^{m} f(\xi^j)\psi_j(x).$$

For the function class $W$ define

$$\Psi(W, X_m)_p := \sup_{f \in W} \|f - \Psi(f, X_m)\|_p.$$

The main result in this section is the following theorem.

**Theorem 5.1.** For any recovery operator $\Psi(\cdot, X_m)$ with respect to an $(n, n-1)$-net $X_m$ and for $1 \leq q < p < \infty$

$$\Psi(H^r_q, X_m)_p \gg 2^{-n(r-\beta)}n^{(d-1)/p}, \quad \beta := \frac{1}{q} - \frac{1}{p}.$$

Before proceeding to the proof of this theorem we make some comments on the history of this area. The problem of optimal recovery on classes of functions with mixed smoothness is wide open. For a class $W$ set

$$\varrho_m(W)_p := \inf_{X_m; \psi_1, \ldots, \psi_m} \Psi(W, X_m)_p.$$

The correct order of this characteristic is known only in a few cases. It was established in [29] that

$$\varrho_m(W^r_2)_\infty \asymp m^{-r+1/2}(\log m)^{r(d-1)}. \quad (5.1)$$

The upper bound in (5.1) was obtained by recovering using a Smolyak-type operator $T_n$ for appropriate $n$. The operator $T_n$ uses the sparse grid $SG(n+d)$ and $\psi_j \in \mathcal{P}(Q_{n+d})$. The operator $T_n$ and its variants have been studied in many papers. It was proved in [30] that for any $f \in H^r_p$, $1 \leq p \leq \infty$, $r > 1/p$

$$\|f - T_n(f)\|_p \ll 2^{-rn}n^{d-1}. \quad (5.2)$$

The following bound was obtained in [21]; see Ch. 4, §5, Remark 2. For any $f \in H^r_q$, where $1 \leq q < p \leq \infty$ and $r > 1/q$

$$\|f - T_n(f)\|_p \ll 2^{-n(r-\beta)}n^{(d-1)/p}. \quad (5.3)$$

The upper bound (5.2) and the lower bound for the Kolmogorov width in [31]:

$$d_m(H^r_q, L_\infty) \asymp m^{-r}(\log m)^{r+1} \quad \text{for } d = 2,$$

imply that for $d = 2$

$$\varrho_m(H^r_q)_\infty \asymp m^{-r}(\log m)^{r+1}. \quad (5.4)$$

The upper bound (5.3) and known bounds for the Kolmogorov width (see, for instance, [20], Ch. 3): for $1 \leq q \leq 2$ and $r > 1/q$

$$d_m(H^r_q, L_2) \asymp m^{-r+\eta}(\log m)^{(d-1)(r+1-1/q)}, \quad \eta := \frac{1}{q} - \frac{1}{2},$$
imply that for $1 \leq q \leq 2$ and $r > 1/q$
\[ \varrho_m(H^r_q)_2 \sim m^{r-\eta}(\log m)^{(d-1)(r+1-1/q)}. \] (5.5)

As we have already pointed out above, $T_n(f) \in \mathcal{T}(Q_{n+d})$. Thus $T_n$ provides an approximation from the hyperbolic cross $\mathcal{T}(Q_{n+d})$. The upper bound (5.3) and known bounds for the best hyperbolic cross approximation (see, for instance, [20], Ch. 2, Theorem 2.2) show that for $1 \leq q < p < \infty$ and $r > 1/q$ the operator $T_n$ provides the optimal approximation for $H^r_q$ in the sense of order of the rate of approximation in the $L_p$.

**Proof of Theorem 5.1.** We use the polynomials $t^1_s$ constructed in the proof of Theorem 4.2. We also need some further constructions. Let
\[ \mathcal{K}_N(x) := \sum_{|k| \leq N} \left(1 - \frac{|k|}{N}\right) e^{ikx} = \frac{(\sin(Nx/2))^2}{N(\sin(x/2))^2} \]
be a univariate Fejér kernel. The Fejér kernel $\mathcal{K}_N$ is an even nonnegative trigonometric polynomial in $\mathcal{T}(N-1)$. From the relations
\[ \|\mathcal{K}_N\|_1 = 1, \quad \|\mathcal{K}_N\|_\infty = N, \]
which are obvious, the inequality
\[ \|f\|_q \leq \|f\|_1^{1/q} \|f\|_\infty^{1-1/q} \]
and the duality argument we find that
\[ CN^{1-1/q} \leq \|\mathcal{K}_N\|_q \leq N^{1-1/q}, \quad 1 \leq q \leq \infty. \] (5.6)

In the multivariate case set
\[ \mathcal{K}_N(x) := \prod_{j=1}^d \mathcal{K}_{N_j}(x_j), \quad N = (N_1, \ldots, N_d). \]

Then the $\mathcal{K}_N$ are nonnegative trigonometric polynomials from $\mathcal{T}(N-1, d)$ which have the following properties:
\[ \|\mathcal{K}_N\|_1 = 1, \] (5.7)
\[ \|\mathcal{K}_N\|_q \propto \vartheta(N)^{1-1/q}, \quad 1 \leq q \leq \infty. \] (5.8)

For $n$ of the form $n = 4l$, $l \in \mathbb{N}$, define
\[ Y(n, d) := \left\{ s : s = (4l_1, \ldots, 4l_d), l_1 + \cdots + l_d = \frac{n}{4}, l_j \in \mathbb{N}, j = 1, \ldots, d \right\}. \]

For $s \in Y(n, d)$ set
\[ t_s(x) := t^1_s(x) \mathcal{K}_{2s-2}(x - x^*), \]
where $x^*$ is a point of maximum of $|t^1_s(x)|$. Finally, set

$$t(x) := \sum_{s \in Y(n,d)} t_s(x) w(s, x).$$

Then we have

$$|t_s(x^*)| \gg 2^n,$$

and, therefore, by Nikol’skii’s inequality

$$\|t_s\|_2 \gg 2^{n/2}. \tag{5.9}$$

It follows from our definition of $Y(n,d)$ that the polynomials $t_s(x) w(s, x)$, $s \in Y(n,d)$, form an orthogonal system. This implies that

$$\|t\|^2_2 \gg 2^n n^{d-1}. \tag{5.10}$$

Relations (5.9) and (5.10) imply

$$\|t\|_p \gg 2^{n(1-1/p)n(d-1)/p}. \tag{5.11}$$

It is clear that

$$\|t\|_{H^p_q} \ll 2^{n(r+1-1/q)}. \tag{5.12}$$

Now, the bounds (5.11), (5.12) and the fact that $\Psi(t, X_m) = 0$ imply the required estimate in Theorem 5.1.

Theorem 5.1 is proved.

The inequality

$$\|t\|_{B_r^q, \theta} \ll 2^{n(r+1-1/q)n(d-1)/\theta}$$

and (5.11) imply the following result.

**Theorem 5.2.** For any recovery operator $\Psi(\cdot, X_m)$ with respect to an $(n, n-1)$-net $X_m$ and for $1 \leq q < p < \infty$ and $r > \beta$,

$$\Psi(B_r^q, \theta, X_m)_p \gg 2^{-n(r-\beta)n(d-1)(1/p-1/\theta)}, \quad \beta := \frac{1}{q} - \frac{1}{p}.$$
which is much weaker than Theorem 1.1. It would be interesting to obtain good
Lebesgue-type inequalities in the case $1 < p < 2$ for either the WCGA or some
other constructive methods.

The main results of this paper are on $m$-term approximation in the case $2 \leq p \leq \infty$. For $p \in [2, \infty)$ the situation is very good: we have a universal algorithm (WCGA),
which provides optimal (up to an extra $(\log m)^C(r,d)$ factor) $m$-term approximation
for all the classes $W^r_q$, $H^r_q$ and $B^r_{q,\theta}$. Also, there are constructive methods, based on
greedy algorithms, which provide the optimal rate for the above classes. However,
the upper bounds in, say, Theorem 1.4 hold for smoothness $r$ larger than the one
required for the embedding of $W^r_q$ into $L_p$. It would be interesting to find con-
structive methods, which provide the right orders of decay of $\sigma_m(W^r_q)_{p}$, $\sigma_m(H^r_q)_{p}$,
and $\sigma_m(B^r_{q,\theta})_{p}$ for small smoothness.

The case $p = \infty$ (approximation in the uniform norm) is very interesting and
difficult. The space $C(\mathbb{T}^d)$ is not a smooth Banach space. Therefore, the existing
greedy approximation theory does not apply directly to the case of approximation
in $L_{\infty}$. In particular, there is no analog of Theorem 1.1 in the case $p = \infty$. However,
for function classes with mixed smoothness there is a way around this problem.
As we showed in the proof of Theorem 2.4 we can use greedy algorithms in $L_p$ for
large $p$ to obtain bounds on $m$-term approximation in $L_{\infty}$. The price we pay for this
trick is an extra $(\log m)^{1/2}$ factor in the error bound. This extra factor results from the factor $p^{1/2}$ in the error bounds for approximation by greedy algorithms in $L_p$,
$2 \leq p < \infty$ (see Remark 2.1 and Theorem 2.2). An extra $(\log m)^{1/2}$ appears, as
a result of different techniques, in other upper bounds of asymptotic characteristics
of classes of functions with mixed smoothness, when we go from $p < \infty$ to $p = \infty$
(see, for instance, [1], Ch. 3, §3.6). Unfortunately, we do not have matching lower
bounds for our upper bounds for $m$-term approximation in $L_{\infty}$. A very special
case in Theorem 3.3 could be interpreted as a hint that we cannot get rid of this
extra $(\log m)^{1/2}$ for approximation in $L_{\infty}$.

We have discussed isotropic classes of functions with mixed smoothness.
‘Isotropic’ means that all the variables play the same role in the definition of our
smoothness classes. Anisotropic classes of functions with mixed smoothness are
of interest and importance in hyperbolic cross approximation theory. We give the
corresponding definitions.

Let $r = (r_1, \ldots, r_d)$ be such that

$$0 < r_1 = r_2 = \cdots = r_{\nu} < r_{\nu+1} \leq r_{\nu+2} \leq \cdots \leq r_d, \quad 1 \leq \nu \leq d.$$ 

For $x = (x_1, \ldots, x_d)$ set

$$F_r(x) := \prod_{j=1}^d F_{r_j}(x_j), \quad W^r_p := \{ f : f = \varphi * F_r, \|\varphi\|_p \leq 1 \}.$$ 

We now proceed to the classes $H^r_q$ and $B^r_{q,\theta}$. Define

$$\|f\|_{H^r_q} := \sup_s \|\delta_s(f)\|_q 2^{(r,s)}.$$
and for $1 \leq \theta < \infty$ define

$$\|f\|_{B^{r}_{q,\theta}} := \left( \sum_{s} (\|\delta_{s}(f)\|_{q,2}^{r,s})^{\theta} \right)^{1/\theta}.$$ 

We will write $B^{r}_{q,\infty} := H^{r}_{q}$. Denote the corresponding unit ball by

$$B^{r}_{q,\theta} := \{ f : \|f\|_{B^{r}_{q,\theta}} \leq 1 \}.$$

It is known that, in many problems where we want to estimate asymptotic characteristics, anisotropic classes of functions of $d$ variables with mixed smoothness behave in the same way as isotropic classes of functions of $\nu$ variables (see, for instance, [20]). It is clear that the above remark holds for lower bounds. To prove it for upper bounds, in some cases one needs to develop a special technique. The techniques developed in this paper work for the anisotropic classes as well. For instance, the main Lemma 2.1 can be replaced by the following lemma.

**Lemma 6.1.** Set $r := r_{1}$. For $f \in L_{1}$ set

$$f_{l,r} := \sum_{s : rl \leq (r,s) < rl+1} \delta_{s}(f), \quad l \in \mathbb{N}_{0}.$$ 

Consider the class

$$W^{r,a,b}_{A} := \{ f : \|f_{l,r}\|_{A} \leq 2^{-al} l^{(\nu-1)b} \}.$$ 

Then for $2 \leq p < \infty$ and $a > 0$ there is a constructive method, based on greedy algorithms, which provides the bound

$$\sigma_{m}(W^{r,a,b}_{A})_{p} \ll m^{-a-1/2} (\log m)^{(\nu-1)(a+b)}.$$ 

(6.1)

For $p = \infty$

$$\sigma_{m}(W^{r,a,b}_{A})_{\infty} \ll m^{-a-1/2} (\log m)^{(\nu-1)(a+b)+1/2}.$$ 

(6.2)

**Proposition 6.1.** The results in §2 hold for anisotropic classes of functions with mixed smoothness with $r = r_{1}$ and $d$ replaced by $\nu$.

---

**Bibliography**

[1] V. Temlyakov, *Greedy approximation*, Cambridge Monogr. Appl. Comput. Math., vol. 20, Cambridge Univ. Press 2011, xiv+418 pp.

[2] V. N. Temlyakov, “Greedy algorithms in Banach spaces”, *Adv. Comput. Math.* 14:3 (2001), 277–292.

[3] V. N. Temlyakov, “Weak greedy algorithms”, *Adv. Comput. Math.* 12:2–3 (2000), 213–227.

[4] V. N. Temlyakov, “Sparse approximation and recovery by greedy algorithms in Banach spaces”, *Forum Math. Sigma* 2 (2014), e12, 26 pp.; IMI preprint 2013:09, 2013, 27 pp.; 2013, arXiv:1303.6811v1.

[5] S. B. Stechkin, “On absolute convergence of orthogonal series”, *Dokl. Akad. Nauk SSSR* 102 (1955), 37–40. (Russian)
[6] R. S. Ismagilov, “Diameters of sets in normed linear spaces and the approximation of functions by trigonometric polynomials”, *Uspekhi Mat. Nauk* 29:3(177) (1974), 161–178; English transl. in *Russian Math. Surveys* 29:3 (1974), 169–186.

[7] V. E. Maiorov, “Trigonometric diameters of the Sobolev classes $W^r_p$ in the space $L_p$”, *Mat. Zametki* 40:2 (1986), 161–173; English transl. in *Math. Notes* 40:2 (1986), 590–597.

[8] R. A. DeVore and V. N. Temlyakov, “Nonlinear approximation by trigonometric sums”, *J. Fourier Anal. Appl.* 2:1 (1995), 29–48.

[9] E. D. Gluskin, “Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces”, *Mat. Sb. (N.S.)* 136(178):1(5) (1988), 85–96; English transl. in *Math. USSR-Sb.* 64:1 (1989), 85–96.

[10] Y. Makovoz, “On trigonometric $n$-widths and their generalization”, *J. Approx. Theory* 41:4 (1984), 361–366.

[11] É. S. Belinskii, “Approximation by a “floating” system of exponentials on classes of smooth periodic functions”, *Mat. Sb. (N.S.)* 132(174):1 (1987), 20–27; English transl. in *Math. USSR-Sb.* 60:1 (1988), 19–27.

[12] E. S. Belinskii, “Decomposition theorems and approximation by a “floating” system of exponentials”, *Trans. Amer. Math. Soc.* 350:1 (1998), 43–53.

[13] V. N. Temlyakov, “Approximation of periodic functions of several variables by bilinear forms”, *Izv. Akad. Nauk SSSR Ser. Mat.* 50:1 (1986), 137–155; English transl. in *Math. USSR-Izv.* 28:1 (1987), 133–150.

[14] B. S. Kashin and V. N. Temlyakov, “On best $m$-term approximations and the entropy of sets in the space $L_1$”, *Mat. Zametki* 56:5 (1994), 57–86; English transl. in *Math. Notes* 56:5 (1994), 1137–1157.

[15] V. N. Temlyakov, “Nonlinear Kolmogorov widths”, *Mat. Zametki* 63:6 (1998), 891–902; English transl. in *Math. Notes* 63:6 (1998), 785–795.

[16] V. N. Temlyakov, “An inequality for the entropy numbers and its application”, *J. Approx. Theory* 173 (2013), 110–121.

[17] S. J. Dilworth, D. Kutzarova and V. N. Temlyakov, “Convergence of some greedy algorithms in Banach spaces”, *J. Fourier Anal. Appl.* 8:5 (2002), 489–505.

[18] V. N. Temlyakov, “Greedy-type approximation in Banach spaces and applications”, *Constr. Approx.* 21:2 (2005), 257–292.

[19] A. S. Romanyuk, “Best $M$-term trigonometric approximations of Besov classes of periodic functions of several variables”, *Izv. RAN Ser. Mat.* 67:2 (2003), 61–100; English transl. in *Izv. Math.* 67:2 (2003), 265–302.

[20] V. N. Temlyakov, “Approximations of functions with bounded mixed derivative”, Trudy Mat. Inst. Steklov., vol. 178, 1986, pp. 3–113; English transl. in *Proc. Steklov Inst. Math.* 178 (1989), 1–121.

[21] V. N. Temlyakov, *Approximation of periodic functions*, Comput. Math. Anal. Ser., Nova Sci. Publ., Commack, NY 1993, x+419 pp.

[22] V. N. Temlyakov, “Estimates for the asymptotic characteristics of classes of functions with bounded mixed derivative or difference”, *A work collection of the All-Union school on function theory* (Dushanbe, August 1986), Trudy Mat. Inst. Steklov., vol. 189, Nauka, Moscow 1989, pp. 138–168; English transl. in *Proc. Steklov Inst. Math.* 189 (1990), 161–197.

[23] M. J. Donahue, C. Darken, L. Gurvits and E. Sontag, “Rate of convex approximation in non-Hilbert spaces”, *Constr. Approx.* 13:2 (1997), 187–220.

[24] É. S. Belinskii, “Approximation of functions of several variables by trigonometric polynomials with given number of harmonics, and estimates of $\varepsilon$-entropy”, *Anal. Math.* 15:2 (1989), 67–74.
[25] V. N. Temlyakov, “An inequality for trigonometric polynomials and its application for estimating the entropy numbers”, *J. Complexity* **11**:2 (1995), 293–307.

[26] V. N. Temlyakov, “Cubature formulas, discrepancy, and nonlinear approximation”, *J. Complexity* **19**:3 (2003), 352–391.

[27] V. N. Temlyakov, “On a way of obtaining lower estimates for the errors of quadrature formulas”, *Mat. Sb.* **181**:10 (1990), 1403–1413; English transl. in *Math. USSR-Sb.* **71**:1 (1992), 247–257.

[28] Dinh Dũng and T. Ullrich, “Lower bounds for the integration error for multivariate functions with mixed smoothness and optimal Fibonacci cubature for functions on the square”, *Math. Nachr.* **288**:7 (2015), 743–762.

[29] V. N. Temlyakov, “On approximate recovery of functions with bounded mixed derivative”, *J. Complexity* **9**:1 (1993), 41–59.

[30] V. N. Temlyakov, “Approximate recovery of periodic functions of several variables”, *Mat. Sb. (N.S.)* **128(170)**:2(10) (1985), 256–268; English transl. in *Math. USSR-Sb.* **56**:1 (1987), 249–261.

[31] V. N. Temlyakov, “An inequality for trigonometric polynomials and its application for estimating the Kolmogorov widths”, *East J. Approx.* **2**:2 (1996), 253–262.

Vladimir N. Temlyakov
University of South Carolina, Columbia, SC, USA;
Steklov Mathematical Institute
of Russian Academy of Sciences, Moscow, Russia

E-mail: temlyak@math.sc.edu

Received 31/DEC/14