Constr. Approx. (2002) 18: 529–550
DOI: 10.1007/s00365-002-0514-1

CONSTRUCTIVE APPROXIMATION

Universal Bases and Greedy Algorithms for Anisotropic Function Classes

V. N. Temlyakov

Abstract. We suggest a three-step strategy to find a good basis (dictionary) for nonlinear m-term approximation. The first step consists of solving an optimization problem of finding a near best basis for a given function class F, when we optimize over a collection $\mathbf{D}$ of bases (dictionaries). The second step is devoted to finding a universal basis (dictionary) $\mathcal{D}_u \in \mathbf{D}$ for a given pair $(\mathcal{F}, \mathbf{D})$ of collections: $\mathcal{F}$ of function classes and $\mathbf{D}$ of bases (dictionaries). This means that $\mathcal{D}_u$ provides near optimal approximation for each class F from a collection $\mathcal{F}$. The third step deals with constructing a theoretical algorithm that realizes near best m-term approximation with regard to $\mathcal{D}_u$ for function classes from $\mathcal{F}$.

In this paper we work this strategy out in the model case of anisotropic function classes and the set of orthogonal bases. The results are positive. We construct a natural tensor–product–wavelet-type basis and prove that it is universal. Moreover, we prove that a greedy algorithm realizes near best m-term approximation with regard to this basis for all anisotropic function classes.

1. Introduction

In this paper we discuss a general approach of how to choose a good basis (dictionary) for approximation. This approach consists of several steps. We have worked it out in the case of multivariate anisotropic function classes. We concentrate here on nonlinear approximation and compare realizations of this approach for linear and nonlinear approximations. The first step in this approach is an optimization problem. In both cases (linear and nonlinear), we begin with a function class F in a given Banach space X with a norm $\|\cdot\| := \|\cdot\|_X$. A classical example of the optimization problem in the linear case is the problem of finding (estimating) the Kolmogorov width

$$d_m(F, X) := \inf_{\varphi_1,\dots,\varphi_m} \sup_{f\in F} \inf_{c_1,\dots,c_m} \left\| f - \sum_{j=1}^{m} c_j\varphi_j \right\|.$$

This concept allows us to choose among various Chebyshev methods (best approximation) having the same dimension of the approximating subspace, the one which

Date received: August 10, 2000. Date revised: September 14, 2001. Date accepted: May 6, 2002. Communicated by Ronald A. DeVore.
AMS classification: 41A25, 41A46.
Key words and phrases: Best m-term approximation, Greedy algorithms, Thresholding, Optimization, Universality.

has the best accuracy. The asymptotic behavior (in the sense of order) of the sequence $\{d_m(F,X)\}_{m=1}^{\infty}$ is known for a number of function classes and Banach spaces. It turns out that in many cases, for instance, in the case where $F = W_p^r$ is a standard Sobolev class and $X = L_p$, the optimal (in the sense of order) m-dimensional subspaces can be formed as subspaces spanned by m elements from one orthogonal system. We describe this for the multivariate periodic Hölder–Nikol'skii classes NH_q^R. We define these classes in the following way. The class NH_q^R, $R = (R_1, \dots, R_d)$ and $1 \le q \le \infty$, is the set of periodic functions $f \in L_q([0, 2\pi]^d)$ such that for each $l_j = [R_j] + 1$, $j = 1, \dots, d$, the following relations hold

$$\|f\|_q \le 1, \qquad \|\Delta_t^{l_j,j} f\|_q \le |t|^{R_j}, \qquad j = 1, \dots, d, \tag{1.1}$$

where $\Delta_t^{l,j}$ is the lth difference with step t in the variable x_j. In the case $d = 1$, NH_q^R coincides with the standard Hölder class H_q^R. It is known (see, for instance, [13]) that

$$d_m(NH_q^R, L_q) \asymp m^{-g(R)}, \qquad 1 \le q \le \infty, \tag{1.2}$$

where

$$g(R) := \left(\sum_{j=1}^{d} R_j^{-1} \right)^{-1}.$$

It is also known that the subspaces of trigonometric polynomials $\mathcal{T}(R, l)$ with frequencies k satisfying the inequalities

$$|k_j| \le 2^{g(R)l/R_j}, \qquad j = 1, \dots, d,$$

can be chosen to realize (1.2). In this case l is set to be the largest satisfying inequality $\dim \mathcal{T}(R, l) \le m$. We stress here that optimal (in the sense of order) subspaces $\mathcal{T}(R, l)$ are different for different R and formed from the same (trigonometric) system.

A nonlinear analog of the Kolmogorov m-width setting was discussed in [15]. In [15] we replace the Chebyshev method of best approximation from a linear subspace of dimension m by best m-term approximation with regard to a given orthogonal basis and optimize over all orthogonal bases. Thus, in the nonlinear case we formulate an optimization problem in a Banach space X for a pair of function class F and collection $\mathbf{D}$ of bases (dictionaries) $\mathcal{D}$:

$$\begin{aligned}
\sigma_m(f, \mathcal{D})_X &:= \inf_{g_i \in \mathcal{D}, c_i, i=1,\dots,m} \left\| f - \sum_{i=1}^{m} c_i g_i \right\|_X, \\
\sigma_m(F, \mathcal{D})_X &:= \sup_{f \in F} \sigma_m(f, \mathcal{D})_X, \\
\sigma_m(F, \mathbf{D})_X &:= \inf_{\mathcal{D} \in \mathbf{D}} \sigma_m(F, \mathcal{D})_X.
\end{aligned}$$

In this paper we consider only the case $\mathbf{D} = \mathbf{O}$—the set of all orthogonal bases on a given domain. In Section 3 we prove that

$$\sigma_m(NH_q^R, \mathbf{O})_{L_p} \asymp m^{-g(R)} \tag{1.3}$$

for

$$1 < q < \infty, \quad 2 \le p < \infty, \qquad g(R) > (1/q - 1/p)_+.$$

It is interesting to remark that we cannot prove anything like (1.3) for L_p with $p < 2$. We proved (see [6]) that there exists $\Phi \in \mathbf{O}$ such that for any $f \in L_1(0, 1)$ we have $\sigma_1(f, \Phi)_{L_1} = 0$. The proof from [6] also works for L_p, $p < 2$, instead of L_1. The following remark has been made in [15].

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

This remark means that to obtain nontrivial lower bounds for $\sigma_m(f, \Phi)_{L_p}$, $p < 2$, we need to impose additional restrictions on $\Phi \in \mathbf{O}$.

It is important to remark that the basis U^d studied in [15] realizes (1.3) for all R (see the definition of U^d in Section 3). We introduce the following definition of a universal dictionary:

Definition 1.1. Let two collections $\mathcal{F}$ (of function classes) and $\mathbf{D}$ (of dictionaries) be given. We say that $\mathcal{D} \in \mathbf{D}$ is universal for the pair $(\mathcal{F}, \mathbf{D})$ if there exists a constant C which may depend only on $\mathcal{F}$, $\mathbf{D}$, and X such that for any $F \in \mathcal{F}$ we have

$$\sigma_m(F, \mathcal{D})_X \le C\sigma_m(F, \mathbf{D})_X.$$

This is a new concept in nonlinear approximation. The following observation motivates our interest in this setting. In practice we often do not know the exact smoothness class F where our input function (signal, image) comes from. Instead, we often know that our function comes from a class of certain structure, for instance, an anisotropic Sobolev class. This is exactly the situation we are dealing with in the universal dictionary setting. So, if for a collection $\mathcal{F}$ there exists a universal dictionary $\mathcal{D}_u \in \mathbf{D}$, it is an ideal situation. We can use this universal dictionary $\mathcal{D}_u$ in all cases and we know that it adjusts automatically to the best smoothness class $F \in \mathcal{F}$ which contains a function under approximation. Next, if a pair $(\mathcal{F}, \mathbf{D})$ does not allow a universal dictionary we have a trade-off between universality and accuracy.

The second step in our approach is to look for a universal basis (dictionary) for approximation. The above-mentioned result on the basis U^d means that U^d is universal for the pair $(\mathcal{F}_q([A, B]), \mathbf{O})$ and the space $X = L_p([0, 2\pi]^d)$ for $A, B \in \mathbf{Z}^d_+$ such that $g(A) > (1/q - 1/p)_+$, $1 < q < \infty$, $2 \le p < \infty$, where

$$\mathcal{F}_q([A, B]) := \{NH^R_q : 0 < A_j \le R_j \le B_j < \infty, j = 1, \dots, d\}.$$

It is interesting to compare this result on a universal basis in nonlinear approximation with the corresponding result in the linear setting. We define the index $\kappa(m, \mathcal{F}, X)$ of universality for a collection $\mathcal{F}$ with respect to the Kolmogorov width in X:

$$\kappa(m, \mathcal{F}, X) := L(m, \mathcal{F}, X)/m,$$

where $L(m, \mathcal{F}, X)$ is the smallest number among those L for which there is a system of functions $\{\varphi_i\}_{i=1}^L$ such that for each $F \in \mathcal{F}$ we have

$$\sup_{f\in F} \inf_{c_1,\dots,c_L} \left\| f - \sum_{i=1}^L c_i\varphi_i \right\| \le d_m(F, X).$$

It is proved in [12] (see also [13, Ch. 3, S.5]) that for any $A, B \in \mathbf{Z}_+^d$ such that $B_j > A_j$, $j = 1, \dots, d$, we have

$$\kappa(m, \mathcal{F}_p([A, B]), L_p) \asymp (\log m)^{d-1}, \qquad 1 < p < \infty. \tag{1.4}$$

The estimate (1.4) implies that there is no Chebyshev methods universal for a nontrivial collection of anisotropic function classes. Thus, from the point of view of the existence of universal methods the nonlinear setting has an advantage over the linear setting.

After two steps of realizing our approach in the nonlinear approximation we get a universal dictionary $\mathcal{D}_u$ for a collection of function classes $\mathcal{F}$, say, U^d for $\mathcal{F}_q([A, B])$. This means that the dictionary $\mathcal{D}_u$ is well-designed for best m-term approximation of functions from function classes in the given collection. The third step is to find an algorithm (theoretical first) to realize best (near best) m-term approximation with regard to $\mathcal{D}_u$. It turns out that in the model case of $\mathcal{F}_q([A, B])$ and the basis U^d there is a simple algorithm which realizes near best m-term approximation for classes NH_q^R. This is a thresholding or greedy-type algorithm. We give the definition of a greedy algorithm for a general basis. Let $\Psi := \{\psi_k\}_{k=1}^\infty$ be a basis for X. Represent $f \in X$ in the form

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

Then $\|c_k(f, \Psi)\psi_k\| \to 0$ as $k \to \infty$. We enumerate the summands in decreasing order

$$\|c_{k_1}(f, \Psi)\psi_{k_1}\| \ge \|c_{k_2}(f, \Psi)\psi_{k_2}\| \ge \cdots$$

and define the mth greedy approximant as

$$G_m^X(f, \Psi) := \sum_{i=1}^m c_{k_i}(f, \Psi)\psi_{k_i}.$$

We prove in Sections 2 and 3 that (1.3) can be realized by the greedy algorithm $G_m^{L_p}(f, U^d)$. Namely,

$$\sup_{f\in NH_q^R} \|f - G_m^{L_p}(f, U^d)\|_{L_p} \asymp m^{-g(R)}, \tag{1.5}$$

$$\text{for} \quad 1 < q, p < \infty, \quad g(R) > (1/q - 1/p)_+.$$

In this paper we realize three steps of our approach in the model case of periodic anisotropic function classes NH_q^R. However, we present the results in sufficiently general form to include wavelet-type bases.

Section 4 is devoted to one more good property of the basis U^d. We prove there that a soft thresholding algorithm with regard to an unconditional basis is a mapping from the Lipschitz class.

Let us agree to denote by C various positive absolute constants and by C, with arguments or indexes ($C(q, p)$, C_r, and so on), positive numbers which depend on the arguments indicated. For two nonnegative sequences $a = \{a_n\}_{n=1}^{\infty}$ and $b = \{b_n\}_{n=1}^{\infty}$ the relation (order inequality) $a_n \ll b_n$ means that there is a number $C(a, b)$ such that for all n we have $a_n \le C(a, b)b_n$; and the relation $a_n \asymp b_n$ means that $a_n \ll b_n$ and $b_n \ll a_n$. The sign $\ll$ will be used for the sake of brevity in estimates of the various characteristics of functions.

2. The Upper Estimates for Anisotropic Function Classes

We consider in this section a basis $\Psi := \{\psi_I\}_{I\in D}$, enumerated by dyadic intervals I of $[0, 1]^d$, $I = I_1 \times \cdots \times I_d$, I_j is a dyadic interval of $[0, 1]$, $j = 1, \dots, d$, which satisfies certain properties (see (2.1)–(2.4) below). Let $L_p := L_p(\Omega)$, $\Omega = [0, 1]^d$, $\mathbf{T}^d$, and alike with normalized Lebesgue measure on Ω, $|\Omega| = 1$. First of all we assume that for all $1 < q, p < \infty$, and $I \in D$, $D := D([0, 1]^d)$ is the set of all dyadic intervals of $[0, 1]^d$, we have

$$\|\psi_I\|_p \asymp \|\psi_I\|_q |I|^{1/p-1/q}, \tag{2.1}$$

with constants independent of I. This property can be easily checked for a given basis.

Next, assume that for any $s = (s_1, \dots, s_d) \in \mathbf{Z}^d$, $s_j \ge 0$, $j = 1, \dots, d$, and any $\{c_I\}$ we have, for $1 < p < \infty$,

$$\left\| \sum_{I\in D_s} c_I \psi_I \right\|_p^p \asymp \sum_{I\in D_s} \|c_I\psi_I\|_p^p, \tag{2.2}$$

where

$$D_s := \{I = I_1 \times \cdots \times I_d \in D : |I_j| = 2^{-s_j}, j = 1, \dots, d\}.$$

This assumption allows us to estimate the L_p-norm of a dyadic block in terms of the coefficients $\{c_I\}_{I\in D_s}$.

The third assumption is that Ψ is a basis satisfying the Littlewood–Paley inequality. This means the following. Let $1 < p < \infty$ and $f \in L_p$ has an expansion

$$f = \sum_I f_I \psi_I.$$

We assume that

$$\lim_{\min_j \mu_j \to \infty} \left\| f - \sum_{s_j \le \mu_j, j=1,\dots,d} \sum_{I\in D_s} f_I\psi_I \right\|_p = 0, \tag{2.3}$$

and

$$\|f\|_p \asymp \left\| \left(\sum_s \left| \sum_{I\in D_s} f_I\psi_I \right|^2 \right)^{1/2} \right\|_p. \tag{2.4}$$

Let $\mu \in \mathbf{Z}^d$, $\mu_j \geq 0$, $j = 1, \dots, d$. Denote by $\Psi(\mu)$ the subspace of polynomials of the form

$$\psi = \sum_{s_j \leq \mu_j, j=1,\dots,d} \sum_{I \in D_s} c_I \psi_I.$$

Our primary goal is to study the wavelet and wavelet-type bases Ψ. The above-described framework of studying bases satisfying (2.1)–(2.4) should be considered as a convenient way to work simultaneously with wavelet bases and also bases like the basis U defined below (see Section 3) that is a wavelet-type basis. We begin studying the approximative properties of Ψ satisfying (2.1)–(2.4) by two lemmas.

Lemma 2.1. *Let $1 < q < p < \infty$. Then for any $f \in \Psi(\mu)$ and $h > 0$ we have*

$$\#\{I : \|f_I \psi_I\|_p \geq h\} \ll \|f\|_q^q h^{-q} 2^{(1-q/p)\|\mu\|_1}, \tag{2.5}$$

with a constant independent of f, h, μ.

Proof. Denote

$$A(f,h) := \{I : \|f_I \psi_I\|_p \geq h\}, \qquad N(f,h) := \#A(f,h),$$

and

$$A_s(f,h) := A(f,h) \cap D_s, \qquad N_s(f,h) := \#A_s(f,h).$$

We estimate first $N_s(f,h)$. Denote

$$\delta_s(f) := \sum_{I \in D_s} f_I \psi_I.$$

By (2.2) and (2.1) we have

$$\begin{aligned} \|\delta_s(f)\|_q^q &= \left\| \sum_{I \in D_s} f_I \psi_I \right\|_q^q \asymp \sum_{I \in D_s} \|f_I \psi_I\|_q^q \\ &\geq \sum_{I \in A_s(f,h)} \|f_I \psi_I\|_q^q \gg \sum_{I \in A_s(f,h)} \|f_I \psi_I\|_p^q 2^{(q/p-1)\|s\|_1} \\ &\geq h^q 2^{(q/p-1)\|s\|_1} N_s(f,h). \end{aligned}$$

Thus,

$$N_s(f,h) \ll \|\delta_s(f)\|_q^q h^{-q} 2^{(1-q/p)\|s\|_1}. \tag{2.6}$$

In order to derive the estimate (2.5) from (2.6) we need the following two inequalities:

$$\left(\sum_s \|\delta_s(f)\|_p^{p_l} \right)^{1/p_l} \ll \|f\|_p \ll \left(\sum_s \|\delta_s(f)\|_p^{p_u} \right)^{1/p_u} \tag{2.7}$$

with $p_l := \max(2, p)$ and $p_u := \min(2, p)$.

The relation (2.7) is a corollary of the Littlewood–Paley inequalities (2.4) and the following known inequalities (see, for instance, [8, p. 73]). We will give a proof of these useful inequalities for completeness. ∎

Lemma 2.2. *For any finite collection $\{f_s\}$ of functions in L_p, $1 \le p \le \infty$, we have*

$$\left(\sum_s \|f_s\|_p^{p_l}\right)^{1/p_l} \le \left\|\left(\sum_s |f_s|^2\right)^{1/2}\right\|_p \le \left(\sum_s \|f_s\|_p^{p_u}\right)^{1/p_u}.$$

Proof. We prove first the upper estimate. For $p = \infty$ it is obvious. Let $2 \le p < \infty$, then

$$\left\|\left(\sum_s |f_s|^2\right)^{1/2}\right\|_p = \left\|\sum_s |f_s|^2\right\|_{p/2}^{1/2} \le \left(\sum_s \||f_s|^2\|_{p/2}\right)^{1/2} = \left(\sum_s \|f_s\|_p^2\right)^{1/2}.$$

Let now $1 \le p \le 2$. Then

$$\left\|\left(\sum_s |f_s|^2\right)^{1/2}\right\|_p = \left(\int_\Omega \left(\sum_s |f_s|^2\right)^{p/2}\right)^{1/p}$$
$$\le \left(\int_\Omega \sum_s |f_s|^p\right)^{1/p} = \left(\sum_s \|f_s\|_p^p\right)^{1/p}.$$

We proceed now to the lower estimate. Again for $p = \infty$ it is obvious. Let $2 \le p < \infty$. Then we have

$$\left\|\left(\sum_s |f_s|^2\right)^{1/2}\right\|_p \ge \left\|\left(\sum_s |f_s|^p\right)^{1/p}\right\|_p = \left(\sum_s \|f_s\|_p^p\right)^{1/p}.$$

For $1 \le p \le 2$, we have

$$\|\{\|f_s\|_p\}\|_{l_2} = \left\|\left\{\int_\Omega |f_s|^p\right\}\right\|_{l_{2/p}}^{1/p}$$
$$\le \left(\int_\Omega \|\{|f_s|^p\}\|_{l_{2/p}}\right)^{1/p} = \left\|\left(\sum_s |f_s|^2\right)^{1/2}\right\|_p.$$

Lemma 2.2 is now proved. ■

We return to the proof of Lemma 2.1. Using (2.6) and (2.7) we obtain, in the case $q < 2$,

$$N(f, h) = \sum_{s \le \mu} N_s(f, h) = \sum_{s \le \mu} N_s(f, h) 2^{-(1-q/p)\|s\|_1} 2^{(1-q/p)\|s\|_1}$$
$$\le \left(\sum_{s \le \mu} (N_s(f, h) 2^{-(1-q/p)\|s\|_1})^{q_l/q}\right)^{q/q_l} \left(\sum_{s \le \mu} 2^{(1-q/p)\|s\|_1 (1-q/q_l)^{-1}}\right)^{1-q/q_l}$$
$$\ll h^{-q} \left(\sum_{s \le \mu} \|\delta_s(f)\|_q^{q_l}\right)^{q/q_l} 2^{(1-q/p)\|\mu\|_1} \ll h^{-q} \|f\|_q^q 2^{(1-q/p)\|\mu\|_1}.$$

In the case $2 \le q < \infty$, we similarly obtain

$$N(f,h) \le \left(\sum_{s\le\mu} N_s(f,h)2^{-(1-q/p)\|s\|_1}\right)2^{(1-q/p)\|\mu\|_1}$$

$$\ll h^{-q}\left(\sum_{s\le\mu}\|\delta_s(f)\|_q^q\right)2^{(1-q/p)\|\mu\|_1} \ll h^{-q}\|f\|_q^q 2^{(1-q/p)\|\mu\|_1}.$$

This completes the proof of Lemma 2.1. ■

Denote by T_h^p the thresholding mapping:

$$T_h^p(f) := \sum_{I:\|f_I\psi_I\|_p\ge h} f_I\psi_I.$$

Lemma 2.3. *Let $1 < q < p < \infty$. Then for each $f \in \Psi(\mu)$ we have*

$$\|f - T_h^p(f)\|_p \ll h^{1-q/p}(\|f\|_q 2^{(1/q-1/p)\|\mu\|_1})^{q/p}.$$

Proof. We estimate first $\|\delta_s(f - T_h^p(f))\|_p$. We have

$$\|\delta_s(f - T_h^p(f))\|_p^p \ll \sum_{I\in D_s\setminus A_s(f,h)}\|f_I\psi_I\|_p^p \le h^{p-q}\sum_{I\in D_s}\|f_I\psi_I\|_p^q$$

$$\ll h^{p-q}\sum_{I\in D_s}\|f_I\psi_I\|_q^q 2^{(1-q/p)\|s\|_1} \ll h^{p-q}\|\delta_s(f)\|_q^q 2^{(1-q/p)\|s\|_1}.$$

Therefore,

$$\|\delta_s(f - T_h^p(f))\|_p \ll h^{1-q/p}(\|\delta_s(f)\|_q 2^{(1/q-1/p)\|s\|_1})^{q/p}. \tag{2.8}$$

Next, using the Hölder inequality with a parameter $pq_l/p_u q$, we get from (2.7) and (2.8) that

$$\|f - T_h^p(f)\|_p \ll \left(\sum_{s\le\mu}\|\delta_s(f - T_h^p(f))\|_p^{p_u}\right)^{1/p_u}$$

$$\ll h^{1-q/p}\left(\sum_{s\le\mu}\|\delta_s(f)\|_q^{qp_u/p}2^{(1/q-1/p)\|s\|_1 qp_u/p}\right)^{1/p_u}$$

$$\ll h^{1-q/p}\left(\sum_{s\le\mu}\|\delta_s(f)\|_q^{q_l}\right)^{q/(pq_l)}(2^{(1/q-1/p)\|\mu\|_1})^{q/p}$$

$$\ll h^{1-q/p}(\|f\|_q 2^{(1/q-1/p)\|\mu\|_1})^{q/p}.$$

This completes the proof of Lemma 2.3. ■

Remark 2.1. Let $h > 0$ be given. Denote

$$A^{=}(f, h) := \{I : \|f_I \psi_I\|_p = h\}.$$

Take any subset $Y \subseteq A^{=}(f, h)$ and denote

$$T^p_{h,Y}(f) := \sum_{I \in A(f,h) \setminus Y} f_I \psi_I.$$

It is not difficult to see that Lemma 2.3 holds with T^p_h replaced by $T^p_{h,Y}$ with any $Y \subseteq A^{=}(f, h)$ and the constant in the estimate does not depend on Y.

We now define a function class. Let $R = (R_1, \dots, R_d)$, $R_j > 0$, $j = 1, \dots, d$, and as above

$$g(R) := \left(\sum_{j=1}^{d} R_j^{-1}\right)^{-1}.$$

For natural numbers l denote

$$\Psi(R, l) := \Psi(\mu), \qquad \mu_j = [g(R)l/R_j], \qquad j = 1, \dots, d.$$

We define the class $H^R_q(\Psi)$ as the set of functions $f \in L_q$ representable in the form

$$f = \sum_{l=1}^{\infty} t_l, \quad t_l \in \Psi(R, l), \qquad \|t_l\|_q \le 2^{-g(R)l},$$

and denote

$$H^R_q(\Psi)C := \{f : f/C \in H^R_q(\Psi)\}.$$

Theorem 2.1. *Let* $1 < q, p < \infty$ *and* $g(R) > (1/q - 1/p)_+$. *Then for* Ψ *satisfying* (2.1)–(2.4) *we have*

$$\sup_{f \in H^R_q(\Psi)} \|f - G^{L_p}_m(f, \Psi)\|_p \ll m^{-g(R)}.$$

Proof. We need some simple properties of the expansions of functions in $H^R_q(\Psi)$. Denote

$$S(f, R, l) := \sum_{s_j \le [g(R)l/R_j], j=1,\dots,d} \delta_s(f),$$

$$f_{R,l} := S(f, R, l+1) - S(f, R, l).$$

It is easy to derive from the definition of $H^R_q(\Psi)$ that

$$\|f - S(f, R, l)\|_q \ll 2^{-g(R)l} \quad \text{and} \quad \|f_{R,l}\|_q \ll 2^{-g(R)l}. \tag{2.9}$$

We consider first the case $q < p$. Take $h > 0$ and specify n such that

$$2^{-(n+1)(g(R)+1/p)} < h \le 2^{-n(g(R)+1/p)}.$$

Then, for a function $f \in H_q^R(\Psi)$ by (2.9) and Lemma 2.1, we obtain

$$\#\{I : \|f_I\psi_I\|_p \ge h\} \ll 2^n + h^{-q}\sum_{l\ge n} 2^{-g(R)lq+(1-q/p)l} \ll 2^n. \tag{2.10}$$

We now estimate the L_p-norm of

$$f_h := f - T_h^p(f).$$

We have

$$\|f_h\|_p \le \|S(f_h, R, n)\|_p + \sum_{l\ge n}\|S(f_h, R, l+1) - S(f_h, R, l)\|_p. \tag{2.11}$$

By (2.9) and Lemma 2.3 we get

$$\|S(f_h, R, l+1) - S(f_h, R, l)\|_p \ll h^{1-q/p}2^{(-g(R)l+(1/q-1/p)l)q/p}. \tag{2.12}$$

For $S(f_h, R, n)$ we have

$$\begin{aligned} \|S(f_h, R, n)\|_p &\le \sum_{s_j\le[g(R)n/R_j],\, j=1,\dots,d} \|\delta_s(f_h)\|_p \\ &\ll \sum_{s_j\le[g(R)n/R_j],\, j=1,\dots,d} h2^{\|s\|_1/p} \ll h2^{n/p}. \end{aligned} \tag{2.13}$$

Combining (2.12) and (2.13) we get, from (2.11),

$$\|f_h\|_p \ll 2^{-g(R)n}. \tag{2.14}$$

Taking into account (2.10) and Remark 2.1 we obtain from here the estimate in Theorem 2.1 for $q < p$. It is clear this implies the general case $1 < q, p < \infty$. ∎

3. Approximation of Anisotropic Hölder–Nikol'skii Classes

Here we study m-term approximation in the L_p-norm of functions from classes NH_q^R with regard to the basis $U^d := U \times \cdots \times U$.

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

$$\begin{aligned} U_n^+(x) &:= \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x}-1}{e^{ix}-1}, \qquad && n = 0, 1, 2, \dots, \\ U_{n,k}^+(x) &:= e^{i2^n x}U_n^+(x - 2\pi k2^{-n}), && k = 0, 1, \dots, 2^n - 1, \\ U_{n,k}^-(x) &:= e^{-i2^n x}U_n^+(-x + 2\pi k2^{-n}), && k = 0, 1, \dots, 2^n - 1. \end{aligned}$$

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

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

and

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

Denote

$$D^+_n := \{I : I = [(k+\tfrac{1}{2})2^{-n}, (k+1)2^{-n}), k = 0, 1, \dots, 2^n - 1\}$$

and

$$D^-_n := \{I : I = [k2^{-n}, (k+1/2)2^{-n}), k = 0, 1, \dots, 2^n - 1\},$$
$$D_0 := [0, 1), \qquad D := \bigcup_{n \ge 0} (D^+_n \cup D^-_n) \cup D_0.$$

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

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

We use the notations, for $f \in L_1$,

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

and

$$\delta^+_s(f) := \sum_{k=2^s}^{2^{s+1}-1} \hat{f}(k)e^{ikx}, \qquad \delta^-_s(f) := \sum_{k=-2^{s+1}+1}^{-2^s} \hat{f}(k)e^{ikx}, \qquad \delta_0(f) := \hat{f}(0).$$

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

$$\delta^+_s(f) = \sum_{I \in D^+_s} f_I U_I, \qquad \delta^-_s(f) = \sum_{I \in D^-_s} f_I U_I, \qquad \delta_0(f) = f_{[0,1)}.$$

Moreover, the following analog of Marcinkiewicz's theorem holds

$$\|\delta^+_s(f)\|_p^p \asymp \sum_{I \in D^+_s} \|f_I U_I\|_p^p, \qquad \|\delta^-_s(f)\|_p^p \asymp \sum_{I \in D^-_s} \|f_I U_I\|_p^p, \tag{3.1}$$

for $1 < p < \infty$ with constants depending only on p. We note that (3.1) and the boundedness of operators δ_s^+, δ_s^-, as operators from L_p into L_p, $1 < p < \infty$, imply

$$\|\delta_s^+(f) + \delta_s^-(f)\|_p^p \asymp \sum_{I \in D_s^+ \cup D_s^-} \|f_I U_I\|_p^p,$$

that is the property (2.2) from Section 2. Indeed, we have on the one hand $\|\delta_s^+(f) + \delta_s^-(f)\|_p \le \|\delta_s^+(f)\|_p + \|\delta_s^-(f)\|_p$ and, on the other hand, we have

$$\|\delta_s^+(f)\|_p = \|\delta_s^+(\delta_s^+(f) + \delta_s^-(f))\|_p \le C_p \|\delta_s^+(f) + \delta_s^-(f)\|_p$$

and the same inequality for $\|\delta_s^-(f)\|_p$ that gives the lower estimate.

We remark that

$$\|U_I\|_p \asymp |I|^{1/p-1/2}, \qquad 1 < p \le \infty, \tag{3.2}$$

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

$$\|U_I\|_p \asymp \|U_I\|_q |I|^{1/p-1/q}. \tag{3.3}$$

This relation gives the property (2.1) from Section 2. In the multivariate case of $x = (x_1, \dots, x_d)$ we define the system U^d as the tensor product of the univariate systems U. Let $I = I_1 \times \cdots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

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

For $s = (s_1, \dots, s_d)$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$ with $\varepsilon_j = +$ or $-$ if $s_j > 0$ and $\varepsilon_j = +, -$ or 0 if $s_j = 0$ denote

$$D_s^\varepsilon := \{I : I = I_1 \times \cdots \times I_d, I_j \in D_{s_j}^{\varepsilon_j}, D_0^0 := D_0, j = 1, \dots, d\}.$$

It is easy to see that (3.2) and (3.3) are also true in the multivariate case. It is not difficult to derive from (3.1) that for any ε we have

$$\|\delta_s^\varepsilon(f)\|_p^p \asymp \sum_{I \in D_s^\varepsilon} \|f_I U_I\|_p^p,$$

and

$$\left\|\sum_\varepsilon \delta_s^\varepsilon(f)\right\|_p^p \asymp \sum_{I \in \cup_\varepsilon D_s^\varepsilon} \|f_I U_I\|_p^p, \qquad 1 < p < \infty, \tag{3.4}$$

with constants depending only on p and d. Here we denote

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

where

$$\rho(s, \varepsilon) := \varepsilon_1[2^{s_1}, 2^{s_1+1} - 1) \times \cdots \times \varepsilon_d[2^{s_d}, 2^{s_d+1} - 1).$$

The convergence

$$\lim_{\min_j \mu_j \to \infty} \left\| f - \sum_{s_j \le \mu_j, j=1,\dots,d} \sum_{\varepsilon} \delta_s^\varepsilon(f) \right\|_p = 0, \qquad 1 < p < \infty, \tag{3.5}$$

and the Littlewood–Paley inequalities

$$\|f\|_p \asymp \left\| \left(\sum_s \left| \sum_\varepsilon \delta_s^\varepsilon(f) \right|^2 \right)^{1/2} \right\|_p \tag{3.6}$$

are well-known. Thus U^d satisfies the properties (2.1)–(2.4) from Section 2.

Denote, for given R and n,

$$E_n^R(f)_p := \inf_{t \in \mathcal{T}(R,n)} \|f - t\|_p,$$

where the trigonometric subspaces $\mathcal{T}(R, n)$ are defined in Section 1. It is known (see [13, Ch. 2, S.3]) that for $f \in NH_q^R$ we have

$$E_n^R(f)_q \ll 2^{-g(R)n}. \tag{3.7}$$

This implies that, for some $C > 0$, we have

$$NH_q^R \subset H_q^R(U^d)C := \{f : f/C \in H_q^R(U^d)\}.$$

Therefore, Theorem 2.1 gives

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

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

We now discuss a question of what other systems Ψ satisfying (2.1)–(2.4) are also good for approximating the classes NH_q^R. The following lemma combined with Theorem 2.1 gives a sufficient condition.

Lemma 3.1. *Let $R = (R_1, \dots, R_d) \in \mathbf{Z}_+^d$, and let $1 \le q \le \infty$ be given, and let A be a number such that $A > R_j$, $j = 1, \dots, d$. Assume that a basis $\Phi := \{\varphi_I\}_{I \in D([0,1])}$ of functions on a single variable has the following approximative property. For any $0 < r < A$ we have*

$$E_n(H_q^r, \Phi)_q := \sup_{f \in H_q^r} \inf_{c_I} \left\| f - \sum_{|I| \ge 2^{-n}} c_I \varphi_I \right\|_q \ll 2^{-rn}.$$

Then for $\Phi^d := \Phi \times \cdots \times \Phi$ we have

$$E_n^R(NH_q^R, \Phi^d)_q := \sup_{f \in NH_q^R} \inf_{c_I} \left\| f - \sum_{|I_j| \ge 2^{-ng(R)/R_j}, j=1,\dots,d} c_I \varphi_I \right\|_q \ll 2^{-g(R)n}$$

and, for some constant C,

$$NH_q^R \subset H_q^R(\Phi)C.$$

Proof. We get from (3.7) that each $f \in NH_q^R$ has a representation

$$f = \sum_{l=1}^{\infty} t_l, \quad t_l \in \mathcal{T}(R, l), \qquad \|t_l\|_q \ll 2^{-g(R)l}. \tag{3.8}$$

We study first the multivariate analogs of approximation of functions on a single variable. Fix $1 \leq j \leq d$ and define

$$E_{n,j}(f, \Phi)_q := \inf_{c_I(x^j)} \left\| f(x) - \sum_{|I| \geq 2^{-n}} c_I(x^j)\varphi_I(x_j) \right\|_q,$$

where $x^j := (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d)$.

For any trigonometric polynomial $t \in \mathcal{T}(R, l)$, we can estimate $E_{n,j}(t, \Phi)_q$ using the following two arguments. The first one is trivial

$$E_{n,j}(t, \Phi)_q \leq \|t\|_q. \tag{3.9}$$

The second one uses the Bernstein inequality: for any x^j we have

$$\|D_{x_j}^r t(\cdot, x^j)\|_q \leq C\|t(\cdot, x^j)\|_q 2^{rg(R)l/R_j}, \tag{3.10}$$

where the L_q-norm is taken only in the variable x_j. The inequality (3.10) implies that for any x^j we have

$$t(\cdot, x^j) \in H_q^r C' 2^{rg(R)l/R_j} \|t(\cdot, x^j)\|_q$$

and, therefore, for $0 < r < A$ we obtain, by our assumption, that

$$E_{n,j}(t, \Phi)_q \ll 2^{r(g(R)l/R_j - n)} \|t\|_q. \tag{3.11}$$

We now take $f \in NH_q^R$ and prove that

$$E_{n,j}(f, \Phi)_q \ll 2^{-nR_j}.$$

We choose a number r such that $R_j < r < A$ and use the representation (3.8). Applying (3.11) for $l \leq L := nR_j/g(R)$ and (3.9) for $l > L$ we obtain

$$E_{n,j}(f, \Phi)_q \ll \sum_{l \leq L} 2^{-rn} 2^{g(R)(r/R_j - 1)l} + \sum_{l > L} 2^{-g(R)l} \ll 2^{-nR_j}. \tag{3.12}$$

We now use the following inequality

$$E_n^R(f, \Phi^d)_q \leq C(p, d) \sum_{j=1}^{d} E_{g(R)n/R_j, j}(f, \Phi)_q, \tag{3.13}$$

which is an analog of Bernstein's theorem (see [1] and also [9], [16], [11]). We prove here Theorem 3.2 that is a more general inequality than (3.13). The inequality (3.13) combined with (3.12) implies for $f \in NH_q^R$ that

$$E_n^R(f, \Phi^d)_q \ll 2^{-g(R)n}.$$

This completes the proof of Lemma 3.1. ■

We now prove an analog of the Bernstein theorem mentioned above. Let X be a Banach space and let Ψ be a basis for X. For a given set G of indices denote

$$E_G(f)_X := \inf_{c_i, i \in G} \left\| f - \sum_{i \in G} c_i \psi_i \right\|_X .$$

Consider a projector S_G which maps a function $f \in X$ to

$$S_G(f) := \sum_{i \in G} c_i(f)\psi_i \qquad \text{where} \quad f = \sum_i c_i(f)\psi_i .$$

If G is finite then the operator S_G is a bounded operator from X onto

$$X_G := \{\psi_i\}_{i \in G}.$$

In the case of infinite G we define

$$X_G := \overline{\operatorname{span}}\{\psi_i\}_{i \in G}$$

and assume that G is such that the operator S_G is a bounded operator from X onto X_G. In the case of the unconditional basis Ψ the operator S_G is bounded for all G with the norm bound independent of G.

Theorem 3.2. *Let two sets G_1 and G_2 of indices be such that*

(3.14) $$\|S_{G_j}\|_{X \to X} \le B, \qquad j = 1, 2.$$

Denote $G := G_1 \cap G_2$. Then for any $f \in X$ we have

$$E_G(f)_X \le \tfrac{1}{2}(B+1)^2(E_{G_1}(f)_X + E_{G_2}(f)_X).$$

Proof. We estimate $\|f - S_G(f)\|_X$. Let us represent

$$S_G(f) = S_{G_1}(f) - S_{G_1 \setminus G_2}(f)$$

and estimate

(3.15) $$\|f - S_G(f)\|_X \le \|f - S_{G_1}(f)\|_X + \|S_{G_1 \setminus G_2}(f)\|_X.$$

By the assumption (3.14) we get

(3.16) $$\|f - S_{G_1}(f)\|_X \le (B+1)E_{G_1}(f)_X.$$

Next, we have

$$\|S_{G_1 \setminus G_2}(f)\|_X \le B(B+1)E_{G_2}(f)_X. \tag{3.17}$$

Combining (3.16) and (3.17) we obtain, from (3.15),

$$\|f - S_G(f)\|_X \le (B+1)(E_{G_1}(f)_X + BE_{G_2}(f)_X). \tag{3.18}$$

Changing the roles of G_1 and G_2 we get in the same way

$$\|f - S_G(f)\|_X \le (B+1)(BE_{G_1}(f)_X + E_{G_2}(f)_X). \tag{3.19}$$

Adding (3.18) and (3.19) we obtain the required inequality. ∎

We now prove the lower estimates in best m-term approximation. These proofs are similar to the corresponding ones from [15, S.4].

Theorem 3.3. *Let $1 < q, p < \infty$. Then for R such that $g(R) > (1/q - 1/p)_+$ we have*

$$\sigma_m(NH_q^R, U^d)_p \gg m^{-g(R)}.$$

Proof. We need a concept of the entropy numbers. For a bounded set F in a Banach space X we denote, for integer m,

$$\varepsilon_m(F, X) := \inf\left\{\varepsilon : \exists f_1, \dots, f_{2^m} \in X : F \subset \bigcup_{j=1}^{2^m}(f_j + \varepsilon B(X))\right\},$$

where $B(X)$ is the unit ball of Banach space X and $f_j + \varepsilon B(X)$ is the ball of radius ε with the center at f_j.

In this proof we use the following estimates:

$$\varepsilon_m(NH_q^R, L_p) \asymp m^{-g(R)}, \qquad 1 \le q, p, \le \infty, \quad g(R) > (1/q - 1/p)_+. \tag{3.20}$$

These estimates should be considered known and can be derived, for instance, from the finite-demensional results (see [10]) by the standard arguments of discretization. The estimates (3.20) will be used in the general method which, roughly speaking, states that m-term approximations with regard to any reasonable basis are bounded from below by the entropy numbers. We now formulate one result from [14, see Th. 4 with $b = 0$]. ∎

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

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

$$\begin{aligned} V_k(\psi_j) &= \lambda_{k,j}\psi_j, \qquad \lambda_{k,j} = 1 \text{ for } j = 1, \dots, n_k, \\ \lambda_{k,j} &= 0 \text{ for } j > A_2 n_k, \\ \|V_k\|_{X \to X} &\le A_3, \qquad k = 1, 2, \dots. \end{aligned} \tag{3.21}$$

Theorem 3.4. *Assume that for some $a > 0$ we have*

$$\varepsilon_m(F, X) \geq C_1 m^{-a}, \qquad m = 2, 3, \ldots.$$

Then if a system Ψ satisfies condition (VP) *and also satisfies the following condition*:

$$(3.22) \qquad E_n(F, \Psi) := \sup_{f \in F} \inf_{c_1,\ldots,c_n} \left\| f - \sum_{j=1}^{n} c_j \psi_j \right\|_X \leq C_2 n^{-a}, \qquad n = 1, 2, \ldots,$$

we have

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

We use this theorem with $\Psi = U^d$ and $X = L_p$. As a sequence of operators V_n we take

$$V_n(f) := \sum_{|I_j| \geq 2^{-ng(R)/R_j}, j=1,\ldots,d} f_I U_I = \sum_{s_j \leq 2^{ng(R)/R_j}-1, j=1,\ldots,d} \sum_{\varepsilon} \delta_s^{\varepsilon}(f).$$

It is well-known that, for any $1 < p < \infty$,

$$\|V_n\|_{L_p \to L_p} \leq C(p, d).$$

The relation (3.22) follows from (3.7). Thus, Theorem 3.3 follows from Theorem 3.4 and the estimates (3.20).

Theorem 3.5. *For any orthogonal basis Φ we have, for R such that $g(R) > (1/q - \frac{1}{2})_+$,*

$$\sigma_m(NH_q^R, \Phi)_2 \gg m^{-g(R)}, \qquad 1 \leq q \leq \infty.$$

Proof. The proof of this theorem is similar to the univariate case (see [5]). We shall not carry it out here and formulate only the key lemma of the proof (see [5, Corollary 2]). ■

Lemma 3.2. *There exists an absolute constant $C_0 > 0$ such that for any orthonormal basis Φ and any N-dimensional cube*

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

we have

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

if $m \leq C_0 N$.

4. Soft Thresholding Is a Lipschitz Mapping

In this section we assume that a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$ is an unconditional normalized ($\|\psi_k\| = 1, k = 1, 2, \ldots$) basis for X.

Definition 4.1. A basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$ of a Banach space X is said to be unconditional if for every choice of signs $\theta = \{\theta_k\}_{k=1}^{\infty}$, $\theta_k = 1$ or -1, $k = 1, 2, \ldots$, the linear operator M_θ, defined by

$$M_\theta\left(\sum_{k=1}^{\infty} a_k\psi_k\right) = \sum_{k=1}^{\infty} a_k\theta_k\psi_k,$$

is a bounded operator from X into X.

The uniform boundedness principle implies that the unconditional constant

$$K := K(X, \Psi) := \sup_\theta \|M_\theta\|$$

is finite.

The following theorem is a well-known fact about unconditional bases (see [8, p. 19]).

Theorem 4.1. *Let Ψ be an unconditional basis for X. Then for every choice of bounded scalars $\{\lambda_k\}_{k=1}^{\infty}$, we have*

$$\left\|\sum_{k=1}^{\infty} \lambda_k a_k\psi_k\right\| \leq 2K \sup_k |\lambda_k| \left\|\sum_{k=1}^{\infty} a_k\psi_k\right\|$$

(in the case of a real Banach space X we can take K instead of $2K$).

In the numerical implementation of nonlinear m-term approximation one usually prefers to employ the strategy known as thresholding (see [2, S.7.8]) instead of a greedy algorithm. We define and study here the soft thresholding. Let a real function $v(x)$ defined for $x \geq 0$ satisfy the following relations:

$$v(x) = \begin{cases} 1 & \text{for } x \geq 1, \\ 0 & \text{for } 0 \leq x \leq \frac{1}{2}, \end{cases} \tag{4.1}$$

$$|v(x)| \leq A, \qquad x \in [0, 1], \tag{4.2}$$

there is a constant C_L such that for any $x, y \in [0, \infty)$ we have

$$|v(x) - v(y)| \leq C_L|x - y|. \tag{4.3}$$

Let

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

We define a soft thresholding mapping $T_{\varepsilon,v}$ as follows. Take $\varepsilon > 0$ and set

$$T_{\varepsilon,v}(f) := \sum_k v(|c_k(f)|/\varepsilon)c_k(f)\psi_k.$$

Theorem 4.1 implies that

$$\|T_{\varepsilon,v}(f)\| \le 2KA\|f\|. \tag{4.4}$$

We now prove that the mapping $T_{\varepsilon,v}$ satisfies the Lipschitz condition with a constant independent of ε.

Theorem 4.2. *For any ε and any functions $f, g \in X$, we have*

$$\|T_{\varepsilon,v}(f) - T_{\varepsilon,v}(g)\| \le (3A + 2C_L)2K\|f - g\|.$$

Proof. Let $\varepsilon > 0$ be fixed. We use for simplicity the following abbreviated notations:

$$v_k(f) := v(|c_k(f)|/\varepsilon), \qquad v_k(g) := v(|c_k(g)|/\varepsilon).$$

Then we have

$$\begin{aligned} T_{\varepsilon,v}(f) - T_{\varepsilon,v}(g) &= \sum_k (c_k(f)v_k(f) - c_k(g)v_k(g))\psi_k \\ &= \sum_k (c_k(f) - c_k(g))v_k(f)\psi_k \\ &\quad + \sum_k c_k(g)(v_k(f) - v_k(g))\psi_k =: \Sigma_1 + \Sigma_2. \end{aligned} \tag{4.5}$$

For the first sum we have, by Theorem 4.1,

$$\|\Sigma_1\| \le 2K\|\{v_k(f)\}\|_{l_\infty}\|f - g\| \le 2KA\|f - g\|. \tag{4.6}$$

In order to estimate the second sum, we introduce the set

$$\Lambda(g, \varepsilon) := \{k : |c_k(g)| \ge \varepsilon\}$$

and write

$$\Sigma_2 = \sum_{k\in\Lambda(g,\varepsilon)} c_k(g)(v_k(f) - v_k(g))\psi_k + \sum_{k\notin\Lambda(g,\varepsilon)} c_k(g)(v_k(f) - v_k(g))\psi_k =: \Sigma_2' + \Sigma_2''.$$

Let us first estimate Σ_2''. We have

$$|c_k(g)(v_k(f) - v_k(g))| \le \varepsilon C_L ||c_k(f)| - |c_k(g)||/\varepsilon \le C_L|c_k(f) - c_k(g)|.$$

We get from here, by Theorem 4.1,

$$\|\Sigma_2''\| \le 2KC_L\|f - g\|. \tag{4.7}$$

We proceed to estimating Σ_2'. Represent Σ_2' in the form

$$\Sigma_2' = \sum_{k\in\Lambda(g,\varepsilon)\cap\Lambda(f,\varepsilon)} c_k(g)(v_k(f) - v_k(g))\psi_k + \sum_{k\in\Lambda(g,\varepsilon)\setminus\Lambda(f,\varepsilon)} c_k(g)(v_k(f) - v_k(g))\psi_k =: \Sigma_{2,1}' + \Sigma_{2,2}'.$$

For k in the first sum we have from the definition of v that $v_k(f) = v_k(g) = 1$ and thus $\Sigma_{2,1}' = 0$. Let us now estimate the following sum:

$$\Sigma := \sum_{k\in\Lambda(g,\varepsilon)\setminus\Lambda(f,\varepsilon)} c_k(f)(v_k(f) - v_k(g))\psi_k.$$

Similarly to Σ_2'' we get

$$\|\Sigma\| \le 2KC_L\|f - g\|. \tag{4.8}$$

Next, we have

$$\|\Sigma_2' - \Sigma\| = \left\| \sum_{k\in\Lambda(g,\varepsilon)\setminus\Lambda(f,\varepsilon)} (c_k(g) - c_k(f))(v_k(f) - v_k(g))\psi_k \right\| \le 2K2A\|f - g\|. \tag{4.9}$$

Combining the estimates (4.5)–(4.9), we complete the proof of Theorem 4.2. ■

Theorem 4.2 provides a way of constructing greedy-type algorithms which have the Lipschitz property. For instance, one can use a soft thresholding algorithm with regard to U^d to approximate functions from classes NH_q^R and also from classes MW_q^r and MH_q^r with a bounded mixed derivative or difference (see [15]). A problem of constructing a continuous mapping in m-term approximation was discussed in [4]. The following remarks will be useful in this regard:

(1) The system U^d is an unconditional basis for L_p, $1 < p < \infty$. See [17] for $d = 1$. The general case $d > 1$ follows from the case $d = 1$ by standard arguments (see, for instance, [3]).

(2) Denote by T_ε the thresholding algorithm, i.e., $T_\varepsilon := T_{\varepsilon,u}$ with $u(x) = 1$ for $x \ge 1$ and $u(x) = 0$ otherwise. Then by Theorem 4.1 we have

$$\|f - T_{\varepsilon,v}(f)\| \le 2KA\|f - T_\varepsilon(f)\|. \tag{4.10}$$

Thus, if we have upper estimates for a thresholding algorithm T_ε (a greedy algorithm), then we can derive from them the corresponding upper estimates for the soft thresholding algorithm.

In [7] we studied the concept of a greedy basis, i.e., a basis Ψ such that for each $f \in X$ we have

$$\|f - G_m(f,\Psi)\| \le G\sigma_m(f,\Psi), \qquad m = 1, 2, \dots,$$

with a constant G independent of f and m. Denote

$$m(f, \varepsilon) := \#\{k : |c_k(f)| \geq \varepsilon\}.$$

Then (4.10) implies that if Ψ is a greedy basis, we have, for each $f \in X$,

$$\|f - T_{\varepsilon,v}(f)\| \leq G' \sigma_{m(f,\varepsilon)}(f, \Psi). \tag{4.11}$$

We note here that (4.11) implies that Ψ is a greedy basis. The proof of this statement uses the arguments from [7] and can be carried out as follows. We have proved in [7] (see Theorem 1) that Ψ is greedy if and only if Ψ is unconditional and democratic.

Definition 4.2. We say that a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$ is a democratic basis if, for any two finite sets of indices P and Q with the same cardinality $\#P = \#Q$, we have

$$\left\| \sum_{k \in P} \psi_k \right\| \leq D \left\| \sum_{k \in Q} \psi_k \right\|$$

with a constant $D := D(X, \Psi)$ independent of P and Q.

Using the arguments from [7], which were used to prove that a greedy basis is unconditional and democratic, we prove that (4.11) implies that Ψ is unconditional and democratic. It remains to apply Theorem 1 from [7] to complete the proof. ■

Acknowledgment. This research was supported by the National Science Foundation Grant DMS 9622925, by ONR Grant N00014-91-5-1076, and by DAAG55-98-10002.

References

1. S. N. Bernstein (1951): *On the best approximation of functions of several variables by means of polynomials of trigonometric sums*. Trudy Mat. Inst. Steklov, **38**:24–29 (Russian).
2. R. A. DeVore (1998): *Nonlinear approximation*. Acta Numerica, 51–150.
3. R. A. DeVore, S. V. Konyagin, V. N. Temlyakov (1998): *Hyperbolic wavelet approximation*. Constr. Approx., **14**:1–26.
4. Dinh Dung (1998): *On best continuous methods in n-term approximation*. Vietnam J. Math., **26**:367–371.
5. B. S. Kashin (1985): *On approximation properties of complete orthonormal systems*. Trudy Mat. Inst. Steklov, **172**:187–191; English transl. in Proc. Steklov Inst. Math., **1987**, no. 3, 207–211.
6. B. S. Kashin, V. N. Temlyakov (1994): *On best m-term approximation and the entropy of sets in the space L^1*. Mat. Zametki, **56**:57–86; English transl. in Math. Notes, **56**:1137–1157.
7. S. V. Konyagin, V. N. Temlyakov (1999): *A remark on greedy approximation in Banach spaces*. East J. Approx., **5**:1–15.
8. J. Lindenstrauss, L. Tzafriri (1977): Classical Banach Spaces I. Berlin: Springer-Verlag.
9. M. K. Potapov (1957): *Imbedding theorems for analytic functions of several variables*. Dokl. Akad. Nauk SSSR, **112**:591–594 (Russian).
10. C. Schütt (1984): *Entropy numbers of diagonal operators between symmetric Banach spaces*. J. Approx. Theory, **40**:121–128.
11. V. N. Temlyakov (1982): *Approximation of functions with bounded mixed difference by trigonometric polynomials, and the widths of some classes of functions*. Math. USSR-Izv., **46**:171–186; English transl. in Math. USSR-Izv., **20** (1983):173–187.

12. V. N. TEMLYAKOV (1988): *Approximation by elements of a finite-dimensional subspace of functions from various Sobolev or Nikol'skii spaces*. Mat. Zametki, **43**:770–786; English transl. in Math. Notes, **43**:444–454.
13. V. N. TEMLYAKOV (1993): Approximation of Periodic Functions. New York: Nova Science.
14. V. N. TEMLYAKOV (1998): *Nonlinear Kolmogorov's widths*. Mat. Zametki, **63**:891–902.
15. V. N. TEMLYAKOV (2000): *Greedy algorithms with regard to multivariate systems with special structure*. Constr. Approx., **16**:399–425.
16. M. F. TIMAN (1957): *Interrelation between total and partial best approximation in the mean of functions of several variables*. Dokl. Akad. Nauk SSSR, **112**:24–26 (Russian).
17. P. WOJTASZCZYK (1997): *On unconditional polynomial bases in L_p and Bergman spaces*. Constr. Approx., **13**:1–15.

V. N. Temlyakov
Department of Mathematics
University of South Carolina
Columbia, SC 29208
USA