Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT•

Journal of Approximation Theory 121 (2003) 158–197

JOURNAL OF Approximation Theory

http://www.elsevier.com/locate/jat

Multivariate n-term rational and piecewise polynomial approximation ☆

Pencho Petrushev*

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

Received 17 May 2001; accepted in revised form 22 November 2002
Communicated by Peter Oswald

Abstract

We study nonlinear approximation in $L_p(\mathbb{R}^d)$ $(0<p<\infty,\ d>1)$ from (a) n-term rational functions, and (b) piecewise polynomials generated by different anisotropic dyadic partitions of $\mathbb{R}^d$. To characterize the rates of each such piecewise polynomial approximation we introduce a family of smoothness spaces (B-spaces) which can be viewed as an anisotropic variation of Besov spaces. We use the B-spaces to prove Jackson and Bernstein estimates and then characterize the piecewise polynomial approximation by interpolation. Our main estimate relates n-term rational approximation with piecewise polynomial approximation in $L_p(\mathbb{R}^d)$. This result enables us to obtain a direct estimate for n-term rational approximation in terms of a minimal B-norm (over all dyadic partitions). We also show that the Haar bases associated with anisotropic dyadic partitions of $\mathbb{R}^d$ can be successfully utilized for nonlinear approximation. We give an effective algorithm for best Haar basis or best B-space selection.

Keywords: Nonlinear approximation; Multivariate piecewise polynomial approximation; Rational approximation

1. Introduction

The theory of univariate rational approximation on $\mathbb{R}$ is a relatively well developed area in approximation theory (see, e.g., [20]). At the same time, the theory of multivariate rational approximation is virtually not existing yet. A reason for this is that it is extremely hard to deal with rational functions of the form $R := P/Q$,

☆ This research was supported by ONR/ARO-DEPSCoR Research Contract DAAG55-98-1-0002.
*Corresponding author. Fax: 803-777-6527.
E-mail address: pencho@math.sc.edu.

0021-9045/03/$ - see front matter
doi:10.1016/S0021-9045(02)00060-6

where P and Q are algebraic polynomial in d variables ($d>1$). Very little is known about this type of rational functions. It seems natural to consider approximation from the smaller set of *n-term rational functions or atomic rational functions* that is the set of all rational functions of the form

$$R = \sum_{j=1}^{n} r_j \quad \text{with } r_j \text{ of the form } r(x) = \prod_{k=1}^{d} \frac{a_k x_k + b_k}{(x_k - \alpha_k)^2 + \beta_k^2}. \tag{1.1}$$

As it will be shown in this article, this is a powerful tool for approximation and at the same time it is more tangible than the former.

It is also interesting to consider approximation from multivariate rational functions of the form $R = \sum_{j=1}^{n} r_j$, where r_j are dilates and shifts of a single radial partial fraction such as $r(x) = 1/(1+|x|^2)^k$. In [12], we consider such approximation and prove a direct estimate in terms of the usual Besov norm (exactly the same as the one used in nonlinear approximation from wavelets or regular splines). To prove this result, we first constructed good bases consisting of dyadic shifts and dilates of a single rational function and then utilized them to nonlinear approximation (see also [19]).

In this article, we take a different approach to the problem. We prove an estimate that relates the multivariate n-term rational approximation to a broad class of nonlinear piecewise polynomial approximation in $L_p(\mathbb{R}^d)$ $(0<p<\infty)$. In particular, this result relates the n-term rational approximation to nonlinear approximation from piecewise polynomials generated by any anisotropic dyadic partition of $\mathbb{R}^d$. Then we utilize this relationship to obtain an estimate for n-term rational approximation in terms of the minimal smoothness norm (over all dyadic partitions). These estimates extend to the multivariate case results from [15,17].

As a consequence of our approach, a substantial part of this article is devoted to nonlinear approximation from piecewise polynomials over dyadic partitions which is interesting in its own right. To the best of our knowledge this problem was first posed explicitly in [14, Section 5.4.3]. Note that we consider not one but a collection of approximation processes each of them determined by a dyadic partition of $\mathbb{R}^d$. The ultimate goal of the theory of any approximation scheme is to characterize the rates of approximation in terms of certain smoothness conditions. To characterize the rates of piecewise polynomial approximation generated by an arbitrary dyadic partition, we introduce a family of new smoothness spaces (B-spaces) which can be viewed as an anisotropic variation of Besov spaces. We use the B-spaces to prove Jackson and Bernstein estimates and then characterize the approximation by interpolation. In [18], we proved that in the univariate case a scale of Besov spaces governs the rates of nonlinear piecewise polynomial approximation. Similar Besov spaces have also been used for characterization of multivariate nonlinear (regular) spline L_p-approximation in [5] $(1\leqslant p<\infty)$ and [7] $(p=\infty)$, see also [11]. Here we extend and refine these results.

In addition to this, we consider the library of anisotropic Haar bases which are naturally associated with anisotropic dyadic partitions of $\mathbb{R}^d$. Since every anisotropic Haar basis is an unconditional basis in L_p $(1<p<\infty)$ and characterizes the corresponding B-spaces (see Section 5), it provides an effective tool for nonlinear approximation from piecewise constants. Moreover, as we show in Section 5, in a natural discrete setting, there is a practically feasible algorithm for best Haar basis or best B-space selection for any given function. In this way, the approximation procedure can effectively be completed.

A leading idea in this article is that the classical smoothness spaces are not suitable for measuring the smoothness of the functions in highly nonlinear approximation such as multivariate rational or piecewise polynomial approximation. More sophisticated means of measuring the smoothness are needed. We believe that, in some cases, the smoothness should be measured by means of a collection of smoothness space scales (like the B-spaces).

The outline of the article is the following. In Section 2, we introduce the B-spaces and establish some of their basic properties. In Section 3, we prove Jackson and Bernstein estimates and then characterize the nonlinear piecewise polynomial approximation generated by an arbitrary anisotropic dyadic partition of $\mathbb{R}^d$. In Section 4, we prove an estimate that relates the n-term rational approximation to nonlinear piecewise polynomial approximation and, as a consequence, we obtain a direct estimate for rational approximation in terms of the minimal B-norm. Section 5 is devoted to the anisotropic Haar bases. We give an algorithm for best Haar basis or best B-space selection. In Section 6, we present our view point on some of the principle questions concerning nonlinear approximation and pose some open problems. Section 7 is an appendix, where we give the proofs of some auxiliary statements from Section 2 and the lengthy proof of an interpolation result from Section 3.

Throughout this article, the positive constants are denoted by $c, c_1, \ldots$ and they may vary at every occurrence, $A \approx B$ means $c_1 B \leqslant A \leqslant c_2 B$; Π_k denotes the set of all algebraic polynomials in d variables of total degree $<k$. For a set $E \subset \mathbb{R}^d$, $\mathbb{1}_E$ denotes the characteristic function of E, and $|E|$ denotes the Lebesgue measure of E. Since we systematically work with quasi-normed spaces such as L_p, $0<p<1$, "norm" will stand for "norm" or "quasi-norm".

2. B-spaces

In this section, we introduce a family of smoothness spaces (B-spaces) which will be used for the characterization of nonlinear piecewise polynomial approximation (Sections 3 and 5) and in n-term rational approximation (Section 4). These spaces can be defined on $\mathbb{R}^d$ $(d>1)$ or on an arbitrary box Ω in $\mathbb{R}^d$. For convenience, we shall only consider the case when $|\Omega| = 1$ and Ω is with sides parallel to the coordinate aces. We shall define the B-spaces by using local polynomial approximation over boxes from nested anisotropic dyadic partitions of $\mathbb{R}^d$ or Ω.

Anisotropic dyadic partitions of $\mathbb{R}^d$ *or* Ω: We call

$$\mathscr{P} = \bigcup_{m \in \mathbb{Z}} \mathscr{P}_m$$

a dyadic partition of $\mathbb{R}^d$ with levels $\{\mathscr{P}_m\}$ if the following conditions are fulfilled:

(a) Every level $\mathscr{P}_m$ is a partition of $\mathbb{R}^d$: $\mathbb{R}^d = \bigcup_{I \in \mathscr{P}_m} I$ and $\mathscr{P}_m$ consists of disjoint dyadic boxes of the form $I = \mathscr{I}_1 \times \cdots \times \mathscr{I}_d$, where each $\mathscr{I}_j$ is a semi-open dyadic interval ($\mathscr{I}_j = [(v-1)2^\mu, v2^\mu)$), and $|I| = 2^{-m}$.
(b) The levels of $\mathscr{P}$ are nested, i.e., $\mathscr{P}_{m+1}$ is a refinement of $\mathscr{P}_m$. Thus each $I \in \mathscr{P}_m$ has two children, say, $J_1, J_2 \in \mathscr{P}_{m+1}$ such that $I = J_1 \cup J_2$ and $J_1 \cap J_2 = \emptyset$.
(c) For any boxes $I', I'' \in \mathscr{P}$ there exists a box $I \in \mathscr{P}$ such that $I' \cup I'' \subset I$.

Also, we call $\mathscr{P} = \bigcup_{m \geqslant 0} \mathscr{P}_m$ a dyadic partition of Ω ($|\Omega| = 1$) if $\mathscr{P}_0 := \{\Omega\}$ and the levels $\{\mathscr{P}_m\}_{m \geqslant 1}$ satisfy conditions (a) and (b) from above with $\mathbb{R}^d$ replaced by Ω.

The next few remarks will help to understand better the nature of dyadic partitions. First, condition (c) above is not very restrictive but it prevents $\mathscr{P}_m$ from possible deteriorations as $m \to -\infty$. This condition implies that in each dyadic partition $\mathscr{P}$ of $\mathbb{R}^d$ there is a single tree structure with set inclusion as the order relation.

We note that the two children, say, $J_1, J_2 \in \mathscr{P}_{m+1}$ of any $I \in \mathscr{P}_m$ can be obtain by splitting I in two equal subboxes in d ($d > 1$) different ways. Therefore, there is a huge variety of anisotropic dyadic partitions $\mathscr{P}$ of $\mathbb{R}^d$ or Ω.

A dyadic partition of any box can easily be obtained inductively (by successive subdividing). For instance, suppose we want to subdivide Ω. Assume that the levels $\{\mathscr{P}_j\}_{0 \leqslant j \leqslant m}$ have already been defined. We now subdivide each box $I \in \mathscr{P}_m$ by "halving" I in one of the d coordinate directions, thus obtaining two new dyadic boxes which we include in $\mathscr{P}_{m+1}$. We process in the same way all boxes from $\mathscr{P}_m$ and as a result obtain the next level $\mathscr{P}_{m+1}$ of dyadic boxes.

To construct an anisotropic partition $\mathscr{P}$ of $\mathbb{R}^d$, one can proceed as follows: First, cover $\mathbb{R}^d$ by a growing sequence of dyadic boxes $I_0 \subset I_1 \subset \cdots$, $|I_j| = 2^j$, $\mathbb{R}^d = \bigcup_{j \geqslant 0} I_j$, starting from an arbitrary dyadic box I_0 and growing the consecutive boxes infinitely many times in all four directions. Second, subdivide each box I_j and its sibling (contained in I_{j+1}) as above.

A typical property of the anisotropic dyadic partitions is that each level $\mathscr{P}_m$ of such a partition $\mathscr{P}$ consists of dyadic boxes I with $|I| = 2^{-m}$ and at the same time there could be extremely (uncontrollably) long and narrow boxes in $\mathscr{P}_m$.

Local polynomial approximation: Fix a box $I \subset \mathbb{R}^d$ and let $f \in L_p(I)$. Then

$$E_k(f, I)_p := \inf_{P \in \Pi_k} \|f - P\|_{L_p(I)} \tag{2.1}$$

is the error of $L_p(I)$ approximation to f from Π_k, the set of all algebraic polynomials of degree $< k$. The local modulus of smoothness $\omega_k(f, I)_p$ is defined as usual by

$$\omega_k(f,I)_p := \sup_{h\in\mathbb{R}^d} \|\Delta_h^k(f,\cdot)\|_{L_p(I)}, \tag{2.2}$$

where $\Delta_h^k(f,x)$ is the kth difference with step $h\in\mathbb{R}^d$ and $\Delta_h^k(f,x) := 0$ if the segment $[x, x+kh]$ is not entirely contained in I.

We shall need the fact that $E_k(f,I)_p$ and $\omega_k(f,I)_p$ are equivalent:

$$E_k(f,I)_p \approx \omega_k(f,I)_p \tag{2.3}$$

with constants of equivalence depending only on p, k, and d. Equivalence (2.3) follows from the case when $I=[0,1)^d$ by a simple change of variables; the upper estimate is Whitney's theorem (see [2] if $p\geqslant 1$ and [22] if $0<p\leqslant 1$) and the lower estimate follows by the fact that $\Delta_h^k(P,x)=0$ if $P\in\Pi_k$.

We shall often use the following lemma which establishes the equivalence of different norms of polynomials over different sets.

Lemma 2.1. *Suppose $R := I\backslash J$, where $J\subset I$ and I, J are dyadic boxes in $\mathbb{R}^d$ or $J=\emptyset$. Let $I'\subset R$ be also a dyadic box with $|I'|=|I|/2$. Then, for each polynomial $P\in\Pi_k$ and $0<\tau,p\leqslant\infty$,*

$$\|P\|_{L_p(I)} \approx \|P\|_{L_p(R)} \approx \|P\|_{L_p(I')} \tag{2.4}$$

and

$$\|P\|_{L_\tau(R)} \approx |R|^{1/\tau-1/p}\|P\|_{L_p(R)} \tag{2.5}$$

with constants of equivalence depending only on p, τ, k, and d.

Proof. This lemma follows immediately from the obvious case $I=[0,1)^d$ (all norms of a polynomial are equivalent) by change of variables. □

We find useful the concept of near best approximation which we borrowed from [8]. A polynomial $Q\in\Pi_k$ is said to be a near best $L_p(I)$ approximation to f from Π_k with constant A if

$$\|f-Q\|_{L_p(I)} \leqslant A E_k(f,I)_p.$$

Note that if $p\geqslant 1$, then a near best $L_p(I)$ approximation $Q := Q_I(f)$ from Π_k can be realized by a linear projector.

Lemma 2.2. *Let $0<q<p$ and let Q_I be a near best $L_q(I)$ approximation to f from Π_k. Then Q_I is a near best $L_p(I)$ approximation to f from Π_k.*

Proof. See [8]. □

Definition of B-spaces on $\mathbb{R}^d$ Let $\mathscr{P}$ be an arbitrary anisotropic dyadic partition of $\mathbb{R}^d$ $(d>1)$, $\alpha>0$, $0<p,q\leqslant\infty$, and $k\geqslant 1$. We define the B-space $B_{pq}^{\alpha k}(\mathscr{P})$ as the set of

all functions $f \in L_p(\mathbb{R}^d)$ such that

$$\|f\|_{B^{\alpha k}_{pq}(\mathscr{P})} := \left(\sum_{m \in \mathbb{Z}} \left[\sum_{I \in \mathscr{P}_m} (|I|^{-\alpha} \omega_k(f, I)_p)^p \right]^{q/p} \right)^{1/q}$$
$$= \left(\sum_{m \in \mathbb{Z}} \left[2^{m\alpha} \left(\sum_{I \in \mathscr{P}_m} \omega_k(f, I)_p^p \right)^{1/p} \right]^q \right)^{1/q} \tag{2.6}$$

is finite, where the ℓ_q-norm is replaced by the sup-norm if $q = \infty$ as usual. From (2.3), it follows that

$$\|f\|_{B^{\alpha k}_{pq}(\mathscr{P})} \approx N_1(f, \mathscr{P}) := \left(\sum_{m \in \mathbb{Z}} \left[\sum_{I \in \mathscr{P}_m} (|I|^{-\alpha} E_k(f, I)_p)^p \right]^{q/p} \right)^{1/q}. \tag{2.7}$$

Evidently, if $f \in B^{\alpha k}_{pq}(\mathscr{P})$ and $\|f\|_{B^{\alpha k}_{pq}(\mathscr{P})} = 0$, then $E_k(f, I)_p = 0$ for all $I \in \mathscr{P}$, which together with the fact that $f \in L_p(\mathbb{R}^d)$ and condition (c) on dyadic partitions implies that $f = 0$ a.e. (see also the proof of Theorem 2.4 in Appendix A). Therefore, $\|\cdot\|_{B^{\alpha k}_{pq}(\mathscr{P})}$ is a norm if $p, q \geqslant 1$ and a quasi-norm otherwise.

We now introduce the linear piecewise polynomial approximation generated by $\mathscr{P}$. Let $\mathscr{S}^k_m := \mathscr{S}^k_m(\mathscr{P})$ be the set of all piecewise polynomials of degree $<k$ on boxes $I \in \mathscr{P}_m$, that is, $S \in \mathscr{S}^k_m$ if $S = \sum_{I \in \mathscr{P}_m} \mathbb{1}_I \cdot P_I$, where $P_I \in \Pi_k$. Evidently, $\cdots \subset \mathscr{S}^k_{-1} \subset \mathscr{S}^k_0 \subset \mathscr{S}^k_1 \subset \cdots$. We denote

$$\mathbb{L}_p := \mathbb{L}_p(\mathscr{P}, k) := \overline{\bigcup_{m \in \mathbb{Z}} \mathscr{S}^k_m},$$

where the closure is taken in $L_p(\mathbb{R}^d)$. Evidently, $\mathbb{L}_p$ is a subspace of L_p and

$$\mathbb{L}_p = \text{span } \{\mathbb{1}_I \cdot P_I : P_I \in \Pi_k, I \in \mathscr{P}\},$$

where “span” means “closed span in L_p”. We denote by $S^k_m(f)_p := S^k_m(f, \mathscr{P})_p$ the error of L_p approximation to f from $\mathscr{S}^k_m$, i.e., $S^k_m(f)_p := \inf_{S \in \mathscr{S}^k_m} \|f - S\|_p$. Clearly, if $f \in L_p$, then $f \in \mathbb{L}_p$ if and only if $\lim_{m \to \infty} S^k_m(f)_p = 0$. It may happen that $\mathbb{L}_p(\mathscr{P}, k) \neq L_p$. However, if $\sup\{\text{diam}\,(I) : I \in \mathscr{P}_m\} \to 0$ as $m \to 0$, then $\mathbb{L}_p(\mathscr{P}, k) = L_p$.

Clearly, by (2.7),

$$N_1(f, \mathscr{P}) = \left(\sum_{m \in \mathbb{Z}} (2^{\alpha m} S^k_m(f, \mathscr{P})_p)^q \right)^{1/q}. \tag{2.8}$$

Therefore, the B-spaces $B^{\alpha k}_{pq}(\mathscr{P})$ are approximation spaces generated by $\{S^k_m(f, \mathscr{P})_p\}$.

Let $Q_{I,\eta}(f)$ be a polynomial of near best $L_\eta(I)$ approximation to f from Π_k with some constant A (the same for all $I \in \mathscr{P}$). Note that $Q_{I,\eta}(f)$ can be defined as a linear projector if $\eta \geqslant 1$. Then $T_{m,\eta}(f) := T_{m,\eta}(f, \mathscr{P}) := \sum_{I \in \mathscr{P}_m} \mathbb{1}_I \cdot Q_{I,\eta}$ is a near best L_η

approximation to f from $\mathscr{S}^k_m$. We define

$$t_{m,\eta}(f) := t_{m,\eta}(f,\mathscr{P}) := T_{m,\eta}(f) - T_{m-1,\eta}(f). \tag{2.9}$$

We now introduce a new norm in $B^{\alpha k}_{pq}(\mathscr{P})$ by

$$N_2(f,\mathscr{P}) := \left(\sum_{m\in\mathbb{Z}} (2^{\alpha m}\|t_{m,\eta}(f)\|_p)^q\right)^{1/q}, \quad \text{where } 0<\eta\leqslant p. \tag{2.10}$$

Lemma 2.3. *The norms* $\|\cdot\|_{B^{\alpha k}_{pq}(\mathscr{P})}$, $N_1(\cdot)$, *and* $N_2(\cdot)$ *are equivalent with constants of equivalence independent of* $\mathscr{P}$.

Proof. The equivalence of $\|\cdot\|_{B^{\alpha k}_{pq}(\mathscr{P})}$ and $N_1(\cdot)$ has already been indicated in (2.7).

Now, we show that $N_1(\cdot)\approx N_2(\cdot)$. Let $N_1(f)<\infty$. By Lemma 2.2, $Q_{I,\eta}(f)$ is a near best $L_p(I)$ approximation to f from Π_k and hence $\|f - T_{m,\eta}(f)\|_p \leqslant cS^k_m(f)_p$. Therefore,

$$\|t_{m,\eta}(f)\|_p \leqslant c\|f - T_{m,\eta}(f)\|_p + c\|f - T_{m-1,\eta}(f)\|_p \leqslant cS^k_m(f)_p + cS^k_{m-1}(f)_p.$$

This implies $N_2(f)\leqslant cN_1(f)$.

In the other direction, if $N_2(f)<\infty$, then it is easily seen that

$$S^k_m(f)_p \leqslant \|f - T_{m,\eta}\|_p \leqslant \left(\sum_{j=m+1}^{\infty} \|t_{j,\eta}\|_p^\lambda\right)^{1/\lambda}, \quad \lambda := \min\{p,1\}. \tag{2.11}$$

To complete the proof, we need the following discrete Hardy inequality: If $\{x_m\}_{m\in\mathbb{Z}}$ and $\{y_m\}_{m\in\mathbb{Z}}$ are two sequences of nonnegative numbers such that $y_m \leqslant \left(\sum_{j=m+1}^{\infty} x_j^\lambda\right)^{1/\lambda}$, $\lambda>0$, then

$$\sum_{m\in\mathbb{Z}} (2^{m\alpha}y_m)^q \leqslant c\sum_{m\in\mathbb{Z}} (2^{m\alpha}x_m)^q, \quad \alpha, q>0, \tag{2.12}$$

where $c = c(\lambda,\alpha,q)$. This inequality follows easily by Hölder's inequality. We use (2.8), (2.11), and (2.12) to obtain $N_1(f)\leqslant cN_2(f)$. Therefore, $N_1(f)\approx N_2(f)$. □

The B-spaces $B^{\alpha k}_\tau(\mathscr{P})$ *on* $\mathbb{R}^d$: For the purposes of nonlinear piecewise polynomial and n-term rational approximation, we shall only need a specific class of B-spaces, namely, the spaces $B^{\alpha k}_{\tau\tau}(\mathscr{P})$. Therefore, for the rest of this section, we focus our attention exclusively on these specific B-spaces.

We shall always assume that $0<p<\infty$, $\alpha>0$, $k\geqslant 1$, and τ is defined by $1/\tau := \alpha + 1/p$. We shall briefly denote the B-space $B^{\alpha k}_{\tau\tau}(\mathscr{P})$ by $B^{\alpha k}_\tau(\mathscr{P})$ or simply by B^α_τ. By the definition of B-spaces in (2.6), we have

$$\|f\|_{B^{\alpha k}_\tau(\mathscr{P})} := \left(\sum_{I\in\mathscr{P}} (|I|^{-\alpha}\omega_k(f,I)_\tau)^\tau\right)^{1/\tau} \tag{2.13}$$

and, using Lemma 2.3,

$$\|f\|_{B^{\alpha k}_\tau(\mathscr{P})} \approx N_2(f,\mathscr{P}) := \left(\sum_{I\in\mathscr{P}} (|I|^{-\alpha}\|t_{I,\eta}(f)\|_\tau)^\tau\right)^{1/\tau} \quad \text{if } 0<\eta\leqslant\tau, \tag{2.14}$$

where $t_{I,\eta}(f) := \mathbb{1}_I \cdot t_{m,\eta}(f)$ if $I\in\mathscr{P}_m$, $m\in\mathbb{Z}$.

In some instances, the B^α_τ-norms from (2.13) to (2.14) are not quite convenient since the L_τ-norm which they involve is not very friendly when $\tau<1$. This is the case when the smoothness parameter $\alpha\geqslant 1$. We next show that this drawback of the above norms can be overcome. We introduce the following new B-norms: For $f\in L_\eta$, $0<\eta<p$, we set

$$\mathscr{N}_{\omega,\eta}(f,\mathscr{P}) := \left(\sum_{I\in\mathscr{P}} (|I|^{1/p-1/\eta}\omega_k(f,I)_\eta)^\tau\right)^{1/\tau} \tag{2.15}$$

and

$$\mathscr{N}_{t,\eta}(f,\mathscr{P}) := \left(\sum_{I\in\mathscr{P}} (|I|^{1/p-1/\eta}\|t_{I,\eta}(f)\|_\eta)^\tau\right)^{1/\tau}, \tag{2.16}$$

where $t_{I,\eta}(f) := \mathbb{1}_I \cdot t_{m,\eta}(f,\mathscr{P})$ if $I\in\mathscr{P}_m$, $m\in\mathbb{Z}$ (see (2.9)). Note that $\mathscr{N}_{\omega,\tau}(f,\mathscr{P}) = \|f\|_{B^{\alpha k}_\tau(\mathscr{P})}$. Using (2.5) and the relation $1/\tau = \alpha + 1/p$, we readily obtain

$$\mathscr{N}_{t,\eta}(f,\mathscr{P}) \approx \left(\sum_{I\in\mathscr{P}} \|t_{I,\eta}(f)\|_p^\tau\right)^{1/\tau}. \tag{2.17}$$

The following embedding theorem will be important for our further developments.

Theorem 2.4. *If $f\in L_\eta$, $0<\eta<p<\infty$, and $\mathscr{N}_{t,\eta}(f,\mathscr{P})<\infty$, then*

$$f = \sum_{m\in\mathbb{Z}} t_{m,\eta}(f) \quad \text{a.e. on } \mathbb{R}^d \tag{2.18}$$

with the series converging absolutely a.e., and

$$\|f\|_p \leqslant \left\|\sum_{m\in\mathbb{Z}} |t_{m,\eta}(f)|\right\|_p \leqslant c\mathscr{N}_{t,\eta}(f,\mathscr{P}), \tag{2.19}$$

where $c = c(\alpha,k,p,d,\eta)$.

We shall deduce this theorem from the following more general embedding theorem:

Theorem 2.5. *Let $1\leqslant p<\infty$. Suppose $\{\Phi_m\}$ is a sequence of functions on $\mathbb{R}^d$ with the properties:*

(i) $\Phi_m\in L_\infty$, $\operatorname{supp}\Phi_m\subset E_m$ *with* $0<|E_m|<\infty$ *and*

$$\|\Phi_m\|_\infty \leqslant c_1|E_m|^{-1/p}\|\Phi_m\|_p.$$

(ii) *If* $x \in E_m$, *then*

$$\sum_{E_j \ni x,\ |E_j| \geqslant |E_m|} (|E_m|/|E_j|)^{1/p} \leqslant c_1,$$

where the summation is over all indices j for which E_j *satisfy the indicated conditions. Then we have*

$$\left\| \sum_j |\Phi_j(\cdot)| \right\|_p \leqslant c \left(\sum_j \|\Phi_j\|_p^\tau \right)^{1/\tau}, \quad 0 < \tau < p,$$

where $c = c(p, \tau, c_1)$.

To avoid nonnecessary technicalities at this early stage, we shall give the proofs of Theorems 2.4 and 2.5 as well as the one of the next theorem in the appendix.

Theorem 2.6. *The norms* $\|\cdot\|_{B_\tau^{\alpha k}(\mathscr{P})}$, $\mathscr{N}_{\omega,\eta}(\cdot, \mathscr{P})$ $(0 < \eta < p)$, *and* $\mathscr{N}_{t,\eta}(\cdot, \mathscr{P})$ $(0 < \eta < p)$, *defined in* (2.13), (2.15), and (2.16), *are equivalent with constants of equivalence depending only on* α, k, p, d, *and* η. *Furthermore, the equivalence of* $\|\cdot\|_{B_\tau^{\alpha k}(\mathscr{P})}$ *and* $\mathscr{N}_{\omega,\eta}(\cdot, \mathscr{P})$ *is no longer valid if* $\eta \geqslant p$.

B-spaces on Ω: We shall only define the B-spaces $B_\tau^{\alpha k}(\mathscr{P})$ on Ω which we need in nonlinear piecewise polynomial and rational approximation. The more general B-spaces $B_{pq}^{\alpha k}(\mathscr{P})$ on Ω can be introduced in an obvious way.

We again assume that $0 < p < \infty$, $\alpha > 0$, $k \geqslant 1$, and $1/\tau := \alpha + 1/p$. Let $\mathscr{P} = \bigcup_{m \geqslant 0} \mathscr{P}_m$ be an arbitrary dyadic partition of Ω ($|\Omega| = 1$). We define the space $B_\tau^\alpha := B_\tau^{\alpha k}(\mathscr{P})$ as the set of all $f \in L_\tau(\Omega)$ such that

$$|f|_{B_\tau^{\alpha k}(\mathscr{P})} := \left(\sum_{I \in \mathscr{P}} (|I|^{-\alpha} \omega_k(f, I)_\tau)^\tau \right)^{1/\tau} < \infty. \tag{2.20}$$

Evidently, $|f + P|_{B_\tau^\alpha} = |f|_{B_\tau^\alpha}$ for $P \in \Pi_k$ and hence $|\cdot|_{B_\tau^\alpha}$ is a semi-norm if $\tau \geqslant 1$ and a semi-quasi-norm if $\tau < 1$.

By Theorems 2.7 and 2.8, if $f \in B_\tau^{\alpha k}(\mathscr{P})$ then $f \in L_p(\Omega)$. Therefore, it is natural to define a norm in $B_\tau^{\alpha k}(\mathscr{P})$ by

$$\|f\|_{B_\tau^{\alpha k}(\mathscr{P})} := \|f\|_{L_p(\Omega)} + |f|_{B_\tau^{\alpha k}(\mathscr{P})}. \tag{2.21}$$

Similarly as in (2.8), we have

$$\|f\|_{B_\tau^{\alpha k}(\mathscr{P})} \approx \|f\|_p + \left(\sum_{m \in \mathbb{Z}} (2^{\alpha m} S_m^k(f, \mathscr{P})_\tau)^\tau \right)^{1/\tau}, \tag{2.22}$$

where $S_m^k(f, \mathscr{P})_\tau$ is the error of linear piecewise polynomial approximation, defined similarly as in the case of B-spaces on $\mathbb{R}^d$ (see the definition above (2.8)).

In analogy to (2.15), we introduce a more general norm by

$$\mathcal{N}_{\omega,\eta}(f,\mathcal{P}) := \|f\|_p + \left(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\omega_k(f,I)_\eta)^\tau\right)^{1/\tau}, \quad 0<\eta<p. \tag{2.23}$$

Also, similarly as in the definition of B-norms on $\mathbb{R}^d$ (see (2.9) and (2.14)), we define the operators: $Q_{I,\eta}(f)$, $T_{m,\eta}(f) := T_{m,\eta}(f,\mathcal{P})$, $t_{m,\eta}(f) := t_{m,\eta}(f,\mathcal{P})$ $(m\geqslant 0)$, and $t_{I,\eta}(f)$, $f\in L_\eta(\Omega)$, with the natural modification $T_{-1,\eta}(f) := 0$, i.e., $t_{0,\eta}(f) := T_{0,\eta}(f) := Q_{\Omega,\eta}(f)$. We define another norm by

$$\mathcal{N}_{t,\eta}(f,\mathcal{P}) := \left(\sum_{I\in\mathcal{P}} (|I|^{1/p-1/\eta}\|t_{I,\eta}(f)\|_\eta)^\tau\right)^{1/\tau} \approx \left(\sum_{I\in\mathcal{P}} \|t_{I,\eta}(f)\|_p^\tau\right)^{1/\tau},$$
$$\text{where } 0<\eta<p. \tag{2.24}$$

Theorem 2.4 implies immediately the following analogue of Theorem 2.5:

Theorem 2.7. *If $f\in L_\eta(\Omega)$, $0<\eta<p<\infty$, and $\mathcal{N}_{t,\eta}(f,\mathcal{P})<\infty$, then*

$$f = \sum_{m\geqslant 0} t_{m,\eta}(f) \quad \textit{absolutely a.e. and} \quad \|f\|_p \leqslant \left\|\sum_{m\geqslant 0} |t_{m,\eta}(f)|\right\|_p \leqslant c\mathcal{N}_{t,\eta}(f,\mathcal{P}).$$

We proceed similarly as in the proof of Theorem 2.6 (see Appendix A) to prove the equivalence of the above defined B-norms:

Theorem 2.8. *The norms $\|\cdot\|_{B_\tau^{\alpha k}(\mathcal{P})}$, $\mathcal{N}_{\omega,\eta}(\cdot,\mathcal{P})$ $(0<\eta<p)$, and $\mathcal{N}_{t,\eta}(\cdot,\mathcal{P})$ $(0<\eta<p)$, defined in* (2.21)–(2.24), *are equivalent with constants of equivalence depending only on α, k, p, d, and η.*

Comparison of B-spaces with Besov spaces: We first recall the definition of Besov spaces on $E=\mathbb{R}^d$, $E=[a,b]^d$ or on a Lipschitz domain $E\subset\mathbb{R}^d$ $(d\geqslant 1)$. The *Besov space* $B_q^s(L_p) := B_q^s(L_p(E))$, $s>0$, $1\leqslant p,q\leqslant\infty$, is defined as the set of all functions $f\in L_p(E)$ such that

$$|f|_{B_q^s(L_p)} := \left(\int_0^\infty (t^{-s}\omega_k(f,t)_p)^q \frac{dt}{t}\right)^{1/q} < \infty \tag{2.25}$$

with the L_q-norm replaced by the sup-norm if $q=\infty$, where $k := [s]+1$ and $\omega_k(f,t)_p$ is the k-th modulus of smoothness of f in $L_p(E)$. The norm in $B_q^s(L_p)$ is usually defined by $\|f\|_{B_q^s(L_p)} := \|f\|_p + |f|_{B_q^s(L_p)}$. It is well known that if in (2.25) k is replaced by any other $k>s$, then the resulting space would be the same with an equivalent norm. The point is that, for nontrivial functions f, the maximal rate of convergence of $\omega_k(f,t)_p$ is $O(t^k)$ when $p\geqslant 1$ and it is $O(t^{k-1+1/p})$ when $p<1$ (see, e.g., [20]). This is the reason for introducing k as a parameter of the Besov spaces with the next definition. We define the space

$$B_q^{s,k}(L_p) := B_q^{s,k}(L_p(E)), \quad 0<p,q\leqslant\infty, \ s>0, \ k\geqslant 1, \tag{2.26}$$

as the Besov space $B_q^s(L_p(E))$ from above, where the parameters k and s are already set independent of each other.

For the theory of nonlinear (regular) spline approximation in $L_p(E)$, $0<p<\infty$, one can utilize the Besov space

$$B_\tau^{d\alpha,k}(L_\tau) := B_\tau^{d\alpha,k}(L_\tau(E))$$

with parameters set as elsewhere in this article: $k\geqslant 1$, $\alpha>0$, and $1/\tau := \alpha + 1/p$ (see [18] when $d=1$, and [5,7] when $d>1$). Since $B_\tau^{d\alpha,k}(L_\tau)$ is embedded in L_p, it is natural to define a norm in $B_\tau^{d\alpha,k}(L_\tau)$ by $||f||_{B_\tau^{d\alpha,k}(L_\tau)} := ||f||_p + |f|_{B_\tau^{d\alpha,k}(L_\tau)}$. In the following, we shall restrict our attention to the case $E = \mathbb{R}^d$ $(d>1)$.

We call a dyadic partition $\mathscr{P}$ of $\mathbb{R}^d$ regular if there is a constant $K\geqslant 2$ such that for each box $I =: \mathscr{I}_1 \times \cdots \times \mathscr{I}_d$ from $\mathscr{P}$ we have $K^{-1}\leqslant|\mathscr{I}_\nu|/|\mathscr{I}_\mu|\leqslant K$, $1\leqslant\nu,\mu\leqslant d$.

Now, if $\mathscr{P}$ is a regular dyadic partition of $\mathbb{R}^d$ and $f\in B_\tau^{d\alpha,k}(L_\tau)$, then $f\in B_\tau^{\alpha k}(\mathscr{P})$ and

$$||f||_{B_\tau^{\alpha k}(\mathscr{P})}\leqslant c||f||_{B_\tau^{d\alpha,k}(L_\tau)},$$

which easily follows using the following equivalence:

$$\omega_k(f,I)_\tau^\tau \approx \frac{1}{|I|}\int_{[0,\,\ell(I)]^d}\int_{I_{kh}} |\Delta_h^k(f,x)|^\tau\,dx\,dh, \quad I\in\mathscr{P}, \tag{2.27}$$

where $I_{kh} := \{x\in I\colon\ [x, x+kh]\subset I\}$ and $\ell(I)$ is the maximal side of I or diam (I) (see [20] for the proof of (2.27) in the univariate case; the same proof applies to the multivariate case as well). Notice that the smoothness parameters of B-spaces and Besov spaces above are normalized differently. Thus the B-space $B_\tau^{\alpha k}(\mathscr{P})$ corresponds to the Besov space $B_\tau^{s,k}(L_\tau)$ with $s = d\alpha$.

Using the idea of the proof of Theorem 2.6 in Appendix A, one can easily prove that, for a regular dyadic partition $\mathscr{P}$,

$$B_\tau^{d\alpha,k}(L_\tau(\mathbb{R}^d)) = B_\tau^{\alpha k}(\mathscr{P}), \quad \text{if } 0<\alpha<1/p, \tag{2.28}$$

with equivalent norms, and this is no longer true if $\alpha\geqslant 1/p$, $B_\tau^{\alpha k}(\mathscr{P})$ is much larger than $B_\tau^{d\alpha,k}(L_\tau(\mathbb{R}^d))$ in this case. A key fact here is that, for each $I\in\mathscr{P}$ and $\alpha\geqslant 1/p$, $||\mathbb{1}_I||_{B_\tau^{d\alpha,k}(L_\tau)} = \infty$, while at the same time $||\mathbb{1}_I||_{B_\tau^{\alpha k}(\mathscr{P})}\approx||\mathbb{1}_I||_p$. The same is true if $\mathbb{1}_I$ is replaced by $P\cdot\mathbb{1}_I$, $P\in\Pi_k$, $P\neq 0$.

Suppose now that $\mathscr{P}$ is an arbitrary dyadic partition of $\mathbb{R}^d$. As we mentioned in Section 2, extremely long and narrow boxes may occur at any level and location of $\mathscr{P}$. Straightforward calculations show that, for such a box $I\in\mathscr{P}$ even if $0<\alpha<1/p$ and α is as small as we wish (fixed), $||\mathbb{1}_I||_{B_\tau^{d\alpha,k}(L_\tau)}/||\mathbb{1}_I||_p$ can be enormously (uncontrollably) large, while $||\mathbb{1}_I||_{B_\tau^{\alpha k}(\mathscr{P})}/||\mathbb{1}_I||_p\approx 1$. This is why the Besov spaces are completely unsuitable for the theory of piecewise polynomial approximation generated by anisotropic dyadic partitions (see also the results of Section 3 below).

The situation is quite similar when comparing two B-spaces over completely different dyadic partitions.

3. Nonlinear piecewise polynomial approximation

In this section, we shall use the B-spaces introduced in Section 2 to characterize the nonlinear piecewise polynomial approximation generated by an arbitrary dyadic partition $\mathscr{P}$ of $\mathbb{R}^d$. The same results with almost identical proofs hold on any box Ω.

We let $\Sigma_n^k(\mathscr{P})$ $(k \geqslant 1)$ denote the nonlinear set consisting of all piecewise polynomial functions

$$\varphi = \sum_{I \in \Lambda_n} \mathbb{1}_I \cdot P_I,$$

where $P_I \in \Pi_k$, $\Lambda_n \subset \mathscr{P}$, and $\# \Lambda_n \leqslant n$. We denote by $\sigma_n(f, \mathscr{P})_p := \sigma_n^k(f, \mathscr{P})_p$ the error of L_p approximation to $f \in L_p(\mathbb{R}^d)$ from $\Sigma_n^k(\mathscr{P})$:

$$\sigma_n(f, \mathscr{P})_p := \inf_{\varphi \in \Sigma_n^k(\mathscr{P})} ||f - \varphi||_p.$$

We next prove Jackson and Bernstein estimates for the above nonlinear approximation. Then the desired characterization of the approximation spaces follows immediately by interpolation. Throughout this section, we assume that $\mathscr{P}$ is an arbitrary dyadic partition of $\mathbb{R}^d$, $0 < p < \infty$, $\alpha > 0$, $k \geqslant 1$, and $1/\tau := \alpha + 1/p$.

Theorem 3.1. *If $f \in B_\tau^{\alpha k}(\mathscr{P})$, then*

$$\sigma_n(f, \mathscr{P})_p \leqslant c n^{-\alpha} ||f||_{B_\tau^{\alpha k}(\mathscr{P})}, \quad n = 1, 2, \ldots,$$

with $c = c(\alpha, p, k, d)$.

Proof. By Theorem 2.4, f can be represented in the form

$$f = \sum_{I \in \mathscr{P}} t_I \quad \text{a.e. on } \mathbb{R}^d \tag{3.1}$$

with the series converging absolutely a.e., where $t_I = \mathbb{1}_I \cdot P_I$ with $P_I \in \Pi_k$ ($t_I := \mathbb{1}_I \cdot t_{m,\eta}$ if $I \in \mathscr{P}_m$, $0 < \eta < p$). In addition to this, by Theorem 2.6,

$$||f||_{B_\tau^{\alpha k}(\mathscr{P})} \approx \left(\sum_{I \in \mathscr{P}} ||t_I||_p^\tau \right)^{1/\tau} =: \mathscr{N}(f).$$

Case I: $1 \leqslant p < \infty$. We define $\mathscr{J}_\mu := \{ I \in \mathscr{P} \colon 2^{-\mu} \mathscr{N}(f) \leqslant ||t_I||_p < 2^{-\mu+1} \mathscr{N}(f) \}$. Clearly,

$$\# \mathscr{J}_\mu \leqslant 2^{\mu\tau}. \tag{3.2}$$

We define

$$g_\mu := \sum_{I\in\mathscr{J}_\mu} t_I, \quad g_\mu^\diamond := \sum_{I\in\mathscr{J}_\mu} |t_I|, \quad \text{and} \quad G_m := \sum_{\mu\leqslant m} g_\mu.$$

We have $G_m \in \Sigma_M^k(\mathscr{P})$ with $M := \sum_{\mu\leqslant m} 2^{\mu\tau} = c2^{m\tau}$. We use (3.1), (3.2), and Lemma 7.1 (as in the proof of Theorem 2.5) to obtain

$$\begin{aligned}\sigma_M(f,\mathscr{P})_p &\leqslant \left\| \sum_{I\in\mathscr{P}\setminus\bigcup_{\mu\leqslant m}\mathscr{J}_\mu} |t_I| \right\|_p \leqslant \left\| \sum_{\mu=m+1}^\infty g_\mu^\diamond \right\|_p \leqslant \sum_{\mu=m+1}^\infty \left\| g_\mu^\diamond \right\|_p \\ &\leqslant c \sum_{\mu=m+1}^\infty 2^{-\mu}\mathscr{N}(f)(\#\mathscr{J}_\mu)^{1/p} \leqslant c\mathscr{N}(f) \sum_{\mu=m+1}^\infty 2^{-\mu(1-\tau/p)} \\ &\leqslant c\mathscr{N}(f)2^{-m(1-\tau/p)} = cM^{-1/\tau+1/p}\mathscr{N}(f) = cM^{-\alpha}\mathscr{N}(f)\end{aligned}$$

which implies the theorem in Case I.

Case II: $0<p<1$. We let $\|t_{I_1}\|_p \geqslant \|t_{I_2}\|_p \geqslant \cdots$ be a nonincreasing rearrangement of the sequence $\{\|t_I\|_p\}$ and define

$$\varphi := \sum_{j=1}^n t_{I_j}, \quad \varphi\in\Sigma_n^k(\mathscr{P}).$$

To estimate $\|f-\varphi\|_p$ we shall use the following simple inequality: If $x_1 \geqslant x_2 \geqslant \cdots \geqslant 0$ and $0<\tau<p$, then

$$\left(\sum_{j=n+1}^\infty x_j^p\right)^{1/p} \leqslant n^{1/p-1/\tau}\left(\sum_{j=1}^\infty x_j^\tau\right)^{1/\tau}.$$

We obtain

$$\begin{aligned}\|f-\varphi\|_p &\leqslant \left\| \sum_{j=n+1}^\infty |t_{I_j}| \right\|_p \leqslant \left(\sum_{j=n+1}^\infty \|t_{I_j}\|_p^p\right)^{1/p} \leqslant cn^{1/p-1/\tau}\left(\sum_{j=1}^\infty \|t_{I_j}\|_p^\tau\right)^{1/\tau} \\ &\leqslant cn^{-\alpha}\|f\|_{B_\tau^{\alpha k}(\mathscr{P})}. \qquad \square\end{aligned}$$

Theorem 3.2. *If* $\varphi\in\Sigma_n^k(\mathscr{P})$, *then*

$$\|\varphi\|_{B_\tau^{\alpha k}(\mathscr{P})} \leqslant cn^\alpha\|\varphi\|_p \tag{3.3}$$

with $c = c(\alpha, p, k, d)$.

Proof. Let $\varphi = \sum_{I\in\Lambda} \mathbb{1}_I \cdot P_I$, where $P_I\in\Pi_k$, $\Lambda\subset\mathscr{P}$, $\#\Lambda\leqslant n$, $n\geqslant 1$. To prove (3.3), we shall use the natural tree structure in $\mathscr{P}$ induced by the inclusion relation: Each box $I\in\mathscr{P}$ has two *children* (boxes $J_1, J_2\subset I$ such that $I = J_1\cup J_2$ and $|J_1| = |J_2| = (1/2)|I|$) and one *parent* in $\mathscr{P}$. Let $I_0\in\mathscr{P}$ be the smallest box containing all boxes from Λ and let $\mathscr{T}$ be the minimal binary subtree of $\mathscr{P}$ containing $\Lambda\cup\{I_0\}$. So, $\mathscr{T}$ is the set of all boxes in $\mathscr{P}$ which contain at least one box from Λ and are contained in I_0. We introduce the following subsets of $\mathscr{T}$:

(i) $\mathscr{T}^1$ the set of all *final boxes* in $\mathscr{T}$ (boxes not containing other boxes from $\mathscr{T}$),

(ii) $\mathscr{T}^2$ the set of all *branching boxes* in $\mathscr{T}$ (boxes with both children in $\mathscr{T}$) and, in addition, we include I_0 in $\mathscr{T}^2$,

(iii) $\mathscr{T}^3$ the set of all *children of branching boxes* in $\mathscr{T}$,

(iv) $\mathscr{T}^4$ the set of all *chain boxes* in $\mathscr{T}$ (boxes with exactly one child in $\mathscr{T}$), excluding I_0 if I_0 has only one child in $\mathscr{T}$.

Obviously, $\mathscr{T}^1 \subset \Lambda$ and hence $\#\mathscr{T}^2 \leqslant \#\mathscr{T}^1 \leqslant n$ and $\#\mathscr{T}^3 \leqslant 2n$. Note that $\#\mathscr{T}^4$ can be much larger than $\#\Lambda$.

The sets Λ and $\mathscr{T}$ generate a natural subdivision of I_0 into a union of disjoint rings. By definition, R is a *ring* if $R = I \backslash J$ with $I \in \mathscr{P}$ and $J \in \mathscr{P}$ or $J = \emptyset$. We say that $R = I \backslash J$ is a *maximal ring* if (a) $I \in \mathscr{T}$ and $J \in \mathscr{T}$ or $J = \emptyset$, (b) R does not contain boxes from Λ which are smaller than I, and (c) R is maximal with these two properties (R is not contained in another such). We denote by $\mathscr{R}$ the set of all maximal rings (generated by Λ). For $R \in \mathscr{R}$, we denote by I_R and J_R the defining boxes of R, that is, $R =: I_R \backslash J_R$ with $I_R \in \mathscr{T}$ and $J_R \in \mathscr{T}$ or $J_R = \emptyset$. Going further, we denote $\mathscr{R}_m := \{R \in \mathscr{R}\colon |I_R| = 2^{-m}\}$. Then $\mathscr{R} = \bigcup_{m \in \mathbb{Z}} \mathscr{R}_m$. Clearly, $\mathscr{R}$ consists of disjoint subsets of I_0 and $I_0 = \bigcup_{R \in \mathscr{R}} R$. It is readily seen that for each $R \in \mathscr{R}$, we have $I_R \in \mathscr{T}^1$ or $I_R \in T^3$ or $I_R \in \mathscr{T} \cap \Lambda$ or $I_R = I_0$. Therefore, $\#\mathscr{R} \leqslant \#\mathscr{T}^1 + \#\mathscr{T}^3 + \#\Lambda \leqslant 4n$.

Also, we introduce *subrings* (of maximal rings). Suppose $R \in \mathscr{R}$ and $R = I_R \backslash J_R$ with $I_R \in \mathscr{P}_\ell$, $J_R \in \mathscr{P}_{\ell+\mu}$ $(\mu \geqslant 1)$. Clearly, for each $\ell \leqslant m < \ell + \mu$, there exists a unique $I' \in \mathscr{P}_m$ such that $J_R \subset I' \subset I_R$. We now define the subring $K_{R,m}$ of R by $K_{R,m} := I' \backslash J_R$. In addition, we define $\varphi_R := \mathbb{1}_R \cdot \varphi$ and $\varphi_{R,m} := \mathbb{1}_{K_{R,m}} \cdot \varphi = \mathbb{1}_{K_{R,m}} \cdot \varphi_R$ for $\ell \leqslant m < \ell + \mu$ and $\varphi_{R,m} := 0$ if $m < \ell$ or $m \geqslant \ell + \mu$. Note that φ_R is the restriction on R of a polynomial of degree $< k$ and $\varphi_{R,m}$ is the restriction of the same polynomial on $K_{R,m} \subset R$. Denote $\mathscr{K}_m := \{R \in \mathscr{R} : K_{R,m} \neq \emptyset\}$. It is easily seen that if $I \subset I_0$, $I \in \mathscr{P}_m$ $(m \in \mathbb{Z})$, and φ is not a polynomial on I, then

$$I = \bigcup_{R \in \mathscr{R},\ R \subset I} R \bigcup \bigcup_{R \subset \mathscr{K}_m,\ R \cap I \neq \emptyset} K_{R,m} \quad \text{(disjoint sets)}, \tag{3.4}$$

where the union on the right contains exactly one subring or none.

We need to estimate $\omega_k(\varphi, I)_\tau$ for every $I \in \mathscr{P}$. There are two possibilities for $I \in \mathscr{P}$:

(i) If $I \cap I_0 = \emptyset$ or $I \subset I_0$ but $I \subset R$ for some $R \in \mathscr{R}$, then φ is a polynomial of degree $< k$ on I and hence $\omega_k(\varphi, I)_\tau = 0$.

(ii) If φ is not a polynomial on I and $I \in \mathscr{P}_m$ $(m \in \mathbb{Z})$, then we have, using (3.4),

$$\omega_k(\varphi, I)_\tau^\tau \leqslant c \|\varphi\|_{L_\tau(I)}^\tau \leqslant c \sum_{\nu=m+1}^{\infty} \sum_{R \in \mathscr{R}_\nu,\ R \subset I} \|\varphi_R\|_\tau^\tau + c \sum_{R \in \mathscr{K}_m,\ R \cap I \neq \emptyset} \|\varphi_{R,m}\|_\tau^\tau,$$

where the second sum contains one element or none. We use this estimate to obtain

$$\begin{aligned}||\varphi||^{\tau}_{B^{\alpha k}_{\tau}(\mathscr{P})} &\coloneqq \sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{I\in\mathscr{P}_m} \omega_k(\varphi, I)^{\tau}_{\tau}\\ &\leqslant c\sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{v=m+1}^{\infty} \sum_{R\in\mathscr{R}_v} ||\varphi_R||^{\tau}_{\tau} + c\sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{R\in\mathscr{K}_m} ||\varphi_{R,m}||^{\tau}_{\tau}\\ &=: \Sigma_1 + \Sigma_2.\end{aligned}$$

Applying inequality (2.12) to the first sum above, we find

$$\Sigma_1 \leqslant c\sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{R\in\mathscr{R}_m} ||\varphi_R||^{\tau}_{\tau} \leqslant c\sum_{R\in\mathscr{R}} ||\varphi_R||^{\tau}_{p},$$

where we used that $||\varphi_R||_{\tau} \leqslant |R|^{1/\tau-1/p}||\varphi_R||_p \leqslant 2^{-\alpha m}||\varphi_R||_p$, $R\in\mathscr{R}_m$, by Hölder's inequality.

We shall estimate Σ_2 using the following inequality:

$$\sum_{m\in\mathbb{Z}} ||\varphi_{R,m}||^{\tau}_{p} \leqslant c||\varphi_R||^{\tau}_{p}, \quad R\in\mathscr{R}. \tag{3.5}$$

To prove this inequality, suppose that $R = I_R\backslash J_R$ with $I_R\in\mathscr{P}_{\ell}$ and $J_R\in\mathscr{P}_{\ell+\mu}$. Using Lemma 2.1, we obtain, for $0\leqslant j<\mu$,

$$||\varphi_{R,\ell+j}||_p \leqslant |K_{R,\ell+j}|^{1/p}||\varphi_R||_{\infty} \leqslant c|K_{R,\ell+j}|^{1/p}|R|^{-1/p}||\varphi_R||_p \leqslant c2^{-j/p}||\varphi_R||_p,$$

which implies (3.5).

As above, by Hölder's inequality, $||\varphi_{R,m}||_{\tau} \leqslant 2^{-m\alpha}||\varphi_{R,m}||_p$. This and (3.5) imply

$$\Sigma_2 \leqslant c\sum_{m\in\mathbb{Z}} \sum_{R\in\mathscr{K}_m} ||\varphi_{R,m}||^{\tau}_{p} \leqslant c\sum_{R\in\mathscr{R}} \sum_{m\in\mathbb{Z}} ||\varphi_{R,m}||^{\tau}_{p} \leqslant c\sum_{R\in\mathscr{R}} ||\varphi_R||^{\tau}_{p},$$

where we switched the order of summation. From the above estimates for Σ_1 and Σ_2, we get

$$\begin{aligned}||\varphi||^{\tau}_{B^{\alpha k}_{\tau}(\mathscr{P})} &\leqslant c\sum_{R\in\mathscr{R}} ||\varphi_R||^{\tau}_{p} \leqslant c\left(\sum_{R\in\mathscr{R}} ||\varphi_R||^{p}_{p}\right)^{\tau/p} (\#\mathscr{R})^{1-\tau/p} \leqslant cn^{1-\tau/p}||\varphi||^{\tau}_{p}\\ &= cn^{\alpha\tau}||\varphi_R||^{\tau}_{p},\end{aligned}$$

where we used Hölder's inequality and that I_0 is a disjoint union of all $R\in\mathscr{R}$. □

We define the approximation space $A^{\gamma}_{q} \coloneqq A^{\gamma}_{q}(L_p, \mathscr{P})$ as the set of all functions $f\in\mathbb{L}_p(\mathscr{P}, k)$ such that

$$||f||_{A^{\gamma}_{q}} \coloneqq ||f||_p + \left(\sum_{n=1}^{\infty} (n^{\gamma}\sigma^{k}_{n}(f, \mathscr{P})_p)^q \frac{1}{n}\right)^{1/q} < \infty \tag{3.6}$$

with the ℓ_q-norm replaced by the sup-norm if $q = \infty$ as usual.

We now recall some basic definitions from the real interpolation method. We refer the reader to [1] as a general reference for interpolation theory. Suppose X and B are two quasi-normed spaces and $B\subset X$. The K-functional is defined for each $f\in X$ and

$t > 0$ by

$$K(f,t) := K(f,t;X,B) := \inf_{g \in B} (\|f-g\|_X + t\|g\|_B).$$

The real interpolation space $(X,B)_{\lambda,q}$ with $0<\lambda<1$ and $0<q\leqslant\infty$ is defined as the set of all $f \in X$ such that

$$\|f\|_{(X,B)_{\lambda,q}} := \|f\|_X + \left(\int_0^\infty (t^{-\lambda}K(f,t))^q \frac{dt}{t}\right)^{1/q} < \infty,$$

where the L_q-norm is replaced by the sup-norm if $q=\infty$.

The Jackson and Bernstein inequalities from Theorems 3.1 and 3.2 yield (see [6,20]) the following characterization of the approximation spaces A_q^γ:

Theorem 3.3. *We have, for* $0<\gamma<\alpha$ *and* $0<q\leqslant\infty$,

$$A_q^\gamma(L_p,\mathscr{P}) = (\mathbb{L}_p(\mathscr{P},k), B_\tau^{\alpha k}(\mathscr{P}))_{\gamma/\alpha,q}$$

with equivalent norms.

We next show that in one specific case the interpolation space as well as the corresponding approximation space can be identified as a B-space. The analogue of this result for Besov spaces is well known (see [8]).

Theorem 3.4. *Suppose* $\mathscr{P}$ *is a partition of* $\mathbb{R}^d$, $k\geqslant 1$, $1\leqslant p<\infty$, *and* $1/\tau := \alpha + 1/p$. *Let* $0<\alpha<\beta$ and $1/\lambda := \beta + 1/p$. *We have*

$$(\mathbb{L}_p(\mathscr{P},k), B_\lambda^{\beta k}(\mathscr{P}))_{\alpha/\beta,\tau} = B_\tau^{\alpha k}(\mathscr{P}) = A_\tau^\alpha(L_p,\mathscr{P})$$

with equivalent norms.

This theorem can be proved by using the machinery of interpolation spaces (see [8]). Here we take another route by employing the approximation from piecewise polynomials directly. This approach will enable us to reveal more deeply the intricacies of nonlinear piecewise polynomial approximation. In order to streamline the presentation of our results, we give the proof of this theorem in Appendix A.

Approximation scheme for nonlinear piecewise polynomial approximation: We assume that $f \in L_p(\mathbb{R}^d)$, $0<p<\infty$, and $\mathscr{P}$ is an arbitrary dyadic partition of $\mathbb{R}^d$. The proof of Theorem 3.1 suggests the following approximation procedure:

Step 1: Use the local polynomial approximation to represent f as follows:

$$f = \sum_{m\in\mathbb{Z}} t_{m,\eta}(f,\mathscr{P}) = \sum_{I\in\mathscr{P}} t_{I,\eta}(f),$$

where $t_{I,\eta}(f) = \mathbb{1}_I \cdot t_{m,\eta}(f,\mathscr{P})$ if $I \in \mathscr{P}_m$ and $\eta<p$ (see Theorem 3.1).

Step 2: Order $\{\|t_{I,\eta}(f)\|_p\}_{I\in\mathscr{P}}$ in a nonincreasing sequence $\|t_{I_1,\eta}(f)\|_p \geqslant \| t_{I_2,\eta}(f)\|_p \geqslant \cdots$ and then define the algorithm by

$$\mathscr{A}_n(f,\mathscr{P})_p := \sum_{j=1}^{n} t_{I_j,\eta}(f).$$

By Theorem 3.1 and its proof, it follows that

$$\|f - \mathscr{A}_n(f)_p\|_p \leqslant cn^{-\alpha}\|f\|_{B_\tau^{\alpha k}(\mathscr{P})}, \quad \text{for } f\in B_\tau^{\alpha k}(\mathscr{P}).$$

Using this result, one can show that $\mathscr{A}_n(f,\mathscr{P})_p$ achieves the rate of the best n-term piecewise polynomial approximation generated by $\mathscr{P}$.

Nonlinear approximation from the library $\{\Sigma_n^k(\mathscr{P})\}_{\mathscr{P}}$: We denote

$$\sigma_n(f)_p := \inf_{\mathscr{P}} \sigma_n(f,\mathscr{P})_p, \tag{3.7}$$

where the infimum is taken over all dyadic partitions $\mathscr{P}$. The following theorem is immediate from the Jackson estimate in Theorem 3.1:

Theorem 3.5. *If* $\inf_{\mathscr{P}} \|f\|_{B_\tau^{\alpha k}(\mathscr{P})} < \infty$, *then*

$$\sigma_n(f)_p \leqslant cn^{-\alpha} \inf_{\mathscr{P}} \|f\|_{B_\tau^{\alpha k}(\mathscr{P})}$$

with $c = c(\alpha, k, p, d)$.

In Section 5, we shall show that, in a natural discrete setting, there exists an effective algorithm for finding a partition $\mathscr{P}^*$ which minimizes $B_\tau^{\alpha k}(\mathscr{P})$ over all dyadic partitions $\mathscr{P}$.

Remark. There exists another technique that can be employed for the proof of Theorem 3.1. This method is called "splitting and merging" and has been introduced in [4] and used for nonlinear approximation of functions from the space $BV(\mathbb{R}^2)$. It was further used in [11]. Also, the modulus $W(f,t)_{\sigma,p}$, used in [11] which is a generalization of a characteristic from [16] ($d=1$), can be generalized and utilized for anisotropic partitions $\mathscr{P}$.

4. Relation between n-term rational and piecewise polynomial approximation

n-term rational functions: We denote by $\mathscr{R}_n$ the set of all n-term rational functions on $\mathbb{R}^d$ of the form

$$R = \sum_{j=1}^{n} r_j,$$

where each r_j is of the form

$$r(x) = \prod_{k=1}^{d} \frac{a_k x_k + b_k}{(x_k - \alpha_k)^2 + \beta_k^2}, \quad a_k, b_k, \alpha_k, \beta_k \in \mathbb{R}, \ \ \beta_k \neq 0,$$
$$x := (x_1, \dots, x_d) \in \mathbb{R}^d. \tag{4.1}$$

Evidently, every $R \in \mathscr{R}_n$ depends on $\leqslant 4dn$ parameters and $\mathscr{R}_n$ is nonlinear. We denote by $R_n(f)_p$ the error of L_p-approximation to f from $\mathscr{R}_n$:

$$R_n(f)_p := \inf_{R \in \mathscr{R}_n} \|f - R\|_p.$$

Our first goal is to show that the rate of n-term rational approximation in L_p $(0 < p < \infty)$ is not worse than the one of nonlinear n-term approximation from piecewise polynomials over nested box partitions of $\mathbb{R}^d$.

Piecewise polynomials over almost nested families of boxes: We denote by $\mathscr{J}$ the set of all semi-open boxes I in $\mathbb{R}^d$ (not necessarily dyadic) with sides parallel to the coordinate axes $(I = \mathscr{I}_1 \times \cdots \times \mathscr{I}_d)$.

Suppose $\Xi_n \subset \mathscr{J}$, $n = 0, 1, \dots$, is a sequence of sets of boxes which satisfy the following:

(i) $\#\Xi_n \leqslant 2^n$.
(ii) For each $n \geqslant 1$ there exists a set Ω_n consisting of disjoint boxes from $\mathscr{J}$ such that
  (a) $\bigcup\{I\colon I \in \Omega_n\} = \bigcup\{I\colon I \in \Xi_n \cup \Xi_{n-1}\}$,
  (b) for each $I \in \Omega_n$ and $J \in \Xi_n \cup \Xi_{n-1}$ either $I \subset J$ or $I \cap J = \emptyset$, and
  (c) $\#\Omega_n \leqslant c_1 2^n$.

Thus Ω_n is a set of "small" disjoint boxes which cover the boxes from $\Xi_n \cup \Xi_{n-1}$. Now, we denote by $\mathscr{S}^k(\Xi_n)$ the set of all piecewise polynomials of degree $< k$ on the boxes from Ξ_n, i.e., $\phi \in \mathscr{S}^k(\Xi_n)$ if $\phi = \sum_{I \in \Xi_n} \mathbb{1}_I \cdot P_I, P_I \in \Pi_k$. We denote by $\mathbb{S}^k_{2^n}(f)_p$ the error of L_p approximation to $f \in L_p(\mathbb{R}^d)$ from $\mathscr{S}^k(\Xi_n)$, i.e.,

$$\mathbb{S}^k_{2^n}(f)_p := \mathbb{S}^k_{2^n}(f, \Xi_n)_p := \inf_{\phi \in \mathscr{S}^k(\Xi_n)} \|f - \phi\|_p.$$

Main results: Our primary goal in this section is to prove the following theorem that relates the n-term rational approximation to the above described piecewise polynomial approximation:

Theorem 4.1. *Let $f \in L_p(\mathbb{R}^d)$, $0 < p < \infty$, $\alpha > 0$, and $k \geqslant 1$. Then*

$$R_{2^n}(f)_p \leqslant c 2^{-\alpha n} \left(\sum_{v=0}^{n} [2^{\alpha v} \mathbb{S}^k_{2^v}(f)_p]^\mu + \|f\|_p^\mu \right)^{1/\mu}, \quad \mu := \min\{p, 1\}, \tag{4.2}$$

with $c = c(p, k, \alpha, d, c_1)$, where c_1 is from the properties of $\{\Xi_n\}$.

We now apply the result from Theorem 4.1 to the more particular situation of nonlinear n-term piecewise polynomial approximation associated with any dyadic partition $\mathscr{P}$, developed in Section 3.

Theorem 4.2. *Suppose* $f \in L_p(\mathbb{R}^d)$, $0<p<\infty$, $\alpha>0$, $k \geqslant 1$, *and* $\mathscr{P}$ *is any anisotropic dyadic partition of* $\mathbb{R}^d$. *Then*

$$R_n(f)_p \leqslant cn^{-\alpha}\left(\sum_{m=1}^{n} \frac{1}{m}[m^\alpha \sigma_m^k(f,\mathscr{P})_p]^\mu + ||f||_p^\mu\right)^{1/\mu}, \quad \mu := \min\{p,1\}, \tag{4.3}$$

where $c = c(p,k,\alpha,d)$.

Corollary 4.3. *Suppose* $\inf_{\mathscr{P}} ||f||_{B_\tau^{\alpha k}(\mathscr{P})} < \infty$ *with* $\alpha>0$, $k \geqslant 1$, *and* $1/\tau := \alpha + 1/p$, $0<p<\infty$, *where the infimum is taken over all dyadic partitions* $\mathscr{P}$ *of* $\mathbb{R}^d$. *Then*

$$R_n(f)_p \leqslant cn^{-\alpha} \inf_{\mathscr{P}} ||f||_{B_\tau^{\alpha k}(\mathscr{P})},$$

where $c = c(\alpha,p,k,d)$.

Proof of the main results. For the proof of Theorem 4.1, we shall utilize some ideas from [15,17]. We let $\mathscr{S}_n^k(\mathscr{J})$ denote the set of all piecewise polynomials of degree k on n disjoint boxes in $\mathbb{R}^d$, i.e., $\varphi \in \mathscr{S}_n^k(\mathscr{J})$ if $\varphi = \sum_{I \in \Lambda_n} \mathbb{1}_I \cdot P_I$, where Λ_n is any collection of n disjoint boxes from $\mathscr{J}$ and $P_I \in \Pi_k$. The approximation will take place in $L_p(\mathbb{R}^d)$, $0<p<\infty$.

Theorem 4.4. *For each* $\varphi \in \mathscr{S}_m^k(\mathscr{J})$, $m \geqslant 1$, *and* $n \geqslant 1$, *there exists* $R \in \mathscr{R}_n$ *such that*

$$||\varphi - R||_p \leqslant c_2^{-1} \exp(-c_2(n/m)^{1/2d})||\varphi||_p, \tag{4.4}$$

where $c_2 = c_2(p,d,k,c_1) > 0$.

D. Newman [13] proved the remarkable result that the uniform nth degree rational approximation of $|x|$ on $[-1,1]$ is of order $O(n^{-c\sqrt{n}})$. The following lemma rests on Newman's construction.

Lemma 4.5. *For each* $\gamma>0$, $0<\delta<1$, *and* v *a positive integer*, *there exists a univariate rational function* σ *such that* deg $\sigma \leqslant c \ln(e+1/\delta)\ln(e+1/\gamma) + 4v$ *and*

$$0 \leqslant 1 - \sigma(t) < \gamma, \quad \text{if } |t| \leqslant 1-\delta,$$

$$0 \leqslant \sigma(t) < \gamma\left(\frac{1}{1+|t|}\right)^{4v}, \quad \text{if } |t| \geqslant 1,$$

$$0 \leqslant \sigma(t) < 1, \quad t \in (-\infty, \infty),$$

where c *is an absolute constant*. *Moreover*, σ *has only simple poles and*, *evidently*, *if* $\sigma = P/Q$, *then* deg $P <$ deg Q.

Proof. It follows by Lemma 8.3 of [20] (see also [17]) that there exists a rational function σ which satisfies all the conditions of Lemma 4.5 eventually except for the last one (simple poles). Evidently, adding a suitable sufficiently small constant to the denominator of σ in its representation as a quotient of two polynomials will ensure the last condition of the lemma without violating the other conditions. $\square$

For the proof of Theorem 4.4, we shall use the Fefferman-Stein vector valued maximal inequality (see [10] or [21]): *If* $0<p<\infty$, $0<q\leqslant\infty$, *and* $0<s<\min\{p,q\}$, *then for any sequence of functions* $f_1, f_2, \ldots$ *on* $\mathbb{R}^d$

$$\left\|\left(\sum_{j=1}^{\infty}[(\mathcal{M}_s f_j)(\cdot)]^q\right)^{1/q}\right\|_p \leqslant c\left\|\left(\sum_{j=1}^{\infty}|f_j(\cdot)|^q\right)^{1/q}\right\|_p, \tag{4.5}$$

where $c=c(p,q,s,d)$ *and*

$$(\mathcal{M}_s f)(x) := \sup_{I\in\mathcal{J}:\, x\in I}\left(\frac{1}{|I|}\int_I |f(y)|^s\, dy\right)^{1/s}, \quad x\in\mathbb{R}^d.$$

Lemma 4.6. *Suppose* $\varphi := \mathbb{1}_I \cdot P$ *with* $I\in\mathcal{J}$ *and* $P\in\Pi_k$, *and let* $\lambda, s>0$. *Then there exists a rational function* $R\in\mathcal{R}_\ell$ *with* $\ell\leqslant c\ln^{2d}(e+1/\lambda)$ *such that*

$$||\varphi - R||_p \leqslant c\lambda||\varphi||_p$$

and

$$|R(x)|\leqslant c\lambda|I|^{-1/p}||\varphi||_p(\mathcal{M}_s\mathbb{1}_I)(x), \quad x\in\mathbb{R}^d\backslash I,$$

where $c=c(k,p,s,d)$.

Proof. It is easily seen that

$$(\mathcal{M}_s\mathbb{1}_I)(x) = \prod_{i=1}^{d}(\mathcal{M}_s\mathbb{1}_{\mathcal{I}_i})(x_i), \quad I=\mathcal{I}_1\times\cdots\times\mathcal{I}_d \tag{4.6}$$

(product of univariate maximal functions).

We shall prove the lemma in the case when $I=Q:=[-1,1)^d$. The general case follows by change of variables. Let $0<\lambda<1$ (the case $\lambda\geqslant 1$ is obvious). Since $P\in\Pi_k$, then all norms of P are equivalent and this yields

$$|P(x)|\leqslant c||\varphi||_p \Pi_{i=1}^d(1+|x_i|)^k, \quad x\in\mathbb{R}^d\backslash\{\tfrac{1}{2}Q\}, \tag{4.7}$$

where $c=c(p,k,d)$ and $\frac{1}{2}Q:=[-\frac{1}{2},\frac{1}{2})^d$.

Let σ be the univariate rational function from Lemma 4.5, applied with $\gamma:=\lambda$, $\delta:=\min\{\lambda^p, 1/2\}$, and $\nu:=[\frac{1}{4}(k+1/s)]+1$. We define $R:=\kappa P$ with $\kappa(x):=\prod_{i=1}^d \sigma(x_i)$. By Lemma 4.5,

$$\deg\sigma\leqslant c\ln(e+1/\lambda^p)\ln(e+1/\lambda)+4\nu\leqslant c\ln^2(e+1/\lambda), \quad c=c(k,p,s),$$

and σ has only simple poles. Therefore, $R \in \mathscr{R}_\ell$ with $\ell \leqslant c \ln^{2d}(e + 1/\lambda)$. Obviously $0 \leqslant \kappa(x) < 1$, $x \in \mathbb{R}^d$. It is readily seen that

$$0 \leqslant 1 - \kappa(x) \leqslant \sum_{i=1}^{d} (1 - \sigma(x_i)) \leqslant d\lambda \quad \text{for } x \in Q_\delta := [-1 + \delta, 1 - \delta]^d.$$

Therefore,

$$\|\varphi - R\|_{L_p(Q_\delta)} = \|P(1 - \kappa)\|_{L_p(Q_\delta)} \leqslant c\lambda \|\varphi\|_p.$$

and, using (4.7),

$$\|\varphi - R\|_{L_p(Q \setminus Q_\delta)} \leqslant c \|\varphi\|_p |Q \setminus Q_\delta|^{1/p} \leqslant c\delta^{1/p} \|\varphi\|_p \leqslant c\lambda \|\varphi\|_p.$$

Finally, by (4.6) and (4.7), we find, for $x \in \mathbb{R}^d \setminus Q$,

$$\begin{aligned} |R(x)| &\leqslant c\lambda \|\varphi\|_p \prod_{i=1}^{d} \left(\frac{1}{1 + |x_i|} \right)^{4v-k} \\ &\leqslant c\lambda \|\varphi\|_p \prod_{i=1}^{d} (\mathscr{M}_s \mathbb{1}_{[-1,1]})(x_i) = c\lambda \|\varphi\|_p (\mathscr{M}_s \mathbb{1}_Q)(x), \end{aligned}$$

where we used that $4v - k \geqslant 1/s$ and hence

$$(\mathscr{M}_s \mathbb{1}_{[-1,1]})(t) = \left(\frac{2}{2 + |t|} \right)^{1/s} > \left(\frac{1}{1 + |t|} \right)^{4v-k}, \quad |t| \geqslant 1. \qquad \square$$

Proof of Theorem 4.4. Suppose $\varphi \in \mathscr{S}_m^k(\mathscr{J})$ $(m \leqslant n)$ and $\varphi =: \sum_{I \in \Lambda_m} \mathbb{1}_I \cdot P_I$, $\Lambda_m \subset \mathscr{J}$. Let $\lambda := \exp(-(n/m)^{1/2d})$ and $s := \frac{1}{2}\min\{p, 1\}$. We apply Lemma 4.6 to each function $\varphi_I := \mathbb{1}_I \cdot P_I$ to conclude that there exist rational functions $R_I \in \mathscr{R}_\ell$ with $\ell \leqslant c \ln^{2d}(e + 1/\lambda)$ such that

$$\|\varphi_I - R_I\|_p \leqslant c\lambda \|\varphi_I\|_p$$

and

$$|R_I(x)| \leqslant c\lambda \|\varphi_I\|_p |I|^{-1/p} (\mathscr{M}_s \mathbb{1}_I)(x), \quad x \in \mathbb{R}^d \setminus I.$$

We define $R := \sum_{I \in \Lambda_m} R_I$. Obviously, $R \in \mathscr{R}_{m\ell} \subset \mathscr{R}_{cn}$. We have

$$\begin{aligned} \|\varphi - R\|_p &\leqslant c \left(\sum_I \|\varphi_I - R_I\|_{L_p(I)}^p \right)^{1/p} + c\lambda \left\| \sum_I |I|^{-1/p} \|\varphi_I\|_p (\mathscr{M}_s \mathbb{1}_I)(\cdot) \right\|_p \\ &\leqslant c\lambda \left(\sum_I \|\varphi_I\|_p^p \right)^{1/p} + c\lambda \left\| \sum_I \|\varphi_I\|_p |I|^{-1/p} \mathbb{1}_I(\cdot) \right\|_p \leqslant c\lambda \|\varphi\|_p, \end{aligned}$$

where we used (4.5) with $q = 1$ and $s = \frac{1}{2}\min\{p, 1\} < \min\{p, 1\}$. Theorem 4.4 follows. $\square$

Proof of Theorem 4.1. *Case* I: $p \geqslant 1$. Evidently, there exists $\phi_v \in \mathscr{S}^k(\Xi_v)$ such that $\|f - \phi_v\|_p = \mathbb{S}_{2^v}(f)_p$. We define $\varphi_v := \phi_v - \phi_{v-1}$, $v \geqslant 1$, and $\varphi_0 := \phi_0$. Then we have, for $v \geqslant 1$,

$$\|\varphi_v\|_p \leqslant \|f - \phi_v\|_p + \|f - \phi_{v-1}\|_p = \mathbb{S}_{2^v}(f)_p + \mathbb{S}_{2^{v-1}}(f)_p \quad \text{and}$$
$$\|\varphi_0\|_p \leqslant \mathbb{S}_1(f)_p + \|f\|_p.$$

From the properties of $\{\Xi_j\}$, there exists a set of disjoint boxes $\Omega_v \subset \mathscr{J}$ such that $m_v := \#\Omega_v \leqslant c_1 2^v$ and $\varphi_v \in \mathscr{S}^k(\Omega_v)$.

We fix $j \geqslant 0$. Now, for each $v = 0, 1, \dots, j$, we apply Theorem 4.4 with $\varphi := \varphi_v$, $m := m_v$ (from above), and $n := N_v := [A 2^v(\alpha(j-v))^{2d}] + 1$, where $A := c_1(\ln 2/c_2)^{2d}$, c_2 is from Theorem 4.4. We obtain that there exist $R_v \in \mathscr{R}_{N_v}$ such that, for $v \geqslant 1$,

$$\begin{aligned}\|\varphi_v - R_v\|_p &\leqslant c_2^{-1} \exp\left(-c_2\left(\frac{N_v}{c_1 2^v}\right)^{1/2d}\right)\|\varphi_v\|_p \\ &\leqslant c 2^{-\alpha(j-v)}(\mathbb{S}_{2^v}(f)_p + \mathbb{S}_{2^{v-1}}(f)_p)\end{aligned} \tag{4.8}$$

and

$$\|\varphi_0 - R_0\|_p \leqslant c 2^{-\alpha j}\|\varphi_0\|_p \leqslant c 2^{-\alpha j}(\mathbb{S}_1(f)_p + \|f\|_p). \tag{4.9}$$

We define $R := \sum_{v=1}^{j} R_v$. Obviously, $R \in \mathscr{R}_N$ with

$$N = \sum_{v=1}^{j} N_v = \sum_{v=1}^{j} (A\alpha^{2d} 2^j (j-v)^{2d} + 1) \leqslant c_3 2^j, \quad c_3 = c_3(p, k, d, \alpha, c_1).$$

From (4.8) and (4.9), we find

$$\|f - R_N\|_p \leqslant \|f - \phi_j\|_p + \sum_{v=0}^{j} \|\varphi_v - R_v\|_p \leqslant c 2^{-\alpha j}\left(\sum_{v=0}^{j} 2^{\alpha v}\mathbb{S}_{2^v}(f)_p + \|f\|_p\right).$$

Estimate (4.2) follows from above by a suitable selection of j (depending on n).

Case II: $0 < p < 1$. The proof is similar to the one from Case I. The only difference is that, in this case, one should use the p-triangle inequality ($\|\sum g_j\|_p^p \leqslant \sum \|g_j\|_p^p$, $0 < p < 1$) instead of Minkovski's inequality. □

Proof of Theorem 4.2. We may assume that $\varphi_v \in \Sigma_{2^v}^k(\mathscr{P})$ are such that $\|f - \varphi_v\|_p = \sigma_{2^v}(f, \mathscr{P})_p$, $v = 0, 1, \dots$. Suppose $\varphi_v =: \sum_{I \in \Lambda_v} \mathbb{1}_I \cdot P_I$, where $P_I \in \Pi_k$, $\Lambda_v \subset \mathscr{P}$, and $\#\Lambda_v \leqslant 2^v$. From the proofs of Theorems 3.2 and 3.4, it follows that the sequence $\{\#\Lambda_v\}$ satisfies conditions (i) and (ii) of $\{\Xi_v\}$ and, therefore, (4.2) holds with $\mathbb{S}_{2^v}^k(f)_p$ replaced by $\sigma_{2^v}^k(\mathscr{P})$ which implies (4.3). □

Proof of Corollary 4.3. This corollary follows immediately by Theorems 3.1 and 4.2. □

Sharpness of the results: It is rather easy to see that the estimates of this section are sharp with respect to the rate of approximation. For a given $n \geq 1$, consider the function

$$f_n(x) := \left(\prod_{v=1}^{d} \sin \pi x_v\right) \cdot \mathbb{1}_{[0,4n]\times[0,1]^{d-1}}(x), \quad x := (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Since $\sin \pi x_1$ oscillates $4n$ times on $[0, 4n]$ and every n-term rational function can oscillate $\leq 2n$ times on any straight line parallel to the x_1-axes (has no more than $2n-1$ zeros), then $R_n(f_n)_p \geq c\|f_n\|_p \geq cn^{1/p}$, $0 < p < \infty$. On the other hand, evidently, if $\alpha > 0$ and $1/\tau = \alpha + 1/p$, then $\|f_n\|_{B_\tau^{d\alpha,k}(L_\tau)} \leq cn^{1/\tau}$, where $B_\tau^{d\alpha,k}(L_\tau)$ is the Besov space defined in (2.26). Therefore, $\sup_{\|f\|_{B_\tau^{d\alpha,k}(L_\tau)} \leq 1} R_n(f)_p \geq cn^{-\alpha}$ and hence the estimate from Corollary 4.3 is sharp, and similarly for the other estimates.

5. Nonlinear n-term approximation from the library of anisotropic Haar bases and best basis selection

An anisotropic Haar basis is naturally associated with each anisotropic dyadic partition $\mathscr{P}$ of a box Ω in $\mathbb{R}^d$ (or $\mathbb{R}^d$). For the sake of simplicity, we shall consider Haar bases only on a box Ω with sides parallel to the coordinate axes and $|\Omega| = 1$. Then $\mathscr{P} = \bigcup_{m=0}^{\infty} \mathscr{P}_m$. Let $I \in \mathscr{P}$ and $I =: \mathscr{I}_1 \times \cdots \times \mathscr{I}_d$. Suppose I is split (in $\mathscr{P}$) by dividing in half the vth ($1 \leq v \leq d$) side of I. Then we define $H_I := \mathbb{1}_{\mathscr{I}_1} \times \cdots \times H_{\mathscr{I}_v} \times \cdots \times \mathbb{1}_{\mathscr{I}_d}$, where $H_{\mathscr{I}_v}$ is the univariate Haar function supported on $\mathscr{I}_v$ and normalized in L_∞. In other words, if $I \in \mathscr{P}$ and J_1, J_2 are the two children of I in $\mathscr{P}$ (properly ordered), then $H_I := \mathbb{1}_{J_1} - \mathbb{1}_{J_2}$. We need to add the characteristic function of Ω to the collection of the above defined Haar functions. To this end we denote $I^0 := I_0 := \Omega$ and include both I^0 and I_0 in $\mathscr{P}_0$ and $\mathscr{P}$. So, there are two copies of Ω in $\mathscr{P}$. We define $H_{I^0} := \mathbb{1}_{I^0}$ and $\mathscr{P}^{\mathrm{o}} := \mathscr{P} \backslash \{I^0\}$.

Thus $\mathscr{H}_{\mathscr{P}} := \{H_I\colon I \in \mathscr{P}\}$ is the Haar basis associated with $\mathscr{P}$. We let $\mathbb{H} := \{\mathscr{H}_{\mathscr{P}}\}_{\mathscr{P}}$ denote the collection (library) of all anisotropic Haar bases on Ω.

Clearly, the following is valid for a fixed partition $\mathscr{P}$: (i) $\mathscr{H}_{\mathscr{P}}$ is an orthogonal system in $L_2(\Omega)$ and it is an orthogonal basis for $\mathbb{L}_2(\mathscr{P}) := \mathbb{L}_2(\mathscr{P}, 1)$. (ii) The linear space $\mathscr{S}_n^1$ of all piecewise constants over the boxes from $\mathscr{P}_n$ (see Section 2) is spanned by $\{H_I\colon I \in \bigcup_{v=0}^{n} \mathscr{P}_v\}$.

Other anisotropic Haar bases which involve products of Haar functions can easily be constructed, too. We do not consider such constructions in this article since it does not change the essence of the problems.

$\mathscr{H}_{\mathscr{P}}$ is a basis for $\mathbb{L}_p(\mathscr{P})$ and $B_\tau^{\alpha,1}(\mathscr{P})$:

Theorem 5.1. *For each dyadic partition $\mathscr{P}$ of Ω the Haar basis $\mathscr{H}_{\mathscr{P}}$ is an unconditional basis for $\mathbb{L}_p(\mathscr{P})$, $1 < p < \infty$.*

Proof. The proof can be carried out exactly as the proof in the case of the univariate Haar system due to Burkholder (see [24]) and we shall skip it. □

Throughout the rest of this section, we shall assume that $1<p<\infty$, $\alpha>0$, $1/\tau := \alpha + 1/p$, and $\mathscr{P}$ is an arbitrary dyadic partition of Ω. We naturally have (see (2.20) and (2.21))

$$\|f\|_{B_\tau^{\alpha,1}(\mathscr{P})} := \|f\|_{L_p(\Omega)} + \left(\sum_{I\in\mathscr{P}^o} |I|^{-\alpha\tau}\omega_1(f,I)_\tau^\tau\right)^{1/\tau}.$$

We next characterize the B-norm of function in $B_\tau^{\alpha,1}(\mathscr{P})$ by means of its Haar coefficients using $\mathscr{H}_{\mathscr{P}}$.

Theorem 5.2. *Every $f\in B_\tau^{\alpha,1}(\mathscr{P})$ can be represented uniquely in the form*

$$f = \sum_{I\in\mathscr{P}} c_I(f)H_I \quad \text{a.e. on } \Omega \text{ with } c_I(f) := |I|^{-1}\int_I fH_I, \tag{5.1}$$

where the series converging absolutely a.e. and unconditionally in L_p. Moreover,

$$\|f\|_{B_\tau^{\alpha,1}(\mathscr{P})} \approx \mathscr{N}(f,\mathscr{H}_{\mathscr{P}}) := \left(\sum_{I\in\mathscr{P}} |I|^{-\alpha\tau}\|c_I(f)H_I\|_\tau^\tau\right)^{1/\tau}$$
$$= \left(\sum_{I\in\mathscr{P}} |I|^{-\alpha\tau+1}|c_I(f)|^\tau\right)^{1/\tau} = \left(\sum_{I\in\mathscr{P}} \|c_I(f)H_I\|_p^\tau\right)^{1/\tau} \tag{5.2}$$

with constants of equivalence depending only on p, α, and d.

Proof. Let $f\in B_\tau^\alpha$, $B_\tau^\alpha := B_\tau^{\alpha,1}(\mathscr{P})$. By Theorems 2.7 and 2.8, $f\in L_p(\Omega)$ and hence, using Theorem 5.1, f has a unique representation in the form (5.1). We shall next prove that

$$\mathscr{N}(f,\mathscr{H}_{\mathscr{P}}) \leqslant c\|f\|_{B_\tau^\alpha}. \tag{5.3}$$

Case I: $\tau\geqslant 1$. This case is trivial because we obviously have, for $I\neq I^0$,

$$|c_{I^0}(f)| = \left|\int_{I^0} f\right| \leqslant \|f\|_p \quad \text{and} \quad |c_I(f)| = |I|^{-1}\left|\int_I fH_I\right| \leqslant |I|^{-1/\tau}\omega_1(f,I)_\tau,$$

which, in view of (5.2), imply (5.3).

Case II: $0<\tau<1$. Clearly,

$$|I^0|^{-\alpha\tau}\|c_{I^0}(f)H_{I^0}\|_\tau^\tau \leqslant \|f\|_{L_1(\Omega)} \leqslant \|f\|_{L_p(\Omega)} \quad (|I^0|=1).$$

By Theorem 2.7 with $\eta=\tau$ and $k=1$, f can be represented in the form

$$f = T_0 + \sum_{j=1}^\infty t_j = T_0 + \sum_{j=1}^\infty \sum_{I\in\mathscr{P}_j} t_I \quad \text{a.e. on } \Omega$$

with the series converging absolutely a.e., where $t_j := t_{j,\tau}(f) := T_j - T_{j-1}$, $T_j := T_{j,\tau}(f, \mathscr{P})$, and $t_I := \mathbb{1}_I \cdot t_j$ if $I \in \mathscr{P}_j$.

Fix $I \in \mathscr{P}_m$ $(m \geqslant 0)$, $I \neq I^0$. Evidently, $||c_I(f)H_I||_1 = |c_I(f)||I| \leqslant ||f - c||_{L_1(I)}$ for every constant c. Therefore,

$$||c_I(f)H_I||_1 \leqslant ||f - T_m||_{L_1(I)} \leqslant \sum_{j=m+1}^{\infty} ||t_j||_{L_1(I)},$$

which readily implies

$$\begin{aligned} |I|^{-\alpha\tau}||c_I(f)H_I||_\tau^\tau &= |I|^{-\alpha\tau+1-\tau}||c_I(f)H_I||_1^\tau \leqslant |I|^{-\gamma\tau}\left(\sum_{j=m+1}^{\infty} ||t_j||_{L_1(I)}\right)^\tau \\ &\leqslant |I|^{-\gamma\tau} \sum_{j=m+1}^{\infty} \sum_{J \in \mathscr{P}_j,\, J \subset I} ||t_J||_1^\tau \\ &\leqslant \sum_{j=m+1}^{\infty} \sum_{J \in \mathscr{P}_j,\, J \subset I} (|J|/|I|)^{\gamma\tau} ||t_J||_p^\tau, \end{aligned}$$

with $\gamma := \alpha - 1/\tau + 1 = 1 - 1/p > 0$, where we used that $\tau < 1$. We now proceed similarly as in the proof of Theorem 2.6 (see Appendix A). We substitute the above estimates in the definition of $\mathscr{N}(f, \mathscr{H}_\mathscr{P})$ in (5.2) and switch the order of summation to obtain (5.3).

In the other direction, the Haar basis $\mathscr{H}_\mathscr{P}$ obviously satisfies the conditions of Theorem 2.5 and hence

$$\left|\left|\sum_{I \in \mathscr{P}} |c_I(f)H_I(\cdot)|\right|\right|_p \leqslant c\mathscr{N}(f, \mathscr{H}_\mathscr{P}). \tag{5.4}$$

On the other hand, by Theorem 5.1, $\mathscr{H}_\mathscr{P}$ is an unconditional basis for $\mathbb{L}_p(\mathscr{P})$. Therefore,

$$f = \sum_{I \in \mathscr{P}} c_I(f)H_I \quad \text{a.e. on } \Omega$$

with the series converging absolutely a.e. and unconditionally in L_p. Using (5.4), we infer $||f||_p \leqslant c\mathscr{N}(f, \mathscr{H}_\mathscr{P})$. We utilize the above representation of f to obtain

$$\begin{aligned} S_m^1(f)_\tau &\leqslant \left|\left| f - \sum_{|I| \geqslant 2^{-m}} c_I H_I \right|\right|_\tau \leqslant \left(\sum_{j=m}^{\infty} \left|\left| \sum_{I \in \mathscr{P}_j} c_I H_I \right|\right|_\tau^\lambda \right)^{1/\lambda} \\ &= \left(\sum_{j=m}^{\infty} \left(\sum_{I \in \mathscr{P}_j} ||c_I H_I||_\tau^\tau \right)^{\lambda/\tau} \right)^{1/\lambda} \end{aligned}$$

with $\lambda := \min\{\tau, 1\}$. Now, exactly as in the proof of Theorem 2.6 (see Appendix A), we use this in (2.22) and switch the order of summation to obtain $||f||_{B_\tau^\alpha} \leqslant c\mathscr{N}(f, \mathscr{H}_\mathscr{P})$. This completes the proof of the theorem. □

Nonlinear n-term approximation from a single basis $\mathscr{H}_{\mathscr{P}}$: For a given partition $\mathscr{P}$, we denote by $\hat{\Sigma}_n(\mathscr{P})$ the set of all functions φ of the form

$$\varphi = \sum_{I \in \Lambda_n} a_I H_I,$$

where $\Lambda_n \subset \mathscr{P}$ and $\#\Lambda_n \leqslant n$. The error $\hat{\sigma}_n(f, \mathscr{H}_{\mathscr{P}})_p$ of nonlinear n-term L_p-approximation to f from $\mathscr{H}_{\mathscr{P}}$ is defined by

$$\hat{\sigma}_n(f, \mathscr{H}_{\mathscr{P}})_p := \inf_{\varphi \in \hat{\Sigma}_n(\mathscr{P})} ||f - \varphi||_{L_p(\Omega)}.$$

Clearly, $\hat{\Sigma}_n(\mathscr{P}) \subset \Sigma_{2n}(\mathscr{P})$ and hence $\sigma_{2n}(f, \mathscr{P})_p \leqslant \hat{\sigma}_n(f, \mathscr{H}_{\mathscr{P}})_p$. The approximation spaces $\hat{A}_q^\gamma := \hat{A}_q^\gamma(L_p, \mathscr{H}_{\mathscr{P}})$ generated by the n-term approximation from $\mathscr{H}_{\mathscr{P}}$ are defined similarly as the approximation spaces A_q^γ (see (3.6)). The problem again is to characterize the approximation spaces $\hat{A}_q^\gamma$ which reduces to establishing Jackson and Bernstein inequalities and interpolation.

Theorem 5.3. *Suppose* $\mathscr{P}$ *is an arbitrary partition of* Ω *and let* $1 < p < \infty$, $\alpha > 0$, *and* $1/\tau := \alpha + 1/p$. *Then the following Jackson and Bernstein inequalities hold*:

$$\hat{\sigma}_n(f, \mathscr{H}_{\mathscr{P}})_p \leqslant cn^{-\alpha} ||f||_{B_\tau^{\alpha,1}(\mathscr{P})}, \quad f \in B_\tau^{\alpha,1}(\mathscr{P}), \tag{5.5}$$

$$||\varphi||_{B_\tau^{\alpha,1}(\mathscr{P})} \leqslant cn^\alpha ||\varphi||_{L_p(\Omega)}, \quad \varphi \in \hat{\Sigma}_n(\mathscr{P}), \quad c = c(\alpha, p, d). \tag{5.6}$$

Therefore, for $0 < \gamma < \alpha$ *and* $0 < q \leqslant \infty$,

$$\hat{A}_q^\gamma(L_p, \mathscr{H}_{\mathscr{P}}) = (\mathbb{L}_p(\mathscr{P}), B_\tau^{\alpha,1}(\mathscr{P}))_{\gamma/\alpha, q} = A_q^\gamma(L_p, \mathscr{H}_{\mathscr{P}}) \tag{5.7}$$

with equivalent norms (*see Theorem* 3.3).

Proof. The Jackson estimate (5.5) can be proved, using Theorem 5.2, exactly as Theorem 3.1 was proved. The Bernstein inequality (5.6) follows by Theorem 3.2. An easier proof can be given by using that $\mathscr{H}_{\mathscr{P}}$ is an unconditional basis for $\mathbb{L}_p$ $(1 < p < \infty)$. The characterization of $\hat{A}_q^\gamma$ in (5.7) follows by (5.5) and (5.6) (see [6,20]). □

Algorithm for n-term approximation from $\mathscr{H}_{\mathscr{P}}$: We note that a near best n-term L_p-approximation from $\mathscr{H}_{\mathscr{P}}$ $(1 < p < \infty)$ to a given function $f \in \mathbb{L}_p(\mathscr{P})$ can be achieved by simply retaining the biggest (in L_p) n terms from the representation of the function f in $\mathscr{H}_{\mathscr{P}}$ (see [23]). This result suggests the following "threshold" algorithm for n-term L_p-approximation from $\mathscr{H}_{\mathscr{P}}$ $(1 < p < \infty)$:

Step 1: Find the Haar decomposition of f in $\mathscr{H}_{\mathscr{P}}$: $f = \sum_{I \in \mathscr{P}} c_I(f) H_I$.

Step 2: Order the terms of $\{||c_I(f)H_I||_p\}_{I \in \mathscr{P}}$ in a nonincreasing sequence $||c_{I_1}(f)H_{I_1}||_p \geqslant ||c_{I_2}(f)H_{I_2}||_p \geqslant \cdots$ and then define the approximant by

$$\hat{\mathscr{A}}_n(f, \mathscr{P})_p := \sum_{j=1}^n c_{I_j}(f) H_{I_j}.$$

From the above observation, $\hat{\mathscr{A}}_n(f,\mathscr{P})_p$ provides a near best n-term L_p-approximation to f from piecewise constants generated by $\mathscr{P}$.

Nonlinear n-term approximation from the library $\mathbb{H} := \{\mathscr{H}_{\mathscr{P}}\}$: We denote by $\hat{\sigma}_n(f)_p$ the error of n-term approximation of $f \in L_p$ from the best basis in $\mathbb{H}$, i.e.,

$$\hat{\sigma}_n(f)_p := \inf_{\mathscr{P}} \hat{\sigma}_n(f,\mathscr{H}_{\mathscr{P}})_p.$$

The following theorem is immediate from the Jackson estimate (5.5):

Theorem 5.4. *If* $\inf_{\mathscr{P}} \|f\|_{B^{\alpha,1}_\tau(\mathscr{P})} < \infty$, *then*

$$\hat{\sigma}_n(f)_p \leqslant c n^{-\alpha} \inf_{\mathscr{P}} \|f\|_{B^{\alpha,1}_\tau(\mathscr{P})}$$

with $c = c(p,\alpha,d)$.

Our approximation scheme for nonlinear n-term approximation of a given function $f \in L_p(\Omega)$ from the library $\mathbb{H} := \{\mathscr{H}_{\mathscr{P}}\}$ of all anisotropic Haar bases consists of two steps:

(i) Find a basis $\mathscr{H}(f) \in \mathbb{H}$ which minimizes the $B^{\alpha,1}_\tau$-norm of f.
(ii) Run the above threshold algorithm for near best n-term approximation from $\mathscr{H}(f)$.

The most significant fact in this part is that, in a natural discrete setting, there is an effective algorithm for best Haar basis selection, which we present below.

The above approximation scheme requires a priori information about the smoothness $\alpha > 0$ of the function f (which is being approximated) with respect to the optimal $B^{\alpha,1}_\tau$-scale. We do not have an effective solution for this hard problem. Of course, one can get some idea about the optimal smoothness α of a given function experimentally.

Best Haar basis or best B-space selection: We next describe a fast algorithm for best anisotropic Haar basis or best B-space selection in the discrete case of dimension $d = 2$. This algorithm is well known (see, e.g., [9] and the references therein). Also, this algorithm is somewhat related with the algorithm for best basis selection from wavelet packets (see [3]). Both algorithms rest on one and the same principle.

We consider the set $\mathscr{X}_n$ of all functions $f : [0,1)^2 \to \mathbb{R}$ which are constants on each of the $2^n \times 2^n$ "pixels"

$$I = [(i-1)2^{-n}, i2^{-n}) \times [(j-1)2^{-n}, j2^{-n}), \quad 1 \leqslant i,j \leqslant 2^n.$$

Denote by $\mathscr{D}_n$ the set of all such pixels on $[0,1)^2$. We let $\mathbb{P}_n$ denote the set of all dyadic partitions $\mathscr{P}$ of $[0,1)^2$ such that $\mathscr{P}_{2n} = \mathscr{D}_n$ and we shall consider $\mathscr{P}$ terminated at level $2n$. Thus $\mathscr{P} = \bigcup_{m=0}^{2n} \mathscr{P}_m$. Clearly, $\mathscr{X}_n = \mathscr{S}^1_n$ (see Section 2).

Motivated by the result from Theorem 5.4, our next goal is to find, for a given $f \in \mathscr{X}_n$, a dyadic partition $\mathscr{P}^* := \mathscr{P}^*(f) \in \mathbb{P}_n$ which minimizes the B-norm $\mathscr{N}(f,\mathscr{P})$ from (5.2). Evidently, for $\mathscr{P} \in \mathbb{P}_n$, $\mathscr{H}_{\mathscr{P}}$ is an orthogonal basis for the linear space $\mathscr{X}_n$

and, therefore,

$$f = \sum_{I \in \mathscr{P}} c_I(f) H_I \quad \text{with } c_I(f) := |I|^{-1} \int_I f H_I.$$

We briefly denote $d(I, \mathscr{P}) := |I|^{-\alpha\tau+1} |c_I(f)|^{\tau}$. Also, we set $d_0(I) := d(I, \mathscr{P})$ if I is subdivided, say, horizontally in $\mathscr{P}$, and $d_1(I) := d(I, \mathscr{P})$ if I is subdivided vertically in $\mathscr{P}$. Then we have, for the B-norm from (5.2),

$$\mathscr{N}(f, \mathscr{P})^{\tau} = \sum_{I \in \mathscr{P}} d(I, \mathscr{P}) =: D(\mathscr{P}).$$

For a given dyadic box J, we denote by $\mathbb{P}_J$ the set of all dyadic partitions $\mathscr{P}_J$ of J which are subpartitions of partitions from $\mathbb{P}_n$. Similarly as above, we set

$$D(\mathscr{P}_J) := \sum_{I \in \mathscr{P}_J} d(I, \mathscr{P}_J).$$

We next describe a fast algorithm for finding a partition $\mathscr{P}^* \in \mathbb{P}_n$ which minimizes the B-norm $\mathscr{N}(f, \mathscr{P})$. The idea of this construction is to proceed from fine to coarse levels minimizing $D(\mathscr{P}_J)$ for every dyadic box J at every step. More precisely, we use the following recursive procedure. We first consider all boxes J with $|J| = 2^{-2n+1}$. Each box J like this is the union of two adjacent pixels and, hence, it can be subdivided in exactly one way. Thus $\mathscr{P}_J^*$ is uniquely determined. Now, suppose that we have already found all partitions $\mathscr{P}_J^*$ of all dyadic boxes J with $|J| \leqslant 2^{-\mu}$ $(0 < \mu < 2n)$ which minimize $D(\mathscr{P}_J)$ over all partitions $\mathscr{P}_J \in \mathbb{P}_J$. Let J be an arbitrary dyadic box such that $|J| = 2^{-\mu+1}$. There are two cases to be considered.

Case I: One of the sides of J is of length 2^{-n}. Then there is only one way to subdivide J and, hence, $\mathscr{P}_J^*$ and $\min D(\mathscr{P}_J) = D(\mathscr{P}_J^*)$ are uniquely determined.

Case II: Both sides of J are of length $> 2^{-n}$. Then J can be subdivided in two possible ways: horizontally or vertically and, therefore, J has two sets of children. Let us denote by J_1° and J_2° the children of J obtained when dividing J horizontally and J_1' and J_2' the children of J obtained when dividing J vertically. The key observation is that

$$\min_{\mathscr{P}_J} D(\mathscr{P}_J) = \min\{D(\mathscr{P}^*_{J_1^{\circ}}) + D(\mathscr{P}^*_{J_2^{\circ}}) + d_0(I), D(\mathscr{P}^*_{J_1'}) + D(\mathscr{P}^*_{J_2'}) + d_1(I)\}.$$

Therefore, if $\min_{\mathscr{P}_J} D(\mathscr{P}_J)$ is attained when J is first subdivided horizontally, then $\mathscr{P}_J^* = \mathscr{P}^*_{J_1^{\circ}} \cup \mathscr{P}^*_{J_2^{\circ}} \cup \{J\}$ will be an optimal partition of J and $\mathscr{P}_J^* = \mathscr{P}^*_{J_1'} \cup \mathscr{P}^*_{J_2'} \cup \{J\}$ will be optimal in the other case. We process like this every dyadic box of area $2^{-\mu+1}$ and this completes the recursive procedure. After finitely many steps we find a partition $\mathscr{P}^*$ of Ω which minimizes $D(\mathscr{P}) = \mathscr{N}(f, \mathscr{P})^{\tau}$.

Every $f \in X_n$ belongs to any (discrete) space $B_{\tau}^{\alpha,1}(\mathscr{P})$ and we have, by Theorem 5.4,

$$\hat{\sigma}_m(f)_p \leqslant c m^{-\alpha} \inf_{\mathscr{P} \in \mathbb{P}_n} \|f\|_{B_{\tau}^{\alpha,1}(\mathscr{P})}, \quad m = 1, 2, \ldots.$$

Once the smoothness parameter $\alpha > 0$ is fixed, the above algorithm provides a basis which minimizes the B_{τ}^{α}-norm of f. It is a problem to find the optimal smoothness α of f.

Several remarks are in order: (i) For a given function $f \in \mathcal{X}_n$, the number of all coefficients $c_I(f)$ (or Haar functions H_I) that participate in the representations of f in all anisotropic Haar bases is $\leqslant 2N$, where $N := 2^{2n}$ is the number of the pixels. Moreover, these coefficients can be found by $O(N)$ operations.

(ii) For a given function $f \in \mathcal{X}_n$ and fixed indices α and τ, only $O(N)$ operations are needed to find a Haar basis $\mathcal{H}(f)$ which minimizes the $B_\tau^{\alpha,1}$-norm $\mathcal{N}(f, \mathcal{P})$.

(iii) Another $O(N \ln N)$ operations (mainly for ordering the coefficients) are needed for finding a near best n-term approximation to f from the best Haar basis $\mathcal{H}(f)$.

The above idea for best basis selection can be utilized for best B-space selection, namely, for the selection of a partition $\mathcal{P}^*$ which minimizes the B-norm $||f||_{B_\tau^{\alpha k}(\mathcal{P})}$ of a given function f when $k > 1$. Indeed, precisely as above we can find a partition $\mathcal{P}^* \in \mathbb{P}_n$ which minimizes $||f||^\tau_{B_\tau^{\alpha k}(\mathcal{P})}$ or an equivalent norm.

6. Concluding remarks and open problems

Our results from Section 4 show that the set of n-term rational functions is a powerful tool for approximation. The n-term rational functions that we consider, however, depend on the coordinate system. It is natural to consider the more general n-term rational functions of the form $R = \sum_{j=1}^n r_j$, where each r_j is of the form $r(Ax)$ with r from (4.1) and A any affine transform. The set of all such rational functions is independent of the coordinate system. Here we do not consider such more general approximation because our approximation method is limited by the conditions on the maximal inequality we use (see Section 4). We believe that n-term rational approximation should be considered as a special case of the more general n-term approximation from the collection (dictionary) of all functions of the form $\varphi(u_1x_1 + v_1, \dots, u_dx_d + v_d)$, or $\varphi(Ax)$, A an affine transform, where φ is a fixed smooth and well localized function such as $\varphi(x) := e^{-|x|^2}$. The ultimate goal of the theory of n-term rational approximation (of any type) is to characterize the corresponding approximation spaces. This article does not solve that problem but shows that the smoothness spaces which govern n-term rational approximation are fairly sophisticated ones.

We now turn to the fundamental question in nonlinear approximation (and not only there) of how to measure the smoothness of the functions. In [18], we showed that all rates of nonlinear univariate spline approximation are governed by the scale of Besov spaces $B_\tau^{\alpha,k}(L_\tau)$ $(1/\tau := \alpha + 1/p)$. For more sophisticated multivariate nonlinear approximation, however, the Besov spaces are inappropriate. This is clearly the case when the approximation tool contains functions supported on long and narrow regions or have elongated level curves like the piecewise polynomials and rational functions considered in this paper (see the end of Section 2). It is crystal clear to us that for highly nonlinear approximation such as the multivariate piecewise polynomial approximation considered in Sections 3 and 5 there does not exist a single super space scale (like the Besov spaces) suitable for measuring the smoothness. We believe that in many cases the smoothness of the functions should be measured by means of an appropriate collection of space scales which should vary with the approximation process. To illustrate this idea we return to the piecewise polynomial approximation considered in Section 3. For this type of approximation, a function f should naturally be considered smooth of order $\alpha > 0$ if $\inf_{\mathscr{P}} \|f\|_{B_\tau^{\alpha k}(\mathscr{P})} < \infty$, which means that there exists a partition $\mathscr{P}^*$ such that $\|f\|_{B_\tau^{\alpha k}(\mathscr{P}^*)} < \infty$. Then the rate of n-term piecewise polynomial (of degree $< k$) approximation to f is roughly $O(n^{-\alpha})$. It is an *open problem* to characterize the approximation spaces generated by $\{\sigma_n(f)_p\}$ (see (3.7)).

Clearly, in nonlinear piecewise polynomial or rational approximation there is no saturation, which means that the corresponding approximation spaces A_q^γ are nontrivial for all $\gamma > 0$. Therefore, it is highly desirable that the smoothness spaces we use characterize the approximation spaces A_q^γ for all $0 < \gamma < \infty$. This was a guiding principle to us in designing the B-spaces in this article. Notice that all our approximation results from Sections 3 and 5 hold for each $\alpha > 0$. To make this point more transparent, we shall next briefly compare our results with existing ones, which involve Besov spaces. We first note that the situation in the univariate case is quite unique, since the scale of Besov spaces $B_\tau^{\alpha,k}(L_\tau)$ $(1/\tau = \alpha + 1/p)$ governs all rates of nonlinear piecewise polynomial approximation (see [18]). Therefore, there is no reason for introducing B-spaces in dimension $d = 1$. They would be equivalent to the corresponding univariate Besov spaces and hence useless. Besov spaces are also used in dimensions $d > 1$ (see [5,7,11]), but they are not the right smoothness spaces even for nonlinear piecewise polynomial approximation generated by regular partitions. It follows by the discussion at the end of Section 2 (see (2.28)) and by Theorems 3.1–3.3 that the Besov spaces $B_\tau^{d\alpha,k}(L_\tau)$ can do the job when $0 < \alpha < 1/p$ and they fail when $\alpha \geqslant 1/p$. Of course, this range for α is wider when approximating from smoother piecewise polynomials (see [5,7]). In a nutshell, the Besov spaces are the right smoothness spaces for characterization of nonlinear piecewise polynomial approximation in dimensions $d > 1$ only for regular partitions and for a limited range of approximation rates, and they are completely unsuitable in the anisotropic case.

Another important element of our concept is to have, together with the library of spaces, a companion library of bases which are (unconditional) bases for the spaces of interest. Such a library of bases should provide an effective tool for nonlinear approximation. As in this paper, we conveniently have the library of anisotropic Haar bases $\{\mathscr{H}_{\mathscr{P}}\}_{\mathscr{P}}$ which are unconditional bases for $\{\mathbb{L}_p(\mathscr{P})\}_{\mathscr{P}}$ and characterize the $B_\tau^{\alpha,1}(\mathscr{P})$-spaces.

An *open problem* for bases is to construct libraries of anisotropic bases consisting of smooth functions.

Next, we pose some more delicate *problems* about the library of anisotropic Haar bases $\mathbb{H}$: *The ultimate problem is to characterize the approximation spaces generated by* $\{\hat{\sigma}_n(f)_p\}$. The difficulty of this problem stems from the highly nonlinear nature of the approximation from the library $\mathbb{H}$. This problem is intimately connected to the problem for existence of a near best (or best) basis: *For a given function* $f \in L_p$, *does there exist a single Haar basis* $\mathscr{H}(f) \in \mathbb{H}$ *such that*

$$\hat{\sigma}_{cn}(f, \mathscr{H}(f))_p \leqslant c \inf_{\mathscr{H} \in \mathbb{H}} \hat{\sigma}_n(f, \mathscr{H})_p$$

for all $n \geqslant 1$ *with a constant* c

independent of f *and* n?

The answer of this question is not known even for $p = 2$. If the answer of the latter question is "Yes", then the approximation of any $f \in L_p$ from the library of anisotropic Haar bases $\mathbb{H}$ could be realized by approximation from a single basis $\mathscr{H}(f)$ and characterized by the interpolation spaces generated by $B^\alpha_\tau(\mathscr{P}^*)$, where $\mathscr{P}^*$ is determined from $\mathscr{H}_{\mathscr{P}^*} = \mathscr{H}(f)$.

Acknowledgments

The author is indebted to the referees for their constructive suggestions for improvements.

Appendix A

A.1. Proof of Theorems 2.4–2.6

For the proof of Theorem 2.5, we need the following lemma:

Lemma A.1. *Suppose* $\{\Phi_m\}$ *satisfies conditions* (i) *and* (ii) *of Theorem* 2.5 *and* $p \geqslant 1$. *Let* $F := \sum_{j \in \mathscr{J}_n} |\Phi_j|$, *where* $\#\mathscr{J}_n \leqslant n$, *and* $||\Phi_j||_p \leqslant A$ *for* $j \in \mathscr{J}_n$. *Then*

$$||F||_p \leqslant cAn^{1/p} \quad \text{with } c = c(c_1).$$

Proof. Using (i), we have

$$||F||_p \leqslant \left\| \sum_{j \in \mathscr{J}_n} ||\Phi_j||_\infty \mathbb{1}_{E_j}(\cdot) \right\|_p \leqslant c_1 A \left\| \sum_{j \in \mathscr{J}_n} |E_j|^{-1/p} \mathbb{1}_{E_j}(\cdot) \right\|_p.$$

We define

$$E := \bigcup_{j \in \mathscr{J}_n} E_j \quad \text{and} \quad \lambda(x) := \min\{|E_j| \colon j \in \mathscr{J}_n \text{ and } E_j \ni x\}, \quad x \in E.$$

Evidently, property (ii) yields $\sum_{j\in\mathcal{J}_n} |E_j|^{-1/p}\mathbb{1}_{E_j}(x)\leqslant c_1\lambda(x)^{-1/p}$, $x\in\mathbb{R}^d$. Therefore,

$$\|F\|_p\leqslant cA\|\lambda(\cdot)^{-1/p}\|_{L_p} = cA\left(\int_E \lambda(x)^{-1}\,dx\right)^{1/p}$$
$$\leqslant cA\left(\sum_{j\in\mathcal{J}_n}|E_j|^{-1}\int_{\mathbb{R}^d}\mathbb{1}_{E_j}(x)\,dx\right)^{1/p} = cA(\#\mathcal{J}_n)^{1/p}\leqslant cAn^{1/p}. \qquad \square$$

Proof of Theorem 2.4. The theorem is trivial if $0<\tau\leqslant 1$. Let $\tau>1$. Then $p>1$. Let $\{\Phi_j^*\}_{j=1}^\infty$ be a rearrangement of the sequence $\{\Phi_j\}$ so that $\|\Phi_1^*\|_p\geqslant\|\Phi_2^*\|_p\geqslant\cdots$. Obviously,

$$\|\Phi_j^*\|_p\leqslant j^{-1/\tau}\mathcal{N}, \quad \text{where } \mathcal{N} := \left(\sum_j \|\Phi_j\|_p^\tau\right)^{1/\tau}. \tag{A.1}$$

We define $\mathcal{J}_m := \{j\colon 2^{-m}\mathcal{N}\leqslant\|\Phi_j\|_p<2^{-m+1}\mathcal{N}\}$. Then $\bigcup_{\mu\leqslant m}\mathcal{J}_\mu = \{j\colon \|\Phi_j\|_p\geqslant 2^{-m}\mathcal{N}\}$ and hence, using (A.1),

$$\#\mathcal{J}_m\leqslant\#\left(\sum_{\mu\leqslant m}\mathcal{J}_\mu\right)\leqslant 2^{m\tau}. \tag{A.2}$$

We denote $F_m := \sum_{j\in\mathcal{J}_m}|\Phi_j|$. Using Lemma A.1 and (A.2), we obtain

$$\left\|\sum_j |\Phi_j(\cdot)|\right\|_p \leqslant \sum_{m=0}^\infty \|F_m\|_p\leqslant c\sum_{m=0}^\infty (\#\mathcal{J}_m)^{1/p}2^{-m}\mathcal{N}$$
$$= c\mathcal{N}\sum_{m=0}^\infty 2^{-m\tau(1/\tau-1/p)}\leqslant c\mathcal{N}. \qquad \square$$

Proof of Theorem 2.5. *Case* I: $1\leqslant p<\infty$. We introduce the following abbreviated notation: $T_m := T_{m,\eta}(f)$, $t_m := t_{m,\eta}(f)$, and $t_I := \mathbb{1}_I\cdot t_m$ if $I\in\mathcal{P}_m$, $m\in\mathbb{Z}$ (see (2.9)). By (2.17), we have

$$\mathcal{N}_{t,\eta}(f,\mathcal{P})\approx\left(\sum_{I\in\mathcal{P}}\|t_I\|_p^\tau\right)^{1/\tau} =: \mathcal{N}(f). \tag{A.3}$$

Clearly, the sequence $\{t_I\}_{I\in\mathcal{P}}$ satisfies the conditions of Theorem 2.5 and hence

$$\left\|\sum_{j\in\mathbb{Z}}|t_j(\cdot)|\right\|_p\leqslant c\mathcal{N}(f). \tag{A.4}$$

We define $g(x) := T_0(x)+\sum_{j=1}^\infty t_j(x)$, $x\in\mathbb{R}^d$. By (A.4), $\sum_{j\in\mathbb{Z}}|t_j(x)|<\infty$ for almost all $x\in\mathbb{R}^d$ and hence g is well defined. Clearly, $g := T_m+\sum_{j=m+1}^\infty t_j$ a.e. on $\mathbb{R}^d$, for each $m\in\mathbb{Z}$, with the series converging absolutely a.e. From this and (A.4), we infer $\|g-T_m\|_p\leqslant\|\sum_{j=m+1}^\infty |t_j(\cdot)|\|_p\to 0$ as $m\to\infty$. On the other hand, since $f\in L_\eta$,

$||f - T_m||_{L_\eta(I)} \to 0$ as $m \to \infty$ for each $I \in \mathscr{P}$. Therefore, $f = g$ a.e. and hence

$$f - T_m = \sum_{j=m+1}^{\infty} t_j \quad \text{a.e. on } \mathbb{R}^d, \ \ m \in \mathbb{Z}, \tag{A.5}$$

where the series converges absolutely a.e., and in addition to this $f \in \mathbb{L}_p(\mathscr{P}, k)$.

We shall next show that there exists a polynomial $P \in \Pi_k$ such that

$$T_m - P = \sum_{j=-\infty}^{m} t_j \quad \text{in } L_\infty(\mathbb{R}^d), \ \ m \in \mathbb{Z}. \tag{A.6}$$

Indeed, using Lemma 2.1 and (A.4), we obtain

$$||t_j||_{L_\infty(I)} \leqslant c|I|^{-1/p} ||t_j||_{L_p(I)} \leqslant c2^{j/p} ||t_j||_{L_p(I)} \leqslant c2^{j/p} \mathscr{N}(f), \quad I \in \mathscr{P}_j,$$

and hence $||t_j||_{L_\infty(\mathbb{R}^d)} \leqslant c2^{j/p}\mathscr{N}(f)$. Therefore,

$$\sum_{j=-\infty}^{m} ||t_j||_{L_\infty(\mathbb{R}^d)} < \infty, \quad m \in \mathbb{Z}. \tag{A.7}$$

Fix $I \in \mathscr{P}$. If $-m$ is sufficiently large and $\mu \leqslant -1$, then $T_m - T_{m+\mu}$ is an algebraic polynomial of degree $<k$ on I and

$$||T_m - T_{m+\mu}||_{L_\infty(I)} = \left|\left| \sum_{j=m+\mu+1}^{m} t_j \right|\right|_{L_\infty(I)} \leqslant \sum_{j=m+\mu+1}^{m} ||t_j||_{L_\infty(I)} \to 0 \quad \text{as } m \to -\infty,$$

where we used (A.7). Therefore, there exists $Q_I \in \Pi_k$ such that

$$\lim_{m \to -\infty} ||T_m - Q_I||_{L_\infty(I)} = 0.$$

From this and (A.7), it readily follows that there exists a unique polynomial $P \in \Pi_k$ such that $\lim_{m \to -\infty} ||T_m - P||_{L_\infty(\mathbb{R}^d)} = 0$. This and (A.7) imply (A.6). In going further, (A.4)–(A.6) yield

$$f - P = \sum_{m \in \mathbb{Z}} t_j \quad \text{a.e. on } \mathbb{R}^d \tag{A.8}$$

with the series converging absolutely a.e., and

$$||f - P||_p \leqslant \left|\left| \sum_{j \in \mathbb{Z}} |t_j(\cdot)| \right|\right|_p \leqslant c\mathscr{N}_{t,\eta}(f, \mathscr{P}) < \infty. \tag{A.9}$$

Now, since $f \in L_p(\mathbb{R}^d)$ and $f - P \in L_p(\mathbb{R}^d)$, then $P \equiv 0$, and (A.8) and (A.9) imply Theorem 2.5 in Case I.

Case II: $0 < p < 1$. Since $p < 1$ and $\tau/p < 1$, we immediately obtain

$$\left|\left| \sum_{j \in \mathbb{Z}} |t_j(\cdot)| \right|\right|_p^p = \left|\left| \sum_{I \in \mathscr{P}} |t_I(\cdot)| \right|\right|_p^p \leqslant \sum_{I \in \mathscr{P}} ||t_I||_p^p \leqslant \left(\sum_{I \in \mathscr{P}} ||t_I||_p^\tau \right)^{p/\tau} \leqslant c||f||_{B_\tau^\alpha}^p.$$

This replaces (A.4) and everything else is the same as in Case 1. We shall skip the details. □

Proof of Theorem 2.6. The equivalence of $\mathcal{N}_{\omega,\eta}(\cdot,\mathcal{P})$ and $\mathcal{N}_{t,\eta}(\cdot,\mathcal{P})$ can be proved exactly as Lemma 2.3 was proved and we skip its proof. If $0<\eta\leqslant\tau$, then the equivalence of $\|\cdot\|_{B_\tau^{\alpha k}(\mathcal{P})}$ and $\mathcal{N}_{t,\eta}(\cdot,\mathcal{P})$ follows by (2.14).

The estimate $\|f\|_{B_\tau^{\alpha k}(\mathcal{P})}\leqslant\mathcal{N}_{\omega,\eta}(f,\mathcal{P})$, for $\tau<\eta<p$, is immediate by applying Hölder's inequality. It remains to prove that, for $f\in B_\tau^{\alpha k}(\mathcal{P})$,

$$\mathcal{N}_{\omega,\eta}(f,\mathcal{P})\leqslant c\mathcal{N}_{t,\tau}(f,\mathcal{P})\approx\|f\|_{B_\tau^{\alpha k}(\mathcal{P})},\quad \text{if } \tau<\eta<p. \tag{A.10}$$

Since $f\in B_\tau^{\alpha k}(\mathcal{P})$, by Theorem 2.4 (with $\eta=\tau$), f can be represented in the form

$$f=\sum_{j\in\mathbb{Z}} t_j=:\sum_{j\in\mathbb{Z}}\sum_{I\in\mathcal{P}_j} t_I \quad \text{a.e. on } \mathbb{R}^d \tag{A.11}$$

with the series converging absolutely a.e., where $P\in\Pi_k$, $t_j:=t_{j,\tau}(f)$, and $t_I:=\mathbb{1}_I\cdot t_j$, if $I\in\mathcal{P}_j$, and

$$\mathcal{N}_{t,\tau}(f,\mathcal{P})^\tau=\sum_{I\in\mathcal{P}}|I|^{-\alpha\tau}\|t_I\|_\tau^\tau.$$

Evidently, $\omega_k(t_j,J)_\eta=0$ for $J\in\mathcal{P}_m$ and $j\leqslant m$. We use Lemma 2.1 to obtain, for $J\in\mathcal{P}_m$ and $j>m$,

$$\omega_k(t_j,J)_\eta^\eta\leqslant c\|t_j\|_{L_\eta(J)}^\eta\leqslant c\sum_{I\in\mathcal{P}_j,\ I\subset J}\|t_I\|_\eta^\eta\leqslant c\sum_{I\in\mathcal{P}_j,\ I\subset J}\|t_I\|_\tau^\eta|I|^{\eta(1/\eta-1/\tau)}.$$

Set $\lambda:=\min\{\eta,1\}$. Using (A.11), we have, for $J\in\mathcal{P}_m$,

$$\omega_k(f,J)_\eta\leqslant\left(\sum_{j=m+1}^\infty\omega_k(t_j,J)_\eta^\lambda\right)^{1/\lambda}\leqslant c\left(\sum_{j=m+1}^\infty\left[\sum_{I\in\mathcal{P}_j,\ I\subset J}\|t_I\|_\tau^\eta|I|^{\eta(\frac{1}{\eta}-\frac{1}{\tau})}\right]^{\lambda/\eta}\right)^{1/\lambda}.$$

Therefore,

$$\begin{aligned}
\mathcal{N}_{\omega,\eta}(f,\mathcal{P})^\tau &:=\sum_{J\in\mathcal{P}}\left(|J|^{-\alpha+\frac{1}{\tau}-\frac{1}{\eta}}\omega_k(f,J)_\eta\right)^\tau\\
&\leqslant c\sum_{m\in\mathbb{Z}}\sum_{J\in\mathcal{P}_m}|J|^{(-\alpha+\frac{1}{\tau}-\frac{1}{\eta})\tau}\left[\sum_{j=m+1}^\infty\left(\sum_{I\in\mathcal{P}_j,\ I\subset J}\|t_I\|_\tau^\eta|I|^{\eta(\frac{1}{\eta}-\frac{1}{\tau})}\right)^{\lambda/\eta}\right]^{\tau/\lambda}\\
&=c\sum_{m\in\mathbb{Z}}\sum_{J\in\mathcal{P}_m}\left[\sum_{j=m+1}^\infty\left[\sum_{I\in\mathcal{P}_j,\ I\subset J}(|I|^{-\alpha}\|t_I\|_\tau)^\eta(|I|/|J|)^{(\alpha+\frac{1}{\eta}-\frac{1}{\tau})\eta}\right]^{\lambda/\eta}\right]^{\tau/\lambda}\\
&=c\sum_{m\in\mathbb{Z}}\sum_{J\in\mathcal{P}_m}\left[\sum_{j=m+1}^\infty\left(\sum_{I\in\mathcal{P}_j,\ I\subset J}A_I^\eta 2^{-\gamma(j-m)\eta}\right)^{\lambda/\eta}\right]^{\tau/\lambda}\\
&=:c\sum_{m\in\mathbb{Z}}\sum_{J\in\mathcal{P}_m}[S_{m,J}]^{\tau/\lambda},
\end{aligned}$$

where $A_I := |I|^{-\alpha}||t_I||_\tau$ and $\gamma := \alpha + \frac{1}{\eta} - \frac{1}{\tau} = \frac{1}{\eta} - \frac{1}{p} > 0$. We now want to shift the order of summation. So, this is a Hardy inequality type situation. We first estimate $S_{m,J}$ by using Hölder's inequality. Choose $\gamma_1, \gamma_2 > 0$ such that $\gamma_1 + \gamma_2 = \gamma$ and set $s := \eta/\lambda$, $1/s' := 1 - 1/s$. We obtain

$$\begin{aligned}
S_{m,J} &= \sum_{j=m+1}^{\infty} 2^{-\gamma_1(j-m)\lambda} 2^{-\gamma_2(j-m)\lambda} \left(\sum_{I \in \mathscr{P}_j,\ I \subset J} A_I^\eta \right)^{\lambda/\eta} \\
&\leqslant \left[\sum_{j=m+1}^{\infty} \left(2^{-\gamma_1(j-m)\lambda}\right)^{s'} \right]^{1/s'} \left[\sum_{j=m+1}^{\infty} \left(2^{-\gamma_2(j-m)\lambda} \left(\sum_{I \in \mathscr{P}_j,\ I \subset J} A_I^\eta \right)^{\lambda/\eta} \right)^s \right]^{1/s} \\
&\leqslant c \left(\sum_{j=m+1}^{\infty} 2^{-\gamma_2(j-m)\eta} \sum_{I \in \mathscr{P}_j,\ I \subset J} A_I^\eta \right)^{\lambda/\eta} \\
&\leqslant c \left(\sum_{j=m+1}^{\infty} 2^{-\gamma_2(j-m)\tau} \sum_{I \in \mathscr{P}_j,\ I \subset J} A_I^\tau \right)^{\lambda/\tau},
\end{aligned}$$

where we used that $\tau \leqslant \eta$. Combining this result with the previous estimates, we obtain

$$\begin{aligned}
\mathscr{N}_{\omega,\eta}(f, \mathscr{P})^\tau &\leqslant c \sum_{m \in \mathbb{Z}} \sum_{J \in \mathscr{P}_m} \sum_{j=m+1}^{\infty} 2^{-\gamma_2(j-m)\tau} \sum_{I \in \mathscr{P}_j,\ I \subset J} A_I^\tau \\
&\leqslant c \sum_{j \in \mathbb{Z}} \sum_{I \in \mathscr{P}_j} A_I^\tau \sum_{m=-\infty}^{j-1} 2^{-\gamma_2(j-m)\tau} \leqslant c \sum_{j \in \mathbb{Z}} \sum_{I \in \mathscr{P}_j} A_I^\tau = c\mathscr{N}_{t,\tau}(f, \mathscr{P})^\tau,
\end{aligned}$$

where we switched the order of summation. Thus (A.10) is proved.

The following simple example shows that the equivalence of $||\cdot||_{B_\tau^{\alpha k}(\mathscr{P})}$ and $\mathscr{N}_{\omega,\eta}(\cdot, \mathscr{P})$ is not valid if $\eta \geqslant p$. Let $f := \mathbb{1}_I$ for some $I \in \mathscr{P}$. It is readily seen that $||f||_{B_\tau^{\alpha k}(\mathscr{P})} \approx |I|^{1/p} \approx ||f||_p$ and at the same time $\mathscr{N}_{\omega,\eta}(f, \mathscr{P}) = \infty$ if $\eta \geqslant p$. □

A.2. Proof of Theorem 3.4

We first prove that, for $f \in B_\tau^\alpha$, $B_\tau^\alpha := B_\tau^{\alpha k}(\mathscr{P})$,

$$||f||_{A_\tau^\alpha} \leqslant c||f||_{B_\tau^\alpha}. \tag{A.12}$$

By Theorem 2.6 and (2.17), $||f||_{B_\tau^\alpha}^\tau \approx \sum_{I \in \mathscr{P}} ||t_I||_p^\tau$ with $t_I := t_{I,\eta}(f) := \mathbb{1}_I \cdot t_{m,\eta}(f)$ if $I \in \mathscr{P}_m$ $(0 < \eta < p)$. Therefore, if $||t_{I_1}||_p \geqslant ||t_{I_2}||_p \geqslant \cdots$ is a nonincreasing rearrangement of the sequence $\{\|t_I\|_p\}$, then

$$\|f\|_{B_\tau^\alpha}^\tau \approx \sum_{\nu=0}^{\infty} 2^\nu \|t_{I_{2^\nu}}\|_p^\tau.$$

On the other hand, Theorem 2.4 implies ($\|f\|_p < \infty$)

$$\sigma_m(f,\mathscr{P})_p \leqslant c \left\| \sum_{j=m+1}^{\infty} |t_{I_j}| \right\|_p.$$

Evidently, the sequence $\{t_I\}_{I\in\mathscr{P}}$ satisfies the conditions of Theorem 2.5 and, therefore, we can apply Lemma A.1 to obtain

$$\sigma_{2^\nu}(f,\mathscr{P})_p \leqslant c \sum_{j=\nu}^{\infty} \left\| \sum_{\ell=2^j+1}^{2^{j+1}} |t_{I_\ell}| \right\|_p \leqslant c \sum_{j=\nu}^{\infty} 2^{j/p} \|t_{I_{2^j}}\|_p, \quad \text{if } 1 \leqslant p < \infty. \tag{A.13}$$

Clearly,

$$\sigma_{2^\nu}(f,\mathscr{P})_p^p \leqslant \sum_{\ell=2^\nu+1}^{\infty} \|t_{I_\ell}\|_p^p \leqslant c \sum_{j=\nu}^{\infty} 2^j \|t_{I_{2^j}}\|_p^p, \quad \text{if } 0 \leqslant p < 1. \tag{A.14}$$

We insert (A.13) or (A.14), respectively, in the definition of $\|f\|_{A_\tau^\alpha}$ (see (3.6)) and apply inequality (2.12) to obtain (A.12).

We next prove that if $f \in A_\tau^\alpha$, then $f \in B_\tau^\alpha$ and

$$\|f\|_{B_\tau^\alpha} \leqslant c \|f\|_{A_\tau^\alpha}. \tag{A.15}$$

Case I: $\tau \leqslant 1$. We may assume that $\varphi_m \in \Sigma_m^k(\mathscr{P})$ are such that $\|f - \varphi_m\|_p = \sigma_m(f,\mathscr{P})_p$. Since $f \in A_\tau^\alpha(L_p,\mathscr{P})$, then $\sigma_m(f,\mathscr{P})_p \to 0$ and hence

$$f = \varphi_1 + \sum_{\nu=1}^{\infty} (\varphi_{2^\nu} - \varphi_{2^{\nu-1}}) \quad \text{in } L_p. \tag{A.16}$$

On the other hand, since $\|\varphi_{2^\nu} - \varphi_{2^{\nu-1}}\|_p \leqslant c\sigma_{2^{\nu-1}}(f)_p$,

$$\begin{aligned} \left\| |\varphi_1| + \sum_{\nu=1}^{\infty} |\varphi_{2^\nu} - \varphi_{2^{\nu-1}}| \right\|_p^\mu &\leqslant \|f\|_p^\mu + \|f - \varphi_1\|_p^\mu + \sum_{\nu=1}^{\infty} \|\varphi_{2^\nu} - \varphi_{2^{\nu-1}}\|_p^\mu \\ &\leqslant \|f\|_p^\mu + c \sum_{\nu=0}^{\infty} \sigma_{2^\nu}(f,\mathscr{P})_p^\mu \leqslant \|f\|_p^\tau + c \sum_{\nu=0}^{\infty} \sigma_{2^\nu}(f,\mathscr{P})_p^\tau \leqslant c\|f\|_{A_\tau^\alpha}^\tau < \infty \end{aligned}$$

with $\mu := \min\{p, 1\}$, where we used that $\tau \leqslant \mu$. Therefore, the series in (A.16) converges absolutely a.e. on $\mathbb{R}^d$ as well. From this, we readily obtain ($\tau \leqslant 1$)

$$\|f\|_{B_\tau^\alpha}^\tau \leqslant \|\varphi_1\|_{B_\tau^\alpha}^\tau + \sum_{\nu=1}^{\infty} \|\varphi_{2^\nu} - \varphi_{2^{\nu-1}}\|_{B_\tau^\alpha}^\tau.$$

Applying the Bernstein inequality from Theorem 3.2 to each term above, we get

$$\|f\|_{B_\tau^\alpha}^\tau \leqslant c\|\varphi_1\|_p^\tau + c\sum_{v=1}^{\infty} (2^{v\alpha}\|\varphi_{2^v} - \varphi_{2^{v-1}}\|_p)^\tau$$
$$\leqslant c\|f\|_p^\tau + c\sum_{v=0}^{\infty} (2^{v\alpha}\sigma_{2^v}(f,\mathscr{P})_p)^\tau \leqslant c\|f\|_{A_\tau^\alpha}^\tau.$$

This completes the proof of (A.15) in Case I.

Case II: $\tau > 1$. Then $p > 1$. This case is more complicated and will require more careful analysis. We may assume that $\varphi_m \in \Sigma_m^k(\mathscr{P})$ are such that $\|f - \varphi_m\|_p = \sigma_m(f,\mathscr{P})_p$. Let

$$\varphi_m =: \sum_{I\in\Lambda_m} \mathbb{1}_I \cdot P_{m,I}, \quad \text{where } \Lambda_m \subset \mathscr{P},\ \#\Lambda_m \leqslant m, \text{ and } P_{m,I}\in\Pi_k.$$

Set $\Lambda_{2^v}^* := \bigcup_{j=0}^{v} \Lambda_{2^j}$. We have

$$\Lambda_{2^v}^* \subset \Lambda_{2^{v+1}}^* \quad \text{and} \quad \#\Lambda_{2^v}^* \leqslant \sum_{j=0}^{v} 2^j = 2^{v+1} - 1 \quad \text{for } v = 1, 2, \ldots .$$

In this part, our construction is quite similar to the one from the proof of Theorem 3.2. Let $I_{v,0}\in\mathscr{P}$ be the smallest box containing all boxes from $\Lambda_{2^v}^*$ and let $\mathscr{T}_v^*$ be the minimal binary subtree of $\mathscr{P}$ containing $\Lambda_{2^v}^* \cup \{I_{v,0}\}$. The set $\Lambda_{2^v}^*$ induces a natural subdivision of $\mathbb{R}^d$ into a union of disjoint maximal rings. By definition, R is a *ring* if $R = I\backslash J$, where $I\in\mathscr{P}$ or $I = \mathbb{R}^d$ and $J\in\mathscr{P}$ or $J = \emptyset$. We say that $R = I\backslash J$ is a *maximal ring* generated by $\Lambda_{2^v}^*$ if (a) $I\in\mathscr{T}_v^*$ or $I = \mathbb{R}^d$ and $J\in\mathscr{T}_v^*$ or $J=\emptyset$, (b) R does not contain a box smaller than I from $\Lambda_{2^v}^*$, and (c) R is maximal with these two properties. We let ρ_v^* denote the set of all maximal rings generated by $\Lambda_{2^v}^*$. We have the following possibilities for a ring $R\in\rho_v^*$ with $R =: I\backslash J$: (i) I is a final box in $\mathscr{T}_v^*$ and $J = \emptyset$; (ii) $J\in\Lambda_{2^v}^*$ or J is a branching box in $\mathscr{T}_v^*$; (iii) $I = \mathbb{R}^d$ and $J = I_{v,0}$. Therefore, $\#\rho_v^* \leqslant 3\#\Lambda_{2^v}^* + 1 < 6\cdot 2^v$. Note that ρ_v^* is a collection of disjoint rings such that

$$\mathbb{R}^d = \bigcup_{R\in\rho_v^*} R.$$

Also, since $\Lambda_{2^v}^* \subset \Lambda_{2^{v+1}}^*$, for each $R\in\rho_{v+1}^*$, we have either $R\in\rho_v^*$ or $R\subset K$ for some $K\in\rho_v^*$. Thus $\{\rho_v^*\}$ is a sequence of nested rings.

For each ring $R\in\rho_v^*$, we denote by I_R (the mother box of R) the smallest box from $\mathscr{P}$ containing R and by I_R' the largest box from $\mathscr{P}$ contained in R. Note that I_R' is uniquely determined and is one of the two children of I_R in $\mathscr{P}$. Also, we define $P_R\in\Pi_k$ by the identity

$$\|f - P_R\|_{L_p(I_R')} = \inf_{P\in\Pi_k} \|f - P\|_{L_p(I_R')} =: E_k(f, I_R')_p.$$

It is easily seen (using Lemma 2.1) that

$$\|f - P_R\|_{L_p(R)} \leqslant cE_k(f,R)_p. \tag{A.17}$$

Now, we set $\varphi^*_{2^\nu} := \sum_{R\in\rho^*_\nu} \mathbb{1}_R \cdot P_R$. It follows, from $\Lambda_{2^\nu} \subset \Lambda^*_{2^\nu}$ and (A.17),

$$\|f - \varphi^*_{2^\nu}\|_p \leqslant c\|f - \varphi_{2^\nu}\|_p = c\sigma_{2^\nu}(f, \mathscr{P})_p. \tag{A.18}$$

By the definition of $\varphi^*_{2^\nu}$, if $R \in \rho^*_\nu$ and $K \in \rho^*_{\nu-1}$ with $I_R = I_K$, then $R \subset K$ and $\varphi^*_{2^\nu} \equiv \varphi^*_{2^{\nu-1}}$ on R. We let $\rho^\diamondsuit_\nu$ $(\nu \geqslant 1)$ denote the set of all rings from $\rho^*_\nu \backslash \rho^*_{\nu-1}$ which do not share mother boxes with rings from $\rho^*_{\nu-1}$ and set $\rho^\diamondsuit_0 := \rho^*_0$. Note that $\rho^\diamondsuit_\nu$ is a collection of disjoint rings. From the above arguments, every two sets from the sequence $\{\rho^\diamondsuit_\nu\}_{\nu=0}^\infty$ are disjoint and

$$\varphi^*_{2^\nu} - \varphi^*_{2^{\nu-1}} = \sum_{R\in\rho^\diamondsuit_\nu} \mathbb{1}_R \cdot P_R =: \sum_{R\in\rho^\diamondsuit_\nu} \Phi_R, \quad \nu \geqslant 1. \tag{A.19}$$

Note that, using (A.18),

$$\left\| \sum_{R\in\rho^\diamondsuit_\nu} \Phi_R \right\|_p = \|\varphi^*_{2^\nu} - \varphi^*_{2^{\nu-1}}\|_p \leqslant c\sigma_{2^{\nu-1}}(f, \mathscr{P})_p, \quad \nu \geqslant 1. \tag{A.20}$$

Let $R \in \bigcup_{\nu=0}^\infty \rho^\diamondsuit_\nu$ and $R =: I \backslash J$ with $I \in \mathscr{P}_\ell$ and $J \in \mathscr{P}_{\ell+\mu}$ for some $\ell \in \mathbb{Z}$ and $\mu \geqslant 1$. For $\ell \leqslant m < \ell + \mu$, there is a unique $I^\natural \in \mathscr{P}_m$ such that $J \subset I^\natural \subset I$. We define $\Phi_{R,m} := \mathbb{1}_{I^\natural} \cdot \Phi_R$ and set $\Phi_{R,m} := 0$ if $m < \ell$ or $m \geqslant \ell + \mu$.

Since $\|f - \varphi^*_{2^\nu}\|_p \leqslant c\|f - \varphi_{2^\nu}\|_p \to 0$ and $f \in A^\gamma_q(L_p, \mathscr{P})$, similarly as in Case I (see (A.16)) we have

$$f = \varphi^*_1 + \sum_{\nu=1}^\infty (\varphi^*_{2^\nu} - \varphi^*_{2^{\nu-1}}) \quad \text{in } L_p \tag{A.21}$$

with the series converging absolutely almost everywhere as well.

We denote by $\mathscr{R}_m$ the set of all rings $R \in \rho^\diamondsuit := \bigcup_{\nu=0}^\infty \rho^\diamondsuit_\nu$ such that $I_R \in \mathscr{P}_m$ and let $\mathscr{K}_m$ be the set of all rings $R \in \rho^\diamondsuit$ with $R =: I \backslash J$ such that $|J| \leqslant 2^{-m} < |I|$. Clearly, $\mathscr{R}_m \cup \mathscr{K}_m$ is a set of disjoint rings. From this, (A.19), and (A.21), it readily follows that $(\tau \geqslant 1)$

$$\begin{aligned}\sum_{I\in\mathscr{P}_m} \omega_k(f, I)^\tau_\tau &\leqslant c\left[\sum_{\mu=m+1}^\infty \left\| \sum_{R\in\mathscr{R}_\mu} \Phi_R \right\|_\tau\right]^\tau + c\left\| \sum_{R\in\mathscr{K}_m} \Phi_{R,m} \right\|_\tau^\tau \\ &= c\left[\sum_{\mu=m+1}^\infty \left(\sum_{R\in\mathscr{R}_\mu} \|\Phi_R\|_\tau^\tau \right)^{\frac{1}{\tau}}\right]^\tau + c\sum_{R\in\mathscr{K}_m} \|\Phi_{R,m}\|_\tau^\tau.\end{aligned}$$

Therefore,

$$\|f\|_{B_\tau^\alpha}^\tau := \sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{I\in\mathscr{P}_m} \omega_k(f,I)_\tau^\tau$$
$$\leqslant c \sum_{m\in\mathbb{Z}} \left[2^{\alpha m} \sum_{\mu=m+1}^{\infty} \left(\sum_{R\in\mathscr{R}_\mu} \|\Phi_R\|_\tau^\tau \right)^{\frac{1}{\tau}} \right]^\tau + c \sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{R\in\mathscr{K}_m} \|\Phi_{R,m}\|_\tau^\tau$$
$$= : \Sigma_1 + \Sigma_2.$$

We apply inequality (2.12) to the first sum above to obtain

$$\Sigma_1 \leqslant c \sum_{m\in\mathbb{Z}} 2^{\alpha m\tau} \sum_{R\in\mathscr{R}_m} \|\Phi_R\|_\tau^\tau \leqslant c \sum_{R\in\rho^\diamond} \|\Phi_R\|_p^\tau,$$

where we used that $\|\Phi_R\|_\tau \leqslant |I_R|^{1/\tau-1/p}\|\Phi_R\|_p = 2^{-\alpha m}\|\Phi_R\|_p$, $R\in\mathscr{R}_m$, by Hölder's inequality.

We shall estimate Σ_2 by using the inequalities: (a) $\|\Phi_{R,m}\|_\tau \leqslant 2^{-\alpha m}\|\Phi_{R,m}\|_p$ which follows by Hölder's inequality as above, and (b) $\sum_{m\in\mathbb{Z}} \|\Phi_{R,m}\|_p^\tau \leqslant c\|\Phi_R\|_p^\tau$, $\quad R\in\rho^\diamond$, which can be proved exactly as (3.5) was proved. Applying these inequalities, we find

$$\Sigma_2 \leqslant c \sum_{m\in\mathbb{Z}} \sum_{R\in\mathscr{K}_m} \|\Phi_{R,m}\|_p^\tau \leqslant c \sum_{R\in\rho^\diamond} \sum_{m\in\mathbb{Z}} \|\Phi_{R,m}\|_p^\tau \leqslant c \sum_{R\in\rho^\diamond} \|\Phi_R\|_p^\tau,$$

where we switched the order of summation.

Combining the above estimates for Σ_1 and Σ_2, we obtain

$$\|f\|_{B_\tau^\alpha}^\tau \leqslant c \sum_{R\in\rho^\diamond} \|\Phi_R\|_p^\tau \leqslant c \sum_{v=0}^{\infty} \sum_{R\in\rho_v^\diamond} \|\Phi_R\|_p^\tau$$
$$\leqslant c \sum_{v=0}^{\infty} \left(\sum_{R\in\rho_v^\diamond} \|\Phi_R\|_p^p \right)^{\tau/p} (\#\rho_{2^v}^\diamond)^{1-\tau/p}$$
$$\leqslant c\|\varphi_1^*\|_p^\tau + c \sum_{v=1}^{\infty} 2^{v\alpha\tau} \|\varphi_{2^v}^* - \varphi_{2^{v-1}}^*\|_p^\tau$$
$$\leqslant c\|f\|_p^\tau + c \sum_{v=0}^{\infty} 2^{v\alpha\tau} \sigma_{2^v}(f,\mathscr{P})_p^\tau \leqslant c\|f\|_{A_\tau^\alpha}^\tau,$$

where we used (A.20) and Hölder's inequality. This completes the proof of (A.15) in Case II. □

References

[1] J. Bergh, J. Löfström, Interpolation spaces: An introduction, in: Grundlehren der Mathematischen Wissenschaften, Vol. 223, Springer, Berlin, New York, 1976.

[2] Yu. Brudnyi, Approximation of functions of *n*-variables by quasi-polynomials, Math. USSR Izv. 4 (1970) 568–586.

[3] R. Coifman, M. Wickerhauser, Entropy based algorithms for best basis selection, IEEE Trans. Inform. Theory 32 (1992) 712–718.

[4] A. Cohen, R. DeVore, P. Petrushev, H. Xu, Nonlinear, Approximation and the space $BV(\mathbb{R}^2)$, Amer. J. Math. 121 (1999) 587–628.

[5] R. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions, Amer. J. Math. 114 (1992) 737–785.

[6] R. DeVore, G. Lorentz, Constructive Approximation, Vol. 303, Springer Grundlehren, Heidelberg, 1993.

[7] R. DeVore, P. Petrushev, X. Yu, Nonlinear wavelet approximation in the space $C(R^d)$, in: A.A. Gonchar, E.B. Saff (Eds.), Progress in Approximation Theory, Springer, New York, 1992, pp. 261–283.

[8] R. DeVore, V.A. Popov, Interpolation of Besov spaces, Trans. Amer. Math. Soc. 305 (1998) 297–314.

[9] D. Donoho, CART and best-ortho-basis: a connection, Ann. Statist. 25 (1997) 1870–1911.

[10] C. Fefferman, E. Stein, Some maximal inequalities, Amer. J. Math. 93 (1971) 107–115.

[11] Y. Hu, K. Kopotun, X. Yu, On multivariate adaptive approximation, Constr. Approx. 16 (2000) 449–474.

[12] G. Kyriazis, P. Petrushev, New bases for Triebel–Lizorkin and Besov spaces, Trans. Amer. Math. Soc. 354 (2002) 749–776.

[13] D. Newman, Rational approximation to $|x|$, Michigan Math. J. 11 (1964) 11–14.

[14] P. Oswald, Multilevel Finite Element Approximation: Theory and Applications, Teubner Skripten zur Numerik, Teubner, Stuttgart, 1994.

[15] A. Pekarskii, Relations between best rational and piecewise polynomial approximations, Vestsi Acad. Navuk. BSSR Ser. Fiz.-Mat. Navuk (1986) No. 5, 36–39 (in Russian).

[16] A. Pekarskii, Estimates for the derivatives of rational functions in $L_p[-1,1]$, Mat. Zametki 39 (1986) 388–394 (English translation in Math. Notes, 39 (1986), 212–216).

[17] P. Petrushev, Relations between rational and spline approximations in L_p metric, J. Approx. Theory 50 (1987) 141–159.

[18] P. Petrushev, Direct and converse theorems for spline and rational approximation and Besov spaces, in: M. Cwikel, et al., (Eds.), Function Spaces and Applications, Lecture Notes in Mathematics, Vol. 1302, Springer, Berlin, 1988, pp. 363–377.

[19] P. Petrushev, Bases consisting of rational functions of uniformly bounded degrees or more general functions, J. Funct. Anal. 174 (2000) 18–75.

[20] P. Petrushev, V. Popov, Rational Approximation of Real Functions, Cambridge University Press, Cambridge, 1987.

[21] E. Stein, Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, NJ, 1993.

[22] E. Storozhenko, P. Oswald, Jackson's theorem in the space $L_p(\mathbb{R}^k)$, $0<p<1$, Siberian Math. J. 19 (1978) 630–639.

[23] V. Temlyakov, The best m-term approximation and greedy algorithms, Adv. Comput. Math. 8 (1998) 249–265.

[24] P. Wojtaszczyk, Banach Spaces for Analysts, in: Cambridge Studies in Advanced Mathematics, Vol. 25, Cambridge University Press, Cambridge, 1991.