Constr. Approx. (2000) 16: 449–474
DOI: 10.1007/s003659910019

CONSTRUCTIVE APPROXIMATION

On Multivariate Adaptive Approximation

Y. K. Hu, K. A. Kopotun, and X. M. Yu

Abstract. Recently, A. Cohen, R. A. DeVore, P. Petrushev, and H. Xu investigated nonlinear approximation in the space $BV(\mathbf{R}^2)$. They modified the classical adaptive algorithm to solve related extremal problems. In this paper, we further study the modified adaptive approximation and obtain results on some extremal problems related to the spaces $V^r_{\sigma,p}(\mathbf{R}^d)$ of functions of "Bounded Variation" and Besov spaces $B^\alpha(\mathbf{R}^d)$.

1. Introduction

Nonlinear approximation has been investigated extensively in recent years. In the univariate case, because of the simplicity of the real line topology, free knot spline approximation is widely used in numerical computations. But in the multivariate case, generating good free spline approximants is a more complicated and difficult task and is still under research. However, there is the so-called Adaptive Approximation that works well in multidimensional spaces and is practically easy to implement. Its main disadvantage is that it gives a slightly lower than the best approximation order. Recently, A. Cohen, R. DeVore, P. Petrushev, and H. Xu [7] successfully introduced a splitting and merging method to modify adaptive approximation, and showed that their method produces near-minimizers to the extremal problems related to the space $BV(\mathbf{R}^2)$. In this paper, we shall explore their method to show that this new modified adaptive approximation generates near-minimizers to some extremal problems in the spaces $V^r_{\sigma,p}(\mathbf{R}^d)$ of functions of "bounded variation" and Besov spaces $B^\alpha(\mathbf{R}^d)$.

If X_0 and X_1 are quasi-normed spaces continuously embedded in a Hausdorff space X, then the K-functional for all $f \in X_0 + X_1$ is defined as

$$K(f, t, X_0, X_1) := \inf_{f=f_0+f_1} \left\{ \| f_0\|_{X_0} + t | f_1|_{X_1} \right\},$$

where $\|\cdot\|_{X_0}$ is a (quasi)norm in X_0, and $|\cdot|_{X_1}$ is a (semi)norm or (semiquasi)norm in X_1. The extremal problem we are interested in is as follows. For a given $f \in X_0 + X_1$, and a parameter $t > 0$, find a function $f_1 \in X_1$ with $f_0 := f - f_1 \in X_0$ which attains the

Date received: November 23, 1998. Date revised: June 25, 1999. Date accepted: September 13, 1999. Communicated by Ronald A. DeVore.
AMS classification: 41A15, 41A17, 41A25, 41A27, 41A46, 41A63.
Key words and phrases: Nonlinear approximation, Approximation spaces, Adaptive algorithms, Piecewise polynomials, Splines, Besov spaces, Degree of approximation, Modulus of smoothness, K-Functionals.

infimum in the definition of $K(f, t, X_0, X_1)$. Such a function f_1 is called a minimizer. A function $g \in X_1$ with $f - g \in X_0$ is called a near minimizer if

$$\|f - g\|_{X_0} + t|g|_{X_1} \le C \inf_{f=f_0+f_1} \left\{\|f_0\|_{X_0} + t|f_1|_{X_1}\right\}.$$

The problem of finding a minimizer or a near minimizer is closely related to the characterization of K-functionals, interpolation spaces, and approximation spaces.

The interpolation space $(X_0, X_1)_{\theta,q}$, $0 < \theta < 1$, $0 < q \le \infty$, consists of all functions $f \in X_0 + X_1$ such that $|f|_{(X_0,X_1)_{\theta,q}} < \infty$, where

$$|f|_{(X_0,X_1)_{\theta,q}} := \begin{cases} \left(\displaystyle\int_0^\infty \left[t^{-\theta} K(f, t, X_0, X_1)\right]^q \frac{dt}{t}\right)^{1/q}, & 0 < q < \infty, \\ \displaystyle\sup_{t>0} t^{-\theta} K(f, t, X_0, X_1), & q = \infty. \end{cases}$$

The approximation space $A_q^\alpha(X, \{M_n\}_{n\in\mathbf{N}})$, $\alpha > 0$, $0 < q \le \infty$, consists of all $f \in X$ such that

$$|f|_{A_q^\alpha(X,\{M_n\}_{n\in\mathbf{N}})} := \begin{cases} \left(\displaystyle\sum_{n=1}^\infty \left[n^\alpha E(f, M_n)_X\right]^q \frac{1}{n}\right)^{1/q}, & 0 < q < \infty, \\ \displaystyle\sup_{n\ge1} n^\alpha E(f, M_n)_X, & q = \infty, \end{cases}$$

is finite. Here $E(f, M_n)_X = \inf_{g\in M_n} \|f - g\|_X$ is the error for approximation from the manifold $M_n \subset X$, $n \in \mathbf{N}$. These M_n are usually required to satisfy the assumptions:

(i) $M_0 = \{0\}$;
(ii) $M_n \subset M_{n+1}$;
(iii) $aM_n = M_n$ for any $a \ne 0$;
(iv) $M_n + M_n \subset M_{cn}$ with $c := c(\{M_n\})$;
(v) $\bigcup_{n=0}^\infty M_n$ is dense in X; and
(vi) any $f \in X$ has a best approximation from each M_n.

The following result is due to DeVore and Popov [17], and shows that if the Jackson and Bernstein inequalities are satisfied for the spaces X and Y ($Y \subseteq X$), then the approximation spaces $A_q^\alpha(X, \{M_n\}_{n\in\mathbf{N}})$ can be characterized as interpolation spaces between X and Y.

Theorem A. *Suppose that for a pair of spaces X, Y we have*

$$E(f, M_n)_X := \inf_{g\in M_n} \|f - g\|_X \le Cn^{-\lambda}|f|_Y, \qquad f \in Y \quad (\textit{Jackson inequality})$$

and

$$|g|_Y \le Cn^\lambda \|g\|_X, \qquad g \in M_n, \quad n \in \mathbf{N} \quad (\textit{Bernstein inequality}).$$

Then for $0 < \alpha < \lambda$ and $0 < q \le \infty$:

$$A_q^\alpha(X, \{M_n\}_{n\in\mathbf{N}}) = (X, Y)_{\alpha/\lambda,q}.$$

In this paper, we deal with $X = L_p$ and $Y = V^r_{\sigma,p}$ or B^α. After reviewing the space B^α, we shall introduce in this section some notation for piecewise polynomial functions on dyadic rings. These piecewise polynomial functions are used for M_n. In the next section, we define the spaces $V^r_{\sigma,p}$ and the moduli of smoothness $\mathcal{W}_r$, and discuss some of their basic properties. In Section 3 we study the approximation on a ring, which is the key step to obtaining a Jackson inequality. After presenting briefly the description of the modified adaptive algorithm in Section 4, and discussing the properties of multivariate free splines on ring partitions in Section 5, we establish the Jackson inequality and the Bernstein inequality for $V^r_{\sigma,p}$ (in Section 6) and for B^α (in Section 7). In the last section, we provide a brief discussion of wavelet decompositions and related approximation spaces.

If $0 < \alpha < r$ and $0 < p, q \le \infty$, then the Besov space $B^\alpha_q(L_p, \Omega)$, $\Omega \subseteq \mathbf{R}^d$, is the set of all functions $f \in L_p(\Omega)$ such that

$$|f|_{B^\alpha_q(L_p,\Omega)} := \begin{cases} \left(\displaystyle\int_0^\infty \left[t^{-\alpha}\omega_r(f,t,\Omega)_p\right]^q \frac{dt}{t}\right)^{1/q}, & 0 < q < \infty, \\ \sup\limits_{t>0} t^{-\alpha}\omega_r(f,t,\Omega)_p, & q = \infty, \end{cases}$$

is finite. (Here, ω_r is the usual rth modulus of smoothness.) The quantity $|\cdot|_{B^\alpha_q(L_p,\Omega)}$ is a semi-(quasi)norm for $B^\alpha_q(L_p,\Omega)$, and the (quasi)norm for $B^\alpha_q(L_p,\Omega)$ is defined by

$$\|\cdot\|_{B^\alpha_q(L_p,\Omega)} := \|\cdot\|_{L_p(\Omega)} + |\cdot|_{B^\alpha_q(L_p,\Omega)}.$$

Also, for $\alpha > 0$ and $0 < p \le \infty$, we define

$$B^\alpha(\Omega) := B^{\alpha,p}(\Omega) := B^\alpha_\sigma(L_\sigma,\Omega) \qquad \text{with} \quad 1/\sigma = \alpha/d + 1/p,$$

and $B^\alpha := B^\alpha([0,\ 1)^d)$. DeVore and Popov [16] proved

Theorem B. *If* $0 < \alpha < \beta$, $0 < p < \infty$, *and* $\sigma = (\alpha/d + 1/p)^{-1}$, *then* $(L_p, B^\beta)_{\alpha/\beta,\sigma} = B^\alpha$.

Besov spaces have the following properties:

$$B^\alpha_q(L_p) \subseteq B^{\alpha'}_{q'}(L_{p'}) \qquad \text{if} \begin{cases} \alpha > \alpha', & p = p', \\ \alpha = \alpha', & q < q', \quad p = p', \\ \alpha = \alpha', & q = q', \quad p > p'. \end{cases}$$

It also follows from Theorem B that $B^\alpha \subseteq B^{\alpha'}$ if $\alpha > \alpha'$.

Let $\mathbf{D}_k(\mathbf{R}^d)$, $k \in \mathbf{Z}$, denote the collection of all dyadic cubes in $\mathbf{R}^d$ of side length 2^{-k}, i.e.,

$$\mathbf{D}_k(\mathbf{R}^d) := \left\{[l_1 2^{-k}, (l_1+1)2^{-k}) \times \cdots \times [l_d 2^{-k}, (l_d+1)2^{-k}) \mid l_i \in \mathbf{Z},\ 1 \le i \le d\right\},$$

and let $\mathbf{D}(\mathbf{R}^d)$ be the collection of *all* dyadic cubes in $\mathbf{R}^d$,

$$\mathbf{D}(\mathbf{R}^d) := \bigcup_{k\in\mathbf{Z}} \mathbf{D}_k(\mathbf{R}^d).$$

It will also be convenient to denote

$$\mathbf{D}_k(\Omega) := \{I \in \mathbf{D}_k(\mathbf{R}^d) \mid I \subseteq \Omega\}$$

and

$$\mathbf{D}(\Omega) := \bigcup_{k \in \mathbf{Z}} \mathbf{D}_k(\Omega) = \left\{I \in \mathbf{D}(\mathbf{R}^d) \mid I \subseteq \Omega\right\}.$$

We will say that $R = I \backslash J$ is a *dyadic ring* if $I, J \in \mathbf{D}(\mathbf{R}^d)$ (with J possibly empty) and $J \subsetneq I$. The set of all dyadic rings will be denoted by

$$\overline{\mathbf{D}}(\mathbf{R}^d) := \left\{R = I \backslash J \mid I, J \in \mathbf{D}(\mathbf{R}^d), \text{ and } J \subsetneq I\right\},$$

and

$$\overline{\mathbf{D}}(\Omega) := \left\{R \in \overline{\mathbf{D}}(\mathbf{R}^d) \mid R \subseteq \Omega\right\}.$$

Given a dyadic ring $R = I \backslash J$ we denote $R_{\boxplus} := I$ and $R_{\square} := J$. This notation turns out to be rather convenient later in the paper. Also, we note that a dyadic ring R is a cube if and only if $R = R_{\boxplus}$, and will say that dyadic cubes are *degenerate* dyadic rings.

It will also be convenient to denote the class of all (finite) partitions of $\Omega \in \mathbf{D}(\mathbf{R}^d)$ into dyadic rings by $\pi_r(\Omega)$, i.e.,

$$\pi_r(\Omega) := \left\{ \{R_i\}_{i \in \Lambda} \mid |\Lambda| < \infty, R_i \in \overline{\mathbf{D}}(\Omega), R_i \cap R_j = \emptyset \text{ if } i \neq j, \bigcup_{i \in \Lambda} R_i = \Omega \right\}.$$

Also, let Π_{r-1} denote the set of all algebraic polynomials of total degree $< r$, and let $\Sigma_{n,r}(\Omega)$ be the set of all piecewise polynomial functions S of order $\leq r$ on partitions in $\pi_r(\Omega)$ consisting of not more than n dyadic rings, i.e.,

$$\Sigma_{n,r}(\Omega) := \left\{ S \,\middle|\, S(x) = \sum_{i=1}^{\nu} p_i(x) \chi_{R_i}(x), \ \{R_i\}_{i=1}^{\nu} \in \pi_r(\Omega), \ \nu \leq n, \ p_i \in \Pi_{r-1} \right\}. \tag{1.1}$$

In all of the above, if Ω is equal to $[0, 1)^d$ then it will be omitted. For example, $\mathbf{D} := \mathbf{D}([0, 1)^d)$, $\pi_r := \pi_r([0, 1)^d)$, $\Sigma_{n,r} := \Sigma_{n,r}([0, 1)^d)$, etc.

2. Spaces of Functions of "Bounded Variation"

Given a cube I, let $\mathrm{len}(I)$ denote its sidelength. For example, if $I \in \mathbf{D}_k$, then $\mathrm{len}(I) = 2^{-k}$. Then, for any cube I, $\mathrm{vol}(I) = \mathrm{len}(I)^d$, where, as usual, $\mathrm{vol}(\Omega)$ denotes the measure of Ω. If R is a dyadic ring ($R \in \overline{D}$), then we define $\mathrm{len}(R) := \mathrm{len}(R_{\boxplus})$, and hence $\mathrm{vol}(R) \sim \mathrm{vol}(R_{\boxplus}) = \mathrm{len}(R_{\boxplus})^d$. In fact, $\mathrm{vol}(R_{\boxplus}) \geq \mathrm{vol}(R) \geq (1 - 2^{-d}) \mathrm{vol}(R_{\boxplus})$.

For $f \in L_p([0, 1)^d)$, $0 < \sigma < p$, and $r \in \mathbf{N}$, let $V^r_{\sigma,p}$ denote the set of functions $f \in L_p([0, 1)^d)$, for which the "variation over rings"

$$|f|_{V^r_{\sigma,p}} := \sup_{\{R_i\}_{i \in \Lambda} \in \pi_r} \left(\sum_{i \in \Lambda} \omega_r(f, \mathrm{len}(R_i), R_i)_p^{\sigma} \right)^{1/\sigma}$$

is finite. We note that, if $\sigma \geq p$, then $V^r_{\sigma,p} = L_p$. This is the reason for the restriction $\sigma < p$.

Also, for $f \in L_p([0,\ 1)^d)$, $0 < \sigma < p$, $\beta := 1/\sigma - 1/p$, and $r \in \mathbf{N}$, we define a modulus

$$\mathcal{W}_r(f,t)_{\sigma,p} := \sup_{0<h\leq t} \sup_{\{R_i\}_{i=1}^n \in \pi_r^h} h^\beta \left(\sum_{i=1}^n \omega_r(f, \operatorname{len}(R_i), R_i)_p^\sigma \right)^{1/\sigma},$$

where the second sup is taken over all partitions of $[0,\ 1)^d$ consisting "of about $1/h$ rings":

$$\pi_r^h := \left\{ \{R_i\}_{i=1}^n \in \pi_r \mid n \leq [1/h] + 1 \right\}.$$

This modulus is a generalization of the univariate modulus $\Omega(f,t)_{\sigma,p}$ of R. A. DeVore and X. M. Yu [19]. It is also a modification and generalization of the univariate characteristic $\kappa_{s,p}(n,\ f)$ of Pekarskii [23].

Let $E_f(R)_p$ denote the rate of L_p-approximation of a function f defined in a region R (which can be a cube, a ring, a rectangular solid in $\mathbf{R}^d$, etc.) by polynomials of total degree $< r$, i.e., $E_f(R)_p := E(f, \Pi_{r-1})_{L_p(R)} = \inf_{P\in\Pi_{r-1}} \| f - P \|_{L_p(R)}$.

Remark. Because of the equivalence (see Lemma 1)

$$E_f(R)_p \sim \omega_r(f, \operatorname{len}(R), R)_p$$

for any dyadic ring R, the moduli ω_r in the above definitions can be replaced by $E_f(R)_p$. For example,

$$\mathcal{W}_r(f,t)_{\sigma,p} \sim \sup_{0<h\leq t} \sup_{\{R_i\}_{i=1}^n \in \pi_r^h} h^\beta \left(\sum_{i=1}^n E_f(R_i)_p^\sigma \right)^{1/\sigma}.$$

We now mention some of the important properties of the modulus $\mathcal{W}_r(f,t)_{\sigma,p}$ which are used later in this paper. First of all, it immediately follows from the definition that $\mathcal{W}_r(f,t)_{\sigma,p}$ is a nondecreasing function of t such that

$$\mathcal{W}_r(f+g,t)_{\sigma,p} \leq C \left(\mathcal{W}_r(f,t)_{\sigma,p} + \mathcal{W}_r(g,t)_{\sigma,p} \right)$$

and

$$\mathcal{W}_r(f, Mt)_{\sigma,p} \leq M^{1/\sigma - 1/p} \mathcal{W}_r(f,t)_{\sigma,p}, \qquad M \geq 1.$$

Also,

$$\mathcal{W}_r(f,t)_{\sigma,p} \leq C \| f \|_{L_p} \qquad \text{for all} \quad f \in L_p([0,\ 1)^d). \tag{2.1}$$

Indeed, using Hölder's inequality, we have

$$\begin{aligned} \mathcal{W}_r(f,t)_{\sigma,p} &\leq C \sup_{0<h\leq t} \sup_{\{R_i\}_{i=1}^n \in \pi_r^h} h^\beta \left(\sum_{i=1}^n \| f \|_{L_p(R_i)}^\sigma \right)^{1/\sigma} \\ &\leq C \sup_{0<h\leq t} \sup_{\{R_i\}_{i=1}^n \in \pi_r^h} h^\beta n^{1/\sigma - 1/p} \left(\sum_{i=1}^n \| f \|_{L_p(R_i)}^p \right)^{1/p} \\ &\leq C \| f \|_{L_p}. \end{aligned}$$

Also, for any $g \in V^r_{\sigma,p}$:

$$\mathcal{W}_r(g,t)_{\sigma,p} \le t^\beta |g|_{V^r_{\sigma,p}}, \tag{2.2}$$

since

$$\begin{aligned}
\mathcal{W}_r(g,t)_{\sigma,p} &= \sup_{0<h\le t} \sup_{\{R_i\}_{i=1}^n \in \pi_r^h} h^\beta \left(\sum_{i=1}^n \omega_r(g, \operatorname{len}(R_i), R_i)_p^\sigma \right)^{1/\sigma} \\
&\le t^\beta \sup_{\{R_i\}_{i=1}^n \in \pi_r} \left(\sum_{i=1}^n \omega_r(g, \operatorname{len}(R_i), R_i)_p^\sigma \right)^{1/\sigma} \\
&= t^\beta |g|_{V^r_{\sigma,p}}.
\end{aligned}$$

3. Approximation on a Ring

We say that a rectangular solid is regular if the ratio of lengths of any two of its sides is between $\frac{1}{4}$ and 4. Let R be a dyadic ring. We shall show that R can be represented as a union of regular dyadic rectangular solids, $R = \bigcup_{i=1}^{\mu} r_i$, $\mu = \mu(R)$, such that:

1. $1 \le \mu \le 2d$.
2. One of the children of $R_\boxplus$ (see Section 4 for the definition of a child) is contained in all those r_i which satisfy $\operatorname{vol}(r_i) \ge \frac{1}{2}\operatorname{vol}(R_\boxplus)$. We denote this child by r_0 and call it "the root of the union." All those r_i containing it are called branches.
3. Any r_i ($i = 1, \dots, \mu$) in the union is either a branch or is intersecting a branch r_j with $\operatorname{vol}(r_i \cap r_j) \ge \frac{1}{2}\operatorname{vol}(r_i)$.

Lemma 1. *Let R be a dyadic ring in $\mathbf{R}^d$. Then there exists a representation of R as a union of regular dyadic rectangular solids, $R = \bigcup_{i=1}^{\mu} r_i$, such that the above three conditions are satisfied. Moreover, for all $0 < p \le \infty$:*

$$E_f(R)_p \le C \sum_{i=1}^{\mu} \omega_r\left(f, \sqrt[d]{\operatorname{vol}(r_i)}, r_i\right)_p \le C\omega_r\left(f, \sqrt[d]{\operatorname{vol}(R)}, R\right)_p,$$

where C is a constant which depends only on r, d, and p and is independent of R.

Proof. Without loss of generality, we may assume that $R_\boxplus = [0,1)^d$ and that $R_\square = [a_1, a_1 + 2^{-k}) \times [a_2, a_2 + 2^{-k}) \times \cdots \times [a_d, a_d + 2^{-k})$ with $a_i \ge 0$ and $a_i + 2^{-k} \le \frac{1}{2}$, $i = 1, 2, \dots, d$. By rotating the coordinate axes if necessary, we may also assume that $0 \le a_1 \le a_2 \le \cdots \le a_d$. Notice that for each $i = 1, 2, \dots, d$, we have either $a_i = 0$ or $a_i \ge 2^{-k}$.

Define $r_i := [0,1) \times \cdots \times [0,1) \times [a_i + 2^{-k}, 1) \times [0,1) \times \cdots \times [0,1)$ for $i = 1, 2, \dots, d$, and $r_{d+i} := [0, c_{i,1}) \times \cdots \times [0, c_{i,i-1}) \times [0, a_i) \times [a_{i+1}, a_{i+1} + c_{i,i+1}) \times \cdots \times [a_d, a_d + c_{i,d})$ for $i = 1, 2, \dots, d$, where

$$c_{i,j} := \begin{cases} \max\left(2(a_j + 2^{-k}), a_i\right), & j < i, \\ \max(2^{-k+1}, a_i), & j > i. \end{cases}$$

Note that if $a_i = 0$, then $r_{d+\nu} = \emptyset$ for all $\nu = 1, \dots, i$. In particular, if $k = 1$, then $r_{d+i} = \emptyset$ for all $i = 1, \dots, d$. Keeping that in mind, everywhere below we assume that $k \geq 2$ (the case $k = 1$ can be treated similarly and, in fact, is much simpler).

We are going to prove that the union $\bigcup_{i=1}^{2d} r_i$ is a desired expression of R. It is apparent that we have the first two conditions satisfied with $r_0 = [\frac{1}{2}, 1) \times \cdots \times [\frac{1}{2}, 1)$ as its root and r_i, $i = 1, 2, \dots, d$, as its branches. From the definitions of r_i and r_{d+i}, we also see that all of them are regular dyadic rectangular solids.

We now show that $R = \bigcup_{i=1}^{2d} r_i$. Let $x = (x_1, x_2, \dots, x_d) \in R$. Then there exists at least one i such that either $a_i + 2^{-k} \leq x_i < 1$, or $0 \leq x_i < a_i$ and $0 \leq x_j < a_j + 2^{-k}$ for all $j \neq i$. In the former case, we have $x \in r_i$. For the latter case, let $i_0 := \max\{i \mid 0 \leq x_i < a_i \text{ and } 0 \leq x_j < a_j + 2^{-k} \text{ for all } j \neq i\}$. By the definition of r_{d+i} $(i = 1, 2, \dots, d)$, we have $x \in r_{d+i_0}$ because $0 \leq x_{i_0} < a_{i_0}$, $a_j \leq x_j < a_j + 2^{-k}$ for all $j = i_0 + 1, \dots, d$ and $c_{i_0,j} \geq a_j + 2^{-k}$ for all $j < i_0$. It is easy to see, on the other hand, that all r_i $(i = 1, 2, \dots, 2d)$ are subsets of R (note that all $c_{i,j}$ for $j < i$ do not exceed 1 by the assumption $a_i + 2^{-k} \leq \frac{1}{2}$ and all $c_{i,j}$ for $j > i$ do not exceed $\frac{1}{2}$). Therefore, we have $R = \bigcup_{i=1}^{2d} r_i$.

It remains to show that for each nonempty rectangular solid r_{d+i}, $i = 1, \dots, d$, there exists a branch r_j with $\mathrm{vol}(r_{d+i} \cap r_j) \geq \frac{1}{2}\,\mathrm{vol}(r_{d+i})$. (It is easy to see that the solid r_{d+i}, $i = 1, \dots, d$ is not a branch since $\mathrm{vol}(r_{d+i}) = a_i \prod_{\nu=1}^{i-1} c_{i,\nu} \prod_{\nu=i+1}^{d} c_{i,\nu} \leq a_i < \frac{1}{2}$.)

Assume that $r_{d+i} \neq \emptyset$. From the definitions of r_j and r_{d+i}, we have $\mathrm{vol}(r_{d+i} \cap r_j) \geq \frac{1}{2}\,\mathrm{vol}(r_{d+i})$ if $j \neq i$. Indeed, we have $r_{d+i} \cap r_j = [0, c_{i,1}) \times \cdots \times [0, c_{i,i-1}) \times [0, a_i) \times [a_{i+1}, a_{i+1} + c_{i,i+1}) \times \cdots \times [a_j + 2^{-k}, a_j + c_{i,j}) \times \cdots \times [a_d, a_d + c_{i,d})$, if $j > i$; and $r_{d+i} \cap r_j = [0, c_{i,1}) \times \cdots \times [a_j + 2^{-k}, c_{i,j}) \times \cdots \times [0, c_{i,i-1}) \times [0, a_i) \times [a_{i+1}, a_{i+1} + c_{i,i+1}) \times \cdots \times [a_d, a_d + c_{i,d})$, if $j < i$. Since $c_{i,j} \geq 2^{-k+1}$ for $j > i$ and $c_{i,j} \geq 2(a_j + 2^{-k})$ for $j < i$, so $\mathrm{vol}(r_{d+i} \cap r_j) \geq \frac{1}{2}\,\mathrm{vol}(r_{d+i})$ and then the third condition is satisfied.

Now let us estimate $E_f(R)_p$ using the above decomposition of R into r_i. The case $p = \infty$ is trivial, and so we assume that $0 < p < \infty$. Let $P_{r_i} \in \Pi_{r-1}$ denote a polynomial of best L_p-approximation to f on r_i, i.e., $E_f(r_i)_p = \| f - P_{r_i} \|_{L_p(r_i)}$.

Since we have

$$
\begin{aligned}
E_f(R)_p^p &\leq \| f - P_{r_0} \|_{L_p(R)}^p \\
&\leq \sum_{i=1}^{\mu} \| f - P_{r_0} \|_{L_p(r_i)}^p,
\end{aligned}
$$

if we can show for each i $(i = 1, 2, \dots, \mu)$ that

$$
\| f - P_{r_0} \|_{L_p(r_i)} \leq C \sum_{j=1}^{\mu} E_f(r_j)_p,
$$

then

$$
\begin{aligned}
E_f(R)_p^p &\leq C \sum_{i=1}^{\mu} E_f(r_i)_p^p \\
&\leq C \sum_{i=1}^{\mu} \omega_r(f, \mathrm{vol}(r_i)^{1/d}, r_i)_p^p.
\end{aligned}
$$

The inequality $E_f(r_i)_p \le C\omega_r(f, \mathrm{vol}(r_i)^{1/d}, r_i)_p$ (used in the last estimate) can easily be obtained from the multidimensional Whitney theorem for a unit cube using the affine transformation $T\colon r_i \to [0,\ 1)^d$ and taking into account that r_i is a regular solid.

Now, if r_i is a branch, then

$$\begin{aligned} \|f - P_{r_0}\|_{L_p(r_i)} &\le C\big[\|P_{r_0} - P_{r_i}\|_{L_p(r_i)} + \|f - P_{r_i}\|_{L_p(r_i)}\big] \\ &\le C\big[\|P_{r_0} - P_{r_i}\|_{L_p(r_0)} + E_f(r_i)_p\big] \\ &\le C\big[\|f - P_{r_0}\|_{L_p(r_0)} + \|f - P_{r_i}\|_{L_p(r_0)} + E_f(r_i)_p\big] \\ &\le C\big[E_f(r_0)_p + E_f(r_i)_p\big] \le CE_f(r_i)_p, \end{aligned}$$

because $r_0 \subset r_i$ and $\mathrm{vol}(r_0) = 2^{-d} \ge 2^{-d}\,\mathrm{vol}(r_i)$.

If r_i is not a branch, by what is proved, there exists a branch r_j such that $\mathrm{vol}(r_i \cap r_j) \ge \frac{1}{2}\,\mathrm{vol}(r_i)$. Then, we have

$$\begin{aligned} \|f - P_{r_0}\|_{L_p(r_i)} &\le C\big[\|P_{r_0} - P_{r_i}\|_{L_p(r_i)} + \|f - P_{r_i}\|_{L_p(r_i)}\big] \\ &\le C\big[\|P_{r_0} - P_{r_i}\|_{L_p(r_i\cap r_j)} + E_f(r_i)_p\big] \\ &\le C\big[\|f - P_{r_0}\|_{L_p(r_i\cap r_j)} + \|f - P_{r_i}\|_{L_p(r_i\cap r_j)} + E_f(r_i)_p\big] \\ &\le C\big[\|f - P_{r_0}\|_{L_p(r_j)} + E_f(r_i)_p\big] \le C\left[E_f(r_j)_p + E_f(r_i)_p\right]. \end{aligned}$$

The proof is now completed. ■

To emphasize that r_i and μ correspond to a particular ring R we will use the notation r_i^R and μ^R. Also, if R is a cube (degenerate ring) we denote $\mu^R := 1$ and $r_1^R := R$.

Corollary 2. *Let R be a dyadic ring. Then, for $0 < \alpha < r$ and $0 < p < \infty$, we have*

$$E_f(R)_p \le C \sum_{i=1}^{\mu^R} E_f(r_i^R)_p \le C \sum_{i=1}^{\mu^R} |f|_{B^\alpha(r_i^R)} \le C|f|_{B^\alpha(R)}.$$

Corollary 2 immediately follows from Lemma 1 and the following Lemma C:

Lemma C (DeVore and Popov [16], see also [9]). *Let $1/\sigma = \alpha/d + 1/p$, then $B_p^\alpha(L_\sigma) \hookrightarrow L_p$, that is, for all $f \in B_p^\alpha(L_\sigma)$, $\|f\|_{L_p} \le C\|f\|_{B_p^\alpha(L_\sigma)}$. In particular, $\|f\|_{L_p} \le C\|f\|_{B_q^\alpha(L_\sigma)}$ for all $0 < q \le p$. Moreover, for each $I \in \mathbf{D}(\mathbf{R}^d)$ and $f \in B^\alpha(I)$:*

$$E_f(I)_p \le C|f|_{B^\alpha(I)}, \qquad 0 < \alpha < r.$$

4. Description of Algorithm

We will use the algorithm developed by Cohen, DeVore, Petrushev, and Xu [7] to construct piecewise polynomial functions satisfying Jackson inequalities in the spaces $V_{\sigma,p}^r$ and B^α (see Sections 6 and 7). For completeness, we describe the algorithm in this section. First, we recall some definitions. Let I be a dyadic cube ($I \in \mathbf{D}$). Then:

- If $I \in \mathbf{D}_k$, then J is the parent of I if and only if $J \in \mathbf{D}_{k-1}$ and $I \subseteq J$.

- J is a child of I if and only if I is the parent of J.
- I and J are brothers if and only if they have the same parent.
- J is a descendent of I if and only if $J \in \mathbf{D}$ and $J \subsetneq I$.
- J is an ancestor of I if and only if I is a descendent of J.

Let Φ denote a nonnegative set function defined on the algebra $\mathcal{A}$ that consists of all subsets of $[0,\ 1)^d$ formed by finite unions and intersections of rings from $\overline{\mathbf{D}}$ and their complements. We assume that Φ has the following properties:

(i) Φ is subadditive: $\Phi(R_1) + \Phi(R_2) \le \Phi(R_1 \cup R_2)$ for $R_1,\ R_2 \in \mathcal{A}$ such that $R_1 \cap R_2 = \emptyset$;
(ii) $\Phi(R) \to 0$ uniformly as $\mathrm{vol}(R) \to 0$.

The set function $\Phi(R)$ usually depends on the error of approximation of f on R. For instance, we can choose $\Phi(R)$ to be $E_f(R)_p^p$.

Given a parameter $\varepsilon > 0$, we define Υ_ε to be the set of cubes $I \in \mathbf{D}$ such that $\Phi(I) > \varepsilon$, and we call Υ_ε a tree which means that whenever $I \in \Upsilon_\varepsilon$ and $I \neq [0,\ 1)^d$, then its parent also belongs to Υ_ε. Note that Υ_ε has finite cardinality because of the second condition on Φ. In Υ_ε, we have three different types of cubes:

(i) The set $\mathcal{F}_\varepsilon$ of *final cubes* consists of the elements $I \in \Upsilon_\varepsilon$ with no child in Υ_ε.
(ii) The set $\mathcal{N}_\varepsilon$ of *branching cubes* consists of the elements $I \in \Upsilon_\varepsilon$ with more than one child in Υ_ε.
(iii) The set $\mathcal{C}_\varepsilon$ of *chaining cubes* consists of the elements $I \in \Upsilon_\varepsilon$ with exactly one child in Υ_ε.

Moreover, the set $\mathcal{C}_\varepsilon$ can be divided into *n maximal chains* C_k such that $\mathcal{C}_\varepsilon = \bigcup_{k=1}^n C_k$ with $C_k = \{I_0, \ldots, I_{m-1}\}$, $m = m(k)$, where each cube I_{i+1} is a child of I_i, $i = 0, \ldots, m-2$, I_0 is not a child of a chaining cube, and I_{m-1} is not a parent of a chaining cube. For a set S, let $|S|$ denote its cardinality. We have (see [7])

$$|\mathcal{N}_\varepsilon| \le |\mathcal{F}_\varepsilon| - 1 \qquad \text{and} \qquad n \le 2|\mathcal{F}_\varepsilon| - 1.$$

Now, let us describe how to construct a partition $\mathcal{P}_\varepsilon$ of $[0,\ 1)^d$ into rings R with $\Phi(R) \le \varepsilon$. In fact, we have $\mathcal{P}_\varepsilon = \mathcal{P}_\varepsilon^1 \cup \mathcal{P}_\varepsilon^2 \cup \mathcal{P}_\varepsilon^3$, where $\mathcal{P}_\varepsilon^1$ is the collection of all children J of the final cubes $I \in \mathcal{F}_\varepsilon$ and $\mathcal{P}_\varepsilon^2$ is the collection of the children J of the branching cubes $I \in \mathcal{N}_\varepsilon$ such that $J \notin \Upsilon_\varepsilon$. The collection $\mathcal{P}_\varepsilon^3$ consists of good rings (or cubes) generated from the maximal chains $C_k = \{I_0, \ldots, I_{m-1}\}$, $m = m(k)$, $k = 1, \ldots, n$, by the following recursion algorithm. Note that the last cube I_{m-1} of C_k always contains exactly one child I_m from Υ_ε which is in either $\mathcal{F}_\varepsilon$ or $\mathcal{N}_\varepsilon$. So, for each C_k, we first define $0 = j_0 < j_1 < \cdots < j_p = m$ in such a way that, assuming $j_i < m$ is chosen, we choose j_{i+1} as follows:

(i) if $\Phi(I_{j_i} \backslash I_m) \le \varepsilon$, then $j_{i+1} := m$ and the algorithm terminates;
(ii) if $\Phi(I_{j_i} \backslash I_{j_i+1}) > \varepsilon$, then $j_{i+1} := j_i + 1$; and
(iii) otherwise, j_{i+1} is chosen such that $\Phi(I_{j_i} \backslash I_{j_{i+1}}) \le \varepsilon$ and $\Phi(I_{j_i} \backslash I_{j_{i+1}+1}) > \varepsilon$.

Then the set $\mathcal{P}_\varepsilon^3$ consists of, from each maximal chain C_k:

(i) all rings $I_{j_i} \backslash I_{j_{i+1}}$ such that $\Phi(I_{j_i} \backslash I_{j_{i+1}}) \le \varepsilon$;

(ii) the children of I_{j_i} that differ from $I_{j_{i+1}}$ for all i such that $\Phi(I_{j_i} \setminus I_{j_{i+1}}) > \varepsilon$ (which is possible only in the case $j_{i+1} := j_i + 1$).

It is shown in [7] that this $\mathcal{P}_\varepsilon$ is a partition of $[0,\ 1)^d$ satisfying the following theorem.

Theorem D. *Let $\varepsilon > 0$ be such that $\Upsilon_\varepsilon \neq \emptyset$. Then, there exist a partition $\mathcal{P}_\varepsilon$ of $[0,\ 1)^d$ into disjoint rings such that $\Phi(K) \leq \varepsilon$ if $K \in \mathcal{P}_\varepsilon$, and a set $\widetilde{\mathcal{P}_\varepsilon}$ of pairwise disjoint rings such that $\Phi(K) > \varepsilon$ if $K \in \widetilde{\mathcal{P}_\varepsilon}$, and $|\mathcal{P}_\varepsilon| \leq 8|\widetilde{\mathcal{P}_\varepsilon}|$. (In particular, we can choose $\mathcal{P}_\varepsilon$ to be the partition produced by the algorithm above.)*

This theorem not only gives a partition of $[0,\ 1)^d$ into disjoint good rings, but it also shows that the number of good rings in the partition is controlled by the number of certain bad rings that are pairwise disjoint. The latter is more important to the estimation of the approximation degree.

5. Multivariate Free Splines on Ring Partitions

It is easy to see that $\Sigma_{n,r}$ are nonlinear manifolds, i.e., the sum of two functions from $\Sigma_{n,r}$ does not have to lie in $\Sigma_{n,r}$. However, as shown in the two lemmas below, the sum belongs to $\Sigma_{Cn,r}$. This will be shown via $S_{n,r}(I)$, the set of all S such that $S = \sum_{i=1}^n p_i \chi_{I_i}$, where $I_i \subseteq I$ are dyadic cubes (which are not necessarily disjoint) and $p_i \in \Pi_{r-1}$.

Lemma 3. *If $n \in \mathbf{N}$ and $I \in \mathbf{D}(\mathbf{R}^d)$, then $S_{n,r}(I) \subseteq \Sigma_{N,r}(I)$ for some $N \leq (n-1)2^d + n + 1 < n(2^d+1)$. That is, if $S(x) = \sum_{i=1}^n p_i(x)\chi_{I_i}(x)$, where $\{I_i\}_{i=1}^n$ is a collection of n (distinct but not necessarily disjoint) dyadic cubes contained in I, and $p_i \in \Pi_{r-1}$, then it can be represented as*

$$S(x) = \sum_{j=1}^{N} q_j(x)\chi_{R_j}(x), \qquad \{R_j\}_{j=1}^N \in \pi_r(I), \qquad q_j \in \Pi_{r-1}. \tag{5.1}$$

Proof. We use induction on n. It is trivial to verify the result for $n = 1$, that is, $S_{1,r}(I) \subseteq \Sigma_{2,r}(I)$. Now we suppose the result is valid for all $I \in \mathbf{D}(\mathbf{R}^d)$ and all $\overline{n}$, $1 \leq \overline{n} < n$.

Let $S(x) = \sum_{i=1}^n p_i(x)\chi_{I_i}(x)$, where $\{I_1, \ldots, I_n\}$ is a collection of distinct cubes in I, and let $\widetilde{I}$ be the smallest dyadic cube in $\mathbf{D}(I)$ that contains $\bigcup_i I_i$. The following two cases are possible:

Case 1: $\widetilde{I}$ coincides with one of I_i, say, $\widetilde{I} = I_n$.

Then the cubes $I_1, \ldots, I_{n-1}$ are contained in $\widetilde{I}$ and

$$S(x) - p_n(x) = \sum_{i=1}^{n-1} p_i(x)\chi_{I_i}(x) \qquad \text{for} \quad x \in \widetilde{I}.$$

Thus, we can use the induction hypothesis for $S_{n-1,r}(\widetilde{I})$ to conclude that, for $x \in \widetilde{I}$,

$$S(x) - p_n(x) = \sum_{j=1}^{\widetilde{N}} q_j(x)\chi_{R_j}(x), \qquad \{R_j\}_{j=1}^{\widetilde{N}} \in \pi_r(\widetilde{I}),$$

where $\widetilde{N} \le (n-2)2^d + n$. Now, if $\widetilde{I} \subsetneq I$, then

$$S(x) = 0 \cdot \chi_{I \setminus \widetilde{I}}(x) + \sum_{j=1}^{\widetilde{N}} \big(q_j(x) + p_n(x)\big)\, \chi_{R_j}(x)$$

for any $x \in I$, and $\{R_j\}_{j=1}^{\widetilde{N}} \cup \{I \setminus \widetilde{I}\} \in \pi_r(I)$. If $\widetilde{I} = I$, then

$$S(x) = \sum_{j=1}^{\widetilde{N}} (q_j(x) + p_n(x))\chi_{R_j}(x)$$

for any $x \in I$, and $\{R_j\}_{j=1}^{\widetilde{N}} \in \pi_r(I)$. In both of these cases $N \le \widetilde{N} + 1 \le (n-2)2^d + n + 1 \le (n-1)2^d + n + 1$.

Case 2: $\widetilde{I} \supsetneq I_i$, $i = 1, \dots, n$.

Then each of the cubes I_i, $i = 1, \dots, n$, is contained in one of the children of $\widetilde{I}$. Let $C_1, \dots, C_m$, $2 \le m \le 2^d$, be the children of $\widetilde{I}$ containing at least one of I_i, $i = 1, \dots, n$ (m cannot be 1 since, otherwise, $\widetilde{I}$ would not be the smallest cube containing $\bigcup_i I_i$), and let $C_{m+1}, \dots, C_{2^d}$ be the children of $\widetilde{I}$ having empty intersection with $\bigcup_i I_i$. For each $k = 1, \dots, m$, denote the number of cubes I_i, $1 \le i \le n$, contained in C_k by n_k. Then $\sum_{k=1}^m n_k = n$, and, since $m \ge 2$ and $n_k \ge 1$, we conclude that $n_k \le n-1$ for all $k = 1, \dots, m$. Thus, we can use the induction hypothesis for each of $S_{n_k,r}(C_k)$, $k = 1, \dots, m$, and conclude that

$$\sum_{j=1}^{n} p_j(x)\chi_{I_j}(x) = \sum_{k=1}^{m}\sum_{j=1}^{l_k} q_{k,j}(x)\chi_{R_{k,j}}(x)$$

for any $x \in I$, where $l_k \le (n_k - 1)2^d + n_k + 1$ and $\{R_{k,j}\}_{j=1}^{l_k} \in \pi_r(C_k)$. Now,

$$\widetilde{S}(x) = \sum_{k=1}^{2^d - m} 0 \cdot \chi_{C_{m+k}}(x) + \sum_{k=1}^{m}\sum_{j=1}^{l_k} q_{k,j}(x)\chi_{R_{k,j}}(x)$$

for any $x \in \widetilde{I}$ (actually any $x \in I$). Note that $\bigcup_{k=1}^m \{R_{k,j}\}_{j=1}^{l_k} \cup \bigcup_{k=1}^{2^d-m}\{C_{m+k}\} \in \pi_r(\widetilde{I})$ and, thus,

$$S(x) = \begin{cases} \widetilde{S}(x), & \text{if } \widetilde{I} = I, \\ \widetilde{S}(x) + 0 \cdot \chi_{I \setminus \widetilde{I}}(x), & \text{if } \widetilde{I} \subsetneq I. \end{cases}$$

In the case $\widetilde{I} \subsetneq I$, $\bigcup_{k=1}^m \{R_{k,j}\}_{j=1}^{l_k} \cup \bigcup_{k=1}^{2^d-m}\{C_{m+k}\} \cup \{I \setminus \widetilde{I}\} \in \pi_r(I)$. Thus,

$$\begin{aligned} N &\le 2^d - m + 1 + \sum_{k=1}^{m} l_k \\ &\le 2^d - m + 1 + \sum_{k=1}^{m}((n_k - 1)2^d + n_k + 1) \\ &= 2^d - m + 1 + (n-m)2^d + n + m \\ &= (n - m + 1)2^d + n + 1 \le (n-1)2^d + n + 1, \end{aligned}$$

which completes the proof of the lemma. ∎

Lemma 4. *If $m, n, r \in \mathbf{N}$, and $I \in \mathbf{D}(\mathbf{R}^d)$, then $\Sigma_{m,r}(I)+\Sigma_{n,r}(I) \subseteq \Sigma_{(2^d+1)(m+n),r}(I)$.*

Proof. It can be readily verified that $\Sigma_{n,r} \subseteq S_{n,r}$ and $S_{n,r} + S_{m,r} \subseteq S_{n+m,r}$. These inclusions and Lemma 3 immediately imply that

$$\Sigma_{n,r} + \Sigma_{m,r} \subseteq S_{n,r} + S_{m,r} \subseteq S_{n+m,r} \subseteq \Sigma_{(2^d+1)(n+m),r}. \qquad \blacksquare$$

The following lemma shows that any function $f \in L_p(I)$, $I \in \mathbf{D}$, has a best approximant from the manifold $\Sigma_{n,r}$. Its proof is the same as that of Lemma 6.1 of [7].

Lemma E. *For each $f \in L_p(I)$ and $n \in \mathbf{N}$ there exists $g \in \Sigma_{n,r}(I)$ such that $\|f - g\|_{L_p(I)} = E(f, \Sigma_{n,r})_{L_p}$.*

6. Adaptive Approximation in $V^r_{\sigma,p}$

Theorem 5. *Let $0 < \sigma < p < \infty$, $\beta := 1/\sigma - 1/p$, and $r \in \mathbf{N}$. Then for $f \in L_p([0,\ 1)^d)$ and $t > 0$ we have*

$$K(f, t^\beta, L_p, V^r_{\sigma,p}) = \inf_{g \in V^r_{\sigma,p}} \left\{\|f - g\|_{L_p} + t^\beta |g|_{V^r_{\sigma,p}}\right\} \sim \mathcal{W}_r(f, t)_{\sigma,p}.$$

In the case $d = 1$, a version of this theorem was proved in [19]. Note that, if $d = 1$, then a partition of $[0, 1]$ into rings is just a regular partition into intervals. Theorem 5 is an immediate corollary of the following result:

Theorem 6. *Let $0 < \sigma < p < \infty$, $\beta := 1/\sigma - 1/p$, and $r \in \mathbf{N}$. Then*:

(i) *for any $f \in L_p([0,\ 1)^d)$ and $n \in \mathbf{N}$ there exists $g \in \Sigma_{n,r}$ such that*

$$\|f - g\|_{L_p} \le C\mathcal{W}_r(f, 1/n)_{\sigma,p} \quad (\textit{Jackson inequality}); \tag{6.1}$$

(ii) *for any $g \in \Sigma_{n,r}$ we have*

$$|g|_{V^r_{\sigma,p}} \le Cn^\beta \mathcal{W}_r(g, 1/n)_{\sigma,p} \quad (\textit{Bernstein inequality}). \tag{6.2}$$

Theorem 6 will be proved in Sections 6.1 and 6.2.

Proof of Theorem 5. First of all, for any $g \in V^r_{\sigma,p}$, using (2.1) and (2.2) we have

$$\begin{aligned} \mathcal{W}_r(f, t)_{\sigma,p} &\le C\left(\mathcal{W}_r(f - g, t)_{\sigma,p} + \mathcal{W}_r(g, t)_{\sigma,p}\right) \\ &\le C\left(\|f - g\|_{L_p} + t^\beta |g|_{V^r_{\sigma,p}}\right). \end{aligned}$$

Hence, taking infimum over all $g \in V^r_{\sigma,p}$, we obtain

$$\mathcal{W}_r(f, t)_{\sigma,p} \le CK(f, t^\beta, L_p, V^r_{\sigma,p}).$$

Now, suppose that Theorem 6 is proved. Let $t > 0$ and choose $n := [1/t] + 1$. Then Theorem 6 and inequality (2.1) imply that there exists $g \in \Sigma_{n,r}$ such that

$$\begin{aligned} t^\beta |g|_{V^r_{\sigma,p}} &\le C(tn)^\beta \mathcal{W}_r(g, 1/n)_{\sigma,p} \\ &\le C\mathcal{W}_r(g, 1/n)_{\sigma,p} \\ &\le C\big(\mathcal{W}_r(g - f, 1/n)_{\sigma,p} + \mathcal{W}_r(f, 1/n)_{\sigma,p}\big) \\ &\le C\big(\|g - f\|_{L_p} + \mathcal{W}_r(f, 1/n)_{\sigma,p}\big) \\ &\le C\mathcal{W}_r(f, 1/n)_{\sigma,p}, \end{aligned}$$

and, therefore,

$$\begin{aligned} K(f, t^\beta, L_p, V^r_{\sigma,p}) &\le \|f - g\|_{L_p} + t^\beta |g|_{V^r_{\sigma,p}} \\ &\le C\mathcal{W}_r(f, 1/n)_{\sigma,p} \le C\mathcal{W}_r(f, t)_{\sigma,p}. \end{aligned}$$ ■

Corollary 7. *Let* $0 < \sigma < p < \infty$, $\beta := 1/\sigma - 1/p$, *and* $n, r \in \mathbf{N}$. *Then*:

(i) $E(f, \Sigma_{n,r})_{L_p} \le Cn^{-\beta} |f|_{V^r_{\sigma,p}}$ *for any* $f \in V^r_{\sigma,p}$; *and*
(ii) $|g|_{V^r_{\sigma,p}} \le Cn^\beta \|g\|_{L_p}$ *for any* $g \in \Sigma_{n,r}$.

Proof. The statement (i) immediately follows from (6.1) and (2.2), and (ii) is a consequence of (6.2) and (2.1). ■

The following is a consequence of the above corollary and Theorems A and 5.

Corollary 8. *Let* $0 < \sigma < p < \infty$, $\beta := 1/\sigma - 1/p$, $r \in \mathbf{N}$. *Then, for* $0 < \alpha < \beta$ *and* $0 < q \le \infty$,

$$A^\alpha_q(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = \big(L_p, V^r_{\sigma,p}\big)_{\alpha/\beta,q}. \tag{6.3}$$

Corollary 9. *Let* $0 < \sigma < p < \infty$ *and* $r \in \mathbf{N}$. *Then, for* $0 < \alpha < 1/\sigma - 1/p$ *and* $0 < q < \infty$, *we have*

$$A^\alpha_q(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = \left\{ f \in L_p([0,\ 1)^d) \;\middle|\; \int_0^\infty \big[t^{-\alpha}\mathcal{W}_r(f, t)_{\sigma,p}\big]^q \frac{dt}{t} < \infty \right\}.$$

Corollary 10 ($q = \infty$). *Let* $0 < \sigma < p < \infty$ *and* $r \in \mathbf{N}$. *Then, for* $0 < \alpha < 1/\sigma - 1/p$:

$$E(f, \Sigma_{n,r})_{L_p} = O(n^{-\alpha}) \iff \mathcal{W}_r(f, t)_{\sigma,p} = O(t^\alpha).$$

The following corollary shows that there exists $g \in \Sigma_{n,r}$ (in fact, the piecewise polynomial function g that we construct in Section 6.1 will do), which is a near-minimizer for the K-functional $K(f, n^{-\beta}, L_p, V^r_{\sigma,p})$.

Corollary 11. *Let* $0 < \sigma < p < \infty$, $\beta := 1/\sigma - 1/p$, *and* $r \in \mathbf{N}$. *Then for* $f \in L_p([0,\ 1)^d)$ *and* $n \in \mathbf{N}$ *there exists* $g \in \Sigma_{n,r}$ *such that*

$$\|f - g\|_{L_p} + n^{-\beta}|g|_{V^r_{\sigma,p}} \le CK(f, n^{-\beta}, L_p, V^r_{\sigma,p}).$$

Proof. Using Theorem 5 with $t = 1/n$, Theorem 6, and the properties of $\mathcal{W}_r$ we conclude that there exists $g \in \Sigma_{n,r}$ such that

$$\begin{aligned} \|f-g\|_{L_p} + n^{-\beta}|g|_{V^r_{\sigma,p}} &\le \|f-g\|_{L_p} + C\mathcal{W}_r(g,1/n)_{\sigma,p} \\ &\le \|f-g\|_{L_p} + C\mathcal{W}_r(g-f,1/n)_{\sigma,p} + C\mathcal{W}_r(f,1/n)_{\sigma,p} \\ &\le C\|f-g\|_{L_p} + C\mathcal{W}_r(f,1/n)_{\sigma,p} \\ &\le C\mathcal{W}_r(f,1/n)_{\sigma,p} \le CK(f,n^{-\beta},L_p,V^r_{\sigma,p}). \end{aligned}$$ ■

6.1. *Jackson Inequality for $V^r_{\sigma,p}$*

First of all, we show that any set of (pairwise) disjoint rings and cubes can be complemented to a partition of $[0,\ 1)^d$ in π_r.

Lemma 12. *Let $\mathcal{R}$ be a collection of pairwise disjoint dyadic rings and cubes contained in a dyadic cube I, and let $|\mathcal{R}|$ denote its cardinality. Then there exists a partition $\mathcal{T}_{\mathcal{R}}(I)$ of I into dyadic rings ($\mathcal{T}_{\mathcal{R}}(I) \in \pi_r(I)$) such that $\mathcal{R} \subseteq \mathcal{T}_{\mathcal{R}}(I)$ and $|\mathcal{T}_{\mathcal{R}}(I)| \le (2^d+2)|\mathcal{R}| - 2^d + 1$ (and, therefore, $|\mathcal{T}_{\mathcal{R}}(I)| \le 2^{d+1}|\mathcal{R}|$).*

Proof. The idea of the proof of this lemma is similar to that of Lemma 3. Clearly the statement is true for a cube $I \in \mathbf{D}$ if and only if it is true for $I = [0,\ 1)^d$. We proceed by induction on $|\mathcal{R}|$. If $|\mathcal{R}| = 1$, then $\mathcal{R} = \{K\}$. Now we choose $\mathcal{T}_{\mathcal{R}} := \mathcal{T}_{\mathcal{R}}([0,\ 1)^d) = \left\{[0,\ 1)^d \setminus K, K\right\}$ (if K is a cube, i.e., $K = K_{\boxplus}$), and $\mathcal{T}_{\mathcal{R}} = \left\{[0,\ 1)^d \setminus K_{\boxplus}, K, K_{\square}\right\}$ (if K is a nondegenerate ring, $K = K_{\boxplus} \setminus K_{\square}$). Clearly, $|\mathcal{T}_{\mathcal{R}}| \le 3$.

Now suppose that the lemma is proved for all collections $\mathcal{R}$ such that $|\mathcal{R}| \le N$, and consider $\mathcal{R}$ such that $|\mathcal{R}| = N+1$, $N \ge 1$, and $I = [0,\ 1)^d$. For a set $\Omega \subseteq [0,\ 1)^d$ denote $\mathcal{R}(\Omega) = \{K \in \mathcal{R} \mid \mathrm{int}(K) \cap \mathrm{int}(\Omega) \ne \emptyset\}$. Let Q be the smallest dyadic cube in $\mathbf{D}(I)$ that contains $\bigcup_{R\in\mathcal{R}} R$. We have the following two possible cases similar to those in Lemma 3.

Case 1: Q coincides with $R_{\boxplus}$ for some $R \in \mathcal{R}$.

Since $\mathcal{R}(R_{\square}) = \mathcal{R}\setminus\{R\}$ contains N cubes and rings from $\mathcal{R}$ we can use the induction hypothesis to conclude that there exists a partition $\mathcal{T}_{\mathcal{R}(R_{\square})}(R_{\square})$ of $R_{\square}$ such that

$$|\mathcal{T}_{\mathcal{R}(R_{\square})}(R_{\square})| \le (2^d+2)|\mathcal{R}(R_{\square})| - 2^d + 1 \le (2^d+2)N - 2^d + 1.$$

Now, $\mathcal{T}_{\mathcal{R}} := \mathcal{T}_{\mathcal{R}(R_{\square})}(R_{\square}) \cup \{R\} \cup \{[0,\ 1)^d \setminus R_{\boxplus}\}$ is a desired partition of $[0,\ 1)^d$, and

$$|\mathcal{T}_{\mathcal{R}}| = 2 + |\mathcal{T}_{\mathcal{R}(R_{\square})}(R_{\square})| \le (2^d+2)(N+1) - 2^d + 1.$$

Case 2: $Q \supsetneq R_{\boxplus}$, $\forall\, R \in \mathcal{R}$.

By the assumption of Q, at least two of its 2^d children I_ν, $\nu = 1, 2, \ldots, 2^d$, contain members of $\mathcal{R}$ in this case, therefore $|\mathcal{R}(I_\nu)| \le N$ for any $1 \le \nu \le 2^d$, and we can use the induction hypothesis to construct partitions $\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)$ of I_ν as follows:

- if $\mathcal{R}(I_\nu) \ne \emptyset$, then let $\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)$ be a partition of I_ν containing $\mathcal{R}(I_\nu)$ and such that $|\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)| \le (2^d+2)|\mathcal{R}(I_\nu)| - 2^d + 1$ (its existence is guaranteed by our induction hypothesis); and

- if $\mathcal{R}(I_\nu)=\emptyset$, then we choose $\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu):=\{I_\nu\}$.

Now,

$$\mathcal{T}_{\mathcal{R}}=\left\{[0,\ 1)^d\setminus Q\right\}\cup\bigcup_{\nu=1}^{2^d}\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)$$

is a partition of $[0,\ 1)^d$ containing $\mathcal{R}$. Taking into account that

$$\sum_{I_\nu:\mathcal{R}(I_\nu)\neq\emptyset}|\mathcal{R}(I_\nu)|=|\mathcal{R}|,$$

we have

$$\begin{aligned}
|\mathcal{T}_{\mathcal{R}}| &= 1+\sum_{\nu=1}^{2^d}|\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)|\\
&= 1+|\{I_\nu\mid\mathcal{R}(I_\nu)=\emptyset\}|+\sum_{I_\nu:\mathcal{R}(I_\nu)\neq\emptyset}|\mathcal{T}_{\mathcal{R}(I_\nu)}(I_\nu)|\\
&\le 1+|\{I_\nu\mid\mathcal{R}(I_\nu)=\emptyset\}|+\sum_{I_\nu:\mathcal{R}(I_\nu)\neq\emptyset}\big((2^d+2)|\mathcal{R}(I_\nu)|-2^d+1\big)\\
&\le 1+|\{I_\nu\mid\mathcal{R}(I_\nu)=\emptyset\}|-(2^d-1)\,|\{I_\nu\mid\mathcal{R}(I_\nu)\neq\emptyset\}|+(2^d+2)|\mathcal{R}|\\
&\le (2^d+2)|\mathcal{R}|-2^d+1,
\end{aligned}$$

since $|\{I_\nu\mid\mathcal{R}(I_\nu)\neq\emptyset\}|\ge 2$ and $|\{I_\nu\mid\mathcal{R}(I_\nu)=\emptyset\}|+|\{I_\nu\mid\mathcal{R}(I_\nu)\neq\emptyset\}|=2^d$. ∎

Proof of (6.1). Let $\Phi(R)=E_f(R)_p^p$, $\varepsilon=n^{-1}\mathcal{W}_r(f,1/n)_{\sigma,p}^p$ (without loss of generality we can assume that $\mathcal{W}_r(f,1/n)_{\sigma,p}\neq 0$, since $\mathcal{W}_r(f,t)_{\sigma,p}=0$ only if $f\in\Pi_{r-1}$), and let $\mathcal{P}_\varepsilon$ be a partition produced by the algorithm described in Section 4. We use Theorem D to show that $|\mathcal{P}_\varepsilon|\le Cn$. Suppose that $|\mathcal{P}_\varepsilon|\ge n$ (otherwise, we are done). Since the set $\widetilde{\mathcal{P}_\varepsilon}$ (see Theorem D) consists of pairwise disjoint rings and cubes contained in $[0,\ 1)^d$, we can use Lemma 12 to enlarge $\widetilde{\mathcal{P}_\varepsilon}$ to $\mathcal{T}_{\widetilde{\mathcal{P}_\varepsilon}}$, a partition of $[0,\ 1)^d$ into dyadic rings such that

$$|\widetilde{\mathcal{P}_\varepsilon}|\le|\mathcal{T}_{\widetilde{\mathcal{P}_\varepsilon}}|\le 2^{d+1}|\widetilde{\mathcal{P}_\varepsilon}|.$$

If we denote $|\widetilde{\mathcal{P}_\varepsilon}|=N$ and $|\mathcal{T}_{\widetilde{\mathcal{P}_\varepsilon}}|=\widetilde{N}$, then $N\le\widetilde{N}\le 2^{d+1}N$. Now, note that the inequalities $|\mathcal{P}_\varepsilon|\ge n$ (our assumption above) and $|\mathcal{P}_\varepsilon|\le 8N$ (Theorem D) imply that $\widetilde{N}\ge N\ge n/8$. Recalling that $E_f(Q)_p^p>\varepsilon$ for $Q\in\widetilde{\mathcal{P}_\varepsilon}$, we have

$$\begin{aligned}
N=|\widetilde{\mathcal{P}_\varepsilon}| &\le \varepsilon^{-\sigma/p}\sum_{Q\in\widetilde{\mathcal{P}_\varepsilon}}E_f(Q)_p^\sigma\le\varepsilon^{-\sigma/p}\sum_{Q\in\mathcal{T}_{\widetilde{\mathcal{P}_\varepsilon}}}E_f(Q)_p^\sigma\\
&\le C\varepsilon^{-\sigma/p}N^{\beta\sigma}\sup_{\{Q_i\}_{i=1}^{\widetilde{N}}\in\pi_r}\widetilde{N}^{-\beta\sigma}\sum_{i=1}^{\widetilde{N}}E_f(Q_i)_p^\sigma\\
&\le C\varepsilon^{-\sigma/p}N^{\beta\sigma}\sup_{0<h\le 1/\widetilde{N}}\ \sup_{\{Q_i\}_{i=1}^{n}\in\pi_r^h}h^{\beta\sigma}\sum_{i=1}^{n}E_f(Q_i)_p^\sigma\\
&\le C\varepsilon^{-\sigma/p}N^{\beta\sigma}\mathcal{W}_r(f,1/\widetilde{N})_{\sigma,p}^\sigma\le C\varepsilon^{-\sigma/p}N^{\beta\sigma}\mathcal{W}_r(f,1/n)_{\sigma,p}^\sigma\le Cn^{\sigma/p}N^{\beta\sigma},
\end{aligned}$$

which implies $N\le Cn$, since $\beta=1/\sigma-1/p$.

Define g by

$$g(x) := P_{I_i}(x) \qquad \text{for} \quad x \in I_i,$$

where $I_i \in \mathcal{P}_\varepsilon$, and P_{I_i} are best L_p approximants to f on I_i from polynomials of degree $< r$. Then $g \in \Sigma_{Cn,r}$ and

$$\begin{aligned}\|f-g\|_{L_p}^p &= \sum_{I_i \in \mathcal{P}_\varepsilon} \|f-P_{I_i}\|_{L_p(I_i)}^p \le \varepsilon |\mathcal{P}_\varepsilon| \\ &\le Cn\varepsilon \le C\mathcal{W}_r(f,1/n)_{\sigma,p}^p,\end{aligned}$$

which implies (6.1). ■

6.2. *Bernstein Inequality for $V_{\sigma,p}^r$*

Let $g \in \Sigma_{n,r}$ be a piecewise polynomial function of order $\le r$ defined on a partition $\mathcal{P}_g = \{R_i\}_{i=1}^n \in \pi_r$, i.e., for each $R_i \in \mathcal{P}_g$, $g|_{R_i} \in \Pi_{r-1}$. In the following lemma we show that, for any finite ring partition $\mathcal{P} = \{Q_i\}_{i\in\Lambda} \in \pi_r$ of $[0,\ 1)^d$, if

$$\Lambda_g := \bigl\{ i \in \Lambda \mid E_g(Q_i)_p \ne 0 \bigr\},$$

then $|\Lambda_g| \le 2n$. More precisely, we prove

Lemma 13. *Let $\mathcal{P} \in \pi_r$ be given. For any $Q \in \mathcal{P}$ with $E_g(Q)_p \ne 0$, there exists an $R \in \mathcal{P}_g$ that satisfies one of the two conditions below*:

(i) $Q_\boxplus \supsetneq R_\boxplus$ *and* $Q \cap R \ne \emptyset$; *and*
(ii) $R_\boxplus \supseteq Q_\boxplus \supsetneq R_\square \ne \emptyset$ *and* $Q_\square \not\supseteq R_\square$.

Moreover, any given $R \in \mathcal{P}_g$ corresponds to at most one ring Q satisfying (i), *and at most one Q satisfying* (ii). *Therefore,* $\bigl|\bigl\{ i \mid E_g(Q_i)_p \ne 0,\ \ Q_i \in \mathcal{P}\bigr\}\bigr| \le 2|\mathcal{P}_g|$.

Proof. We shall repeatedly use the fact that two dyadic cubes are either disjoint or one is contained in the other. We fix Q such that $E_g(Q)_p \ne 0$. Because $\mathcal{P}_g \in \pi_r$, there exists $R \in \mathcal{P}_g$ such that $Q_\boxplus \cap R_\boxplus \supseteq Q \cap R \ne \emptyset$ and $Q \not\subseteq R$, thus either $Q_\boxplus \supsetneq R_\boxplus$, which is (i); or $R_\boxplus \supseteq Q_\boxplus$. In the latter case, from $Q \not\subseteq R$ we know $Q_\boxplus \cap R_\square \ne \emptyset$, and in particular $R_\square \ne \emptyset$; from $Q \cap R \ne \emptyset$ we know $Q_\boxplus \not\subseteq R_\square$. Therefore $Q_\boxplus \supsetneq R_\square$. From $Q \not\subseteq R$ again, we know $Q_\square \not\supseteq R_\square$, which gives (ii).

For the uniqueness of the ring Q in (i), we fix $R \in \mathcal{P}_g$ and suppose there are two such rings Q and Q'. Since $Q_\boxplus \cap Q'_\boxplus \supseteq R_\boxplus \ne \emptyset$, one of them contains the other. Without loss of generality, we assume $Q_\boxplus \supseteq Q'_\boxplus$. Since $Q \cap Q' = \emptyset$, we have $Q_\square \supseteq Q'_\boxplus \supsetneq R_\boxplus$; but $Q_\square$ cannot contain $R_\boxplus$, for $Q \cap R \ne \emptyset$. This contradiction shows the ring in (i) is unique.

For the uniqueness of (ii), we suppose there are two rings Q and $Q' \in \mathcal{P}$ both satisfying (ii) for the same ring R. Similar to the above, since $Q_\boxplus \cap Q'_\boxplus \supseteq R_\square \ne \emptyset$, we can assume $Q_\boxplus \supseteq Q'_\boxplus$. From $Q \cap Q' = \emptyset$ we derive

$$R_\boxplus \supseteq Q_\boxplus \supsetneq Q_\square \supseteq Q'_\boxplus \supsetneq R_\square,$$

which contradicts the fact $Q_\square \not\supseteq R_\square$. ■

Proof of (6.2). Now, let $\mathcal{P} = \{Q_i\}_{i\in\Lambda} \in \pi_r$ be an arbitrary finite partition of $[0,\ 1)^d$ into dyadic rings. We consider its subset $\{Q_i\}_{i\in\Lambda_g}$, whose cardinality $|\Lambda_g| \le 2n$ by Lemma 13. According to Lemma 12 there exists a partition $\mathcal{T}_{\mathcal{P},g}$ such that $\{Q_i\}_{i\in\Lambda_g} \subseteq \mathcal{T}_{\mathcal{P},g}$ and

$$|\mathcal{T}_{\mathcal{P},g}| \le 2^{d+1}|\Lambda_g| \le 2^{d+2}n.$$

Now, denoting $h := 1/(2^{d+2}n)$, we have

$$\begin{aligned}
|g|_{V^r_{\sigma,p}} &\sim \sup_{\mathcal{P}\in\pi_r}\left(\sum_{i\in\Lambda} E_g(Q_i)_p^\sigma\right)^{1/\sigma} = \sup_{\mathcal{P}\in\pi_r}\left(\sum_{i\in\Lambda_g} E_g(Q_i)_p^\sigma\right)^{1/\sigma}\\
&\le \sup_{\mathcal{P}\in\pi_r}\left(\sum_{R\in\mathcal{T}_{\mathcal{P},g}} E_g(R)_p^\sigma\right)^{1/\sigma} \le \sup_{\pi\in\pi_r^h}\left(\sum_{R\in\pi} E_g(R)_p^\sigma\right)^{1/\sigma}\\
&\le Cn^\beta \sup_{\pi\in\pi_r^h} h^\beta\left(\sum_{R\in\pi} E_g(R)_p^\sigma\right)^{1/\sigma} \le Cn^\beta \mathcal{W}_r(g,h)_{\sigma,p}\\
&\le Cn^\beta \mathcal{W}_r(g,1/n)_{\sigma,p},
\end{aligned}$$

which is the inequality (6.2). ∎

7. Adaptive Approximation in the Besov Space B^α

In this section, we develop results for the Besov spaces B^α, which were defined on page 451.

Theorem 14 (Jackson Inequality). *For every $f \in B^\alpha$, $0 < \alpha < r$, $0 < p < \infty$,*

$$E(f, \Sigma_{n,r})_{L_p} \le Cn^{-\alpha/d}|f|_{B^\alpha}.$$

Theorem 15 (Bernstein Inequality). *Let $g \in \Sigma_{n,r}$, i.e., $g = \sum_{R\in\mathcal{P}} p_R\chi_R$, where $\mathcal{P} \in \pi_r$ with $|\mathcal{P}| \le n$, and $p_R \in \Pi_{r-1}$. Also, let $0 < p < \infty$ and $0 < \alpha < \min\{r, d/(d-1)p\}$. Then*

$$|g|_{B^\alpha} \le Cn^{\alpha/d}\|g\|_{L_p}.$$

Corollary 16. *Let $0 < p < \infty$ and let $0 < \alpha < \min\{r, d/(d-1)p\}$. Then, for $0 < \gamma < \alpha/d$ and $0 < q \le \infty$,*

$$A_q^\gamma(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = \left(L_p, B^\alpha\right)_{d\gamma/\alpha,q}.$$

This corollary can be restated as follows:

Corollary 16′. *Let $0 < p < \infty$ and let $0 < \beta < \min\{r/d, 1/(d-1)p\}$. Then, for $0 < \gamma < \beta$ and $0 < q \le \infty$,*

$$A_q^\gamma(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = \left(L_p, B^{d\beta}\right)_{\gamma/\beta,q}.$$

We can now characterize the interpolation spaces $\left(L_p, B^\alpha\right)_{\lambda/\alpha,q}$ in terms of the modulus $\mathcal{W}_r(f,t)_{\sigma,p}$. The following result immediately follows from Corollaries 9 and 16, and Theorem B:

Corollary 17. *Let* $0<\sigma<p<\infty$, $r\in\mathbf{N}$, $1/\sigma=\alpha/d+1/p$, *and* $0<\alpha<\min\{r, d/(d-1)p\}$. *Then for* $0<\lambda<\alpha$ *and* $0<q<\infty$:

$$\left(L_p, B^\alpha\right)_{\lambda/\alpha,q}=\left\{f\in L_p([0,\ 1)^d)\ \Big|\ \int_0^\infty\left[t^{-\lambda/d}\mathcal{W}_r(f,t)_{\sigma,p}\right]^q\frac{dt}{t}<\infty\right\}.$$

In particular, if $q=\tau=(\lambda/d+1/p)^{-1}$, *then*

$$\left(L_p, B^\alpha\right)_{\lambda/\alpha,\tau}=B^\lambda=\left\{f\in L_p([0,\ 1)^d)\ \Big|\ \int_0^\infty\left[t^{-\lambda/d}\mathcal{W}_r(f,t)_{\sigma,p}\right]^\tau\frac{dt}{t}<\infty\right\},$$

and, thus,

$$\int_0^\infty\left[t^{-\lambda}\mathcal{W}_r(f,t^d)_{\sigma,p}\right]^\tau\frac{dt}{t}<\infty\quad\Longleftrightarrow\quad\int_0^\infty\left[t^{-\lambda}\omega_r(f,t)_\tau\right]^\tau\frac{dt}{t}<\infty.$$

Using Corollaries 10 and 16 one can also get a similar statement in the case $q=\infty$.

7.1. *Proof of the Jackson Inequality for* B^α

Without loss of generality we can assume that $|f|_{B^\alpha}\neq 0$. Let $\Phi(R)=E_f(R)_p^p$, $\sigma=(\alpha/d+1/p)^{-1}$, and $\varepsilon=n^{-p/\sigma}|f|_{B^\alpha}^p$. Also, let $\mathcal{P}_\varepsilon$ be a partition of $[0,\ 1)^d$ produced by the algorithm of Section 4. Then $|\mathcal{P}_\varepsilon|\le 8|\widetilde{\mathcal{P}_\varepsilon}|$ (see Theorem D), where $\widetilde{\mathcal{P}_\varepsilon}$ is a set of pairwise disjoint dyadic rings such that for all $R\in\widetilde{\mathcal{P}_\varepsilon}$, $\Phi(R)>\varepsilon$.

We now estimate $|\widetilde{\mathcal{P}_\varepsilon}|$. If $R\in\widetilde{\mathcal{P}_\varepsilon}$, then $1<\varepsilon^{-\sigma/p}E_f(R)_p^\sigma$, and using Corollary 2 we have the following estimates:

$$|\widetilde{\mathcal{P}_\varepsilon}|\le\varepsilon^{-\sigma/p}\sum_{R\in\widetilde{\mathcal{P}_\varepsilon}}E_f(R)_p^\sigma\le C\varepsilon^{-\sigma/p}\sum_{R\in\widetilde{\mathcal{P}_\varepsilon}}\sum_{i=1}^{\mu^R}|f|_{B^\alpha(r_i^R)}^\sigma,\tag{7.1}$$

where, if $R\in\widetilde{\mathcal{P}_\varepsilon}$ is a cube, then $\mu^R:=1$ and $r_1^R:=R$.

We will now show that

$$\sum_{R\in\widetilde{\mathcal{P}_\varepsilon}}\sum_{i=1}^{\mu^R}|f|_{B^\alpha(r_i^R)}^\sigma\le C|f|_{B^\alpha}^\sigma,\tag{7.2}$$

which together with (7.1) implies that $|\mathcal{P}_\varepsilon|\le 8|\widetilde{\mathcal{P}_\varepsilon}|\le C\varepsilon^{-\sigma/p}|f|_{B^\alpha}^\sigma\le Cn$.

To prove (7.2) note that if Q is a regular solid then the following holds (see DeVore [8, ineq. (4.5)], for example, keeping in mind that the result for a regular solid follows from that for a cube by a change of variables):

$$\omega_r(f,t,Q)_\sigma^\sigma\sim t^{-d}\int_{[-t,t]^d}\int_{Q(rs)}|\Delta_s^r(f,x)|^\sigma\,dx\,ds,$$

where $Q(h) := \{x \mid [x, x+h] \subseteq Q\}$ and Δ_s^r is the usual rth difference. Thus, taking into account that according to Lemma 1 each $x \in R$ belongs to at most $\mu \le 2d$ regular solids r_i^R, we have

$$\begin{aligned}
\sum_{R\in\widetilde{\mathcal{P}}_\varepsilon} \sum_{i=1}^{\mu^R} |f|^\sigma_{B^\alpha(r_i^R)} &= \sum_{R\in\widetilde{\mathcal{P}}_\varepsilon} \sum_{i=1}^{\mu^R} \int_0^\infty t^{-\alpha\sigma-1}\omega_r(f,t,r_i^R)^\sigma_\sigma\,dt \\
&= \int_0^\infty t^{-\alpha\sigma-1}\left(\sum_{R\in\widetilde{\mathcal{P}}_\varepsilon} \sum_{i=1}^{\mu^R} \omega_r(f,t,r_i^R)^\sigma_\sigma\right)dt \\
&\le C\int_0^\infty t^{-\alpha\sigma-1}\left(\sum_{R\in\widetilde{\mathcal{P}}_\varepsilon} \sum_{i=1}^{\mu^R} t^{-d}\int_{[-t,t]^d}\int_{r_i^R(rs)} |\Delta_s^r(f,x)|^\sigma\,dx\,ds\right)dt \\
&\le C\int_0^\infty t^{-\alpha\sigma-1}\left(t^{-d}\int_{[-t,t]^d}\int_{[0,\,1)^d(rs)} |\Delta_s^r(f,x)|^\sigma\,dx\,ds\right)dt \\
&\le C\int_0^\infty t^{-\alpha\sigma-1}\omega_r(f,t,[0,\,1)^d)^\sigma_\sigma\,dt \le C|f|^\sigma_{B^\alpha}.
\end{aligned}$$

Thus, $|\mathcal{P}_\varepsilon| \le Cn$, and defining $g(x) := P_{I_i}(x)$, for $x \in I_i$, where $I_i \in \mathcal{P}_\varepsilon$, and P_{I_i} are best L_p approximants to f on I_i from polynomials of degree $< r$ (hence, $g \in \Sigma_{Cn,r}$), we have

$$\begin{aligned}
\|f-g\|^p_{L_p} &= \int_{[0,\,1)^d} |f-g|^p = \sum_{R\in\mathcal{P}_\varepsilon}\int_R |f-g|^p = \sum_{R\in\mathcal{P}_\varepsilon} E_f(R)^p_p \\
&\le \sum_{R\in\mathcal{P}_\varepsilon} \varepsilon = \varepsilon|\mathcal{P}_\varepsilon| \le Cn\varepsilon \le Cn^{1-p/\sigma}|f|^p_{B^\alpha}.
\end{aligned}$$

Thus,

$$E(f,\Sigma_{n,r})_{L_p} \le \|f-g\|_{L_p} \le Cn^{-\alpha/d}|f|_{B^\alpha}. \qquad \blacksquare$$

7.2. *Proof of the Bernstein Inequality for B^α*

The following proof is somewhat similar to the proofs that were used in [15] and [9]. We denote $a_R(x) := p_R(x)\chi_R(x)$ and note that

$$\begin{aligned}
\|g\|^p_{L_p} &= \int_{[0,\,1)^d} |g(x)|^p\,dx = \sum_{R\in\mathcal{P}}\int_R |g(x)|^p\,dx \\
&= \sum_{R\in\mathcal{P}}\int_R |a_R(x)|^p\,dx = \sum_{R\in\mathcal{P}} \|a_R\|^p_{L_p(R)}.
\end{aligned}$$

To estimate $|g|^\sigma_{B^\alpha}$ we first show that

$$\omega_r(g,t)^\sigma_\sigma = \omega_r\left(\sum_{R\in\mathcal{P}} a_R,\ t\right)^\sigma_\sigma \le C\sum_{R\in\mathcal{P}} \omega_r(a_R,t)^\sigma_\sigma. \tag{7.3}$$

Note that this inequality is trivial if $\sigma \le 1$, since, in this case, $\left\|\sum f_i\right\|_{L_\sigma}^{\sigma} \le \sum \|f_i\|_{L_\sigma}^{\sigma}$. For general σ, we let $x, h \in \mathbf{R}^d$ be fixed. Then

$$\begin{aligned}
\Delta_h^r(g, x) &= \sum_{i=0}^{r} (-1)^{r-i} \binom{r}{i} g(x + ih) \\
&= \sum_{i=0}^{r} (-1)^{r-i} \binom{r}{i} \sum_{R \in \mathcal{P}} a_R(x + ih) \\
&= \sum_{i=0}^{r} (-1)^{r-i} \binom{r}{i} \sum_{R \in \mathcal{P},\, R \cap \{x+ih:0 \le i \le r\} \ne \emptyset} a_R(x + ih) \\
&= \Delta_h^r(\widetilde{g}, x),
\end{aligned}$$

where $\widetilde{g}(x) = \sum_{R \in \widetilde{\mathcal{P}}} a_R(x)$, $\widetilde{\mathcal{P}} := \{R \in \mathcal{P},\, R \cap \{x + ih : 0 \le i \le r\} \ne \emptyset\}$. Now, using the observation that $\left|\widetilde{\mathcal{P}}\right| \le r + 1$, we have

$$\begin{aligned}
\left|\Delta_h^r(g, x)\right|^\sigma &= \left|\Delta_h^r(\widetilde{g}, x)\right|^\sigma = \left|\sum_{R \in \widetilde{\mathcal{P}}} \Delta_h^r(a_R, x)\right|^\sigma \\
&\le C(r, \sigma) \sum_{R \in \widetilde{\mathcal{P}}} \left|\Delta_h^r(a_R, x)\right|^\sigma \le C \sum_{R \in \mathcal{P}} \left|\Delta_h^r(a_R, x)\right|^\sigma.
\end{aligned}$$

Therefore, integrating the last inequality over $[0,\ 1)^d$, taking the supremum over h: $|h| \le t$, and using the inequality $\sup_h \left(\sum f_i\right) \le \sum \sup_h f_i$, we obtain (7.3).

We now fix $R \in \mathcal{P}$ and $h \in \mathbf{R}^d$, and denote

$$D(R, r, h) := \left\{x \in [0,\ 1)^d \mid \{x, \dots, x + rh\} \cap R \ne \emptyset,\ \{x, \dots, x + rh\} \cap R^C \ne \emptyset\right\},$$

where R^C denotes the complement of R ($R^C = [0,\ 1)^d \setminus R$). Then

$$\operatorname{vol}(D(R, r, h)) \le C \min\{\operatorname{vol}(R), \operatorname{vol}(\partial R)|h|\} \le C \min\{\operatorname{vol}(R), \operatorname{vol}(R)^{1-1/d}|h|\}.$$

Also,

$$\left|\Delta_h^r(a_R, x)\right| \le 2^r \begin{cases} |a_R(x)| + \cdots + |a_R(x + rh)|, & x \in D(R, r, h), \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned}
\omega_r(a_R, t)_\sigma^\sigma &= \sup_{|h| \le t} \int_{[0,\ 1)^d} \left|\Delta_h^r(a_R, x)\right|^\sigma\, dx \le \sup_{|h| \le t} \int_{D(R,r,h)} \left|\Delta_h^r(a_R, x)\right|^\sigma\, dx \\
&\le 2^{r\sigma}(r + 1)^\sigma \sup_{|h| \le t} \int_{D(R,r,h)} \|a_R\|_{L_\infty}^\sigma\, dx \\
&\le C\|a_R\|_{L_\infty}^\sigma \sup_{|h| \le t} \operatorname{vol}(D(R, r, h)) \\
&\le C\|a_R\|_{L_\infty(R)}^\sigma \min\{\operatorname{vol}(R), \operatorname{vol}(R)^{1-1/d} t\}.
\end{aligned}$$

Now, $\operatorname{vol}(R) \sim \operatorname{vol}(R_\boxplus)$, and, hence, for any polynomial p_r of total degree $< r$, $\|p_r\|_{L_p(R)} \sim \|p_r\|_{L_p(R_\boxplus)}$ and

$$\|p_r\|_{L_\infty(R)} \le \|p_r\|_{L_\infty(R_\boxplus)} \le C(\operatorname{vol}(R_\boxplus))^{-1/p}\|p_r\|_{L_p(R_\boxplus)} \le C(\operatorname{vol}(R))^{-1/p}\|p_r\|_{L_p(R)}$$

(see Ditzian [20], for example). Thus,

$$\omega_r(a_R, t)_\sigma^\sigma \le C(\mathrm{vol}(R))^{-\sigma/p} \min\left\{\mathrm{vol}(R), \mathrm{vol}(R)^{1-1/d} t\right\} \|a_R\|_{L_p(R)}^\sigma,$$

and using (7.3) and the fact that $\alpha\sigma < 1$ (this inequality is equivalent to $\alpha < d/(d-1)p$) we have

$$\begin{aligned}
|g|_{B^\alpha}^\sigma &= \int_0^\infty t^{-\alpha\sigma-1}\omega_r(g,t)_\sigma^\sigma\,dt \le C\sum_{R\in\mathcal{P}}\int_0^\infty t^{-\alpha\sigma-1}\omega_r(a_R,t)_\sigma^\sigma\,dt \\
&\le C\sum_{R\in\mathcal{P}}\int_0^\infty t^{-\alpha\sigma-1}(\mathrm{vol}(R))^{-\sigma/p}\min\left\{\mathrm{vol}(R), \mathrm{vol}(R)^{1-1/d}t\right\}\|a_R\|_{L_p(R)}^\sigma\,dt \\
&\le C\sum_{R\in\mathcal{P}}(\mathrm{vol}(R))^{-\sigma/p}\|a_R\|_{L_p(R)}^\sigma \\
&\quad\times\left(\mathrm{vol}(R)^{1-1/d}\int_0^{\mathrm{vol}(R)^{1/d}} t^{-\alpha\sigma}\,dt + \mathrm{vol}(R)\int_{\mathrm{vol}(R)^{1/d}}^\infty t^{-\alpha\sigma-1}\,dt\right) \\
&\le C\sum_{R\in\mathcal{P}}(\mathrm{vol}(R))^{1-\sigma/p-\alpha\sigma/d}\|a_R\|_{L_p(R)}^\sigma = C\sum_{R\in\mathcal{P}}\|a_R\|_{L_p(R)}^\sigma.
\end{aligned}$$

Finally, since $\sigma < p$, and using Hölder's inequality, we have

$$\begin{aligned}
|g|_{B^\alpha} &\le C\left(\sum_{R\in\mathcal{P}}\|a_R\|_{L_p(R)}^\sigma\right)^{1/\sigma} \\
&\le C\,|\mathcal{P}|^{1/\sigma-1/p}\left(\sum_{R\in\mathcal{P}}\|a_R\|_{L_p(R)}^p\right)^{1/p} \le Cn^{\alpha/d}\|g\|_{L_p}. \qquad \blacksquare
\end{aligned}$$

8. Wavelet Decompositions

Let $\phi \in W_\infty^s(\mathbf{R}^d)$ be a compactly supported function which has the following properties:

- ϕ is refinable, i.e.,

(8.1) $$\phi(x) = \sum_{j\in\mathbf{Z}^d} c_j\phi(2x-j).$$

- ϕ satisfies the Strang–Fix conditions of order r:

(8.2) $$\begin{aligned}
\hat\phi(0) &= 1, & \hat\phi(2\pi j) &= 0, & j &\in \mathbf{Z}^d, & j &\ne 0, \\
D^\nu\hat\phi(2\pi j) &= 0, & j &\in \mathbf{Z}^d, & j &\ne 0, & |\nu| &< r.
\end{aligned}$$

- The shifts of ϕ are locally linearly independent, i.e.,

(8.3) $\forall\, Q \in \mathbf{D}$ the functions $\phi(\cdot - j)$,
$j \in \Lambda_Q = \left\{j \in \mathbf{Z}^d \mid \mathrm{meas}(\{x|\phi(x-j)\ne 0\}\cap Q) > 0\right\}$,
are linearly independent over Q.

Define

$$\widetilde{\Sigma}_n(\phi) = \left\{ f \mid f = \sum_{I \in \mathbf{D}} a_I \phi_I \text{ such that } |\{a_I \mid a_I \neq 0\}| \leq n \right\},$$

where we use the usual indexing of the translates of dyadic dilates of ϕ by dyadic cubes, i.e., $\phi_I(x) := \phi(2^k x - j)$, where $I = j2^{-k} + 2^{-k}[0, 1)^d$.

Theorem F (DeVore, Jawerth, and Popov [11]). *If $\phi \in W^s_\infty(\mathbf{R}^d)$ is a compactly supported function satisfying* (8.1)–(8.3), *then, for each $0 < \alpha < \min\{r, s\}$ and $f \in B^\alpha$,*

$$E(f, \widetilde{\Sigma}_n(\phi))_{L_p} = \inf_{g \in \widetilde{\Sigma}_n(\phi)} \|f - g\|_{L_p} \leq C n^{-\alpha/d} |f|_{B^\alpha}, \qquad 0 < p < \infty.$$

Theorem G (Jia [21]). *Let ϕ be a compactly supported function in $W^s_\infty(\mathbf{R}^d)$, $s \in \mathbf{N}$. Suppose that ϕ is refinable, and its shifts are locally linearly independent. Then, for $0 < \alpha < s$ and $0 < p \leq \infty$:*

$$|f|_{B^\alpha} \leq C n^{\alpha/d} \|f\|_{L_p} \qquad \text{for all} \quad f \in \widetilde{\Sigma}_n(\phi),$$

where C is a constant depending only on ϕ and p when p is small.

We remark that the above result was proved in [21] for finitely generated shift-invariant spaces. The following result now follows from Theorems G, F, and A:

Theorem H. *Let $\phi \in W^s_\infty(\mathbf{R}^d)$, $s \in \mathbf{N}$, be a compactly supported function satisfying* (8.1)–(8.3). *Also, let $0 < p < \infty$ and $0 < \alpha < \min\{r, s\}$. Then, for $0 < \gamma < \alpha/d$ and $0 < q \leq \infty$, we have*

$$A^\gamma_q(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n \in \mathbf{N}}) = \left(L_p, B^\alpha\right)_{d\gamma/\alpha, q}.$$

Thus, in particular, for $q = \tau = (\gamma + 1/p)^{-1}$, we have

$$A^\gamma_\tau(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n \in \mathbf{N}}) = \left(L_p, B^\alpha\right)_{d\gamma/\alpha, \tau} = B^{d\gamma}.$$

From Corollaries 8, 16, and Theorem H we obtain the following:

Corollary 18. *Let $\phi \in W^s_\infty(\mathbf{R}^d)$, $s \in \mathbf{N}$, be a compactly supported function satisfying* (8.1)–(8.3) (*i.e., ϕ is a refinable function having locally linearly independent shifts and satisfying the Strang–Fix conditions of order r*). *Also, let $0 < \sigma < p < \infty$, $\beta = 1/\sigma - 1/p$, and $\beta < \min\{r/d, s/d, 1/(d-1)p\}$. Then, for $0 < \gamma < \beta$ and $0 < q \leq \infty$:*

$$A^\gamma_q(L_p, \{\Sigma_{n,r}\}_{n \in \mathbf{N}}) = A^\gamma_q(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n \in \mathbf{N}}) = \left(L_p, B^{d\beta}\right)_{\gamma/\beta, q} = \left(L_p, V^r_{\sigma,p}\right)_{\gamma/\beta, q}.$$

In particular, for $q = \tau = (\gamma + 1/p)^{-1}$, we have

$$A^\gamma_\tau(L_p, \{\Sigma_{n,r}\}_{n \in \mathbf{N}}) = A^\gamma_\tau(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n \in \mathbf{N}}) = \left(L_p, B^{d\beta}\right)_{\gamma/\beta, \tau} = \left(L_p, V^r_{\sigma,p}\right)_{\gamma/\beta, \tau} = B^{d\gamma}.$$

We now restate this corollary in somewhat more standard and easier-to-use form (we choose $\alpha = d\beta$ and $\lambda = d\gamma$).

Corollary 18′. *Let $\phi \in W^s_\infty(\mathbf{R}^d)$, $s \in \mathbf{N}$, be a compactly supported function satisfying* (8.1)–(8.3) (*i.e., ϕ is a refinable function having locally linearly independent shifts and satisfying the Strang–Fix conditions of order r*). *Also, let $0 < \sigma < p < \infty$, $1/\sigma = \alpha/d + 1/p$, and $\alpha < \min\{r, s, d/(d-1)p\}$. Then, for $0 < \lambda < \alpha$ and $0 < q \le \infty$,*

$$A_q^{\lambda/d}(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = A_q^{\lambda/d}(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}) = \left(L_p, B^\alpha\right)_{\lambda/\alpha,q} = \left(L_p, V^r_{\sigma,p}\right)_{\lambda/\alpha,q}.$$

In particular, for $q = \tau = (\lambda/d + 1/p)^{-1}$, we have

$$A_\tau^{\lambda/d}(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}) = A_\tau^{\lambda/d}(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}) = \left(L_p, V^r_{\sigma,p}\right)_{\lambda/\alpha,\tau} = \left(L_p, B^\alpha\right)_{\lambda/\alpha,\tau} = B^\lambda.$$

In other words,

$$\begin{aligned} f \in B^\lambda \quad &\Leftrightarrow \quad \sum_{n=1}^{\infty}\left[n^{\lambda/d}E(f, \Sigma_{n,r})_{L_p}\right]^\tau \frac{1}{n} < \infty, \\ &\Leftrightarrow \quad \sum_{n=1}^{\infty}\left[n^{\lambda/d}E(f, \widetilde{\Sigma}_n(\phi))_{L_p}\right]^\tau \frac{1}{n} < \infty, \\ &\Leftrightarrow \quad \int_0^\infty \left[t^{-\lambda/\alpha}K(f, t, L_p, V^r_{\sigma,p})\right]^\tau \frac{dt}{t} < \infty, \\ &\Leftrightarrow \quad \int_0^\infty \left[t^{-\lambda/d}\mathcal{W}_r(f, t)_{\sigma,p}\right]^\tau \frac{dt}{t} < \infty. \end{aligned}$$

The inequality $\alpha < d/(d-1)p$ is sharp and cannot be removed. The reason for this is that the spaces $(L_p, V^r_{\sigma,p})_{\lambda/\alpha,\tau}$ contain certain piecewise polynomials (though not all of them) if $\alpha > d/(d-1)p$, and, at the same time, for these α, the spaces B^α contain no characteristic functions (see the next section for more details). We also note that this somewhat anomalous situation does not exist in the univariate case (see also [15] for discussions).

8.1. *Remarks*

Lemma 19. *Let $r \in \mathbf{N}$, $0 < \alpha < r$, and $1/\sigma = \alpha/d + 1/p$. Then, for $f \in B^\alpha$,*

$$|f|_{V^r_{\sigma,p}} \le C|f|_{B^\alpha}. \tag{8.4}$$

Proof. For any partition $\{R_i\}_{i\in\Lambda} \in \pi_r$ using Lemma 1, Corollary 2, and (7.2) we have

$$\sum_{i\in\Lambda}\omega_r(f, \operatorname{len}(R_i), R_i)_p^\sigma \le C\sum_{i\in\Lambda}E_f(R_i)_p^\sigma \le C\sum_{i\in\Lambda}\sum_{\nu=1}^{\mu^{R_i}}|f|^\sigma_{B^\alpha(r_\nu^{R_i})} \le C|f|^\sigma_{B^\alpha}.$$

Therefore, taking sup over all partitions in π_r, we obtain (8.4). ∎

The inequality (8.4) immediately implies that

$$K(f, t, L_p, V^r_{\sigma,p}) \le CK(f, t, L_p, B^\alpha),$$

hence,

$$\left(L_p, B^\alpha\right)_{\theta,q} \subseteq \left(L_p, V^r_{\sigma,p}\right)_{\theta,q}$$

for $0 < \theta < 1$ and $0 \le q \le \infty$. In particular, for $0 < \lambda < \alpha$ and $q = \tau = (\lambda/d + 1/p)^{-1}$,

$$\left(L_p, B^\alpha\right)_{\lambda/\alpha,\tau} = B^\lambda \subseteq \left(L_p, V^r_{\sigma,p}\right)_{\lambda/\alpha,\tau}.$$

The proof of the following lemma is rather straightforward and will be omitted.

Lemma I. *Let* $Q \subseteq [0, 1)^d$ *be a cube, and let* $a_Q(x) := p_r(x)\chi_Q(x)$ (*i.e.*, a_Q *is a piecewise polynomial function on* $[0, 1)^d$). *Then*

$$a_Q \in B^\alpha_q(L_\tau) \quad \Leftrightarrow \quad \begin{cases} \alpha\tau < 1, & \text{if } 0 < q < \infty, \\ \alpha\tau \le 1, & \text{if } q = \infty. \end{cases}$$

In particular, for $0 < p < \infty$,

$$a_Q \in B^\alpha \quad \Leftrightarrow \quad \alpha < \frac{d}{(d-1)p}.$$

Obviously, the space $V^r_{\sigma,p}$ contains some piecewise polynomial functions (e.g., $\chi_{[0,\,1/2)^d}(x)$). In fact, it contains all piecewise polynomial functions on dyadic ring partitions.

Thus, it follows from Lemmas 19 and I that, if $1/\sigma = \alpha/d + 1/p$ and $\alpha \ge d/(d-1)p$, then

$$B^\alpha \subsetneq V^r_{\sigma,p}.$$

Also, if $\alpha > d/(d-1)p$, $1/\sigma = \alpha/d + 1/p$, $d/(d-1)p \le \lambda < \alpha$, and $\tau = (\lambda/d + 1/p)^{-1}$, then

$$A^{\lambda/d}_\tau(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}) = B^\lambda \subsetneq \left(L_p, V^r_{\sigma,p}\right)_{\lambda/\alpha,\tau} = A^{\lambda/d}_\tau(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}). \tag{8.5}$$

(This, in particular, shows that the condition $\alpha < d/(d-1)p$ in Corollary 18(18′) is sharp and cannot be removed.)

Thus, on the one hand, we know that

$$A^{\lambda/d}_\tau(L_p, \{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}) \subseteq A^{\lambda/d}_\tau(L_p, \{\Sigma_{n,r}\}_{n\in\mathbf{N}}),$$

which means that the error of approximation from the manifold $\{\Sigma_{n,r}\}_{n\in\mathbf{N}}$ is not worse than the error of approximation from $\{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}$. On the other hand, the error of approximation by elements of $\{\Sigma_{n,r}\}_{n\in\mathbf{N}}$ can be essentially better (smaller) than the error of approximation by elements of $\{\widetilde{\Sigma}_n(\phi)\}_{n\in\mathbf{N}}$ since, in the case $q = \tau$, there exist functions (e.g., characteristic functions of dyadic rings and their liner combinations) such that

$$\sum_{n=1}^{\infty}[n^{\lambda/d}E(f, \widetilde{\Sigma}_n(\phi))_{L_p}]^\tau \frac{1}{n} = \infty \qquad \text{and} \qquad \sum_{n=1}^{\infty}[n^{\lambda/d}E(f, \Sigma_{n,r})_{L_p}]^\tau \frac{1}{n} < \infty.$$

Acknowledgments. The authors are indebted to Ron DeVore for inspirational discussions, which motivated this work, during their visit to the University of South Carolina. We are also grateful to him and to Albert Cohen for their input on the first draft of this paper, and for providing us with copies of the manuscript [7]. Supported by NSF grant DMS 9705638.

References

1. C. Bennett, R. Sharpley (1988): Interpolation of Operators. New York: Academic Press.
2. J. Bergh, J. Peetre (1974): *On the space* V_p $(0 < p \leq \infty)$. Boll. Un. Mat. Ital., **4**:632–648.
3. M. S. Birman, M. Z. Solomjak (1967): *Piecewise polynomial approximations of functions of the classes* W_p^α. Math. USSR–Sb., **2**:295–317.
4. Yu. Brudnyi (1974): *Spline approximation and functions of bounded variation*. Dokl. Akad. Nauk SSSR, **215**:518–521.
5. Yu. A. Brudnyi (1970): *A multidimensional analog of a theorem of Whitney*. Mat. Sb., **82(124)**:157–170.
6. H. G. Burchard, D. F. Hale (1975): *Piecewise polynomial approximation on optimal meshes*. J. Approx. Theory, **14**:128–147.
7. A. Cohen, R. DeVore, P. Petrushev, H. Xu (to appear): *Nonlinear approximation and the space* $BV(\mathbf{R}^2)$. Amer. J. Math.
8. R. A. DeVore (1987): *A note on adaptive approximation*. Approx. Theory Appl., **3:4**:74–78.
9. R. A. DeVore (1989): *Degree of nonlinear approximation*. In: Approx. Theory VI (C. K. Chui, L. L. Schumaker, J. D.Ward, eds.), pp. 175–201.
10. R. A. DeVore, R. Howard, C. A. Micchelli (1989): *Optimal nonlinear approximation*. Manuscripta Math., **63**:469–478.
11. R. A. DeVore, B. Jawerth, V. A. Popov (1992): *Compression of wavelet decompositions*. Amer. J. Math., **114**:737–785.
12. R. A. DeVore, G. Kyriazis, D. Leviatan, V. M. Tikhomirov (1993): *Wavelet compression and nonlinear n-widths*. Adv. in Comput. Math., **1**:197–214.
13. R. A. DeVore, G. G. Lorentz (1993): Constructive Approximation. Berlin: Springer-Verlag.
14. R. A. DeVore, P. Petrushev, X. M. Yu (1992): *Wavelet approximation in the space* $C(\mathbf{R}^d)$. In: Progress in Approximatin Theory (A. A. Gonchar, E. Saff, eds.). New York: Springer-Verlag, pp. 261–283.
15. R. A. DeVore, V. A. Popov (1987): *Free multivariate splines*. Constr. Approx., **3**:239–248.
16. R. A. DeVore, V. A. Popov (1988): *Interpolation of Besov spaces*. Trans. Amer. Math. Soc., **305**:397–414.
17. R. A. DeVore, V. A. Popov (1988): *Interpolation spaces and nonlinear approximation*. In: Functions Spaces and Approximation (M. Cwikel, J. Peetre, Y. Sagher, H. Wallin, eds.), Lecture Notes in Mathematics, New York: Springer-Verlag, pp. 191–205.
18. R. A. DeVore, R. C. Sharpley (1993): *Besov spaces on domains in* $\mathbf{R}^d$. Trans. Amer. Math. Soc., **335(2)**:843–864.
19. R. A. DeVore, X. M. Yu (1991): *K-functional for Besov spaces*. J. Approx. Theory, **67**:38–50.
20. Z. Ditzian (1996): *Polynomial approximation in* $L_p(S)$ for $p > 0$. Constr. Approx., **12**:241–269.
21. R.-Q. Jia (1993): *A Bernstein-type inequality associated with wavelet decomposition*. Constr. Approx., **9**:299–318.
22. P. Oswald (1990): *On the degree of nonlinear spline approximation in Besov-Sobolev spaces*. J. Approx. Theory, **61**:131–157.
23. A. Pekarskii (1986): *Estimates of the derivatives of rational functions in* $L_p[-1, 1]$. Mat. Zametki, **39**:388–394 (English translation in Math. Notes, **39**:212–216).
24. P. Petrushev (1988): *Direct and converse theorems for spline and rational approximation and Besov spaces*. In: Functions Spaces and Approximation (M. Cwikel, J. Peetre, Y. Sagher, H. Wallin, eds.). Lecture Notes in Mathematics, New York: Springer-Verlag, pp. 363–377.

Y. K. Hu
Department of Mathematics and Computer Science
Georgia Southern University
Statesboro, GA 30360
USA

K. A. Kopotun
Department of Mathematics
Vanderbilt University
Nashville, TN 37240
USA

X. M. Yu
Department of Mathematics
Southwest Missouri State University
Springfield, MO 65804
USA