Restricted nonlinear approximation

A. Cohen, R. DeVore and R. Hochmut
1. Introduction.

Approximation by a linear combination of $n$ wavelets is a form of nonlinear approximation that occurs in several applications including image processing, statistical estimation, and the numerical solution of differential equations. In this paper, we shall consider variants of $n$ term approximation which we call restricted approximation. As explained further, we are motivated by certain applications in statistics and by the interpolation of Besov spaces.

To describe our results, we recall the usual setting of multivariate wavelet analysis. Let $\mathcal{D}$ be the set of dyadic cubes in $\mathbb{R}^d$ and for $k \in \mathbb{Z}$, we let $\mathcal{D}_k$ denote the set of those cubes $I \in \mathcal{D}$ at dyadic level $k$, i.e. $|I| = 2^{-kd}$, where we use $|K|$ to denote the Euclidean measure of a set $K \subset \mathbb{R}^d$. We denote by $\Omega := [0,1]^d$ the unit cube in $\mathbb{R}^d$. Each cube $I \in \mathcal{D}_k$ is of the form $I = 2^{-k}(j + \Omega)$ with $j \in \mathbb{Z}^d$. We identify $I$ with $(j,k)$. If $g$ is any function defined on $\mathbb{R}^d$, we define

$$g_{I,p}(x) := 2^{kd/p}g(2^{k}x - j).$$

In the case $g \in L_p$, then $\|g_{I,p}\|_{L_p} = \|g\|_{L_p}$. Here and throughout this paper all function spaces and all norms are taken over $\mathbb{R}^d$ unless explicitly stated otherwise. In order to streamline notation, we shall often simply write $g_I$ in place of $g_{I,p}$; however, it will always be clear from the text what is the value of $p$ in the normalization.

Wavelet theory generates a set $\Psi \subset L_2$ of $2^d - 1$ functions whose shifted dilates form a Riesz basis for $L_2$ as follows. We begin with univariate scaling function $\phi$ and an associated univariate wavelet function $\psi$ and define $\psi^0 := \phi$ and $\psi^1 := \psi$. Let $E$ denote the set of nonzero vertices of $\Omega$ and define

$$(1.1) \quad \psi^e(x_1, \ldots, x_d) := \prod_{i=1}^{d} \psi^{e_i}(x_i), \quad e \in E.$$ 

Then, $\Psi := \{\psi^e : e \in E\}$ is such a set.

We shall restrict ourselves in this paper to the case of compactly supported biorthogonal wavelets. This means that the family of functions $\psi^e$, $e \in E$, are assumed to have been generated by a compactly scaling function $\phi$ with a dual scaling function $\tilde{\phi}$ which also has compact support. The wavelet function $\psi$ also has compact support and has associated to it a compactly supported dual wavelet $\tilde{\psi}$ (see [CDF] or [Da, Chapter 8] for the definition and properties of biorthogonal wavelets). We remark that all of our theorems hold in more generality. In particular compact support can be replaced by suitable decay conditions. However, by imposing these

---

1This research was supported by the Office of Naval Research Contract N0014-91-J1343 and the Deutsche Forschungsgemeinschaft under grant Ho 1846/1-1. This research was completed while the third author was visiting the University of South Carolina.
additional assumptions, our development will be more simple and hopefully more clear.

The set of functions given in (1.1) generates by shifts and dyadic dilates a Riesz basis for \( L_2 \). This means that each function \( f \in L_2 \) has the unique expansion

\[
f = \sum_{I \in \mathcal{D}} A_I(f), \quad A_I(f) := \sum_{e \in E} a^e_I(f) \psi^e_I,
\]

with the wavelet functions \( \psi^e_I = \psi^e_{I,2} \) normalized in \( L_2(\mathbb{R}^d) \). Moreover, we have

\[
\|f\|_{L_2(\mathbb{R}^d)}^2 \cong \sum_{I \in \mathcal{D}} a_I(f)^2, \quad a_I(f) := \left( \sum_{e \in E} |a^e_I(f)|^2 \right)^{1/2}.
\]

The set of functions \( \{\psi^e_I\}_{I \in \mathcal{D}, e \in E} \) is also an unconditional basis for \( L_p(\mathbb{R}^d) \), \( 1 < p < \infty \), and for many other function spaces such as the Hardy spaces and the Besov spaces. We shall discuss this in more detail in the following section. For now, we want to turn to the formulation of the nonlinear approximation problem that we shall study in this paper.

Let \(-\infty < \alpha < 1\) and define for each set \( \Lambda \subset \mathcal{D} \),

\[
\Phi(\Lambda) := \Phi_{\alpha}(\Lambda) := \sum_{I \in \Lambda} |I|^\alpha.
\]

Thus, \( \Phi \) is a measure defined on the subsets of the discrete space \( \mathcal{D} \). For each \( t > 0 \), we define the space \( \Sigma_t \) as the set of all

\[
S = \sum_{I \in \Lambda} A_I(S), \quad \Phi(\Lambda) \leq t.
\]

Since the set \( \Lambda \) is possibly infinite, some sense of convergence must be attached to the series in (1.4). We postpone a discussion of this until §2 when we formulate the restricted approximation problem in more detail. One should note in any case that \( \Sigma_t \) is not a linear space. For example, the sum of two elements from \( \Sigma_t \) is generally not in \( \Sigma_t \) although it is in \( \Sigma_{2t} \).

We shall consider approximation in the Hardy space \( H_p, 0 < p < \infty \), by the elements of \( \Sigma_t \). We recall that \( H_p = L_p, 1 < p < \infty \). Given \( f \) we define

\[
\sigma(f, t)_p := \sigma(f, t)_{H_p} := \inf_{S \in \Sigma_t} \|f - S\|_{H_p}.
\]

We remark that we do not necessarily assume that \( f \in H_p \) in the definition (1.5); however, this situation will only appear in our results when dealing with the case \( \alpha > 0 \). In this case, it can happen that (1.5) is finite even when \( f \) is not in \( H_p \). In the case \( \alpha = 0 \) and \( t = n \) is a positive integer, the space \( \Sigma_n \) consists of all functions \( S \) which are a linear combination of \( n \) wavelets. Then, (1.5) is the error of \( n \) term approximation in \( H_p \).

We shall be interested in this paper in describing the functions \( f \) for which \( \sigma(f, t)_p \) has a prescribed asymptotic behavior as \( t \to \infty \) and \( t \to 0 \). For \( 0 < p < \infty \),
0 < q ≤ ∞ and γ > 0, we define the approximation class \( A_q^\gamma(H_p) \) to be the set of all \( f \) such that

\[
|f|_{A_q^\gamma(H_p)} := \begin{cases} 
\left( \int_0^\infty [t^\gamma \sigma(f,t)_p]^q \, dt \right)^{1/q}, & 0 < q < \infty \\
\sup_{t > 0} t^\gamma \sigma(f,t)_p, & q = \infty.
\end{cases}
\]

From the monotonicity of \( \sigma(f,t)_p \), it follow that (1.6) is equivalent to a discrete norm

\[
|f|_{A_q^\gamma(H_p)} \leq \begin{cases} 
\left( \sum_{j \in \mathbb{Z}} [2^j \gamma \sigma(f,2^j)_p]^q \right)^{1/q}, & 0 < q < \infty, \\
\sup_{j \in \mathbb{Z}} 2^j \gamma \sigma(f,2^j)_p, & q = \infty.
\end{cases}
\]

In the case \( \alpha \leq 0 \), one can actually restrict \( t \) in (1.6) to be \( \geq 1 \) \( (j \) in (1.7) to be \( \geq 0 \)) without changing the space \( A_q^\gamma(H_p) \). However, in order to treat all cases of \( \alpha \) simultaneously we need the full range of \( t \geq 0 \).

Our main results characterize the spaces \( A_q^\gamma(H_p) \) in several ways: in terms of interpolation spaces; in terms of wavelet coefficients; and in terms of smoothness spaces (Besov spaces). Consider for example the case \( 1 < p < \infty \) and \( \alpha \leq 0 \) and let \( \beta := 1 - \alpha \) so that \( \beta \geq 1 \). For \( s > 0 \), let \( B^s_q(L_\tau) \) denote the Besov space of smoothness order \( s \) in \( L_\tau \) and auxiliary parameter \( q \) (a fuller discussion of Besov spaces is given in §2). For spaces \( X,Y \) we also denote by \( (X, Y)_{\alpha, q} \) the interpolation spaces generated by the real method of interpolation \( (K\)-functional) with parameters \( 0 < \theta < 1, 0 < q \leq \infty \) (see §2). We show that for each \( 1 < p < \infty \) and \( q > 0 \), we have

\[
A_q^\gamma(L_p) = (L_p, B^s_q(L_\tau))_{\gamma/s, q} \quad 0 < \gamma < s, \; 0 < q \leq \infty,
\]

for a certain range of \( s \) which depends on the wavelets \( \psi_0, \tilde{\psi} \) and with \( \tau \) defined by \( s = \beta d(1/\tau - 1/p) \). It is well known that for each such \( \gamma \) and for \( q \) defined by \( \gamma = \beta d(1/q - 1/p) \), the interpolation space on the right side of (1.8) is the Besov space \( B^s_q(L_\tau) \). This has a simple geometrical description given in Figure 1. In this figure, the \( x \)-axis corresponds to \( L_q \) spaces with \( x \) identified with \( 1/q \). The \( y \)-axis corresponds to the smoothness order. Thus, the point \( (1/q, \gamma) \) corresponds to the smoothness space \( B^s_q(L_\tau) \). Then, (1.8) says that the approximation space \( A_q^\gamma(L_p) \) corresponds to the point \( (1/q, \gamma) \) on the line segment connecting \( (1/p, 0) \) (corresponding to \( L_p \)) to \( (1/\tau, s) \) (corresponding to \( B^s_q(L_\tau) \)). This line segment has slope \( \beta d \).

Our results also serve to prove theorems about the interpolation of Besov spaces on the line with slope \( \beta d \) in Figure 1. While these interpolation theorems are known, wavelet methods provide simple proofs and also allow ways to realize the \( K\)-functional between \( H_p \) and one of these Besov spaces.

Another way to describe the space \( A_q^\gamma(H_p) \) is through thresholding and wavelet coefficients. It turns out that restricted approximation is intimately connected to thresholding coefficients in the \( L_r \) norm with \( r := p/\beta \). Let \( a_I(f) \) be defined as in (1.3) except we now take the wavelets normalized in \( L_r \). We can create a good approximation to \( f \) from \( \Lambda_I \) by a sum of the form

\[
S = \sum_{I \in \Lambda(\epsilon, f)} A_I(f)
\]
with $\Lambda(\epsilon, f) := \{I : a_I(f) > \epsilon\}$. The proper choice of $\epsilon$ gives an element of $\Sigma_t$. Using these ideas, we can characterize the approximation space $A_q(H_p)$ as the set of all $f$ for which the sequence $(a_I(f))_{I \in D}$ is in the weighted Lorentz space $\ell_{\mu,q}(w)$ with $\mu$ related to $\gamma$ by $\gamma = \beta d(1/\mu - 1/p)$, and $w(I) := |I|^{\alpha}$, $I \in D$.

The study of the $L^p$ error resulting from a thresholding of the wavelet expansion in $L^r$ norm with $r \neq p$ is motivated by problems of statistical estimation: in a white noise model, one is required to threshold the noisy signal in $L^2$, even when interested in minimizing the estimation error in $L^p$ for $p \neq 2$ (see [DJKP] for a general review of wavelet thresholding techniques for statistical estimation and [CDKP] for the application of our results in this context).

In order to prove our main results, we shall introduce new techniques for nonlinear wavelet approximation which apply even to the case of $n$-term approximation. These new proofs for $n$-term wavelet approximation are somewhat simpler than those given in [DJP].

An outline of this paper is as follows. In §2, we discuss wavelet characterizations of spaces and define the smoothness spaces (in terms of wavelet coefficients) which we shall use in the characterization of approximation order. In §3, we discuss certain fundamental relations between approximation spaces and interpolation spaces which we shall use in our characterization of approximation spaces. In particular, we discuss the role of Jackson and Bernstein inequalities in these matters. We also prove some general results on when approximation methods can realize the $K$-functional for a pair of spaces. In §4, we consider $n$-term wavelet approximation corresponding to the particular choice $\alpha = 0$. Most results of this section are already known, but the way of proof is somehow simpler than in the existing literature. In §5, we consider the general case of restricted nonlinear approximation.
as described above and prove the corresponding Jackson and Bernstein inequalities for this type of approximation. In §6, we characterize the approximation spaces for restricted nonlinear approximation as noted above. In §7, we show that restricted approximation can be achieved through simple thresholding procedure of the wavelet expansion. For the sake of simplicity, all our results are stated for spaces of functions defined on the whole of \( \mathbb{R}^d \), using the whole range of scales \( k \in \mathbb{Z} \) in the wavelet decomposition. In §8, we make some concluding remarks on the adaptation of our results to the approximation of functions defined on a bounded domain, using the scales \( k \geq 0 \) together with a layer of scaling functions at the coarsest resolution.

2. Wavelet decompositions and characterization of function spaces.

We shall describe in this section the properties of wavelet decompositions which we shall use in this paper. Let \( E \) be the non-zero vertices of \( \Omega \) as introduced earlier and let \( \psi^e_I, \ e \in D, \ I \in \mathcal{D} \), be the biorthogonal wavelet basis obtained from the compactly supported scaling function \( \phi \) and compactly supported univariate wavelet \( \psi \) as described in (1.1). This basis will be fixed throughout this paper. We denote by \( \tilde{\psi}^e_I \) the functions in the dual basis. If \( f \) is a tempered distribution, the wavelet coefficients

\[
(2.1) \quad a^e_{I,p}(f) := \langle f, \tilde{\psi}^e_{I,p} \rangle, \quad I \in \mathcal{D}, e \in E,
\]

with the dual wavelets \( \tilde{\psi}^e_I \) normalized in \( L_{p'} \), \( 1/p + 1/p' = 1 \), are defined whenever the order of \( f \) is sufficiently small compared to the smoothness of \( \hat{\phi}, \hat{\psi} \). For example, they are defined if \( \hat{\phi} \) and \( \hat{\psi} \) are in \( C^r \) with \( r \) exceeding the order of the distribution \( f \). Thus, for example, they are defined whenever \( f \in L_p \), \( 1 \leq p \leq \infty \) and whenever \( f \in H_p, 0 < p \leq 1 \), provided the dual wavelets are in \( C^r \) with \( r \geq \lfloor d(1/p - 1) \rfloor \).

We continue with the notation of the introduction and in particular define

\[
(2.2) \quad a_{I,p}(f) := \left( \sum_{e \in E} a^e_{I,p}(f)^2 \right)^{1/2}, \quad I \in E.
\]

We shall frequently use the following formula for changing between normalizations:

\[
(2.3) \quad |I|^{-1/p} a_{I,p}(f) = |I|^{-1/q} a_{I,q}(f), \quad |I|^{-1/p} a_{I,p}(f) = |I|^{-1/q} a_{I,q}(f),
\]

which holds for any \( 0 < p, q \leq \infty \).

It is well known (see [Da]) that \( (\psi^e_I)_{I \in \mathcal{D}, e \in E} \) is an unconditional basis for \( L_p \), \( 1 < p < \infty \). Each \( f \in L_p \) has a unique decomposition

\[
(2.4) \quad f = \sum_{I \in \mathcal{D}} A_I(f), \quad A_I(f) := \sum_{e \in E} a^e_I(f) \psi_I.
\]

We can compute \( L_p \) norms of functions \( f \) from their wavelet decompositions using the square function which is defined by

\[
(2.5) \quad S(f) := \left( \sum_{I \in \mathcal{D}} a_{I,2}(f)^2 |I|^{-1} \chi_I \right)^{1/2} = \left( \sum_{I \in \mathcal{D}} a_{I,p}(f)^2 |I|^{-2/p} \chi_I \right)^{1/2}.
\]
Namely, for $1 < p < \infty$,

\[(2.6a) \quad \|f\|_{L_p} \asymp \|S(f)\|_{L_p}.
\]

The equivalence (2.6a) follows from general results in Littlewood Paley theory (see [Me] or [FJ]).

When $p \leq 1$ the right side of (2.6a) gives the norm in the real Hardy space $H_p$ (see [FS] for the definition and properties of $H_p$) for a certain range of $p$ which depends on the univariate wavelets $\psi$, $\tilde{\psi}$. We shall say that $p \leq 1$ is admissible if

\[(2.6b) \quad \|f\|_{H_p} \asymp \|S(f)\|_{L_p}.
\]

Wavelet coefficients also can be used to characterize smoothness spaces. We shall use wavelet coefficients to define a class of spaces $B_{q,p}^s$ for $0 < q, p \leq \infty$, $s \geq 0$. For certain values of these parameters, these spaces will coincide with the Besov spaces as we shall explain. If $p$ is admissible then the space $B_{q,p}^s$ is defined as the set of all distributions in $H_p$ for which the following (quasi-semi-)norm is finite:

\[(2.7) \quad |f|_{B_{q,p}^s} := \|(2^{ks}\|a_{I,p}(f)\|_{L_p})_{I \in \mathcal{D}_k}\|_{\ell_p(\mathbb{D})}.
\]

There are many other forms for the right side of (2.7) obtained by using different normalization of the wavelets $\psi_I$ and the fact that $|I| = 2^{-kd}$ for $I \in \mathcal{D}_k$. For example, when $q = p$, we can rewrite (2.7) as

\[(2.8) \quad |f|_{B_{p,p}^s} := \|(|I|^{-s/d}a_{I,p}(f))_{I \in \mathcal{D}}\|_{\ell_p(\mathbb{D})}.
\]

We shall use the abbreviated notation $B_p^s := B_{p,p}^s$ throughout. The case $s = 0$ in (2.8) will be important in this paper. We shall denote $B_p^0$ simply as $B_p$. Thus,

\[(2.9) \quad |f|_{B_p} := \|(a_{I,p}(f))_{I \in \mathcal{D}}\|_{\ell_p(\mathbb{D})}.
\]

The space $B_p$ can be viewed as a substitute for $L_p$; it has a simpler structure in terms of its wavelet decomposition.

The spaces $B_{q,p}^s$ are the same as Besov spaces for a certain range of $s$ which depends on the smoothness of $\psi$ and the number of vanishing moments of $\tilde{\psi}$ as we shall now describe. Consider first the case $1 \leq p \leq \infty$ in which case the $B_{q,p}^s$ are related to the Besov spaces $B_q^s(L_p)$ defined by moduli of smoothness in $L_p$ (see [DJP]). Let $r(p)$ be a real number such that $\psi$ is in $B_r^p(p)(L_p)$ and all moment of $\tilde{\psi}$ of order $< r(p)$ vanish. Then, $B_{q,p}^s$ is the same as the Besov space $B_q^s(L_p)$ for all $0 < q \leq \infty$ and $0 < s < r(p)$. When $0 < p < 1$, we use the Besov spaces $B_q^s(H_p)$ which can be defined in several ways (Fourier transforms, Littlewood-Paley theory, or $H_p$ moduli of smoothness) as is thoroughly discussed in [K]. If $p$ is admissible and $r(p)$ is defined as before (with $H_p$ in place of $L_p$), then $B_{q,p}^s = B_q^s(H_p)$, for all $0 < s < r(p)$, and $0 < q \leq \infty$. Finally, it is known that $B_q^s(L_p) = B_q^s(H_p)$ whenever $s > d(1/p - 1)_+$ (see [K]), in which case these spaces are embedded in $L_1$.

In summary, the spaces $B_{q,p}^s$ are defined by the size properties of wavelet coefficients for the full range $s \geq 0$, whereas the Besov spaces are characterized by these wavelet coefficients for a smaller range of $s$. 

Approximation spaces and interpolation spaces are intimately connected; each can be characterized in terms of the other. In this section, we wish to recall some of these connections and add a little to this theory.

Let $X, Y$ be a pair of spaces which are embedded in some Hausdorff space $X$. Then, one can form the space $X + Y$ which consists of all functions $f$ which can be written as $h + g$ with $h \in X, g \in Y$. We define the norm on $X + Y$ by

$$\|f\|_{X+Y} := \inf_{f = h + g} \|h\|_X + \|g\|_Y.$$  

More generally, for any $t > 0$, we define the K-functional

$$K(f, t) := K(f, t; X, Y) := \inf_{f = h + g} \|h\|_X + t \|g\|_Y.$$

In this definition, we may also replace norms by semi-norms.

K-functionals have many uses. They were originally introduced as a means of generating interpolation spaces. We recall that if $0 < \theta < 1$ and $0 < q \leq \infty$ then the interpolation space $(X, Y)_{\theta, q}$ is defined as the set of all functions $f \in X + Y$ such that

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

is finite.

We next describe the usual vehicle for characterizing approximation spaces and connecting them to interpolation spaces as described in DeVore and Lorentz [DL §9 of Chapter 7]. We suppose that $X$ and $Y$ are as above and for each $t > 0$, $X_t$ is a (possibly nonlinear) subspace of $X + Y$. The usual setting for approximation takes $t = n, n = 1, 2, \ldots$ and $Y \subset X$ but the results are the same (and the proofs almost identical) in this more general setting. We let

$$\sigma(f, t)_X := \inf_{S \in \Sigma_t} \|f - g\|_X$$

measure the approximation error for this family and define the approximation spaces $A^q(X)$ as in (1.6) with $H_p$ replaced by $X$.

We assume in addition that $X_t \subset X_u$ if $t \leq u$ and that the nonlinearity of the family $X_t$ is controlled in the following sense: there exists a constant $a$ such that $X_t + X_u \subset X_{a(t+u)}$.

We can characterize the approximation spaces if for some $r > 0$, we can establish the following two inequalities:

**Jackson inequality:** $\sigma(f, t)_X \leq C t^{-r} \|f\|_Y, \quad f \in Y, \quad t > 0.$

**Bernstein inequality:** $|S|_Y \leq C t^r \|S\|_X, \quad S \in \Sigma_t, \quad t > 0.$

From the Jackson inequality, one can derive a comparison between $\sigma$ and $K$ as follows. Let $\epsilon > 0$ be arbitrary and let $g$ be such that the decomposition $f = f - g + g$ gives the K-functional to within $\epsilon$:

$$\|f - g\|_X + t^{-r} \|g\|_Y = K(f, t^{-r}) + \epsilon.$$
If $S$ is a best approximation to $g$ from $X_t$ (when best approximation is not known to exist then one adds another $\epsilon$ in the following derivation with the same end result), then

$$\sigma(f, t)_X \leq \|f - S\|_X \leq \|f - g\|_X + \|g - S\|_X \leq K(f, t^{-\gamma}) + \epsilon + Ct^{-\gamma}\|g\|_Y \leq CK(f, t^{-\gamma}) + \epsilon.$$  

Since $\epsilon$ is arbitrary, we have

$$\sigma(f, t)_X \leq CK(f, t^{-\gamma}).$$  

(3.3)

The Bernstein inequality provides a weak inverse inequality to (3.3) which we do not give (see Theorem 5.1 of Chapter 7 in [DL]). From this one derives the following relation between approximation spaces and interpolation spaces.

**Theorem 3.1.** If the Jackson and Bernstein inequalities are valid, then for each $0 < \gamma < r$ and $0 < q \leq \infty$ the following relation holds between approximation spaces and interpolation spaces

$$A^\gamma_q(X) = (X, Y)_{\gamma/r, q},$$  

with equivalent norms.

*Proof.* See Theorem 9.1 of Chapter 7 in [DL] where the Theorem is proved under the additional assumption that $Y$ is embedded in $X$: a simple modification of that proof gives (3.4) in the general case.

There is a further connection between approximation and interpolation. In certain cases, we can realize the K-functional by an approximation process. We continue with the above setting. We say a family $(A_t)$, $t > 0$, of (possibly nonlinear) operators, with $A_t$ mapping $X$ into $X_t$, provides near best approximation if there is an absolute constant $C > 0$ such that

$$\|f - A_t f\|_X \leq C\sigma(f, t)_X, \quad t > 0.$$  

(3.5)

We say this family is stable on $Y$ if

$$\|A_t f\|_Y \leq C\|f\|_Y, \quad t > 0,$$

with an absolute constant $C > 0$.

**Theorem 3.2.** Let $X$, $Y$, $X_t$ be as above and suppose that $X_t$ satisfies the Jackson and Bernstein inequalities. Suppose further that the family of operators $A_t$, $t > 0$, provides near best approximation and is stable on $Y$ then $A_t$ realizes the $K$-functional, in the sense that

$$\|f - A_t f\|_X + t^{-\gamma}\|A_t f\|_Y \leq CK(f, t^{-\gamma}, X, Y),$$  

(3.6)

with an absolute constant $C$.

*Proof.* We fix $t > 0$ and let $g \in Y$ be a function which realizes $K(f, t^{-\gamma})$, i.e

$$\|f - g\|_X + t^{-\gamma}\|g\|_Y \leq K(f, t^{-\gamma}).$$  

(3.7)
When \( g \) is not known to exist, we add an \( \epsilon > 0 \) as above. From the near best assumption, we have

\[
\|f - A_t f\|_X \leq C \sigma(f, t)_X \\
\leq C \|f - A_t g\|_X \\
\leq C (\|f - g\|_X + \|g - A_t g\|_X) \\
\leq C (\|f - g\|_X + \sigma(g, t)_X) \\
\leq C (\|f - g\|_X + t^{-r} \|g\|_Y) \\
\leq C K(f, t^{-r}),
\]

(3.8)

where we have used the Jackson inequality.

Moreover, using the \( Y \)-stability of \( A_t \) and the Bernstein inequality, we obtain

\[
t^{-r} \|A_t f\|_Y \leq Ct^{-r}(\|A_t f - A_t g\|_Y + \|A_t g\|_Y) \\
\leq C (\|A_t f - A_t g\|_X + t^{-r} \|g\|_Y) \\
\leq C (\|f - a_t f\|_X + \|f - g\|_X + t^{-r} \|g\|_Y) \\
\leq C (\|f - A_t f\|_X + K(f, t^{-r})).
\]

This combined with (3.8) shows that \( A_t f \) realizes the \( K \)-functional. \( \Box \)

**Remark 3.1.** If in place of near best approximation we assume only that

\[
\|f - A_t f\|_X \leq C \sigma(f, at)_X,
\]

with absolute constant \( a \leq 1 \), then (3.6) is still valid.

Indeed, the same proof gives (3.8) with \( t^{-r} \) replaced by \((at)^{-r}\) and the remark follows because

\[
K(f, (at)^{-r}) \leq a^{-r} K(f, t^{-r}).
\]

In practice, the stability of an approximation operator \( B_t \) mapping \( X \) to \( X_t \) is not always easy to check directly, but it can be derived from the stability of one particular approximation operator \( A_t \) combined with Jackson and Bernstein estimates, as shown by the following result. We say that \( A_t \) provides a Jackson inequality if

\[
\|g - A_t g\|_X \leq C t^{-r} \|g\|_Y
\]

holds for all \( g \in Y \) with an absolute constant \( C \).

**Theorem 3.3.** Let \( X, Y, X_t \) be as above and suppose that \( X_t \) satisfies the Jackson and Bernstein inequalities. Let \( A_t, B_t \) provide the Jackson inequality and suppose that \( A_t \) is stable on \( Y \). Then, \( B_t \) is also stable on \( Y \).

**Proof.** Let \( g \in Y \). Then,

\[
\|B_t g\|_Y \leq C(\|A_t g - B_t g\|_Y + \|A_t g\|_Y) \leq C(t^r \|A_t g - B_t g\|_X + \|g\|_Y) \\
\leq C(t^r \|g - A_t g\|_X + t^r \|g - B_t g\|_X + \|g\|_Y) \leq C \|g\|_Y. \Box
\]
4. $n$-term wavelet approximation in $H_p$, $0 < p < \infty$.

In this section, we shall treat the case $\alpha = 0$ in restricted nonlinear approximation. This case corresponds to the standard $n$-term wavelet approximation. While the results of this section are for the most part known (see [DJP]), we shall give new and simpler techniques for their proof. We shall later use these same ideas to obtain the corresponding theory for restricted nonlinear approximation. The main new ingredient here is the use of the interval $I(x)$ defined below for a set $\Lambda$ of intervals and a point $x \in \mathbb{R}^d$. The interval $I(x)$ can be used to replace the role of maximal functions used in the original proofs of Jackson and Bernstein inequalities for $n$-term approximation given in [DJP].

We take $\alpha = 0$ throughout this section. In this case, it is enough to consider approximation from $\Sigma_t$ only in the case $t = n$ with $n$ a natural number. We shall use the notation

$$\sigma_n(f)_p := \sigma(f, n)_p$$

in this section.

Let $\Lambda$ be any finite set of dyadic intervals. For each $x \in \bigcup_{I \in \Lambda} I$, we define $I(x)$ to be the smallest interval in $\Lambda$ which contains $x$. We use the notation for wavelet decompositions given in §3.1-2. We shall frequently make use of the following observation of Temlyakov [T2] which holds for any finite set $\Lambda \subset D$.

**Lemma 4.1.** Let $0 < p < \infty$ be admissible and let $\Lambda$ be a finite subset of $D$. If $f$ has the wavelet decomposition

$$f = \sum_{I \in \Lambda} A_I(f)$$

with $a_{I,p}(f) \leq M$, for all $I \in \Lambda$, then

$$\|f\|_{H_p} \leq C_1 M \# \Lambda^{1/p}$$

with $C_1$ depending only on $p$. Similarly, if $a_{I,p}(f) \geq M$, for all $I \in \Lambda$, then

$$\|f\|_{H_p} \geq C_2 M \# \Lambda^{1/p}$$

with $C_2$ depending only on $p$.

**Remark 4.1.** Recall that $H_p = L_p$ with equivalent norms when $p > 1$.

**Proof.** We let $a_I := a_{I,p}(f)$. Then, for (4.2a), we use the square function (2.5-6) to find

$$\|f\|_{H_p} \leq C \|S(f)\|_{L_p} = C \left( \sum_{I \in \Lambda} a_I^2 |I|^{-2/p} \chi_I \right)^{1/2} \|L_p\|_{L_p} \leq CM \left( \sum_{I \in \Lambda} |I|^{-2/p} \chi_I \right)^{1/2} \|L_p\|_{L_p} \leq CM \|\|I(x)|^{-1/p}\|_{L_p}$$

where $|I(x)|^{-1}$ is defined to be 0 when $x \notin \bigcup_{I \in \Lambda} I$. If $J \in \Lambda$ then the set $\tilde{J} := \{x : I(x) = J\}$ is a subset of $J$, and we have $\bigcup_{I \in \Lambda} I = \bigcup_{J \in \Lambda} \tilde{J}$. It follows that

$$\|f\|_{H_p}^p \leq CM^p \int_{\mathbb{R}^d} |I(x)|^{-1} dx \leq CM^p \sum_{J \in \Lambda} \int_{J} |J|^{-1} \leq CM^p \# \Lambda,$$
which proves (4.2a).

For the proof of (4.2b), we have

$$S(f, x) \geq M \left( \sum_{I \in \Lambda} |I|^{-2/p} \chi_I(x) \right)^{1/2} \geq M |I(x)|^{-1/p}.$$  

Also, $|I(x)|^{-1} \geq C \sum_{I \in \Lambda} |I|^{-1} \chi_I(x)$. Hence,

$$\|f\|_{H_p}^p \geq C \|S(f)\|_{L_p}^p \geq CM^p \int_{\mathbb{R}^d} \sum_{I \in \Lambda} |I|^{-1} \chi_I(x) = CM^p \# \Lambda. \quad \Box$$

As a consequence of the Lemma we will prove the following interesting theorem of Temlyakov[T2]. We fix an admissible value of $p$ with $0 < p < \infty$ and let

$$B_n f = \sum_{I \in \Lambda_b} \sum_{c \in E} b_I^c \psi_I^c, \quad \# \Lambda_b \leq n,$$

be a best $H_p$-approximation to $f$ from $\Sigma_n$ (the existence of best $m$-term approximations was proved in [T1]). We modify $B_n f$ by replacing $b_I^c$ by $a_I^c(f)$ to get $\tilde{B}_n f$ which is also in $\Sigma_n$. It follows that $S(f - \tilde{B}_n f) \leq S(f - B_n f)$ with $S$ the square function of (2.5). Therefore, from (2.6b) we obtain

$$\|f - \tilde{B}_n f\|_{H_p} \leq C \|f - B_n f\|_{H_p} \leq C \sigma_n(f)_p,$$

We also introduce the thresholding operator $\mathcal{T}_n f := \sum_{I \in \Lambda_t} A_I(f)$ where $\Lambda_t$ consists of the $n$ cubes $I$ for which $a_{I,p}(f)$ is largest (with ties handled in an arbitrary way).

**Theorem 4.1.** For any admissible $p$ with $0 < p < \infty$ and for all $n = 1, 2, \ldots$, $\mathcal{T}_n f$ is a near best approximation to $f$ from $\Sigma_n$, i.e.

$$\|f - \mathcal{T}_n f\|_{H_p} \leq C \sigma_n(f)_p.$$  

**Proof.** It is enough to estimate

$$\tilde{B}_n f - \mathcal{T}_n f = - \sum_{I \in \Lambda_t \setminus \Lambda_b} A_I(f) + \sum_{I \in \Lambda_b \setminus \Lambda_t} A_I(f) =: f_0 + f_1.$$  

Using the square function, we have

$$\|f_0\|_{H_p} \leq C \|f - \tilde{B}_n f\|_{H_p} \leq C \sigma_n(f)_p.$$  

If $M$ is the smallest of the values $a_{I,p}(f)$, $I \in \Lambda_t$, then for all $I \in \Lambda_b \setminus \Lambda_t$ we have $a_{I,p}(f) \leq M$. Hence, from Lemma 4.1, we have

$$\|f_1\|_{H_p} \leq CM \#(\Lambda_b \setminus \Lambda_t)^{1/p}.$$  

On the other hand, for all $I \in \Lambda_t \setminus \Lambda_b$ we have $a_{I,p}(f) \geq M$ and hence

$$\|f_0\|_{H_p} \geq CM \#(\Lambda_t \setminus \Lambda_b)^{1/p}.$$  

Since $\#(\Lambda_t \setminus \Lambda_b) = \#(\Lambda_b \setminus \Lambda_t)$, we have $\|f_1\|_{H_p} \leq C\|f_0\|_{H_p} \leq C \sigma_n(f)_p$ which completes the proof.  \( \Box \)
4.1. The Jackson inequality for \( n \)-term wavelet approximation.

Recall that for \( 0 < \tau < \infty \), a sequence \( (a_n) \) of real numbers is in the Lorentz space \( w\ell_{\tau} := \ell_{\tau,\infty} \) (called weak \( \ell_{\tau} \)) if

\[
\#\{n : |a_n| > \epsilon\} \leq M^\tau \epsilon^{-\tau},
\]

for all \( \epsilon > 0 \). The norm \( \|(a_n)\|_{w\ell_{\tau}} \) is the smallest value of \( M \) such that (4.3) holds. Also,

\[
\|(a_n)\|_{w\ell_{\tau}} \leq \|(a_n)\|_{\ell_{\tau}}.
\]

**Theorem 4.2.** Let \( p \) be admissible with \( 0 < p < \infty \), and \( s > 0 \), and let \( f \in H_p \) and \( a_I := a_I(f) := a_{I,p}(f) \), \( I \in \mathcal{D} \), be such that \( (a_I)_{I \in \mathcal{D}} \) is in \( w\ell_{\tau}, 1/\tau = s + 1/p \). Then, we have

\[
\sigma_n(f)_p \leq Cn^{-s}\|(a_I(f))\|_{w\ell_{\tau}},
\]

with the constant \( C \) depending only on \( p \) and \( s \).

**Proof.** We have

\[
\#\{I : a_I > \epsilon\} \leq M^\tau \epsilon^{-\tau},
\]

for all \( \epsilon > 0 \) with \( M := \|(a_I)\|_{w\ell_{\tau}} \). Let \( \Lambda_j := \{I : 2^{-j} < |a_I| \leq 2^{-j+1}\} \). Then, for each \( k = 1, 2, \ldots \), we have

\[
\sum_{j=-\infty}^{k} \#\Lambda_j \leq CM^\tau 2^{k\tau},
\]

with \( C \) depending only on \( \tau \).

Let \( S_j := \sum_{I \in \Lambda_j} A_I(f) \) and \( T_k := \sum_{j=-\infty}^{k} S_j \). Then \( T_k \in \Sigma_N \) with \( N = CM^\tau 2^{k\tau} \). We finish the proof in the case \( 1 \leq p \leq \infty \) (the case \( 0 < p < 1 \) is handled similarly but with \( \|\cdot\|_{H_p} \) used in place of \( \|\cdot\|_{H_p} \)). We have

\[
\|f - T_k\|_{H_p} \leq \sum_{j=k+1}^{\infty} \|S_j\|_{H_p},
\]

We fix \( j > k \) and estimate \( \|S_j\|_{H_p} \). Since \( a_I \leq 2^{-j+1} \) for all \( I \in \Lambda_j \), we have from Lemma 4.1 and (4.4),

\[
\|S_j\|_{H_p} \leq C2^{-j} \#\Lambda_j^{1/p} \leq CM^{\tau/p} 2^{j(\tau/p-1)}.
\]

We therefore conclude from (4.5) that

\[
\|f - T_k\|_{H_p} \leq CM^{\tau/p} \sum_{j=k+1}^{\infty} 2^{j(\tau/p-1)} \leq CM(M2^k)^{\tau/p-1},
\]

because \( \tau/p - 1 < 0 \). In otherwords, for \( N \approx M^\tau 2^{k\tau} \), we have

\[
\sigma_N(f)_p \leq CMN^{1/p-1/\tau} = CMN^{-s}.
\]

\( \square \)
Corollary 4.1. Let $p$ be admissible with $0 < p < \infty$, let $s > 0$ and let $f \in B^s_r$, $1/\tau = s/d + 1/p$, with $\tau$ admissible. Then,

$$\sigma_n(f)_p \leq C |f|_{B^s_r} n^{-s/d},$$

with $C$ depending only on $p$ and $s$.

Proof. We have $a_{I,\tau} = a_{I,p}|I|^{1/\tau-1/p} = a_{I,p}|I|^{s/d}$. Thus, from the definition (2.8) we find

$$|f|_{B^s_r} = \|(a_I)\|_{\ell_\tau} \geq \|(a_I)\|_{\omega\ell_\tau}.$$ 

Hence (4.6) follows from Theorem 4.2 with $s$ replaced by $s/d$.

Remark 4.1. As noted in §2, the space $B^s_r$ coincides with the Besov space $B^s_r(H_\tau)$ for a certain range of $s$ and this space coincides with $B^s_r(L_\tau)$ if $s > d/\tau - d$.

Remark 4.2. Theorem 4.2 also holds with $H_p$ replaced by $B_p$ with a simpler proof. This is proved for restricted nonlinear approximation in §5.4.

4.2. The Bernstein inequality for $n$-term wavelet approximation.

We shall next prove the Bernstein inequality which is the companion to (4.6).

Theorem 4.3. Let $p$ be admissible with $0 < p < \infty$, and let $s > 0$. If $f = \sum_{I \in \Lambda} a_I(f)$ with $\#\Lambda \leq n$, we have

$$\|f\|_{B^s_r} \leq C n^{s/d} \|f\|_{L_p},$$

with $1/\tau = s/d + 1/p$ whenever $\tau$ is admissible.

Proof. Case 1: $p \geq 2$. With $a_I := a_{I,p}(f)$, we have from (4.7)

$$|f|_{B^s_r} = \left(\sum_{I \in \Lambda} a_I^2\right)^{1/2} \leq n^{1/\tau-1/p} \left(\sum_{I \in \Lambda} |a_I|^p\right)^{1/p}.$$ 

On the other hand,

$$\|S(f)\|_{L_p} = \|(\sum_{I \in \Lambda} a_I^2 |I|^{-2/p} \chi_I)^{1/2}\|_{L_p} \geq \|(\sum_{I \in \Lambda} a_I^p |I|^{-1} \chi_I)^{1/p}\|_{L_p} = \left(\sum_{I \in \Lambda} a_I^p\right)^{1/p},$$

which in view of (2.6b) completes the proof in this case.

Case 2: $p \leq 2$. With $I(x)$ defined as the smallest interval in $\Lambda$ that contains $x$, we have

$$|f|_{B^s_r} = \int_{\mathbb{R}^d} \sum_{I \in \Lambda} |a_I|^2 |I|^{-1} \chi_I = \int_{\mathbb{R}^d} \sum_{I \in \Lambda} a_I^2 |I|^{-\tau/p} \chi_I |I|^{-1+\tau/p} \chi_I$$

$$\leq C \int_{\mathbb{R}^d} S(f, x)^\tau |I(x)|^{-1+\tau/p} \leq C \left(\int_{\mathbb{R}^d} S(f, x)^p\right)^{\tau/p} \left(\int_{\mathbb{R}^d} |I(x)|^{-1}\right)^{1-\tau/p}$$

$$\leq C n^{1-\tau/p} \|S(f)\|_{L_p} \leq C n^{1-\tau/p} \|f\|_{H_p},$$

where the second to last inequality follows as in the proof of Lemma 4.1. □
Remark 4.3. The Bernstein inequality of Theorem 4.3 also holds with \( H_p \) replaced by \( B_p \). The proof is simply Hölder’s inequality (as in the first line of the above proof).

4.3. Approximation spaces for \( n \)-term wavelet approximation.

In this section, we state without much elaboration the conclusions that can be drawn from the Jackson and Bernstein inequalities for \( n \)-term approximation via the characterization of approximation spaces. A similar development with more details is given in §6 for restricted nonlinear approximation (which includes the results of this section as a particular case).

Let \( p \) be admissible with \( 0 < p < \infty \). Let \( s > 0 \) and \( 1/\tau := s/d + 1/p \) with \( \tau \) admissible. We denote by \( K(f, t) \) the K-functional for the pair \( H_p, \, B^{s}_p \) with the semi-norm of \( B^{s}_p \) used in the definition of \( K \). It follows (see Theorem 3.1) from the Jackson and Bernstein inequalities that for any \( 0 < \gamma < s \) and any \( 0 < q \leq \infty \),

\[
\begin{align*}
A^{\gamma/d} \left( H_p \right) & = (H_p, B^{s}_\tau)_{\gamma/q, q}, \\
A^{\gamma/d} \left( B_p \right) & = (B_p, B^{s}_\tau)_{\gamma/q, q}.
\end{align*}
\]

(4.8)

The interpolation spaces on the right side of (4.8) are in fact identical and can be described in two ways. First of all they can be described by a condition on the wavelet coefficients. Namely, a function is in this space if and only if \((a_{\tau, f}(f))_{\tau \in \mathcal{D}}\) is in the Lorentz space \( \ell_{\mu, q} \) where \( 1/\mu := \gamma/d + 1/p \) and in fact, we have

\[
|f|_{A^{\gamma/d} \left( H_p \right)} \lesssim \| (a_{\tau, f}(f)) \|_{\ell_{\mu, q}}.
\]

(4.9)

Secondly, in the case that \( q = \mu \), then \( A^{\gamma/d} \left( H_p \right) = B^{\gamma}_\mu \) with equivalent norms. Thus, as noted before, for a certain range of \( \gamma \) these spaces are the Besov spaces \( B^{\gamma}_\mu \).

There is a further connection between \( n \) term approximation and interpolation that we wish to bring out. Let \( p, \, s, \) and \( \tau \) have the same meaning as above. We recall the thresholding operator \( \mathcal{T}_n \) of Theorem 4.1. It follows from Theorem 4.1 and Theorem 3.2 that for each \( n = 1, 2, \ldots \), we have

\[
K(f, n^{-s}, H_p, B^{s}_\tau) \asymp \| f - \mathcal{T}_n f \|_{H_p} + n^{-s} |\mathcal{T}_n f|_{B^{s}_\tau}.
\]

In other words, \( \mathcal{T}_n f \) realizes the K-functional at \( t = n^{-s} \).

5. Restricted approximation in \( H_p \).

For the remainder of this paper, we shall consider the general problem of restricted nonlinear approximation. Since we have already treated the case \( \alpha = 0 \) (the case of \( n \)-term approximation) in the previous section, for convenience, we shall exclude that case in the following development. We fix \( \alpha \) and let \( \Phi := \Phi_\alpha \) throughout this section.

We fix an admissible value of \( p \) with \( 0 < p < \infty \) and a value of \( s > 0 \) and let \( \tau \) be defined by the equation \( s = d\beta(1/\tau - 1/p) \) where \( \beta := 1 - \alpha \). We shall prove Jackson and Bernstein inequalities for restricted nonlinear approximation in \( H_p \) using for \( Y \) (as in §3) the space

\[
B^{s}_{\tau} := B^{s}_{\tau, \tau},
\]
in the case that \( \tau \) is admissible. This scale of spaces is depicted in Figure 1. They lie on the line with slope \( \beta d \) which passes through the point \((1/p,0)\) corresponding to the space \( H_p \).

For each \( t > 0 \), we define the space \( \Sigma_t \) as the set of all \( S \in H_p + B^s_\tau \) for which (1.4) holds. In particular, the wavelet coefficients of \( S \) are defined and (1.4) converges in the sense of \( H_p + H_\tau \).

If \( f \in H_p + B^s_\tau \), we define

\[
\sigma(f,t)_p := \inf_{S \in \Sigma_t} \| f - S \|_{H_p}.
\]

It will follow from the discussion in §5.1 that \( \sigma(f,t)_p \) is finite for each \( t > 0 \).

The case \( \beta = 1 \) is the usual case of nonlinear approximation. If \( \beta > 1 \), the restricted nonlinear approximation will follow the same lines as the usual nonlinear approximation since \( B^s_\tau \) is embedded in \( H_p \). However, in the case \( \beta < 1 \) (i.e. \( 0 < \alpha < 1 \)) several new ingredients appear. First of all the space \( B^s_\tau \) is not embedded in \( H_p \). This means that in the theory of K-functionals we need to consider the full range of \( t > 0 \) (not just \( 0 < t \leq 1 \)). Correspondingly, we need the full range of \( t \) in \( \sigma(f,t)_p \), not just \( t \geq 1 \).

As we have seen in §4, a near best \( n \)-term wavelet approximation in \( H_p \) can be obtained by thresholding the \( L_p \) normalized wavelet coefficients. We shall see in §7 that restricted approximation is intimately connected with thresholding the normalized wavelet coefficients \( a_{I,r}(f) \) with \( r := p/\beta \).

The development given below is similar to that in §4 except that we use \( \Phi := \Phi_\alpha \) to count the number of cubes and we use the different thresholding. Let \( \Lambda \subset D \) be a set of cubes for which \( \Phi(\Lambda) \) is finite. As earlier, we define \( I(x) \) as the smallest interval from \( \Lambda \) which contains \( x \). In the case \( 0 < \alpha < 1 \), for certain \( x \), there may not be a smallest \( I(x) \) since there may be cubes of arbitrary small measure in \( \Lambda \). However, it is easy to see that the set \( E \) of such \( x \) has measure zero. Indeed, if \( E_k := \Lambda \cap D_k \), then \( E \subset \bigcup_{k \geq m} \bigcup_{I \in E_k} I \) for each \( m > 0 \). Hence,

\[
|E| \leq \sum_{k \geq m} \sum_{I \in E_k} |I| = \sum_{k \geq m} \sum_{I \in E_k} |I|^\alpha |I|^{1-\alpha} \leq \Phi(\Lambda) \sum_{k \geq m} 2^{(\alpha-1)dk} \leq C \Phi(\Lambda) 2^{(\alpha-1)dm},
\]

and the right side tends to zero as \( m \to \infty \).

We shall use the following analogue of Lemma 4.1.

**Lemma 5.1.** Let \( p, s, \tau, r \) be as above. If \( f \in H_p + B^s_\tau \) has the wavelet decomposition

\[
f = \sum_{I \in \Lambda} A_I(f),
\]

with \( \Phi(\Lambda) \) finite. If \( a_{I,r}(f) \leq M \), for all \( I \in \Lambda \), then

\[
\| f \|_{H_p} \leq C_1 M \Phi(\Lambda)^{1/p}, \tag{5.2a}
\]

with \( C_1 \) depending only on \( p \). Similarly, if \( a_{I,r}(f) \geq M \), for all \( I \in \Lambda \), then

\[
\| f \|_{H_p} \geq C_2 M \Phi(\Lambda)^{1/p}, \tag{5.2b}
\]
with $C_2$ depending only on $p$.

Proof. We first note that the square function (2.5) satisfies

$$S(f, x)^2 = \sum_{i \in \Lambda} |I|^{-2/r} a_{I, r}(f)^2 \chi_I(x) \leq CM^2 |I(x)|^{-2/r}, \quad x \in \mathbb{R}^d,$$

where we define $|I(x)|^{-2/r} := 0$ if $x \notin \bigcup_{I \in \Lambda} I$. Hence,

$$\|f\|_{H_p}^p \leq C\|S(f)\|_{L_p}^p \leq C M^p \|I(x)|^{-1/r}\|_{L_p}^p \leq C M^p \sum_{I \in \Lambda} |I|^{1-p/r} = CM^p \Phi(\Lambda),$$

which is (5.2a).

For the proof of (5.2b), we have

$$S(f, x)^2 \geq M^2 \sum_{I \in \Lambda} |I|^{-2/r} \chi_I(x) \geq M^2 |I(x)|^{-2/r}.$$

Also, $|I(x)|^{-p/r} \geq C \sum_{I \in \Lambda} |I|^{-p/r} \chi_I(x)$. Hence,

$$\|f\|_{H_p}^p \geq C \|S(f)\|_{L_p}^p \geq C M^p \int_{\mathbb{R}^d} |I(x)|^{-p/r} dx \geq C M^p \sum_{I \in \Lambda} |I|^{1-p/r}$$

$$= CM^p \sum_{I \in \Lambda} |I|^\alpha = CM^p \Phi(\Lambda). \quad \square$$

5.1. A Jackson inequality for restricted nonlinear approximation.

We fix $f \in B^s_\tau$ and let $a_I := a_I(f) := a_{I, r}(f), I \in D$, and for $j \in \mathbb{Z}$ define

$$\Lambda_j := \Lambda_j(f) := \{I : 2^{-j} \leq a_{I, r}(f) < 2^{-j+1}\},$$

and the operators

$$S_j f := \sum_{I \in \Lambda_j(f)} A_I(f),$$

and

$$T_k f := \sum_{j = -\infty}^k S_j.$$

**Theorem 5.1.** Let $s > 0$, and $p$ and $\tau$ be admissible, $0 < \tau < p < \infty$, and satisfy $s = \beta d(1/\tau - 1/p)$. If $f \in B^s_\tau$, then for each $j, k \in \mathbb{Z}$, we have

(i) $\Phi(\Lambda_j) \leq C \|f\|_{B^s_\tau}^2 2^{j\tau},$

(ii) $\Phi(\cup_{j \leq k} \Lambda_j) \leq C \|f\|_{B^s_\tau} 2^{k\tau},$

(iii) $\|f - T_k f\|_{H_p} \leq C 2^{k(\tau/p - 1)} \|f\|_{B^s_\tau}^{\tau/p}$.

In addition, for each real number $t > 0$, we have

$$\sigma(f, t)_p \leq C \|f\|_{B^s_\tau} t^{-s/\beta d}. \quad (5.3)$$
Proof. The proof is similar to that in §4.1.

(i) Since $a_{I,r}(f) = |I|^{1/\tau - 1/r} a_{I,r}(f)$, the assumption $f \in B^s_\tau$ implies that

$$C|f|_{B^s_\tau} = \sum_{j \in \mathbb{Z}} \sum_{I \subseteq \Lambda_j} [|I|^{-s/d} a_{I,r}(f)]^\tau = \sum_{j \in \mathbb{Z}} \sum_{I \subseteq \Lambda_j} [|I|^{-s/d + 1/\tau - 1/r} a_{I,r}(f)]^\tau$$

$$= \sum_{I \subseteq \Lambda_j} [|I|^\beta(1/p-1/\tau) + 1/\tau - \beta/p a_{I,r}(f)]^\tau \geq C \sum_{j \in \mathbb{Z}} 2^{-j\tau} \sum_{I \subseteq \Lambda_j} |I|^\alpha$$

$$= C \sum_{j \in \mathbb{Z}} 2^{-j\tau} \Phi(\Lambda_j).$$

It follows therefore that

$$\Phi(\Lambda_j) \leq C|f|_{B^s_\tau} 2^{j\tau},$$

which is (i).

(ii) We obtain (ii) by summing the inequalities in (i).

(iii) From Lemma 5.1, we have

$$\|S_j\|_{H^p_{\mathbb{H}}} \leq C 2^{-j/p} \Phi(\Lambda_j) \leq C 2^{-j(p-\tau)} |f|_{B^s_\tau}, \quad j = 1, 2, \ldots.$$

We complete the proof in the case $p \leq 1$ (the case $p > 1$ being handled in a similar way). We have

$$\|f - T_k\|_{H^p_{\mathbb{H}}} \leq \sum_{j > k} \|S_j\|_{H^p_{\mathbb{H}}} \leq C \sum_{j > k} 2^{-j(p-\tau)} |f|_{B^s_\tau} \leq C 2^{-k(p-\tau)} |f|_{B^s_\tau}$$

which is (iii).

From (ii) and (iii), we have that for $t \propto |f|_{B^s_\tau} 2^{k\tau}$,

$$\sigma(f, t)_p \leq C|f|_{B^s_\tau} (|f|_{B^s_\tau} 2^{k\tau})^{1/p-1/\tau} = C|f|_{B^s_\tau} t^{-s/\beta d}.$$

From the monotonicity of $\sigma(f, t)_p$ we obtain (5.3) for all real numbers $t > 0$ which is (5.3).

Corollary 5.1. For each $t > 0$, we have

$$\sigma(f, t)_p \leq K(f, t^{-s/\beta d})$$

where $K$ is the $K$-functional for the pair $H^p_{p}$ and $B^s_\tau$.

Proof. This follows from the Jackson inequality and (3.3) for the pair $X = H^p_{p}$, $Y = B^s_\tau$.

5.2. The Bernstein inequality for restricted approximation.

We shall prove next the Bernstein inequality which is the companion of the Jackson inequality in §5.1. We continue with the notation of §5.1.
Theorem 5.2. Let \( s > 0 \), and \( p \) and \( \tau \) be admissible and satisfy \( s = \beta d(1/\tau - 1/p) \).
If \( f \in H_p \) has the wavelet expansion \( f = \sum_{I \in \Lambda} A_I(f) \) with \( \Phi(\Lambda) \leq t \), then
\[
\|f\|_{B^*_p} \leq Ct^{s/\beta d}\|f\|_{H_p},
\]
with \( C \) depending only on \( p \) and \( s \).

Proof. Case 1: \( p \geq 2 \). We have
\[
\|f\|_{B^*_p} = \sum_{I \in \Lambda} |I|^{-s\tau/d} a_{I,\tau}(f) = \sum_{I \in \Lambda} a_{I,p}(f)^\tau |I|^{-s\tau/d + 1-\tau/p} = \sum_{I \in \Lambda} a_{I,p}(f)^\tau |I|^\alpha(1-\tau/p)
\]
\[
\leq \left( \sum_{I \in \Lambda} a_{I,p}(f)^p \right)^{\tau/p} \left( \sum_{I \in \Lambda} |I|^\alpha \right)^{1-\tau/p} \leq t^{s\tau/d} \left( \sum_{I \in \Lambda} a_{I,p}(f)^p \right)^{\tau/p}.
\]

On the other hand, as in the proof of Theorem 4.3, we have
\[
\|S(f)\|_{L_p} \geq \left( \sum_{I \in \Lambda} a_{I,p}(f)^2 |I|^{-2/p} \chi_I \right)^{1/2} \|L_p \geq \left( \sum_{I \in \Lambda} a_{I,p}(f)^p \right)^{1/p},
\]
which completes the proof in this case.

Case 2: \( p \leq 2 \). With \( I(x) \) defined as the smallest interval in \( \Lambda \) that contains \( x \), we have with \( 1/\mu := 1 - \tau/p \),
\[
\|f\|_{B^*_p} = \sum_{I \in \Lambda} |I|^{-s\tau/d} a_{I,\tau}(f) = \int_{\mathbb{R}^d} \sum_{I \in \Lambda} a_{I,\tau}(f)^\tau |I|^{-1-s\tau/d} \chi_I
\]
\[
\leq C \int_{\mathbb{R}^d} S(f,x)^\tau |I(x)|^{-s\tau/d} dx \leq C \left( \int_{\mathbb{R}^d} S(f,x)^p dx \right)^{\tau/p} \left( \int_{\mathbb{R}^d} |I(x)|^{-s\mu/d} dx \right)^{1/\mu}
\]
\[
\leq C \left( \int_{\mathbb{R}^d} S(f,x)^p dx \right)^{\tau/p} \left( \sum_{I \in \Lambda} |I|^{-s\mu/d} \right)^{1/\mu} = C \|S(f)\|_{L_p} \Phi(\Lambda)^{1/\mu}
\]
\[
\leq Ct^{1-\tau/p}\|f\|_{H_p} = Ct^{s\tau/d}\|f\|_{H_p},
\]
because \( 1 - s\tau\mu/d = \alpha \). \( \square \)

5.3. An analogue of Temlyakov’s result for restricted approximation.

We shall prove an analogue of the theorem of Temlyakov for restricted approximation. We assume that \( \beta \neq 1 (\alpha \neq 0) \) since the case \( \beta = 1 \) is already covered in §4. We continue with the same notation as in the previous sections on restricted approximation. Let \( f \in H_p + B^*_p \) and for \( t > 0 \), let \( B_t f = \sum_{I \in \Lambda_t} A_I(B_t) \in \Sigma_t \) satisfy
\[
\|f - B_t\|_{H_p} \leq 2\sigma(f,t)_p.
\]

The set \( \Lambda_t \) thus satisfies \( \Phi(\Lambda_t) \leq t \). By adding small (in the case \( \alpha > 0 \)) or large (in the case \( \alpha < 0 \)) cubes to \( \Lambda_t \) (and putting coefficients equal to 0 for the new cubes), we can assume that \( \Phi(\Lambda_t) = t \)
We modify $B_t f$ by replacing $A_t(B_t)$ by the exact components $A_t(f)$ of $f$ to get $B_t^* f := \sum_{I \in \Lambda_t} A_t(f)$ which is also in $\Sigma_t$. We also introduce operators associated with thresholding. Given $\epsilon > 0$, let $\Lambda_\epsilon := \{ I : a_{I, r}(f) > \epsilon \}$ where as before $1/r := \beta/p$. We let $t := t_\epsilon := \Phi(\Lambda_\epsilon)$ and define $\tilde{B}_t f := \sum_{I \in \Lambda_\epsilon} A_t(f)$.

While the results that follow in this section include statements for $\tilde{B}_t$, they are not completely satisfactory because $\tilde{B}_t f$ is not necessarily defined for a given value of $t > 0$. We shall discuss thresholding operators in more detail in §7

**Theorem 5.3.** Let $s > 0$, and $p$ and $\tau$ be admissible and satisfy $s = \beta d(1/\tau - 1/p)$. For each $t > 0$ and $f \in H_p + B_t^*$ the functions $B_t f$ and $B_t^* f$ are near best $H_p$ approximations to $f$ from $\Sigma_t$. Similarly, for each $t$ for which $\tilde{B}_t f$ is defined, it is also a near best approximation to $f$ from $\Sigma_n$. In other words,

\[(5.5) \quad \| f - A_t f \|_{H_p} \leq C \sigma(f, t)_p,\]

for $A_t f = B_t f$ or $B_t^* f$, and for $A_t f = \tilde{B}_t f$ when the latter is defined, with a constant $C \geq 1$ depending only on $p$.

**Proof.** The conclusions of the theorem for $B_t f$ are obvious in view of its definition (5.4). For $B_t^* f$, we have from the square function

\[\| f - B_t^* f \|_{H_p} \leq C \| f - B_t f \|_{H_p} \leq 2C\sigma(f, t)_p.\]

Finally, we prove the theorem for $\tilde{B}_t f$. Let $\Lambda_\epsilon$ be the set associated with $\tilde{B}_t f$ and let $\Lambda_t$ be the set associated with $B_t^*$. It is enough to show that

\[(5.6) \quad \| B_t^* f - \tilde{B}_t f \|_{H_p} \leq C \sigma(f, t)_p.\]

We have

\[B_t^* f - \tilde{B}_t f = \sum_{I \in \Lambda_\epsilon \setminus \Lambda_t} A_I(f) + \sum_{I \in \Lambda_t \setminus \Lambda_\epsilon} A_I(f) =: f_0 + f_1\]

Using the square function, we see that

\[\| f_0 \|_{H_p} \leq C \| f - B_t^* f \|_{H_p} \leq C\sigma(f, t)_p.\]

Now, for all $I \in \Lambda_t \setminus \Lambda_\epsilon$ we have $a_{I, r}(f) \leq \epsilon$. Hence, from Lemma 5.1, we have

\[\| f_1 \|_{H_p} \leq C \epsilon \Phi(\Lambda_t \setminus \Lambda_\epsilon)^{1/p}.\]

On the other hand, for all $I \in \Lambda_\epsilon \setminus \Lambda_t$ we have $a_{I, r}(f) \geq \epsilon$ and hence

\[\| f_0 \|_{H_p} \geq C \epsilon \Phi(\Lambda_\epsilon \setminus \Lambda_t)^{1/p}.\]

Since $\Phi(\Lambda_\epsilon \setminus \Lambda_t) = \Phi(\Lambda_t \setminus \Lambda_\epsilon)$, we have $\| f_1 \|_{H_p} \leq C \| f_0 \|_{H_p} \leq C\sigma(f, t)_p$ which completes the proof. □
Corollary 5.2. Let $s > 0$, and $p$ and $\tau$ be admissible and satisfy $s = \beta d(1/\tau - 1/p)$. For each $f \in L^p + B^s_\tau$ and each $t > 0$, the function $B^s_\tau f$ realizes the $K$-functional i.e.

$$
\|f - B^s_\tau f\|_{H_p^\tau} + t^{-s/\beta d}\|B^s_\tau f\|_{B^s_\tau} \leq C K(f, t^{-s/\beta d}, H_p, B^s_\tau),
$$

with the constant $C$ depending only on $p$, $s$, and $\beta$. The same result holds for $\tilde{B}_t f$ whenever $\tilde{B}_t f$ is defined.

Proof. This follows from Theorem 3.2. Indeed, both operators $B^s_\tau$ and $\tilde{B}_t$ provide near best approximations as was shown in Theorem 5.3 and both are $B^s_\tau$ stable (with stability constant $C = 1$ since we use the wavelet definition of these spaces). $\square$

5.4. Jackson and Bernstein inequalities for restricted approximation in $B^s_p$.

The proofs of the Jackson and Bernstein inequalities for restricted approximation in $B^s_p$ are somewhat simpler than in $H_p$ and all follow simply by analyzing the sequence of wavelet coefficients. We shall continue to use the spaces $B^s_\tau$ where $s = d\beta(1/\tau - 1/p)$ and the parameter $r := p/\beta$.

To prove the Jackson inequality, we use the notation of §5.1.

Theorem 5.4. Let $s > 0$, and $p$ and $\tau$ be admissible and such that $s = \beta d(1/\tau - 1/p)$. If $f \in B^s_\tau$, then for each $j, k \in \mathbb{Z}$, we have

(i) $\Phi(\Lambda_k) \leq C|f|_{B^s_\tau} 2^{k\tau}$,

(ii) $\Phi(\cup_{j \leq k} \Lambda_j) \leq C|f|_{B^s_\tau} 2^{k\tau}$,

(iii) $\|f - T_k f\|_{B^s_p} \leq C2^{k(\tau/p - 1)}|f|_{B^s_\tau}^{\tau/p}$.

In addition, for each $t > 0$, we have

$$
\sigma(f, t)_{B^s_p} \leq C|f|_{B^s_\tau} t^{-s/\beta d}.
$$

Proof. (i) and (ii) were proved in Theorem 5.1. For the proof of (iii), we have

$$
\|S_j\|_{B^s_p}^p = \sum_{I \in \Lambda_j} a^p_I |I|^{p(1/p - 1/r)} \leq C 2^{-jp} \sum_{I \in \Lambda_j} |I|^{1-p/r} 
$$

$$
= C 2^{-j\tau} \Phi(\Lambda_j) \leq C 2^{-j(\tau/p - 1)}|f|_{B^s_\tau}^{\tau/p},
$$

where we used (i). Therefore, assuming $p \geq 1$ (a simple modification applies when $p < 1$), we have

$$
\|f - T_k f\|_{B^s_p} \leq C \sum_{j > k} \|S_j\|_{B^s_p} \leq \sum_{j > k} 2^{-j(1-q/p)}|f|_{B^s_\tau}^{\tau/p} \leq C 2^{-k(1-\tau/p)}|f|_{B^s_\tau}^{\tau/p},
$$

which is (iii).

From (ii) and (iii), we have that for $t \gg |f|_{B^s_\tau} 2^{k\tau}$,

$$
\sigma(f, t)_{B^s_p} \leq C|f|_{B^s_\tau} (|f|_{B^s_\tau} 2^{k\tau})^{1/p - 1/\tau} \leq C|f|_{B^s_\tau} t^{-s/\beta d},
$$

which is (5.7). $\square$

We shall prove next the Bernstein inequality which is the companion of the Jackson inequality for $B^s_p$. We continue with the previous notation.
Theorem 5.5. Let $s > 0$, and $p$ and $\tau$ be admissible and satisfy $s = \beta d(1/\tau - 1/p)$. If $f \in H_p + B^s_\tau$ has the wavelet expansion $f = \sum_{I \in \Lambda} A_I(f)$ with $\Phi(\Lambda) \leq t$, then

$$
\| f \|_{B^s_\tau} \leq C t^{s/\beta d} \| f \|_{B^s_p}.
$$

Proof. We have

$$
|f|_{B^s_\tau} = \sum_{I \in \Lambda} |I|^{-s\tau/k} d a_{I, \tau}(f) = \sum_{I \in \Lambda} d^{-\tau/k} |I|^{1-\tau/p-\beta a/k} = \sum_{I \in \Lambda} a_{I, \tau}(f) |I|^{\alpha(1-\tau/p)}
$$

$$
\leq \left( \sum_{I \in \Lambda} a^p_{I, \tau} \right)^{\tau/p} \left( \sum_{I \in \Lambda} |I|^\alpha \right)^{1-\tau/p} \leq t^{\tau\beta/k} \left( \sum_{I \in \Lambda} a_{I, \tau}(f)^p \right)^{\tau/p}. \quad \square
$$

6. Approximation spaces for restricted approximation.

The following discussion applies to both the case of restricted approximation and the case of ordinary $n$-term wavelet approximation (the case $\beta = 1$, $\alpha = 0$). We fix an admissible $p$ with $0 < p < \infty$. Further, we let $s, \tau$ be parameters for which the Jackson and Bernstein inequalities hold in §5 and which satisfy $s = d/\beta(1/\tau - 1/p)$ with $\tau$ admissible. We fix $s$ and $\tau$ throughout. We shall use frequently in this section without further mention the fact that the Jackson and Bernstein inequalities also hold for any $0 < \gamma < s$ and $\mu := \mu(\gamma)$ defined by the relation $\gamma = \beta d(1/\mu - 1/p)$.

For any $0 < \gamma$ and $0 < q \leq \infty$, we define the approximation space $A^\gamma_q(H_p)$ by using the quasi-semi-norm $|f|_{A^\gamma_q(H_p)}$ of (1.6) or the equivalent quasi-semi-norm (1.7). We add $\| f \|_{L^p + B^s_\tau}$ to $|f|_{A^\gamma_q(H_p)}$ to obtain the norm $\| f \|_{A^\gamma_q(H_p)}$. We remark that in the case $\beta \geq 1$, we have $B^s_q$ is embedded in $H_p$ and therefore $f \in H_p$. It follows that $\sigma(f, t)_{H_p} \leq \| f \|_{H_p}$. Therefore, the indices in (1.7) can be taken over $k \geq 0$ with an equivalent norm. However, we shall not make any use of this fact in what follows.

The spaces $A^\gamma_q(B_p)$ and their semi-norms and norms are defined in the same way with $H_p$ replaced by $B_p$.

We shall show how the spaces $A^\gamma_q(H_p)$ and $A^\gamma_q(B_p)$ can be characterized by wavelet coefficients. We use the abbreviated notation $a_I := a_I(f) := a_I(r)(f)$ throughout this section with $r = p/\beta$ as introduced and used earlier.


We shall first consider approximation in $B_p$ which is somewhat simpler than approximation in $H_p$. We first note that

$$
\| f \|_{B^s_p}^p := \sum_{I \in D} a_{I, p}(f)^p = \sum_{I \in D} a_I(f)^p |I|^{1-p/r} = \sum_{I \in D} a_I(f)^p |I|^\alpha.
$$

Similarly, for each $0 < \gamma \leq s$, and $\mu := \mu(\gamma)$ defined by $\gamma = d/\beta(1/\mu - 1/p)$, we have

$$
|f|_{B^s_p}^\mu := \sum_{I \in D} \| I \|^{-\gamma/d} a_{I, \mu}(f)^\mu = \sum_{I \in D} a_I(f)^\mu |I|^{1-\gamma |d-\mu|/r} = \sum_{I \in D} a_I(f)^\mu |I|^\alpha.
$$
For $0 < \lambda < \infty$, we let $\ell_{\lambda}(w)$ denote the space of all sequences $(c_I)_{I \in \mathcal{D}}$ with the norm

$$\|(c_I)\|_{\ell_{\lambda}(w)} := \left( \sum_{I \in \mathcal{D}} |I|^\alpha |c_I|^\lambda \right)^{1/\lambda},$$

corresponding to the weight $w(I) := |I|^\alpha$. We similarly define the weighted Lorentz spaces $\ell_{\lambda, q}(w)$ (see Chapter 1, p.8 of [BL]).

The identities (6.1-6.2) say that the linear mapping which takes $f$ into its wavelet coefficients is an isometry between $B_p$ and $\ell_p(w)$ and between $B^s_\tau$ and $\ell_\tau(w)$. It follows therefore that this mapping also gives an isometry between the interpolation spaces $(B_p, B^s_\tau)_{\theta, q}$ and the interpolation spaces $(\ell_p(w), \ell_\tau(w))_{\theta, q}$. The latter are well-known to be weighted Lorentz spaces $\ell_{\mu, q}(w)$ with $1/\mu = (1-\theta)/p + \theta/\tau$ (see Chapter 5, p.109 of [BL]). Therefore, $f \in (B_p, B^s_\tau)_{\theta, q}$, $0 < \theta < 1$, $0 < q \leq \infty$, if and only if

$$\|(a_I(f))_{I \in \mathcal{D}}\|_{\ell_{\mu, q}(w)} \leq \frac{1}{\mu} = \frac{1-\theta}{p} + \frac{\theta}{\tau},$$

and $\|(a_I(f))\|_{\ell_{\mu, q}}$ is an equivalent norm for $(B_p, B^s_\tau)_{\theta, q}$.

In particular, (6.4) (with $q = \mu$) and (6.2) give that for any $0 < \theta < 1$

$$\frac{1}{\mu} = \frac{1-\theta}{p} + \frac{\theta}{\tau},$$

where $\mu$ and $\gamma$ are related as before by $\gamma = \beta d(\frac{1}{\mu} - \frac{1}{p})$. More generally, let $\mu_j$ and $\gamma_j$ be related by $\gamma_j = \beta d(1/\mu_j - 1/p)$, $j = 1, 2$. Then, from the reiteration theorem for interpolation, we obtain

$$\frac{1}{\gamma} = \frac{1-\theta}{\gamma_1} + \frac{\theta}{\gamma_2},$$

where again $\mu$ and $\gamma$ are related by $\gamma = d\beta(\frac{1}{\mu} - \frac{1}{p})$.

**Theorem 6.1.** Let $p$ be admissible with $0 < p < \infty$ and let $s > 0$ and $\tau$ be defined by $s = \beta d(1/\tau - 1/p)$ with $\tau$ admissible. For each $0 < \gamma < s/\beta d$, $0 < q \leq \infty$, we have

$$A^\gamma_q(B_p) = (B_p, B^s_\tau)_{\theta, q}, \quad \theta := \gamma \beta d/s,$$

with equivalent norms.

**Proof.** This follows from the Jackson and Bernstein inequalities of §5.4 and Theorem 3.1. \(\square\)

**Corollary 6.1.** Let $p$ be admissible with $0 < p < \infty$ and let $s > 0$ and $\tau$ be defined by $s = \beta d(1/\tau - 1/p)$ with $\tau$ admissible. For each $0 < \gamma < s$, and $\mu := \mu(\gamma)$ defined by the equation $\gamma = \beta d(1/\mu - 1/p)$, we have

$$A^\gamma_\mu(B_p) = B^\gamma_\mu,$$

with equivalent norms.

**Proof.** This follows from Theorem 6.1 and (6.5). \(\square\)
Corollary 6.2. Let \( p \) be admissible with \( 0 < p < \infty \) and let \( s > 0 \) and \( \tau \) be defined by \( s = \beta d(1/\tau - 1/p) \) with \( \tau \) admissible. Let \( 0 < \gamma < s \) and let \( \mu \) be defined by the relation \( \gamma = \beta d(1/\mu - 1/p) \). Then, for each \( 0 < q \leq \infty \), \( f \in A_{q}^{\gamma/\beta d}(B_{p}) \) if and only if \( (a_{1}(f))_{i \in \mathbb{D}} \in \ell_{\mu, q}(\omega) \), and the two norms \( \|f\|_{A_{q}^{\gamma/\beta d}(B_{p})} \) and \( \|(a_{1}(f))\|_{\ell_{\mu, q}} \) are equivalent.

Proof. This follows by using (6.4) to characterize the interpolation space. \( \square \)

Finally, we observe that, in view of our results, Corollary 5.2 also holds with \( B_{p} \) in place of \( H_{p} \).

6.2. Characterization of \( A_{q}^{\gamma}(H_{p}) \) by interpolation.

We can carry out an analysis similar to that of \( \S 6.1 \) to show that restricted approximation in \( H_{p} \) can also be characterized by interpolation. We use the same notation as in \( \S 6.1 \) except that now \( K \) denotes the K-functional for the pair of spaces \( H_{p} \) and \( B_{p}^{\ast} \). Using the Jackson and Bernstein inequalities for restricted approximation in \( H_{p} \), we derive the following analogue of Theorem 6.1.

Theorem 6.2. Let \( p \) be admissible with \( 0 < p < \infty \) and let \( s > 0 \) and \( \tau \) be defined by \( s = \beta d(1/\tau - 1/p) \) with \( \tau \) admissible. For each \( 0 < \gamma < s/\beta d \), \( 0 < q \leq \infty \), we have

\[
A_{q}^{\gamma}(H_{p}) = (H_{p}, B_{p}^{\ast})_{\theta, q}, \quad \theta := \gamma \beta d/s,
\]

with equivalent norms.

At present, Theorem 6.2 is not quite satisfactory because we still do not know the interpolation spaces appearing on the right side of (6.9). However, the next theorem will show that these interpolation spaces are the same as those for the pair \( B_{p}, B_{p}^{\ast} \) which we have already characterized.

Theorem 6.3. Let \( p \) be admissible with \( 0 < p < \infty \) and let \( s > 0 \) and \( \tau \) be defined by \( s = \beta d(1/\tau - 1/p) \) with \( \tau \) admissible. For each \( 0 < \gamma < s \) and \( 0 < q \leq \infty \), we have

\[
A_{q}^{\gamma/\beta d}(H_{p}) = A_{q}^{\gamma/\beta d}(B_{p}).
\]

Proof. We first note the embeddings

\[
A_{\bar{\mu}}^{\gamma/\beta d}(H_{p}) \subset B_{\mu}^{\gamma} \subset A_{\infty}^{\gamma/\beta d}(H_{p}),
\]

which hold for any \( 0 < \gamma < s \), \( \mu = \mu(\gamma) \) satisfying \( \gamma = \beta d(1/\mu - p) \) and \( \bar{\mu} := \min\{1, \mu\} \). Indeed, the right embedding in (6.11) follows from (5.3) with \( s \) replaced by \( \gamma \). To prove the left embedding in (6.11), we let \( f \in A_{\mu}^{\gamma/\beta d}(H_{p}) \) and let \( S_{k} \in \Sigma_{2^{k}} \) satisfy

\[
\|f - S_{k}\|_{H_{p}} \leq \sigma(f, 2^{k})_{p}, \quad k \in \mathbb{Z}.
\]

Then, we have \( f = \sum_{k=-\infty}^{\infty} (S_{k} - S_{k-1}) \) and therefore

\[
|f|^{\bar{\mu}}_{B_{\mu}^{\gamma}} \leq \sum_{k=-\infty}^{\infty} |S_{k} - S_{k-1}|^{\bar{\mu}}_{B_{\mu}^{\gamma}} \leq C \sum_{k=-\infty}^{\infty} 2^{k\gamma \bar{\mu}/\beta d} \|S_{k} - S_{k-1}\|_{H_{p}}^{\bar{\mu}} \leq \sum_{k=-\infty}^{\infty} 2^{k\gamma \bar{\mu}/\beta d} (\sigma(f, 2^{k})_{p} + \sigma(f, 2^{k-1})_{p})^{\bar{\mu}} \leq C \|f\|_{A_{\mu}^{\gamma/\beta d}}^{\bar{\mu}}.
\]


Here, we have used the subadditivity of $|f|_{B_s^{\gamma \beta}}$ in the first inequality, the Bernstein inequality of Theorem 5.2 (with $s$ replaced by $\gamma$) in the second inequality, and the discrete norm (1.7) in the last inequality.

Let $0 < \gamma_j < s$ and $\mu_j$ be related by $\gamma_j = \beta d(1/\mu_j - 1/p)$, $j = 1, 2$. We assume that $\gamma_1 < \gamma_2$. We recall that both the $A_{\mu_1}^\gamma(H_p)$ and $A_{\mu_2}^\gamma(B_p)$ are interpolation families. Therefore, the reiteration theorem for interpolation together with the embeddings (6.10-6.11) give that for each $0 < \theta < 1$ and $0 < q \leq \infty$, we have (6.12)

$$(A_{\mu_1}^{\gamma_1/\beta d}(H_p), A_{\mu_2}^{\gamma_2/\beta d}(H_p))_{\theta,q} = (B_{\mu_1}^{\gamma_1}, B_{\mu_2}^{\gamma_2})_{\theta,q} = (A_{\mu_1}^{\gamma_1/\beta d}(B_p), A_{\mu_2}^{\gamma_2/\beta d}(B_p))_{\theta,q}.$$ 

The left side of (6.12) is the approximation space $A_{\mu}^{\gamma/\beta d}(H_p)$, $\gamma = (1-\theta)\gamma_1 + \theta \gamma_2$ and the right side of (6.12) is the approximation space $A_{\mu}^{\gamma/\beta d}(B_p)$ with the same parameters. Since $\theta$ and $q$ are arbitrary and $\gamma_1$ can be chosen arbitrarily close to 0 and $\gamma_2$ arbitrarily close to $s$, (6.9) follows.

**Corollary 6.3.** Let $p$ be admissible with $0 < p < \infty$ and let $s > 0$ and $\tau$ be defined by $s = \beta d(1/\tau - 1/p)$ with $\tau$ admissible. For each $0 < \gamma < s$, and $\mu := \mu(\gamma)$ defined by the equation $\gamma = \beta d(1/\mu - 1/p)$, we have (6.13)

$$A_{\mu}^{\gamma/\beta d}(H_p) = B_{\mu}^{\gamma},$$

with equivalent norms.

**Proof.** This follows from Theorem 6.3 and Corollary 6.1.

**Corollary 6.4.** Let $p$ be admissible with $0 < p < \infty$ and let $s > 0$ and $\tau$ be defined by $s = \beta d(1/\tau - 1/p)$ with $\tau$ admissible. Let $0 < \gamma < s$, and let $\mu$ be defined by the relation $\gamma = \beta d(1/\mu - 1/p)$. Then for each $0 < q \leq \infty$, $f \in A_{\mu}^{\gamma/\beta d}(H_p)$ if and only if $(a_1(f))_{I \in \mathcal{D}} \in l_{\mu,q}(w)$, and the two norms $\|f\|_{A_{\mu}^{\gamma/\beta d}(H_p)}$ and $\|a_1(f)\|_{l_{\mu,q}}$ are equivalent.

**Proof.** This follows from Theorem 6.3 and Corollary 6.2.

### 7. Thresholding.

One of the most frequently used numerical methods for generating adaptive wavelet approximations consists in thresholding the coefficients of the function to be approximated. In this section, we shall look more closely at thresholding for restricted approximation. We fix an admissible $p$ with $0 < p < \infty$. Further, we let $s, \tau$ be parameters which satisfy $s = d\beta(1/\tau - 1/p)$ with $\tau$ admissible. We fix $s$ and $\tau$ throughout. For $f \in H_p + B_s^\tau$, we let $a_I := a_I(f) := a_{I,r}(f)$ with $r = p/\beta$ throughout this section.

For each $\epsilon > 0$, we let

$$\Lambda(f, \epsilon) = \{I : a_I(f) \geq \epsilon\}$$

and let

$$T_\epsilon f := \sum_{I \in \Lambda(\epsilon, f)} A_I(f).$$

The next theorem characterizes functions $f$ for which $\|f - T_\epsilon f\|_{L_p}$ has a certain decay. We recall the weighted Lorentz spaces $\ell_{\mu,q}(w)$, $w(I) := |I|^q$ which appeared in the characterization of the approximation spaces for restricted approximation. We shall be especially interested in the case $q = \infty$ where $\ell_{\mu,\infty}(w) = \text{weak-} \ell_{\mu}$. 
**Theorem 7.1.** Let $p$ be admissible with $0 < p < \infty$ and let $s > 0$ and $\tau$ be defined by $s = \beta d(1/\tau - 1/p)$ with $\tau$ admissible. For each $\tau < \mu < p$, a function $f$ satisfies

\begin{equation}
\|f - \mathcal{T}_\epsilon f\|_{H_p} \leq M^{\mu/p} \epsilon^{1-\mu/p}
\end{equation}

if and only if $(a_f)_{1 \in \mathbb{D}} \in l_{\mu, \infty}(w)$ and the smallest $M$ satisfying (7.1) is equivalent to $\|(a_f)\|_{l_{\mu, \infty}(w)}$.

**Proof.** First assume that $(a_f) \in l_{\mu, \infty}(w)$ and let $M := \|(a_f)\|_{l_{\mu, \infty}(w)}$. Let $\epsilon > 0$ and define $k \in \mathbb{Z}$ such that $2^{-k-1} < \epsilon \leq 2^{-k}$. We define the sets $\Lambda_j$ and the function $S_j f$ as in Theorem 5.1. Then, from the definition of the $l_{\mu, \infty}(w)$ norm, we have

$$\Phi(\Lambda_j) \leq M^{\mu} 2^{j\mu}, \quad j \in \mathbb{Z}.$$ 

From Lemma 5.1, we have

$$\|S_j f\|_{H_p} \leq C 2^{-j} \Phi(\Lambda_j)^{1/p} \leq C 2^{-j} M^{\mu/p} 2^{j\mu/p} \leq C M^{\mu/p} 2^{-j(1-\mu/p)}.$$ 

We continue with the case $p \geq 1$ (a similar argument applies when $0 < p < 1$). We have

$$\|f - \mathcal{T}_\epsilon f\|_{H_p} \leq \sum_{j=k+1}^{\infty} \|S_j f\|_{H_p} \leq C M^{\mu/p} \sum_{j=k+1}^{\infty} 2^{-j(1-\mu/p)} \leq C M^{\mu/p} 2^{-k(1-\mu/p)} \leq C M^{\mu/p} \epsilon^{(1-\mu/p)}.$$ 

This proves one of the implications in the theorem.

Conversely, we assume that for each $\epsilon > 0$,

$$\|f - \mathcal{T}_\epsilon f\|_{H_p} \leq M^{\mu/p} \epsilon^{1-\mu/p}.$$ 

With $S_j$ as above, and using the square function, we find

$$\|S_j f\|_{H_p} \leq C \|f - \mathcal{T}_{2^{-j}} f\|_{H_p} \leq C M^{\mu/p} 2^{-j(1-\mu/p)}.$$ 

Hence, using Lemma 5.1 again, we find

$$\Phi(\Lambda_j)^{1/p} 2^{-j} \leq C \|S_j f\|_{H_p} \leq C M^{\mu/p} 2^{-j(1-\mu/p)}.$$ 

That is,

$$\Phi(\Lambda_j) \leq C M^{\mu} 2^{j\mu}.$$ 

Therefore, with $2^{-k-1} < \epsilon \leq 2^{-k}$, we have

$$\Phi(\Lambda(f, \epsilon)) \leq \sum_{j=-\infty}^{k+1} C M^{\mu} 2^{j\mu} \leq C M^{\mu} \epsilon^{-\mu},$$

which proves the other implication in the Theorem. \qed
8. Adaptation to a bounded domain.

Most practical applications of restricted approximation arise in the context of bounded domains, i.e. the function \( f \) to be approximated is defined on an open connected set \( \Omega \subset \mathbb{R}^d \).

With a little more work (see e.g. [D] or [C]) and some reasonable assumptions on the geometry of \( \Omega \), multiscale decompositions into wavelet bases can be adapted to such bounded domains. In such decompositions, the range of scales is only \( k = 0, 1, 2, \cdots \), i.e. functions on \( \Omega \) are decomposed according to

\[
\sum_{I \in \mathcal{D}_+} A_I(f),
\]

with \( \mathcal{D}_+ = \cup_{k \geq 0} \mathcal{D}_k(\Omega) \), and \( \mathcal{D}_k(\Omega) \) a subset of \( \mathcal{D}_k \) that describes the wavelets adapted to \( \Omega \) at scale \( k \). The basis functions in the coarsest layer \( \mathcal{D}_0(\Omega) \) are scaling functions which do not not oscillate (their integrals differ from zero), since they are meant to describe a coarse approximation of \( f \).

We want to discuss here the adaptation of our results to this slightly different setting. A first remark is that all the results of this paper will also hold in this setting, if we formulate them in terms of sequence spaces: we define \( h_p \) and \( b^s_{q,p} \) consisting respectively of those sequences \( a = (a_I)_{I \in \mathcal{D}_+} \) such that

\[
\|a\|_{h_p} := \left[ \int_{\mathbb{R}^d} \left( \sum_{I \in \mathcal{D}_+} |a_I|^2 |I|^{-1} \chi_I \right)^p \right]^{1/p},
\]

and

\[
\|a\|_{b^s_{q,p}} := \|(2^{ks}2^{k(d(1/2 - 1/p))}) (a_I)_{I \in \mathcal{D}_k(\Omega)}\|_{\ell^p,k \geq 0} \|\ell^s\|.
\]

are finite. Replacing \( H_p \) by \( h_p \) and \( B^s_{q,p} \) by \( b^s_{q,p} \), we can utilise the same method of proof and characterize restricted approximation in the \( h_p \) metric.

Accordingly, we thus obtain similar results for restricted approximation if we define \( H_p(\Omega) \) and \( B^s_{q,p}(\Omega) \) to be spaces of distributions \( f \) in \( \Omega \) such that for a fixed wavelet basis, the sequence of coefficients \( a_I(f) = a_{I,2}(f) \) exists and belong to the space \( h_p \) and \( b^s_{q,p} \), with corresponding norms given by (8.2) and (8.3).

In general, the above defined \( H_p(\Omega) \) and \( B^s_{q,p}(\Omega) \) will depend on the particular choice of the wavelet basis, unless we can identify them as classical function spaces. In [C], it is proved that, under general smoothness assumptions on the wavelet basis, \( H_p(\Omega) \) coincides with the usual Lebesgue space \( L^p(\Omega) \) for \( 1 < p < \infty \) and \( B^s_{q,p}(\Omega) \) with the usual Besov space \( B^s_{q,q}(L^p(\Omega)) \) if \( s > d/p - d \) (under minimal smoothness assumptions on the boundary of the domain, the latter can be defined equivalently by restriction of the Besov spaces defined on \( \mathbb{R}^d \) or by their inner description using moduli of smoothness in \( \Omega \)).

Our results can thus be applied to these classical spaces for this range of indices \( s \) and \( p \). For more general indices, we can accept \( H_p(\Omega) \) and \( B^s_{q,p}(\Omega) \) as a definition of Hardy and Besov spaces on domains, having in mind the possible dependence of these spaces upon the choice of the wavelet basis.

REFERENCES


[T1] V. Temlyakov, Best \(m\)-term approximation and greedy algorithms., preprint.

[T2] V. Temlyakov, Nonlinear \(m\)-term approximation with regard to the multivariate Haar system., preprint.


Albert Cohen, Laboratoire d’Analyse Numérique, Université Pierre et Marie Curie, 75252, Paris cedex 05, France
Ronald A. DeVore, Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA
Reinhard Hochmuth, I. Mathematisches Institut, Freie Universität Berlin, Arnimallee 3, 14195 Berlin, Germany