Constr. Approx. (1998) 14: 1–26

CONSTRUCTIVE APPROXIMATION

Hyperbolic Wavelet Approximation

R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov

Abstract. We study the multivariate approximation by certain partial sums (hyperbolic wavelet sums) of wavelet bases formed by tensor products of univariate wavelets. We characterize spaces of functions which have a prescribed approximation error by hyperbolic wavelet sums in terms of a K-functional and interpolation spaces. The results parallel those for hyperbolic trigonometric cross approximation of periodic functions [DPT].

1. Introduction

Let φ be a univariate function that satisfies multiresolution analysis (see, e.g., [Da] for a description of multiresolution analysis). We denote by $\mathcal{S} := \mathcal{S}(\varphi)$ the shift-invariant space which is defined as the $L_2(\mathbf{R})$-closure of finite linear combinations of the shifts $\varphi(\cdot - j)$, $j \in \mathbf{Z}$, of φ. By dilation, we obtain the univariate spaces

$$\mathcal{S}^k := \mathcal{S}^k(\varphi) := \{S(2^k \cdot) : S \in \mathcal{S}\}, \qquad k \in \mathbf{Z}.$$

We obtain univariate wavelets ψ by considering projectors P_k from $L_2(\mathbf{R})$ onto $\mathcal{S}^k$. The wavelet space W^{k-1} is defined to be the image of $Q_k := P_k - P_{k-1}$. Wavelet functions ψ are generators of the shift-invariant space $W := W^0$, i.e., $W = \mathcal{S}(\psi)$. We have in mind here the usual orthogonal wavelets in the case the P_k are orthogonal projectors and the biorthogonal wavelets (see [CDF]) obtained when considering certain oblique projectors P_k.

From the univariate wavelet ψ, we can construct efficient bases for $L_2(\mathbf{R})$ and other function spaces by dilation and shifts. For example, the functions

$$\psi_{j,k} := 2^{k/2}\psi(2^k \cdot - j), \qquad j, k \in \mathbf{Z},$$

form a stable basis (orthogonal basis in the case of an orthogonal wavelet ψ) for $L_2(\mathbf{R})$.

It is convenient to use a different indexing for the functions $\psi_{j,k}$. Let $\mathcal{D}(\mathbf{R})$ denote the set of dyadic intervals. Each such interval I is of the form $I = [j2^{-k}, (j+1)2^{-k}]$. We define

$$\psi_I := \psi_{j,k}, \qquad I = [j2^{-k}, (j+1)2^{-k}]. \tag{1.1}$$

Thus the basis $\{\psi_{j,k}\}_{j,k\in\mathbf{Z}}$ is the same as $\{\psi_I\}_{I\in\mathcal{D}(\mathbf{R})}$.

Date received: October 16, 1995. Date revised: August 28, 1996. Communicated by Carl de Boor.
AMS classification: 41A63, 46C99.
Key words and phrases: Hyperbolic wavelets, Multivariate wavelets, Interpolation spaces.

Multivariate wavelets are usually obtained from multiresolution analysis on the tensor product spaces $\mathcal{S}^k \otimes \cdots \otimes S^k$. For example, in the case of bivariate approximation, this leads to the bivariate wavelet basis consisting of all functions

$$\varphi_{j,k}(x)\psi_{j',k}(y) \quad \psi_{j,k}(x)\varphi_{j',k}(y) \quad \psi_{j,k}(x)\psi_{j',k}(y), \tag{1.2}$$

with $j, j', k \in \mathbf{Z}$. The approximation properties of these wavelets are now well understood (see, e.g., [M] or for nonlinear approximation [DJP]).

A second way to construct multivariate wavelet bases is to simply take tensor products of the univariate basis functions $\psi_{j,k}$. If ψ is a univariate wavelet and $d \geq 1$, then the functions

$$\psi_{j_1,k_1}(x_1)\cdots\psi_{j_d,k_d}(x_d), \quad j = (j_1, \ldots, j_d) \in \mathbf{Z}^d, \quad k = (k_1, \ldots, k_d) \in \mathbf{Z}^d, \tag{1.3}$$

are a basis for $L_2(\mathbf{R}^d)$.

There is quite a distinction between these two wavelet bases. The functions in (1.2) have roughly the same support in each coordinate direction while the tensor products of (1.3) have support which is scaled independently in the different coordinate directions. As we shall see, this is also reflected in the approximation properties of the two sets of wavelets.

Again, it is more convenient to use another indexing for the basis functions (1.3). We let $\mathcal{D}(\mathbf{R}^d)$ denote the set of all dyadic rectangles in $\mathbf{R}^d$. Any $I \in \mathcal{D}(\mathbf{R}^d)$ is of the form $I = I_1 \times \cdots \times I_d$ with $I_1, \ldots, I_d \in \mathcal{D}(\mathbf{R})$. We define

$$\psi_I(x_1, \ldots, x_d) := \psi_{I_1}(x_1)\cdots\psi_{I_d}(x_d), \qquad I \in \mathcal{D}(\mathbf{R}^d). \tag{1.4}$$

Therefore, the wavelet basis (1.3) is the same as the set of functions $\{\psi_I\}_{I\in\mathcal{D}(\mathbf{R}^d)}$.

We are interested in the approximation properties of the functions (1.4). For $n = 0, 1, \ldots$ and $0 < p \leq \infty$, let

$$\mathcal{H}_n := \mathcal{H}_n(L_p(\mathbf{R}^d)) := \overline{\text{span}}\{\psi_I : |I| > 2^{-n}\}$$

denote the closed linear span of the finite linear combinations of the functions ψ_I, $|I| > 2^{-n}$, with the closure taken with respect to the $L_p(\mathbf{R}^d)$-(quasi-)norm.

We call the approximation by $\mathcal{H}_n$ hyperbolic wavelet approximation in analogy with the approximation by trigonometric polynomials with frequencies from the hyperbolic cross (see Temlyakov [T] for a discussion of this type of approximation).

We can also describe $\mathcal{H}_n$ in terms of the scaling function φ. Namely, $\mathcal{H}_n$ is the closed linear span of the functions

$$\varphi_I, \quad |I| \geq 2^{-n}.$$

In most wavelet applications, approximation occurs over a compact subset Ω of $\mathbf{R}^d$ and the approximation takes place from a finite-dimensional linear subspace $\tilde{\mathcal{H}}_n$ of $\mathcal{H}_n$.

The present paper is concerned with the approximation efficiency of the spaces $\mathcal{H}_n$. To measure this efficiency, we introduce the following approximation error. For $0 < p \leq \infty$ and $f \in L_p(\mathbf{R}^d)$, we define

$$E_n(f)_p := E(f, \mathcal{H}_n)_p := \inf_{g\in\mathcal{H}_n} \|f - g\|_p$$

with $\|\cdot\|_p$ here and later the $L_p(\mathbf{R}^d)$-norm.

We are interested in characterizing functions which have a given order of approximation. For $\alpha > 0$ and $0 < p, q \le \infty$, we define $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ as the collection of all functions $f \in L_p(\mathbf{R}^d)$ such that

$$|f|_{\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))} := \begin{cases} \left(\sum_{k\ge 0}[2^{k\alpha}E_k(f)_p]^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{k\ge 0} 2^{k\alpha}E_k(f)_p, & q = \infty, \end{cases} \tag{1.5}$$

is finite. We define the norm on the approximation space $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ by

$$\|f\|_{\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))} := \|f\|_p + |f|_{\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))}.$$

The characterization of the approximation classes $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ is tantamount to proving two inequalities (called Jackson and Bernstein inequalities) for the approximation process (see [DL, Chap. 7]). In the case of hyperbolic wavelet approximation, these inequalities involve mixed derivatives.

If r is a positive integer, we define the differential operator

$$\mathbf{D}^r := \frac{\partial^r}{\partial x_1^r} \cdots \frac{\partial^r}{\partial x_d^r}.$$

For $1 \le p \le \infty$, we let $\mathbf{W}^r(L_p(\mathbf{R}^d))$ be the set of all functions f in $L_p(\mathbf{R}^d)$ whose distributional derivative $\mathbf{D}^r f$ is in $L_p(\mathbf{R}^d)$ and define the seminorm on $\mathbf{W}^r(L_p(\mathbf{R}^d))$ by

$$|f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))} := \|\mathbf{D}^r f\|_p.$$

We will show in Section 3, that under certain conditions on the function ψ and for $1 < p < \infty$, we have the following two inequalities:

$$E_n(f)_p \le C2^{-nr}|f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}, \qquad n = 0, 1, \ldots, \qquad f \in \mathbf{W}^r(L_p(\mathbf{R}^d)), \tag{J}$$

with C independent of n and f and

$$|g|_{\mathbf{W}^r(L_p(\mathbf{R}^d))} \le C2^{nr}\|g\|_p, \qquad g \in \mathcal{H}_n, \quad n = 0, 1, \ldots, \tag{B}$$

with C independent of n and g.

From these Jackson and Bernstein inequalities for hyperbolic wavelet approximation it follows that we can characterize the spaces $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ in terms of the K-functional

$$K(f, t) := K(f, t, L_p(\mathbf{R}^d), \mathbf{W}^r(L_p(\mathbf{R}^d))). \tag{1.6}$$

Namely, for $0 < q < \infty$, a function $f \in L_p(\mathbf{R}^d)$ is in $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ if and only if

$$\int_0^\infty [t^{-\alpha}K(f, t^r)]^q \frac{dt}{t} < \infty. \tag{1.7}$$

A similar result holds for $q = \infty$ with the integral replaced by a sup. In the case of periodic functions, we have shown in [DPT] that $K(f, t)$ is equivalent to a certain modulus of smoothness $\Omega_r(f, t)_p$ based on mixed differences. In principle, this result should carry over to the case of approximation on $\mathbf{R}^d$; however, we have not yet carried out the details of this equivalence.

Our proof of the Jackson and Bernstein inequalities for hyperbolic wavelet approximation rests on the characterization of functions f in $L_p(\mathbf{R}^d)$ by a hyperbolic wavelet series

$$f = \sum_{I \in \mathcal{D}(\mathbf{R}^d)} c_I(f)\psi_I$$

and the calculation of its norm by the following *square function*

$$\left\| \sum_{I \in \mathcal{D}} c_I \psi_I \right\|_p \approx \left\| \left(\sum_{I \in \mathcal{D}} [c_I \psi_I]^2 \right)^{1/2} \right\|_p . \tag{1.8}$$

Characterizations (1.8) are at the heart of what is called the Littlewood–Paley theory for wavelets.

Littlewood–Paley theory has a long and important history in harmonic analysis. For the most part, we will utilize known aspects of this theory adapted to the case of hyperbolic wavelet decompositions. We describe the results from Littlewood–Paley theory that we will need and give their adaptation to hyperbolic wavelets in Section 2. There are several results which establish sufficient conditions for the family of functions $\{\psi_I\}_{I \in \mathcal{D}(\mathbf{R}^d)}$ to satisfy (1.8). However, these conditions are not always applicable in wavelet theory since they require smoothness of ψ not met by wavelets or their derivatives. We therefore establish (in Section 4) sufficient conditions which allow only piecewise Lipschitz continuity on the function ψ for the Littlewood–Paley characterization to hold.

Finally, in Section 5 we give an application of our results to the Daubechies wavelets. Other wavelets can be handled in a similar manner. As mentioned earlier, we will restrict our development in this paper to approximation on $\mathbf{R}^d$. We could also give a similar development for the case of approximation on a compact set in $\mathbf{R}^d$ or on the torus $\mathbf{T}^d$. In this way, our approach could be applied to other wavelet-like bases such as the Franklin system and its generalizations. While preparing the present paper, we were sent a preprint by A. Kamont [K] that proves Jackson and Bernstein inequalities for the Franklin system by utilizing Littlewood–Paley theory.

2. The Elements of Littlewood–Paley Theory

Littlewood–Paley theory gives a way of characterizing norms of linear combinations of certain basis functions. Its roots lie in the Littlewood–Paley theorems for Fourier series in which case the basis functions are the complex exponentials

$$e_k(x) := e^{ik\cdot x} = e^{i(k_1x_1+\cdots+k_dx_d)}, \qquad x \in \mathbf{R}^d.$$

However, the theory applies to many other orthogonal and nonorthogonal expansions (see, e.g., [FJ], [FJM], or [M]).

For us, Littlewood–Paley theory will provide a vehicle to prove our results on multivariate wavelet approximation. We begin in this section by introducing various forms of the Littlewood–Paley theory for systems of functions. While the discussion we give is for the most part known, it will enable us to set the framework for this paper and introduce several important results which will be employed later in the paper.

Let $\mathcal{D} = \mathcal{D}(\mathbf{R}^d)$ denote the collection of all dyadic rectangles in $\mathbf{R}^d$. Thus, a rectangle $I \subset \mathbf{R}^d$ is in $\mathcal{D}$ if and only if $I = I_1 \times \cdots \times I_d$ with $I_1, \ldots, I_d$ dyadic intervals in $\mathbf{R}$. We will consider in this section systems of functions $\{\eta(I, \cdot)\}_{I \in \mathcal{D}}$.

In the univariate case, one particular way to obtain such systems is by shifts and dilates of a univariate wavelet. For a univariate function ψ, we use the notation ψ_I of (1.1) to denote its $L_2(\mathbf{R})$-normalized, shifted dilates. The function ψ is an orthogonal wavelet if the collection of functions ψ_I, $I \in \mathcal{D}(\mathbf{R})$, forms a complete orthonormal system for $L_2(\mathbf{R})$. Other cases of interest in wavelet theory are the prewavelets [BDR], spline wavelets [CW], and biorthogonal wavelets [CDF]. In the latter cases, the orthogonality of the family ψ_I, $I \in \mathcal{D}(\mathbf{R})$, is replaced by L_2-stability (Riesz basis).

Given a univariate function ψ, we can obtain a multivariate family of functions by taking tensor products. For rectangles $I \in \mathcal{D}(\mathbf{R}^d)$, we define

$$\psi_I(x_1, \ldots, x_d) := \psi_{I_1}(x_1) \cdots \psi_{I_d}(x_d), \qquad I = I_1 \times \cdots \times I_d. \tag{2.1}$$

We will use the notation ψ_I to denote the family of functions obtained by tensor products of shifted dilates of a univariate function ψ and will use the notation $\eta(I, \cdot)$ to denote families of functions indexed on $I \in \mathcal{D}(\mathbf{R}^d)$ that are not necessarilly obtained by shifted dilates of one function. This notation will distinguish between the space dimension by the Euclidean dimension of the index rectangles.

A particularly important example occurs when we take for ψ the univariate Haar wavelet

$$H(x) := \begin{cases} +1, & 0 \le x \le \frac{1}{2}, \\ -1, & \frac{1}{2} < x \le 1. \end{cases}$$

The Haar function H is the simplest example of a univariate orthogonal wavelet.

If $1 < p < \infty$, we say that a family of real-valued functions $\eta(I, \cdot)$, $I \in \mathcal{D}$, satisfies the *strong Littlewood–Paley property* for p, if for any finite sequence (c_I) of real numbers, we have

$$\left\| \sum_{I \in \mathcal{D}} c_I \eta(I, \cdot) \right\|_p \approx \left\| \left(\sum_{I \in \mathcal{D}} [c_I \eta(I, \cdot)]^2 \right)^{1/2} \right\|_p \tag{2.2}$$

with constants of equivalency depending at most on p and d. Here and later we use the notation $A \approx B$ to mean that there are two constants $C_1, C_2 > 0$ such that

$$C_1 A \le B \le C_2 A.$$

We will indicate what the constants depend on (in the case of (2.2) they may depend on p and d).

Here is another useful remark concerning (2.2). From the validity of (2.2) for finite sequences, we can deduce its validity for infinite sequences by a limiting argument. For example, if $(c_I)_{I \in \mathcal{D}}$ is an infinite sequence for which the sum on the left side of (2.2) converges in $L_p(\mathbf{R}^d)$ with respect to some ordering of the $I \in \mathcal{D}$, then the right side of (2.2) will converge with respect to the same ordering and the right side of (2.2) will be less than a multiple of the left. Likewise, we can reverse the roles of the left- and right-hand sides. Similar remarks hold for other statements like (2.2) made in this paper.

The term *strong Littlewood–Paley inequality* is used to differentiate (2.2) from other possible forms of Littlewood–Paley inequalities. For example, the Littlewood–Paley inequalities for the complex exponentials take a different form (see [Z, Chap. XV]). Another form of interest in our considerations is the following:

$$\left\| \sum_{I\in\mathcal{D}} c_I \eta(I,\cdot) \right\|_p \approx \left\| \left(\sum_{I\in\mathcal{D}} [c_I \chi_I]^2 \right)^{1/2} \right\|_p . \tag{2.3}$$

We use the notation χ for the characteristic function of $[0, 1]$ and χ_I for its $L_2(\mathbf{R}^d)$-normalized, shifted dilates given by (2.1) (with $\psi = \chi$).

The two forms (2.2) and (2.3) are equivalent under very mild conditions on the functions $\eta(I,\cdot)$. To see this, we will use the Hardy–Littlewood maximal operator, which is defined for a locally integrable function g on $\mathbf{R}^d$ by

$$Mg(x) := \sup_{J\ni x} \frac{1}{|J|} \int_J |g(y)|\, dy$$

with the sup taken over all cubes J that contain x. It is well known that M is a bounded operator on $L_p(\mathbf{R}^d)$ for all $1 < p \le \infty$. The Fefferman–Stein inequality [FS] bounds the mapping M on sequences of functions. We shall only need the following special case of this inequality which says that for any functions $\eta(I,\cdot)$ and constants c_I, $I \in \mathcal{D}$, we have for $1 < p \le \infty$,

$$\left\| \left(\sum_{I\in\mathcal{D}} (c_I M\eta(I,\cdot))^2 \right)^{1/2} \right\|_p \le A \left\| \left(\sum_{I\in\mathcal{D}} (c_I \eta(I,\cdot))^2 \right)^{1/2} \right\|_p \tag{2.4}$$

with A a constant depending only on the space dimension d.

Consider now as an example, the equivalence of (2.2) in the univariate case. If the univariate functions $\eta(I,\cdot)$, $I \in \mathcal{D}$, satisfy

$$|\eta(I,x)| \le CM\chi_I(x), \qquad \chi_I(x) \le CM\eta(I,x), \qquad \text{a.e.} \quad x \in \mathbf{R}, \tag{2.5}$$

then using (2.4), we see that (2.2) holds if and only if (2.3) holds. The left inequality in (2.5) is a decay condition on $\eta(I,\cdot)$. For example, if $\eta(I,\cdot)$ is given by the normalized, shifted-dilates of the function ψ, $\eta(I,\cdot) = \psi_I$, then the left inequality in (2.5) holds whenever

$$|\psi(x)| \le C[\max(1, |x|)]^{-\lambda}, \qquad \text{a.e.} \quad x \in \mathbf{R},$$

with $\lambda \ge 1$. The right condition in (2.5) is extremely mild. For example, it is always satisfied in the case that the family $\eta(I,\cdot)$ is generated by the shifted dilates of a nonzero function ψ.

The Littlewood–Paley inequalitites are intimately connected with unconditional bases. Given a family of functions $\{\eta(I,\cdot)\}_{I\in\mathcal{D}}$ from $L_p(\mathbf{R}^d)$, we define its span X in $L_p(\mathbf{R}^d)$ as the $L_p(\mathbf{R}^d)$-closure of the linear space spanned by its finite linear combinations. The ordered family $\{\eta(I,\cdot)\}_{I\in\mathcal{D}}$ is a basis for X if each element $f \in X$ has a unique representation

$$f = \sum_{I\in\mathcal{D}} c_I \eta(I,\cdot). \tag{2.6}$$

In describing the convergence of the series (2.6), we should specify the ordering (i.e., the partial sums). We will only consider unconditional bases (described in a moment) in which order is not important. However, for completeness of the definition of a basis, we will take the partial sums S_n of the series (2.6) to consist of the sum over all rectangles $I = I_1 \times \cdots \times I_d$ such that $2^{-n} \le |I_j| \le 2^n$, $j = 1, \ldots, d$.

We recall that a basis $\eta(I, \cdot)$, $I \in \mathcal{D}(\mathbf{R}^d)$, is said to be unconditional for $L_p(\mathbf{R}^d)$ if for each assignment $\varepsilon_I := \pm 1$, $I \in \mathcal{D}(\mathbf{R}^d)$, and each finite sequence c_I, $I \in \mathcal{D}(\mathbf{R}^d)$, we have

$$\left\| \sum_{I \in \mathcal{D}} \varepsilon_I c_I \eta(I, \cdot) \right\|_p \approx \left\| \sum_{I \in \mathcal{D}} c_I \eta(I, \cdot) \right\|_p \tag{2.7}$$

with constants of equivalency independent of the sequences $(c_I)_{I \in \mathcal{D}}$ and $(\varepsilon_I)_{I \in \mathcal{D}}$.

If a basis is unconditional, then the upper estimate in (2.7) also holds for any sequence (ε_I) taking the values 0, 1. From this, it follows easily that the series (2.6) converges independently of the ordering.

We take for granted the known fact (see, e.g., [KS]) that for each $1 < p < \infty$, the univariate Haar family H_I, $I \in \mathcal{D}(\mathbf{R})$, satisfies the strong Littlewood–Paley property. These functions also form an unconditional basis for $L_p(\mathbf{R})$, for all $1 < p < \infty$; this can be found in [KS] and also follows from Lemma 2.2 below. We want next to conclude from this that the multivariate Haar system H_I, $I \in \mathcal{D}(\mathbf{R}^d)$, also satisfies the strong Littlewood–Paley property.

Let $r_j(t) := \operatorname{sign}(\sin 2^{j+1}\pi t)$, $t \in [0, 1]$, $j = 0, 1, \ldots$, be the univariate Rademacher functions. We take any one-to-one correspondence of the natural numbers with the rectangles of $\mathcal{D}(\mathbf{R})$. This gives an indexing $r(I, \cdot)$, $I \in \mathcal{D}(\mathbf{R})$, of the Rademacher functions. In $\mathbf{R}^d$, we let

$$r(I, (x_1, \ldots, x_d)) := r(I_1, x_1) \cdots r(I_d, x_d), \qquad I = I_1 \times \cdots \times I_d,$$

be the tensor products of the Rademacher functions. We recall Khinchine's inequality (see [KS]) which says that for $1 \le p < \infty$ and for any finite sequence c_I, $I \in \mathcal{D}(\mathbf{R}^d)$, we have

$$\left\| \sum_{I \in \mathcal{D}} c_I r(I, \cdot) \right\|_{L_p([0,1]^d)} \approx \left(\sum_{I \in \mathcal{D}} |c_I|^2 \right)^{1/2}. \tag{2.8}$$

Lemma 2.1. *Let $1 < p < \infty$ and let ψ be a univariate function such that the univariate family ψ_I, $I \in \mathcal{D}(\mathbf{R})$, is an unconditional basis for $L_p(\mathbf{R})$. Then the multivariate family ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, satisfies the Littlewood–Paley property* (2.2) *for this value of p.*

Proof. For notational simplicity, we give the proof only in the case $d = 2$. Let c_I, $I \in \mathcal{D}(\mathbf{R}^2)$, be a sequence with finitely many nonzero terms, and let $I = I_1 \times I_2$. Then, using the unconditionality of the univariate basis, for each $t_1, t_2 \in [0, 1]$ which are not endpoints of dyadic intervals, we have with constants of equivalency depending at most on p and ψ,

$$\begin{aligned} &\int_{\mathbf{R}} \int_{\mathbf{R}} \left| \sum_{I \in \mathcal{D}(\mathbf{R}^2)} c_I r(I, (t_1, t_2)) \psi_I(x_1, x_2) \right|^p dx_1\, dx_2 \\ &\quad = \int_{\mathbf{R}} \int_{\mathbf{R}} \left| \sum_{I_1 \in \mathcal{D}(\mathbf{R})} r(I_1, t_1) \left[\sum_{I_2 \in \mathcal{D}(\mathbf{R})} c_I r(I_2, t_2) \psi_{I_2}(x_2) \right] \psi_{I_1}(x_1) \right|^p dx_1\, dx_2 \end{aligned} \tag{2.9}$$

$$\approx \int_{\mathbf{R}}\int_{\mathbf{R}} \left| \sum_{I_1\in\mathcal{D}(\mathbf{R})} \left[\sum_{I_2\in\mathcal{D}(\mathbf{R})} c_I r(I_2, t_2)\psi_{I_2}(x_2) \right] \psi_{I_1}(x_1) \right|^p dx_1\, dx_2$$
$$= \int_{\mathbf{R}}\int_{\mathbf{R}} \left| \sum_{I_2\in\mathcal{D}(\mathbf{R})} r(I_2, t_2) \left[\sum_{I_1\in\mathcal{D}(\mathbf{R})} c_I \psi_{I_1}(x_1) \right] \psi_{I_2}(x_2) \right|^p dx_1\, dx_2$$
$$\approx \int_{\mathbf{R}}\int_{\mathbf{R}} \left| \sum_{I_2\in\mathcal{D}(\mathbf{R})} \left[\sum_{I_1\in\mathcal{D}(\mathbf{R})} c_I \psi_{I_1}(x_1) \right] \psi_{I_2}(x_2) \right|^p dx_1\, dx_2$$
$$= \int_{\mathbf{R}}\int_{\mathbf{R}} \left| \sum_{I\in\mathcal{D}(\mathbf{R}^2)} c_I \psi_I(x_1, x_2)) \right|^p dx_1\, dx_2.$$

We now integrate (2.9) with respect to $t_1, t_2 \in [0, 1]$ and interchange the order of integration in the first term. By Khinchine's inequalities (2.8) with respect to the norm in t_1, t_2, the first term of (2.9) is equivalent to

$$\int_{\mathbf{R}}\int_{\mathbf{R}} \left(\sum_{I\in\mathcal{D}(\mathbf{R}^2)} |c_I \psi_I(x_1, x_2)|^2 \right)^{p/2} dx_1\, dx_2.$$

Comparing this with the last term in (2.9), we see that we have proved the lemma. ■

It follows, in particular from Lemma 2.1, that the multivariate Haar functions H_I, $I \in \mathcal{D}(\mathbf{R}^d)$, satisfy the strong Littlewood–Paley properties (2.2) and (2.3) (note that $|H_I| = \chi_I$).

Lemma 2.2. *Let* $1 < p < \infty$ *and let* $\eta(I, \cdot)$, $I \in \mathcal{D}(\mathbf{R}^d)$, *be any collection of multivariate functions. Concerning the following statements*:

(i) $(\eta(I, \cdot))_{I\in\mathcal{D}}$ *satisfies the Littlewood–Paley condition* (2.2) *for this value of* p;
(ii) $(\eta(I, \cdot))_{I\in\mathcal{D}}$ *satisfies the Littlewood–Paley condition* (2.3) *for this value of* p;
(iii) $(\eta(I, \cdot))_{I\in\mathcal{D}} \approx (H_I)_{I\in\mathcal{D}}$; *and*
(iv) $(\eta(I, \cdot))_{I\in\mathcal{D}}$ *is an unconditional basis for* $L_p(\mathbf{R}^d)$;

we have that (i) *and* (iv) *are equivalent,* (ii) *and* (iii) *are equivalent, and* (ii) *implies* (iv). *Moreover, if* (2.5) *holds, then all these statements are equivalent.*

Proof. We leave the proof of this lemma to the reader. ■

3. Approximation by Hyperbolic Wavelets

In this section, we will discuss approximation in $L_p(\mathbf{R}^d)$, $1 < p < \infty$, from the hyperbolic wavelet spaces $\mathcal{H}_n := \mathcal{H}_n(L_p(\mathbf{R}^d))$. Let ψ be a univariate function and let ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, be defined as in (2.1) and let $\mathcal{H}_n$ be the closed linear span of the finite linear combinations of the ψ_I with $|I| > 2^{-n}$. For $n = 0, 1, \ldots$, and $f \in L_p(\mathbf{R}^d)$, we define

$$E_n(f)_p := E(f, \mathcal{H}_n)_p := \inf_{g \in \mathcal{H}_n} \|f - g\|_p$$

with $\|\cdot\|_p$ here and later the $L_p(\mathbf{R}^d)$ norm.

Our main interest in this paper is the characterization of the approximation spaces $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ defined by (1.5). We will characterize the spaces $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ by proving Bernstein and Jackson inequalities for $L_p(\mathbf{R}^d)$ and $\mathbf{W}^r(L_p(\mathbf{R}^d))$ (recall the spaces $\mathbf{W}^r(L_p(\mathbf{R}^d))$ and the operator $\mathbf{D}^r$ of the Introduction). We will give two approaches to proving companion Jackson and Bernstein inequalities in this section. The first approach assumes certain conditions on ψ that allow us to characterize $\mathbf{W}^r(L_p(\mathbf{R}^d))$, $1 < p < \infty$, in terms of expansions in the basis ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$. Using this characterization, we will then easily prove the Jackson and Bernstein inequalities. Our second approach will assume weaker (and more easily verifiable conditions on ψ) that still allow the proof of the Jackson and Bernstein inequalities. We begin with the first approach.

If ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, is a Schauder basis, then associated to this basis we have its dual basis. In the case that $1 < p < \infty$, the dual basis is given by linear functionals c_I with

$$c_I(f) = \int_{\mathbf{R}^d} f(x)\lambda(I, x)\, dx$$

and the functions $\lambda(I, \cdot)$ are in $L_{p'}(\mathbf{R}^d)$ with $1/p + 1/p' = 1$.

If $\lambda(x) := \lambda([0, 1], x)$, then it is easy to see (by using shifts and dilations) that we can take $\lambda(I, \cdot) = \lambda_I$, $I \in \mathcal{D}(\mathbf{R}^d)$, with the λ_I defined as in (2.1). We note that in the case that ψ is suitably differentiable, we have $(\mathbf{D}^r\psi)_I = |I|^r \mathbf{D}^r \psi_I$, $I \in \mathcal{D}(\mathbf{R}^d)$. We will make for our first approach the following assumptions about the multivariate basis ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, and its dual basis λ_I, $I \in \mathcal{D}(\mathbf{R}^d)$:

- (A1) ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, span $L_p(\mathbf{R}^d)$, $1 < p < \infty$, and satisfy the Littlewood–Paley inequalities (2.3);
- (A2) $(\mathbf{D}^r\psi)_I$, $I \in \mathcal{D}(\mathbf{R}^d)$, span $L_p(\mathbf{R}^d)$, $1 < p < \infty$, and satisfy the Littlewood–Paley inequalities (2.3);
- (A3) $\int_{\mathbf{R}} x^j \lambda(x)\, dx = 0$, $j = 0, \ldots, r$; and
- (A4) $|\lambda(x)| \le C \max(1, |x|^{-r-1-\varepsilon})$, for some $\varepsilon > 0$.

Because of (A3) and (A4), we can integrate the univariate function λ, r times to find a function $\mu \in L_{p'}(\mathbf{R})$ which satisfies $(-1)^r \mu^{(r)} = \lambda$. It follows that $\mathbf{D}^r \mu_I = (-1)^{rd}|I|^{-r}\lambda_I$, $I \in \mathcal{D}(\mathbf{R}^d)$. Integration by parts then shows that

$$\int_{\mathbf{R}^d} (\mathbf{D}^r\psi)_I \mu_J\, dx = |I|^r |J|^{-r} \int_{\mathbf{R}^d} \psi_I \lambda_J\, dx = \delta(I, J), \qquad I, J \in \mathcal{D}(\mathbf{R}^d),$$

with δ the Kronecker delta. Hence, μ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, is the dual basis for $(\mathbf{D}^r\psi)_I$, $I \in \mathcal{D}(\mathbf{R}^d)$.

Theorem 3.1. *Let r be a positive integer, $1 < p < \infty$, and let ψ be a univariate function which satisfies assumptions* (A1)–(A4). *Then a function $f \in L_p(\mathbf{R}^d)$ is in $\mathbf{W}^r(L_p(\mathbf{R}^d))$ if and only if*

$$f = \sum_{I \in \mathcal{D}(\mathbf{R}^d)} c_I(f)\psi_I \tag{3.1}$$

with

$$\left\| \left(\sum_{I\in\mathcal{D}(\mathbf{R}^d)} [|I|^{-r}|c_I(f)|\chi_I]^2 \right)^{1/2} \right\|_p < \infty. \tag{3.2}$$

Furthermore, the left side of (3.2) *is equivalent to* $|f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}$.

Proof. Suppose first that $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$. Assumption (A1) gives that the functions ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, satisfy the strong Littlewood–Paley inequalities. Hence, these functions are an unconditional basis for $L_p(\mathbf{R}^d)$ and

$$f = \sum_{I\in\mathcal{D}(\mathbf{R}^d)} c_I(f)\psi_I$$

with

$$c_I(f) = \int_{\mathbf{R}^d} f\lambda_I\, dx.$$

Likewise, the functions $(\mathbf{D}^r\psi)_I$, $I \in \mathcal{D}(\mathbf{R}^d)$, are also a basis for $L_p(\mathbf{R}^d)$, and we have

$$\mathbf{D}^r f = \sum_{I\in\mathcal{D}(\mathbf{R}^d)} d_I(f)(\mathbf{D}^r\psi)_I$$

with

$$d_I(f) = \int_{\mathbf{R}^d} \mathbf{D}^r f\mu_I\, dx = (-1)^{rd}\int_{\mathbf{R}^d} f\mathbf{D}^r\mu_I\, dx = |I|^{-r}\int_{\mathbf{R}^d} f\lambda_I\, dx = |I|^{-r}c_I(f). \tag{3.3}$$

We can compute $\|\mathbf{D}^r f\|_p$ from the Littlewood–Paley condition for the basis $(\mathbf{D}^r\psi)_I$, $I \in \mathcal{D}(\mathbf{R}^d)$. This gives that the left side of (3.2) is equivalent to $\|\mathbf{D}^r f\|_p$.

Conversely, assume that $f \in L_p(\mathbf{R}^d)$ is such that (3.2) is finite. Because ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, is an unconditional basis for $L_p(\mathbf{R}^d)$, we have

$$f = \sum_{I\in\mathcal{D}(\mathbf{R}^d)} c_I(f)\psi_I$$

in the sense of $L_p(\mathbf{R}^d)$-convergence. From (3.2) and the fact that $(\mathbf{D}^r\psi)_I$, $I \in \mathcal{D}(\mathbf{R}^d)$, satisfies the Littlewood–Paley inequalities, we find that there is a function $g \in L_p(\mathbf{R}^d)$ with

$$g = \sum_{I\in\mathcal{D}(\mathbf{R}^d)} |I|^{-r}c_I(f)(\mathbf{D}^r\psi)_I$$

again in the sense of $L_p(\mathbf{R}^d)$-convergence. We compute the coefficients of g with respect to the basis $(\mathbf{D}^r\psi)_I)$, $I \in \mathcal{D}(\mathbf{R}^d)$, and find

$$\int_{\mathbf{R}^d} g\mu_I = |I|^{-r}c_I(f) = |I|^{-r}\int_{\mathbf{R}^d} f\lambda_I = (-1)^{rd}\int_{\mathbf{R}^d} f\mathbf{D}^r\mu_I.$$

This shows that g is the distributional derivative $\mathbf{D}^r f$ and therefore $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$. ■

The following theorem gives the Jackson and Bernstein inequalities for approximation by the spaces $\mathcal{H}_n$.

Theorem 3.2. *Let r be a positive integer, $1 < p < \infty$, and let ψ be a univariate function which satisfies assumptions* (A1)–(A4). *If $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$, then*

$$E_n(f)_p \le C2^{-nr}|f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}, \qquad n = 0, 1, \dots, \tag{3.4}$$

with C independent of n and f. If $g \in \mathcal{H}_n(L_p(\mathbf{R}^d))$, then

$$|g|_{\mathbf{W}^r(L_p(\mathbf{R}^d))} \le C2^{nr}\|g\|_p, \qquad n = 0, 1, \dots, \tag{3.5}$$

with C independent of n and g.

Proof. First, let $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$. Then

$$f = \sum_{I \in \mathcal{D}(\mathbf{R}^d)} c_I(f)\psi_I$$

in the sense of $L_p(\mathbf{R}^d)$-convergence and (3.2) is satisfied. We let $g := \sum_{|I|>2^{-n}} c_I(f)\psi_I$ which is a function in $\mathcal{H}_n(L_p(\mathbf{R}^d))$. The remainder $f - g$ is given by

$$f - g = \sum_{|I| \le 2^{-n}} c_I(f)\psi_I.$$

We can estimate $\|f - g\|_p$ by using the Littlewood–Paley inequalities:

$$\begin{aligned}\|f-g\|_p &\le C\left\|\left(\sum_{|I|\le 2^{-n}}[|c_I(f)|\chi_I]^2\right)^{1/2}\right\|_p \le C2^{-nr}\left\|\left(\sum_{|I|\le 2^{-n}}[|I|^{-r}|c_I(f)|\chi_I]^2\right)^{1/2}\right\|_p \\ &\le C2^{-nr}\left\|\left(\sum_{I\in\mathcal{D}(\mathbf{R}^d)}[|I|^{-r}|c_I(f)|\chi_I]^2\right)^{1/2}\right\|_p \le C2^{-nr}\|\mathbf{D}^r f\|_p.\end{aligned}$$

This proves (3.4).

Suppose now that $g \in \mathcal{H}_n(L_p(\mathbf{R}^d))$. Then,

$$g = \sum_{|I|>2^{-n}} c_I(g)\psi_I,$$

and from Theorem 3.1, we have

$$\begin{aligned}\|\mathbf{D}^r g\|_p &\le C\left\|\left(\sum_{|I|>2^{-n}}[|I|^{-r}|c_I(g)|\chi_I]^2\right)^{1/2}\right\|_p \\ &\le C2^{nr}\left\|\left(\sum_{|I|>2^{-n}}[c_I(g)|\chi_I]^2\right)^{1/2}\right\|_p \le C2^{nr}\|g\|_p\end{aligned}$$

because ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, satisfies the Littlewood–Paley inequalities (2.3). This proves (3.5). ∎

As we have mentioned earlier in this section, the Jackson and Bernstein inequalities (3.4) and (3.5) allow the characterization of the approximation spaces $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$. For this, we will use the K-functional

$$
\begin{aligned}
(3.6) \qquad K(f,t) &:= K(f,t,L_p(\mathbf{R}^d),\mathbf{W}^r(L_p(\mathbf{R}^d))) \\
&:= \inf_{g\in\mathbf{W}^r(L_p(\mathbf{R}^d))} \|f-g\|_p + t|g|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}.
\end{aligned}
$$

We recall the interpolation spaces $(L_p(\mathbf{R}^d),\mathbf{W}^r(L_p(\mathbf{R}^d)))_{\theta,q}$ defined by the real method of interpolation (see Chapter 6 of [DL]).

Corollary 3.3. *Let r be a positive integer, $1<p<\infty$, $0<q\le\infty$, and $0<\alpha<r$. Let ψ be a univariate function which satisfies assumptions* (A1)–(A4). *Then, $f\in\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$ if and only if $f\in(L_p(\mathbf{R}^d),\mathbf{W}^r(L_p(\mathbf{R}^d)))_{\alpha/r,q}$ with equivalent norms.*

Proof. This corollary follows from Theorem 3.2 and general facts about approximation and K-functionals that can be found in Chapter 7 of [DL]. ■

While the above approach is simple and direct, the assumption (A2) is too severe for some applications. It is also uncomfortable to make such an assumption for a Jackson inequality since it is unclear why the Jackson inequality should depend on the smoothness of ψ. Recall that in the univariate case of wavelet approximation, Jackson inequalities depend only on conditions (A1), (A3), and (A4). We shall therefore now give a second approach to proving the Jackson and Bernstein inequalities which separates the proof of these inequalities. This approach allows us to prove the Jackson inequality under a much weaker assumption than (A2).

We first consider the Jackson inequality. Suppose that we have in hand two multivariate families $\eta(I,\cdot)$, $\mu(I,\cdot)$, $I\in\mathcal{D}(\mathbf{R}^d)$. We will use the notation $\{\eta(I,\cdot)\}_{I\in\mathcal{D}}\prec\{\mu(I,\cdot)\}_{I\in\mathcal{D}}$, if there is a constant $C>0$ such that

$$
(3.7) \qquad \left\|\sum_{I\in\mathcal{D}} c_I\eta(I,\cdot)\right\|_p \le C\left\|\sum_{I\in\mathcal{D}} c_I\mu(I,\cdot)\right\|_p
$$

holds for all finite sequences $(c_I)_{I\in\mathcal{D}}$ with C independent of the sequence. If $\{\eta(I,\cdot)\}_{I\in\mathcal{D}}\prec\{\mu(I,\cdot)\}_{I\in\mathcal{D}}$ and $\{\mu(I,\cdot)\}_{I\in\mathcal{D}}\prec\{\eta(I,\cdot)\}_{I\in\mathcal{D}}$, then we write $\{\eta(I,\cdot)\}_{I\in\mathcal{D}}\approx\{\mu(I,\cdot)\}_{I\in\mathcal{D}}$.

Given two multidimensional families $\eta(I,\cdot)$, $\mu(I,\cdot)$, $I\in\mathcal{D}(\mathbf{R}^d)$, we define the operator T which maps $\mu(I,\cdot)$ into $\eta(I,\cdot)$ for all $I\in\mathcal{D}$ and we extend T to finite linear combinations of the $\mu(I,\cdot)$ by linearity. Then (3.7) holds if and only if T is a bounded operator with respect to the L_p-norm and $\{\mu(I,\cdot)\}_{I\in\mathcal{D}}\prec\{\eta(I,\cdot)\}_{I\in\mathcal{D}}$ holds if and only if T has a bounded inverse with respect to the L_p-norm.

We recall the Haar basis H_I, $I\in\mathcal{D}(\mathbf{R}^d)$. In place of (A2), we will assume

(A2′) $\{\mu_I\}_{I\in\mathcal{D}(\mathbf{R}^d)}\prec\{H_I\}_{I\in\mathcal{D}(\mathbf{R}^d)}$,

where as before μ satisfies $\mu^{(r)}=(-1)^r\lambda$. It follows from (A2′) that the operator T defined by

$$
Tf := \sum_{I\in\mathcal{D}(\mathbf{R}^d)} \langle f,H_I\rangle\mu_I
$$

is bounded on $L_p(\mathbf{R}^d))$ for each $1 < p < \infty$. Hence, by duality, its adjoint

$$T^* f := \sum_{I \in \mathcal{D}(\mathbf{R}^d)} \langle f, \mu_I \rangle H_I$$

is also bounded on $L_p(\mathbf{R}^d))$, for each $1 < p < \infty$.

Theorem 3.4. *Assume that* (A1), (A2′), (A3), *and* (A4) *hold. If* $1 < p < \infty$, r *is a positive integer and* $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$, *then*

$$E_n(f)_p \le C 2^{-nr} |f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}, \qquad n = 0, 1, \dots, \tag{3.8}$$

with C *independent of* n *and* f.

Proof. Let $f \in \mathbf{W}^r(L_p(\mathbf{R}^d))$. From assumption (A1), we have (see (3.3))

$$f = \sum_{I \in \mathcal{D}(\mathbf{R}^d)} c_I(f) \psi_I, \qquad c_I(f) = \langle f, \lambda_I \rangle = |I|^r \langle \mathbf{D}^r f, \mu_I \rangle,$$

in the sense of $L_p(\mathbf{R}^d)$-convergence. We let $g := \sum_{|I| > 2^{-n}} c_I(f) \psi_I$ which is a function in $\mathcal{H}_n(L_p(\mathbf{R}^d))$. We can estimate the remainder $f - g$ by

$$\begin{aligned}
\|f - g\|_p &= \left\| \sum_{|I| \le 2^{-n}} c_I(f) \psi_I \right\|_p \le C \left\| \sum_{|I| \le 2^{-n}} c_I(f) H_I \right\|_p \\
&= C \left\| \sum_{|I| \le 2^{-n}} |I|^r \langle \mathbf{D}^r f, \mu_I \rangle H_I \right\|_p \le C \left\| \left(\sum_{|I| \le 2^{-n}} [|I|^r \langle \mathbf{D}^r f, \mu_I \rangle \chi_I]^2 \right)^{1/2} \right\|_p \\
&\le C 2^{-nr} \left\| \left(\sum_{|I| \le 2^{-n}} [\langle \mathbf{D}^r f, \mu_I \rangle \chi_I]^2 \right)^{1/2} \right\|_p \le C 2^{-nr} \left\| \sum_{I \in \mathcal{D}(\mathbf{R}^d)} \langle \mathbf{D}^r f, \mu_I \rangle H_I \right\|_p \\
&\le C 2^{-nr} \|\mathbf{D}^r f\|_p,
\end{aligned}$$

where the last inequality uses the boundedness of the adjoint operator T^* (which follows from (A2′)). ■

We next consider the Bernstein inequality. In place of (A2), we will assume that

(A2″) $\{(\mathbf{D}^r \psi)_I\}_{I \in \mathcal{D}(\mathbf{R}^d)} \prec \{H_I\}_{I \in \mathcal{D}(\mathbf{R}^d)}$.

Theorem 3.5. *Assume that* (A1), (A2″), (A3), *and* (A4) *hold. Then, for each positive integer* r *and each* $g \in \mathcal{H}_n(L_p(\mathbf{R}^d))$, $1 < p < \infty$, *we have*

$$\|\mathbf{D}^r g\|_p \le C 2^{nr} \|g\|_p, \qquad n = 0, 1, \dots, \tag{3.9}$$

with C *independent of* n *and* g.

Proof. Because of (A1), we can write

$$g = \sum_{|I|>2^{-n}} c_I(g)\psi_I. \tag{3.10}$$

In the proof of (3.9), it is enough to consider functions g for which the sum in (3.10) has only a finite number of terms; the general case follows by a limiting argument. In this case, we have

$$\begin{aligned}
\|\mathbf{D}^r g\|_p &= \left\| \sum_{|I|>2^{-n}} c_I(g)\mathbf{D}^r\psi_I \right\|_p = \left\| \sum_{|I|>2^{-n}} |I|^{-r} c_I(g)(\mathbf{D}^r\psi)_I \right\|_p \\
&\le C \left\| \sum_{|I|>2^{-n}} |I|^{-r} c_I(g) H_I \right\|_p \le C \left\| \sum_{|I|>2^{-n}} |I|^{-r} c_I(g)\psi_I \right\|_p \\
&\le C2^{nr} \left\| \sum_{|I|>2^{-n}} c_I(g)\psi_I \right\|_p = C2^{nr}\|g\|_p.
\end{aligned}$$

■

4. Sufficient Conditions for the Littlewood–Paley Inequalities

In the previous section, we have characterized the approximation spaces for hyperbolic wavelet approximation under certain assumptions on the univariate functions ψ and $\psi^{(r)}$ relating to Littlewood–Paley theory. For many functions ψ that occur in wavelet theory, it is possible to utilize the existing Littlewood–Paley theory to verify these assumptions. However, in some instances (e.g., for spline wavelets), the application of the existing theory will not give the largest possible value of r because this theory requires global smoothness of ψ (respectively $\psi^{(r)}$). The purpose of the present section is to prove a Littlewood–Paley theorem which does not require global smoothness (rather it is enough to have certain piecewise continuity). We shall also address some related questions associated with Littlewood–Paley theory.

It is possible to formulate our theorems without assuming that the family of functions under consideration are all shifted-dilates of a single function. We shall therefore revert back to our notation $\eta(I,\cdot)$, $I \in \mathcal{D} := \mathcal{D}(\mathbf{R}^1)$, to denote an arbitrary family of functions indexed on dyadic intervals.

The strong Littlewood–Paley inequalities (2.3) are the same as the equivalence $\{\eta(I,\cdot)\} \approx \{H_I\}$. We begin this section by discussing sufficient conditions in order that $\{\eta(I,\cdot)\} \prec \{H_I\}$. Let ξ_I, $I \in \mathcal{D}$, denote the center of the dyadic interval I. We will assume in this section that $\eta(I,\cdot)$, $I \in \mathcal{D}$, is a family of univariate functions that satisfy the following assumptions:

(A5) There is an $\varepsilon > 0$, and a constant C_1 such that for all $t \in \mathbf{R}$ and all $J \in \mathcal{D}$, we have

$$|\eta(J, \xi_J + t|J|)| \le C_1 |J|^{-1/2}(1+|t|)^{-1-\varepsilon}.$$

(A6) There is an $\varepsilon > 0$ and a constant C_2 and a partition of $[-\frac{1}{2},\frac{1}{2}]$ into intervals $J_1,\dots,J_m$ that are dyadic with respect to $[-\frac{1}{2},\frac{1}{2}]$, such that for any $J \in \mathcal{D}$, any $j \in \mathbf{Z}$, and any t_1, t_2 in the interior of the same interval J_k, $k = 1,\dots,m$, we have

$$|\eta(J,\xi_J+j|J|+t_1|J|)-\eta(J,\xi_J+j|J|+t_2|J|)| \le C_2|J|^{-1/2}(1+|j|)^{-1-\varepsilon}|t_2-t_1|^{\varepsilon},$$

(A7) For any $J \in \mathcal{D}$, we have

$$\int_{\mathbf{R}} \eta(J,x)\,dx = 0.$$

In the case that $\eta(J,\cdot) = \psi_J$ for a function ψ, it is enough to check these assumptions for $J = [0,1]$, i.e., for the function ψ alone. They follow for all other dyadic intervals J by dilation and translation.

The condition (A5) is a standard decay assumption and (A7) is the zero moment condition. The condition (A6) requires that the functions $\eta(I,\cdot)$ be piecewise in Lip ε.

Let T be the linear operator which satisfies

$$T\left(\sum_{I\in\mathcal{D}} c_I H_I\right) = \sum_{I\in\mathcal{D}} c_I \eta(I,\cdot) \tag{4.1}$$

for each finite linear combination $\sum_{I\in\mathcal{D}} c_I H_I$ of the H_I. We wish to show that

$$\left\| T\left(\sum_{I\in\mathcal{D}} c_I H_I\right)\right\|_p \le C \left\|\sum_{I\in\mathcal{D}} c_I H_I\right\|_p$$

for each such sum. From this it would follow that T extends (by continuity) to a bounded operator on all of $L_p(\mathbf{R})$ and therefore $\{\eta(I,\cdot)\} \prec \{H_I\}$.

We can expand $\eta(J,\cdot)$ into its Haar decomposition. Let

$$\lambda(I,J) := \int_{\mathbf{R}} \eta(J,x) H_I(x)\,dx, \tag{4.2}$$

so that

$$\eta(J,\cdot) = \sum_{I\in\mathcal{D}} \lambda(I,J) H_I.$$

It follows that

$$T\left(\sum_{J\in\mathcal{D}} c_J H_J\right) = \sum_{I\in\mathcal{D}} \sum_{J\in\mathcal{D}} \lambda(I,J) c_J H_I. \tag{4.3}$$

We see that the mapping T is tied to the bi-infinite matrix $\Lambda := (\lambda(I,J))_{I,J\in\mathcal{D}}$ which maps the sequence $c := (c_J)$ into the sequence

$$(c_I') := \Lambda c.$$

One approach to proving Littlewood–Paley inequalities is to show that the matrix Λ decays sufficiently fast away from the diagonal (see [FJ, §3]). Following [FJ], we say that a matrix $A = (a(I,J))_{I,J\in\mathcal{D}}$ is *almost diagonal* if for some $\varepsilon > 0$, we have

$$|a(I,J)| \le C\omega(I,J) \tag{4.4}$$

with

$$\omega(I,J) := \left(1 + \frac{|\xi_I - \xi_J|}{\max(|I|,|J|)}\right)^{-1-\varepsilon} \left(\min\left(\frac{|I|}{|J|},\frac{|J|}{|I|}\right)\right)^{(1+\varepsilon)/2}. \tag{4.5}$$

We will use the following special case of a theorem of Frazier and Jawerth [FJ, Theorem 3.3] concerning almost diagonal operators.

Theorem 4.1. *If* $(a(I,J))_{I,J\in\mathcal{D}}$ *is an almost diagonal matrix, then the operator* A *defined by*

$$A\left(\sum_{J\in\mathcal{D}} c_J H_J\right) := \sum_{I\in\mathcal{D}}\sum_{J\in\mathcal{D}} a(I,J)c_J H_I \tag{4.6}$$

is bounded on $L_p(\mathbf{R})$ *for each* $1 < p < \infty$.

Proof. For the completeness of the present paper, we give the following proof of this theorem. Let W be the operator defined as in (4.6) with $(\omega(I,J))_{I,J\in\mathcal{D}}$ in place of $(a(I,J))_{I,J\in\mathcal{D}}$. From the Littlewood–Paley inequalities for the Haar functions, we have

$$\begin{aligned}\left\|A\left(\sum_{J\in\mathcal{D}} c_J H_J\right)\right\|_p &\le C\left\|\sum_{I\in\mathcal{D}}\sum_{J\in\mathcal{D}} |a(I,J)||c_J|H_I\right\|_p \le C\left\|\sum_{I\in\mathcal{D}}\sum_{J\in\mathcal{D}} \omega(I,J)|c_J|H_I\right\|_p \\ &\le C\|W\|\left\|\sum_{J\in\mathcal{D}} |c_J|H_J\right\|_p \le C\|W\|\left\|\sum_{J\in\mathcal{D}} c_J H_J\right\|_p\end{aligned}$$

with $\|W\|$ the norm of W as an operator from $L_p(\mathbf{R})$ into itself and with the constants C depending only on p. Thus, it is sufficient to show that $\|W\|$ is finite.

We write

$$\omega(I,J) = \omega^+(I,J) + \omega^-(I,J),$$

with

$$\omega^+(I,J) := \begin{cases}\omega(I,J), & |J| \ge |I|,\\ 0, & |J| < |I|,\end{cases}$$

and we let W^+ and W^- be defined as in (4.6) for these two sequences. We shall show that W^+ and W^- are bounded on $L_p(\mathbf{R})$ which will complete the proof of the theorem. Since the proof of boundedness in these two cases is similar, we shall only consider W^+.

We will employ the following inequality for nonnegative sequences (b_ℓ):

$$\sum_{\ell=1}^{\infty} b_\ell \ell^{-\tau} \le C_\tau \max_{m\ge 1}\frac{1}{m}\sum_{\ell=1}^{m} b_\ell, \qquad \tau > 1,$$

which is easily proved by summation by parts.

To bound W^+, it is enough to consider its action on $\sum_{J\in\mathcal{D}} c_J H_J$ where the c_J are nonnegative and only a finite number of them are nonzero. We let

$$c'_I := \sum_{J\in\mathcal{D}} \omega^+(I,J)c_J.$$

From the above inequality, it follows that for each interval I with $|I| = 2^\nu$ and $\mu \ge \nu$, we have

$$\begin{aligned}\sum_{|J|=2^\mu}\left(1 + \frac{|\xi_I - \xi_J|}{\max(|I|,|J|)}\right)^{-1-\varepsilon} c_J &\le C2^{\mu/2}M\left(\sum_{|J|=2^\mu} c_J\chi_J, x\right) \\ &= C2^{\mu/2}M(f_\mu, x), \qquad x\in I,\end{aligned}$$

with M the Hardy–Littlewood maximal operator, and with

$$f_\mu := \sum_{|J|=2^\mu} c_J \chi_J = \left(\sum_{|J|=2^\mu} c_J^2 \chi_J^2 \right)^{1/2},$$

and with C here and later in this proof depending on ε. Hence, from (4.5)

$$|I|^{-1/2}|c_I'| \le C \sum_{\mu \ge \nu} 2^{-\mu\varepsilon/2} 2^{\nu\varepsilon/2} M(f_\mu, x), \qquad x \in I.$$

Using the Cauchy–Schwartz inequality, we obtain

$$|c_I'|^2 \chi_I^2(x) \le C \sum_{\mu \ge \nu} 2^{(\nu-\mu)\varepsilon/2} (M(f_\mu, x))^2.$$

Since the functions χ_I^2 have disjoint supports, we have

$$\sum_{|I|=2^\nu} |c_I'|^2 \chi_I^2 \le C \sum_{\mu \ge \nu} 2^{(\nu-\mu)\varepsilon/2} (M(f_\mu))^2.$$

Summing over $\nu \in \mathbf{Z}$ gives

$$\sum_{I \in \mathcal{D}} |c_I'|^2 \chi_I^2 \le C \sum_{\nu \in \mathbf{Z}} \sum_{\mu \ge \nu} 2^{(\nu-\mu)\varepsilon/2} (M(f_\mu))^2.$$

We now apply the Littlewood–Paley inequalities for the Haar system and the Fefferman–Stein inequality (2.4) to find

$$\begin{aligned} \left\| W^+ \left(\sum_{I \in \mathcal{D}} c_I H_I \right) \right\|_p &\le C \left\| \left(\sum_{I \in \mathcal{D}} |c_I'|^2 \chi_I^2 \right)^{1/2} \right\|_p \le C \left\| \left(\sum_{\nu \in \mathbf{Z}} \sum_{\mu \ge \nu} 2^{(\nu-\mu)\varepsilon/2} f_\mu^2 \right)^{1/2} \right\|_p \\ &\le C \left\| \left(\sum_{\mu \in \mathbf{Z}} f_\mu^2 \right)^{1/2} \right\|_p \le C \left\| \sum_{I \in \mathcal{D}} c_I H_I \right\|_p, \end{aligned}$$

where the last inequality follows from the Littlewood–Paley inequalities (2.3) for the Haar functions and the definition of the functions f_μ. ■

Theorem 4.2. *If $\eta(I, \cdot)$, $I \in \mathcal{D}$, satisfy assumptions* (A5)–(A7), *then the operator T defined by* (4.1) *is bounded from $L_p(\mathbf{R})$ into itself for each $1 < p < \infty$.*

Proof. For an interval $I \in \mathcal{D}$, let $I_\pm$ denote the dyadic intervals with $|I_\pm| = |I|$, which are immediately to the right and left of I, respectively. We define $I^* := I_- \cup I \cup I_+$. For the $\lambda(I, J)$ of (4.2), we define

$$\lambda'(I, J) := \begin{cases} \lambda(I, J), & |I| \le |J|, \\ 0, & \text{else}, \end{cases}$$

$$\lambda''(I, J) := \begin{cases} \lambda(I, J), & |I| > |J|,\ J \not\subseteq I^*, \\ 0, & \text{else}, \end{cases}$$

$$\lambda'''(I, J) := \begin{cases} \lambda(I, J), & |I| > |J|,\ J \subset I^*, \\ 0, & \text{else}. \end{cases}$$

Then,

$$\lambda(I, J) = \lambda'(I, J) + \lambda''(I, J) + \lambda'''(I, J).$$

We let T_1, T_2, T_3 be the operators defined as in (4.3) with λ replaced by $\lambda', \lambda'', \lambda'''$, respectively. We will show that each of these operators is bounded from $L_p(\mathbf{R})$ into itself.

We first show that T_1 satisfies the conditions of Theorem 4.1. Let $2^{-\alpha}$ be the length of the smallest dyadic interval appearing in the statement of condition (A6). We first consider intervals I such that $|I| \le 2^{-\alpha}|J|$. Then from property (A6), there is a constant a such that

$$|\eta(J, x) - a| \le C_2|J|^{-1/2}\left(1 + \frac{|\xi_I - \xi_J|}{|J|}\right)^{-1-\varepsilon}\left(\frac{|I|}{|J|}\right)^{\varepsilon}, \qquad x \in \text{ int } (I).$$

Indeed, for one of the intervals J_k of assumption (A6) and an appropriately chosen j, any $x \in I$ can be written as $x = \xi_J + j|J| + t|J|$ with $t \in J_k$. Using the last inequality, we obtain

$$\begin{aligned} |\lambda'(I, J)| \ &\le \ \left|\int_I H_I(x)(\eta(J, x) - a)\, dx\right| \\ &\le \ C_2|I|^{1/2}|J|^{-1/2}\left(1 + \frac{|\xi_I - \xi_J|}{|J|}\right)^{-1-\varepsilon}\left(\frac{|I|}{|J|}\right)^{\varepsilon}. \end{aligned}$$

By using (A5), we see that this last inequality is also valid if $|J| \ge |I| > 2^{-\alpha}|J|$ because the term $(|I|/|J|)^{\varepsilon}$ can be replaced by 1. This shows that $(\lambda'(I, J))_{I,J\in\mathcal{D}}$ is almost diagonal and hence the boundedness of T_1 on $L_p(\mathbf{R})$ follows from Theorem 4.1.

A similar calculation shows that $(\lambda''(I, J))_{I,J\in\mathcal{D}}$ is almost diagonal and hence T_2 is also bounded on $L_p(\mathbf{R})$ because of Theorem 4.1.

The proof of the boundedness of T_3 will require more care since this operator is not necessarily almost diagonal. We can decompose this operator into a sum of four operators corresponding to whether J is contained in the left- or right-half of I, or J is contained in I_+ or I_-. We can show that each of these operators is bounded on $L_p(\mathbf{R})$ in a similar way. We consider in detail the operator A_0 corresponding to the left-half of I and show that it is bounded on L_p, $1 < p < \infty$. Later we shall note the modifications that need to be made to handle the remaining three cases.

We denote the left-half of intervals $I \in \mathcal{D}$ by I'. Then A_0 has the associated matrix

$$\lambda_0(I, J) : \begin{cases} \lambda(I, J), & J \subset I', \\ 0, & \text{else.} \end{cases}$$

We can write

$$A_0 = \sum_{m=1}^{\infty} A_m,$$

where A_m has the associated matrix

$$\lambda_m(I, J) := \begin{cases} \lambda(I, J), & |J| = 2^{-m}|I|, \ J \subset I', \\ 0, & \text{else.} \end{cases}$$

We will show that for a certain $\delta > 0$ (depending on p),

$$\|A_m\| \le C2^{-m\delta}, \tag{4.7}$$

with $\|\cdot\|$ denoting the norm of the operator A_m from $L_p(\mathbf{R})$ into itself, and with C independent of m. This will then complete the proof of the theorem.

We recall that an operator on $L_p(\mathbf{R})$ has the same norm as its adjoint on $L_{p'}$, $1/p + 1/p' = 1$. It is therefore enough to show that the adjoint operators A_m^* satisfy (4.7) for each $1 < p < \infty$. The operator A_m^* is defined by

$$A_m^*\left(\sum_{J\in\mathcal{D}} c_J H_J\right) = \sum_{J\in\mathcal{D}} \sum_{\substack{|I|=2^m|J|\\ I'\supset J}} \lambda_m(I, J) c_I H_J.$$

We first estimate $\lambda_m(I, J)$, $|I| = 2^m|J|$, $J \subset I'$. Let $\mathcal{C}(I')$ denote the complement of I'. Using assumptions (A7) and (A5), we have

$$\begin{aligned}
|\lambda_m(I, J)| &= \left|\int_{\mathbf{R}} \eta(J, x) H_I(x)\, dx\right| \\
&\le |I|^{-1/2}\left(\left|\int_{I'} \eta(J, x)\, dx\right| + \left|\int_{I\setminus I'} \eta(J, x)\, dx\right|\right) \\
&\le 2|I|^{-1/2}\int_{\mathcal{C}(I')} |\eta(J, x)|\, dx \\
&\le 2C_1|I|^{-1/2}|J|^{-1/2}\int_{\mathcal{C}(I')}\left(1 + \frac{|x-\xi_J|}{|J|}\right)^{-1-\varepsilon} dx \\
&\le C\frac{|J|^{1/2}}{|I|^{1/2}} k(I, J)^{-\varepsilon},
\end{aligned} \tag{4.8}$$

where $k(I, J) \ge 1$ is the largest integer k such that

$$\operatorname{dist}(\xi_J, \mathcal{C}(I')) \ge k|J|/2.$$

Let c'_J be the H_J-coefficient of $A_m^*(\sum c_I H_I)$. Then, there is at most one interval I' with $J \subset I'$ and $|I| = 2^m|J|$, and for that interval I, we have from (4.8),

$$|c'_J| \le |\lambda_m(I, J)||c_I| \le Ck(I, J)^{-\varepsilon}2^{-m/2}|c_I|. \tag{4.9}$$

This give

$$\sum_{J\subset I'} |c'_J|^2 \le C2^{-m}\sum_{k=1}^{2^m} k^{-2\varepsilon}|c_I|^2 \le C2^{-2\varepsilon m}|c_I|^2.$$

We now sum over all $I \in \mathcal{D}$ to find

$$\sum_{J\in\mathcal{D}} |c'_J|^2 \le C2^{-2\varepsilon m}\sum_{I\in\mathcal{D}} |c_I|^2.$$

Since $(H_I)_{I\in\mathcal{D}}$ is an orthonormal system, it follows that

$$\left\| A_m^* \left(\sum_{J\in\mathcal{D}} c_J H_J \right) \right\|_2 \le C2^{-\varepsilon m} \left\| \sum_{J\in\mathcal{D}} c_J H_J \right\|_2 . \tag{4.10}$$

In other words, the operator A_m^* has norm at most $C2^{-\varepsilon m}$ when acting on $L_2(\mathbf{R})$.

We now consider $1 < q < \infty$ and bound A_m^* on the space $L_q(\mathbf{R})$. We use (4.9) and the fact that $k(I, J) \ge 1$ to find that

$$\sum_{J\subset I} |c'_J|^2 \chi_J^2 \le C2^{-m} |c_I|^2 \sum_{J\subset I} \chi_J^2 = C|c_I|^2 \chi_I^2.$$

Therefore, summing over $I \in \mathcal{D}$, we find

$$\sum_{J\in\mathcal{D}} |c'_J|^2 \chi_J^2 \le C \sum_{I\in\mathcal{D}} |c_I|^2 \chi_I^2. \tag{I}$$

From this, and the strong Littlewood–Paley inequalities for the Haar functions, we obtain

$$\left\| A_m^* \left(\sum_{J\in\mathcal{D}} c_J H_J \right) \right\|_q \le C \left\| \sum_{J\in\mathcal{D}} c_J H_J \right\|_q . \tag{4.11}$$

In other words, the operator A_m^* has norm at most C when acting on $L_q(\mathbf{R})$.

Suppose that $1 < p \le 2$. We choose a value of q with $1 < q < p$. If $0 < \theta < 1$ satisfies $1/p = (1-\theta)/q + \theta/2$, then (4.10), (4.11), and the Riesz–Thorin interpolation theorem for linear operators gives

$$\|A_m^*\|_{L_p\to L_p} \le \|A_m^*\|_{L_q\to L_q}^{1-\theta} \|A_m^*\|_{L_2\to L_2}^{\theta} \le C2^{-\delta m}, \qquad \delta := \theta\varepsilon.$$

A similar argument applies when $2 < p < \infty$. Thus, we have shown that the operator A_0 is bounded from L_p to L_p for all $1 < p < \infty$.

In the same way, we can show that the operator A_0 corresponding to the case J is contained in the right-half of I is bounded on L_p, $1 < p < \infty$.

To show the boundedness of the operator A_0 corresponding to the case $J \subset I_\pm$ we will need a slight modification of the above proof. The same argument as above gives the inequality

$$\lambda_m(I, J) \le |I|^{-1/2} \int_I |\eta(J, x)|\, dx \le C \frac{|J|^{1/2}}{|I|^{1/2}} k(I, J)^{-1/2}$$

with $k(I, J)$ the largest integer such that

$$\operatorname{dist}(\xi_J, I) \ge k|J|/2.$$

As in the previous case, we obtain (4.10) and in place of (I) we obtain

$$\sum_{J\in\mathcal{D}} |c'_J|^2 \chi_J^2 \le C \sum_{I\in\mathcal{D}} |c_I|^2 \chi_{I_\pm}^2.$$

Since $\chi_{I_\pm} \le CM(\chi_I)$ with M the Hardy–Littlewood maximal function, we can use the Fefferman–Stein inequality together with the strong Littlewood–Paley inequalities for Haar functions to arrive at (4.11). The remainder of the proof of the boundedness of A_0 in this case is the same as above. This completes the proof of the theorem. ■

Corollary 4.3. *If* $\eta(I,\cdot)$, $I\in\mathcal{D}$, *satsify the assumptions* (A5)–(A7), *then* $\{\eta(I,\cdot\}_{I\in\mathcal{D}}\prec\{H_I\}_{I\in\mathcal{D}}$.

We can use a duality argument to give sufficient conditions that the operator T of (4.1) is boundedly invertible. For this, we assume that $\eta(I,\cdot)$, $I\in\mathcal{D}$, is a family of functions for which there is a dual family $\lambda(I,\cdot)$, $I\in\mathcal{D}$, that satisfies

$$\langle\eta(I,\cdot),\lambda(J,\cdot)\rangle=\delta(I,J),\qquad I,J\in\mathcal{D}.$$

Theorem 4.4. *If the functions* $\lambda(I,\cdot)$, $I\in\mathcal{D}$, *satisfy the assumptions* (A5)–(A7), *then* $\{H_I\}_{I\in\mathcal{D}}\prec\{\eta(I,\cdot\}_{I\in\mathcal{D}}$.

Proof. We need to show that for each $1<p<\infty$ and each sequence $(c_I)_{I\in\mathcal{D}}$, we have

$$\left\|\sum_{I\in\mathcal{D}}c_IH_I\right\|_p\le C\left\|\sum_{I\in\mathcal{D}}c_I\eta(I,\cdot)\right\|_p \tag{4.12}$$

with a constant C independent of the sequence $(c_I)_{I\in\mathcal{D}}$. We can assume that the sequence $(c_I)_{I\in\mathcal{D}}$ has at most a finite number of nonzero entries. We have

$$\left\|\sum_{I\in\mathcal{D}}c_IH_I\right\|_p=\sup_{(d_I)}\left\langle\sum_{I\in\mathcal{D}}c_IH_I,\sum_{I\in\mathcal{D}}d_IH_I\right\rangle=\sup_{(d_I)}\left\langle\sum_{I\in\mathcal{D}}c_I\eta(I,\cdot),\sum_{I\in\mathcal{D}}d_I\lambda(I,\cdot)\right\rangle$$

with the sup taken over all sequences $(d_I)_{I\in\mathcal{D}}$ with at most a finite number of nonzero entries which satisfy $\|\sum_{I\in\mathcal{D}}d_IH_I\|_{p'}=1$. From Hölder's inequality, we have

$$\begin{aligned}\left\|\sum_{I\in\mathcal{D}}c_IH_I\right\|_p&\le\left\|\sum_{I\in\mathcal{D}}c_I\eta(I,\cdot)\right\|_p\left\|\sum_{I\in\mathcal{D}}d_I\lambda(I,\cdot)\right\|_{p'}\\&\le C\left\|\sum_{I\in\mathcal{D}}c_I\eta(I,\cdot)\right\|_p\left\|\sum_{I\in\mathcal{D}}d_IH_I\right\|_{p'}=C\left\|\sum_{I\in\mathcal{D}}c_I\eta(I,\cdot)\right\|_p,\end{aligned}$$

where in the last inequality we used Theorem 4.2 for the sequence $(\lambda(I,\cdot))_{I\in\mathcal{D}}$. ■

We now apply Theorems 4.2 and 4.4 to the setting of Section 3. Let ψ be a univariate function and let $(\psi_I)_{I\in\mathcal{D}}$ be its univariate shifted-dilates. We also suppose that (ψ_I) has a dual basis $(\lambda_I)_{I\in\mathcal{D}}$ satisfying

$$\int_{\mathbf{R}}\psi_I(x)\lambda_I(x)\,dx=\delta(I,J).$$

Corollary 4.5. *If the functions* ψ *and* λ *satisfy conditions* (A5)–(A7) *in the case* $J=[0,1]$, *then* $(\psi_I)_{I\in\mathcal{D}}$ *satisfies the strong Littlewood–Paley property* (2.3).

Proof. If follows from Theorem 4.2 that $(\psi_I)_{I\in\mathcal{D}}\prec(H_I)_{I\in\mathcal{D}}$ and from Theorem 4.4 that $(H_I)_{I\in\mathcal{D}}\prec(\psi_I)_{I\in\mathcal{D}}$. Thus, $(\psi_I)_{I\in\mathcal{D}}\approx(H_I)_{I\in\mathcal{D}}$ and the theorem follows from Lemma 2.2. ■

We shall use the remainder of this section to give an example which shows that, in a certain sense, the assumption of piecewise Lipschitz ε continuity of the $\eta(I, \cdot)$ in (A6) cannot be removed. Namely, we will show that there is a continuous function ψ supported on [0, 1] with mean value 0 for which the Littlewood–Paley inequalities (2.2) and (2.3) do not hold.

Let X denote the set of all functions in $C[0, 1]$ which vanish at the endpoints ($f(0) = f(1) = 0$) and have mean value zero ($\int_0^1 f(x)\,dx = 0$). This is a closed subspace of $C[0, 1]$. For the formulas that follow, we consider $f(x) := 0$ for $x \in \mathbf{R}\backslash[0, 1]$. Let us consider the operator R defined for any function supported on [0, 1] by

$$Rf(x) := f(2x) + f(2x-1), \qquad x \in [0, 1]. \tag{4.13}$$

Then, R has norm one on $L_2[0, 1]$. We will use the following lemma:

Lemma 4.6. *For each $\varepsilon > 0$, there is a function $f = f_\varepsilon$ in X such that $\|f\|_{L_2[0,1]} \geq 1 - \varepsilon/4$ and*

$$\|f - Rf\|_{L_2[0,1]} < \varepsilon. \tag{4.14}$$

Proof. We choose an integer $N > 1$ such that

$$2^{-N}\binom{2N+1}{N}^{1/2} < \varepsilon/4. \tag{4.15}$$

Each point in $x \in (0, 1)$ has a dyadic expansion with digits $x_1, x_2, \ldots$. We require that infinitely many of the x_i are zero; then these digits are unique. We define

$$\tilde{f}(x) := \begin{cases} 1, & \text{if } \sum_{i=1}^{2N+1} x_i \leq N, \\ -1, & \text{else.} \end{cases}$$

Then $\tilde{f}$ is piecewise constant taking the values ± 1 on dyadic intervals of length 2^{-2N-1}. We have

$$R\tilde{f}(x) = \begin{cases} 1, & \sum_{i=2}^{2N+2} x_i \leq N, \\ -1, & \text{else.} \end{cases}$$

It follows that $R\tilde{f}(x) = \tilde{f}(x)$ except for the set E of points x for which $\sum_{i=1}^{2N+1} x_i = N, N+1$. Since E has measure

$$2^{-2N-1}\left(\binom{2N+1}{N} + \binom{2N+1}{N+1}\right) = 2^{-2N}\binom{2N+1}{N}.$$

We have (using (4.15))

$$\|\tilde{f} - R\tilde{f}\|_{L_2[0,1]} \leq 2^{-N+1}\binom{2N+1}{N}^{1/2} < \varepsilon/2.$$

The function $\tilde{f}$ has a finite number of discontinuities which occur at points $j2^{-2N-1}$, $j = 1, \ldots, 2^{2N+1}$. We can adjust $\tilde{f}$ near its points of discontinuity to obtain a function $f \in X$ with $\|f\|_{C[0,1]} = 1$ and

$$\|f - \tilde{f}\|_{L_2[0,1]} < \varepsilon/4.$$

Then,

$$\|f\|_{L_2[0,1]} \geq \|\tilde{f}\|_{L_2[0,1]} - \varepsilon/4 = 1 - \varepsilon/4.$$

Since the operator R has norm one on $L_2[0,1]$, it follows that

$$\|f - Rf\|_{L_2[0,1]} \leq 2\|f - \tilde{f}\|_{L_2[0,1]} + \|\tilde{f} - R\tilde{f}\|_{L_2[0,1]} < 2\varepsilon/4 + \varepsilon/2 = \varepsilon. \quad \blacksquare$$

Theorem 4.7. *There is a continuous function ψ supported on* $[0,1]$ *with mean value zero such that the Littlewood–Paley inequalities* (2.2) *and* (2.3) *do not hold for ψ_I, $I \in \mathcal{D}$.*

Proof. Consider again the operator R defined by (4.13) and define for each $n = 1, 2, \ldots,$ the operator

$$S_n f := \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} R^k f, \qquad f \in X.$$

Let $\mathcal{D}_n$ be the set of dyadic intervals in $[0,1]$ of length $\geq 2^{-n-1}$, it follows that

$$S_n f = \frac{1}{\sqrt{n}} \sum_{I \in \mathcal{D}_n} |I|^{1/2} f_I.$$

For the Haar function H, we have by orthogonality that $\|S_n H\|_{L_2[0,1]} = 1$. It is therefore enough to show that there is a function $\psi \in X$ such that

$$\sup_{n \geq 1} \|S_n \psi\|_{L_2[0,1]} = \infty.$$

By the Banach–Steinhaus theorem, we need only show that the operators S_n are unbounded as mappings from X into $L_2[0,1]$. To this end, let $\varepsilon > 0$ and let $f = f_\varepsilon$ be the function in X of Lemma 4.6. Since R has norm one on $L_2[0,1]$, we have

$$\|R^k f - f\|_{L_2[0,1]} \leq \sum_{\ell=1}^{k} \|R^\ell f - R^{\ell-1} f\|_{L_2[0,1]} \leq \sum_{\ell=1}^{k} \|f - Rf\|_{L_2[0,1]} \leq k\varepsilon.$$

Therefore,

$$\begin{aligned} \|S_n f - \sqrt{n} f\|_{L_2[0,1]} &= \frac{1}{\sqrt{n}} \left\| \sum_{k=0}^{n-1} (R^k f - f) \right\|_{L_2[0,1]} \\ &\leq \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \|R^k f - f\|_{L_2[0,1]} \leq \frac{\sqrt{n}(n-1)\varepsilon}{2}. \end{aligned}$$

It follows that

$$\begin{aligned} \|S_n f\|_{L_2[0,1]} &\geq \sqrt{n}\|f\|_{L_2[0,1]} - \|S_n f - \sqrt{n} f\|_{L_2[0,1]} \\ &\geq \sqrt{n}\|f\|_{L_2[0,1]} - \frac{\sqrt{n}(n-1)\varepsilon}{2}. \end{aligned}$$

Since $\|f\|_{L_2[0,1]} \geq 1 - \varepsilon/4$, by letting $\varepsilon \to 0$, we see that the norm of S_n is $\geq \sqrt{n}$. $\blacksquare$

5. Further Remarks

The results we have developed in the previous sections can be applied to any of the known wavelets. We shall consider in detail only the example of the Daubechies wavelets. This will indicate how the results of the previous sections are applied. Other examples such as the biorthogonal wavelets of Cohen, Daubechies, and Feauveau [CDF] can be treated in a similar way.

Let $\psi := D_k$, $k > 1$, be the kth univariate, orthogonal wavelet of Daubechies (see §6.4 of [Da]). We define, as usual, the multivariate family $\{\psi_I\}_{I\in\mathcal{D}(\mathbf{R}^d)}$ by (1.4). We first consider how one verifies the Jackson inequality (J) of Section 1 for the space $\mathcal{H}_n$. We do not want to use Theorem 3.2 to establish the Jackson inequality because it requires too much smoothness for ψ. Instead, we will use Theorem 3.4. This requires us to establish conditions (A1), (A2′), (A3), (A4) of Section 3 for the functions $\eta(I,\cdot) = \psi_I$.

The usual theory of multiresolution tells us that the ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, span $L_p(\mathbf{R}^d)$, $1 < p < \infty$. It is also well known (see, e.g., [M]) that ψ_I, $I \in \mathcal{D}(\mathbf{R})$, is an unconditional basis for $L_p(\mathbf{R})$, $1 < p < \infty$ (this also follows from our Section 4). The results in Section 2 then tell us that ψ_I, $I \in \mathcal{D}(\mathbf{R}^d)$, is an unconditional basis for $L_p(\mathbf{R}^d)$ and the Littlewood–Paley relations (2.3) hold for this basis. Hence, condition (A1) is satisfied.

Since $\lambda = \psi$, the moment condition (A3) holds for any $r \le k$. The decay condition (A4) obviously holds for any r because ψ has compact support.

We are left with showing (A2′). We can show that this condition holds for $r = k - 1$. We need to show that $\{\mu_I\}_{I\in\mathcal{D}} \prec \{H_I\}_{I\in\mathcal{D}}$ where μ is the rth integral of ψ. For this, we can use Corollary 4.3. Since D_k has k vanishing moments, the functions μ_I, $I \in \mathcal{D}(\mathbf{R})$, have compact support and therefore satisfy (A5). The function μ is at least Lipschitz continuous since it is the integral of a bounded function. Hence, condition (A6) is valid. Since D_k has k vanishing moments, μ has at least one vanishing moment and so (A7) is satisfied. Theorem 4.3 now implies that (A2′) is satisfied. We therefore have the Jackson inequality for hyperbolic approximation with the Daubechies wavelets

$$E_n(f)_p \le C2^{-nr}|f|_{\mathbf{W}^r(L_p(\mathbf{R}^d))}, \qquad n = 0, 1, \ldots, \quad 1 < p < \infty, \tag{5.1}$$

with $r = k - 1$ and with C independent of n and f.

It is important to contrast the difference between (5.1) in the univariate case and the multivariate case. It is well known that (5.1) holds in the univariate case for $r = k$. The reason for this is that in the univariate case one does not need to assume r moments are zero as in (A3) but only that $r - 1$ moments are zero. However, in the multivariate case, we cannot make this less restrictive assumption as can be seen already for the Haar function D_1. For example, if we define $F(x) := \prod_{j=1}^d f(x_j)$, with f a univariate function from $W^1(L_2(\mathbf{R}))$ which has compact support and for which $f(t) = t$, $t \in (\frac14, \frac34)$. Then for any dyadic rectangle $I \subset (\frac14, \frac34)^d$, we have

$$|\langle f, H_I\rangle| \ge C|I|^{3/2}$$

and therefore

$$E_n(f)_2^2 \ge C \sum_{|I|<2^{-n}} |\langle f, H_I\rangle|^2 \ge C \sum_{|I|<2^{-n}, I\subset(1/4,3/4)^d} |I|^3 \ge C\left(n^{(d-1)/2}2^{-n}\right)^2.$$

This example shows, in particular, that the relation (5.1) is not correct for the case

$r = 1$ and approximation using the Haar system. However, we do not know if (5.1) is correct for the Haar system in the case $0 < r < 1$. Note that the class $W^r(L_p(\mathbf{R}^d))$ can be defined for $0 < r < 1$ using fractional derivatives.

We next discuss the Bernstein inequalities for Daubhechies wavelets. Let again $\psi = D_k$, $k > 1$. We can use Theorem 3.5 to establish a Bernstein inequality. To apply this theorem, we need to verify assumptions (A1), (A2″), (A3), (A4). We have already noted that (A1), (A3), (A4) hold for any $r \le k$. Let $\rho = \rho(k)$ be the maximum of all α such that D_k has Hölder smoothness of order α. There are several papers dealing with the values of ρ. A discussion of this topic can be found, for example, in Chapter 7 of Daubechies book [Da]. It is known that

$$c_0 k \le \rho(k, p) \le c_1 k \tag{5.2}$$

for constants $0 < c_0, c_1 < 1$. For example, it is known that for sufficiently large k, the constant $c_0 = 0.20754$ suffices.

We can again use Corollary 4.3 to show that(A2″) holds for all $r < \rho$, i.e., $\{\mathbf{D}^r\psi_I\}_{I\in\mathcal{D}(\mathbf{R}^d)} \prec \{H_I\}_{I\in\mathcal{D}(\mathbf{R}^d)}$. The decay ssumption (A5) and the moment condition (A7) of that corollary is satisfied because $\mathbf{D}^r\psi$ has compact support. The definition of ρ and $r < \rho$ gives that $\mathbf{D}^r\psi$ has Lipschitz smoothness and therefore (A6) follows. So all the conditions of Corollary 4.3 are satisfied and property (A2″) follows.

In summary, we have shown

Theorem 5.1. *Let $\psi = D_k$, $k = 2, 3, \ldots,$ be one of the Daubechies wavelets. Then*:

(i) *The Jackson inequality* (3.8) *holds for $r = k - 1$.*
(ii) *If $r \le c_0 k$ with c_0 any constant for which* (5.2) *is valid, then the Bernstein inequality* (3.8) *holds for this value of r.*
(iii) *For the r in* (ii) *and for any $0 < \alpha < r$, the approximation class $\mathcal{A}_q^\alpha(L_p(\mathbf{R}^d))$, $1 < p < \infty$, of* (1.5) *is identical with the class of functions satisfying* (1.2) *with $K(f, t)$ the K-functional of* (1.6).

Acknowledgments. The authors wish to thank Professor Pencho Petrushev for several valuable discussions concerning the material in this paper. The authors would also like to thank the referee for many constructive comments about this paper. This research was supported by the Office of Naval Research Contract N0014-91-J1343 and the National Science Foundation Grant EHR 9108772 and was completed while S. V. Konyagin was a visiting scholar at the University of South Carolina. S. V. Konyagin was supported in part by the Russian Fund of Fundamental Investigations Contract N 93-01-0024.

References

[BDR] C. de Boor, R. A. DeVore, A. Ron (1993): *On the construction of multivariate (pre) wavelets*. Const. Approx., **9**:123–166.

[CW] C. K. Chui, J. Z. Wang (1992): *On compactly supported spline wavelets and a duality principle*. Trans. Amer. Math. Soc., **330**:903–915.

[CDF] A. Cohen, I. Daubechies, J.-C. Feauveau (1992): *Biorthogonal bases of compactly supported wavelets*. Comm. Pure Appl. Math., **45**:485–560.

[Da] I. Daubechies (1992): Ten Lectures on Wavelets. CBMS–NSF Regional Conference Series in Applied Mathematics, vol. 61. Philadelphia, PA: SIAM.

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

[DL] R. DeVore, G. Lorentz (1993): Constructive Approximation. Heidelberg: Springer-Verlag.

[DPT] R. DeVore, P. Petrushev, V. Temlyakov (1994): *Multivariate trigonometric approximation with frequencies from the hyperpolic cross*. Mat. Zametki, **56**(3):36–63; English translation in Math. Notes, **56**.

[Di] J. Diestel (1984): Sequences and Series in Banach Spaces. Graduate Series in Mathematics. New York: Springer-Verlag.

[FS] C. Fefferman, E. Stein (1972): *H^p spaces of several variables*. Acta Math., **129**:137–193.

[FJ] M. Frazier, B. Jawerth (1990): *A discrete transform and decomposition of distribution spaces*. J. Funct. Anal., **93**:34–170.

[FJM] M. Frazier, B. Jawerth, G. Weiss (1991): Littlewood–Paley Theory and the Study of Function Spaces. CBMS Regional Conference Series in Mathematics, vol. 79. Providence, RI: American Mathematical Society.

[K] A. Kamont (1996): *On hyperbolic summation and hyperbolic moduli of smoothness*. Const. Approx., **12**:111–125.

[KS] B. S. Kashin, A. A. Saakyan (1989): Orthogonal Series. Providence, RI: American Mathematical Society.

[M] Y. Meyer (1990): Ondelettes et Opérateurs, vols. 1 and 2. Paris: Hermann.

[T] V. Temlyakov (1989): *Approximation of functions with bounded mixed derivative*. Proc. Steklov Inst. Math., **178**:1–122.

[Z] A. Zygmund (1977): Trigonometric Series, vols. I and II. London: Cambridge University Press.

R. A. DeVore
Department of Mathematics
University of South Carolina
Columbia
SC 29208
USA

S. V. Konyagin
Department OPU, Mech.-Math.
Moscow State University
Leninskie Gory
Moscow 117234
Russia

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