NONLINEAR WAVELET APPROXIMATION IN ANISOTROPIC BESOV SPACES

Christopher Leisner

Dartmouth College

Abstract. We introduce new anisotropic wavelet decompositions associated with the smoothness $\boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$, $\beta_1, \dots, \beta_d > 0$ of multivariate functions as measured in anisotropic Besov spaces $B^{\boldsymbol{\beta}}$. We give the rate of nonlinear approximation of functions $f \in B^{\boldsymbol{\beta}}$ by these wavelets. Finally, we prove that, among a general class of anisotropic wavelet decompositions of a function $f \in B^{\boldsymbol{\beta}}$, the anisotropic wavelet decomposition associated with $\boldsymbol{\beta}$ gives the optimal rate of compression of the wavelet decomposition of f.

Chapter 1
Introduction

1.1 Overview. In regular wavelet decompositions of a function f, one approximates f at several resolution levels. For example, if $f \in L^2(\mathbb{R}^2)$, one could find a best constant approximation $P_0 f$ to f, then a best piecewise constant approximation $P_1 f$ to f on a 2×2 grid of subsquares of $[0,1]^2$, then a best piecewise constant approximation $P_2 f$ on a 4×4 grid, etc. If we think of f as representing a grayscale image, then it is easy to see that "smooth" parts of f can be approximated well using characteristic functions of large squares, while less smooth parts of f (e.g., discontinuities) can only be approximated well using characteristic functions of small squares.

Once we have computed $P_0 f$, $P_1 f$, ... , we write

$$f = P_0 f + \sum_{k=1}^{\infty} (P_k f - P_{k-1} f), \tag{1.1.1}$$

interpreting the terms $P_k f - P_{k-1} f$ as the "details" which are added when going from one resolution level to the next. One then writes $P_k f - P_{k-1} f$ as the so-called *wavelet sum*

$$P_k f - P_{k-1} f = \sum_{\mathbf{j} \in \mathbb{Z}^2} d_{\mathbf{j},k} \psi_{\mathbf{j},k} := \sum_{\mathbf{j} \in \mathbb{Z}} d_{\mathbf{j},k} 2^k \psi_{\mathbf{j},k}(2^k \cdot - \mathbf{j}) \tag{1.1.2}$$

1991 *Mathematics Subject Classification.* 41,41A15,41A17,41A25.

Key words and phrases. Wavelets, multiresolution, anisotropic, Besov spaces, B-splines, nonlinear approximation, interpolation spaces.

and $P_0 f$ as a linear combination of translations of a *scaling function*,

$$P_0 f = \sum_{\mathbf{j}\in\mathbb{Z}^2} c_{\mathbf{j}} \phi(\cdot - \mathbf{j}) \tag{1.1.3}$$

so that (1.1.1) can be rewritten

$$f = \sum_{\mathbf{j}\in\mathbb{Z}} c_{\mathbf{j}} \phi(\cdot - \mathbf{j}) + \sum_{k=0}^{\infty} \sum_{\mathbf{j}\in\mathbb{Z}^2} d_{\mathbf{j},k} \psi_{\mathbf{j},k}. \tag{1.1.4}$$

The science of compressing such *wavelet decompositions* of f is to choose, for $n = 1, 2, \ldots$ the n most "significant" terms in (1.1.4) and to form an approximation f_n to f using these terms.

There is a mathematical theory by DeVore, Jawerth and Popov (see [D-J-P]) which quantifies the performance of this type of compression. What is implicit in this theory is that this method assumes that the smoothness of f is the same in each coordinate direction, so that when one goes from one resolution level to the next, it makes sense to double the number of rows and columns in our approximation grid at the same time.

Some types of functions do not have the same smoothness in each coordinate direction. For example, suppose that $f(x, y, t)$ models the footage from a security video camera. Here x and y represent the spatial directions, while t represents time. The smoothness in each frame is quite low (the same for still images), but in time the smoothness is high, in that most of the frames are the same from one frame to the next. To compress this function, it is wasteful to divide both the spatial and temporal approximation "cubes" by 2 each time we pass from one resolution level to the next higher one.

So the author's idea is as follows. First, we use so-called anisotropic Besov spaces to measure the smoothness of multivariate functions. Instead of assigning a single smoothness parameter α to all coordinate directions, each coordinate direction i is assigned its own smoothness parameter α_i. So let us say that f is bivariate and the smoothness in the x_1 direction is α_1 and the smoothness in the x_2 direction is α_2 and that $\alpha_2 > \alpha_1$. We shall prove that the best way to pass from one resolution level to the next finer resolution level is the following: One works with rectangles instead of subsquares and starts again with a function which is constant on $[0, 1]^2$ with which to approximate f. Then to go from one resolution to the next, one always divides the rectangles by two in the x_1 direction, but one divides the rectangles in the x_2 direction only the fraction α_1/α_2 of the time. This way one approximates the function not on smaller and smaller squares, but on smaller and smaller rectangles that are also getting "skinnier". It turns out that compressing the details that result from subtracting approximations to the original f at adjacent pairs of these "anisotropic" resolution levels is the optimal way to compress f.

1.2 Background. As was mentioned in the introduction, some multivariate functions have different smoothness in different coordinate directions. Such smoothness is called *anisotropic*, in contrast to smoothness that is the same in all directions, which is called *isotropic*.

We use the so-called anisotropic Besov spaces to quantify the anisotropic smoothness of functions. Let $f : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$ (typically $\Omega = [0, 1]^d$). We assume that there are smoothness parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$, with $\alpha_i > 0$ for $i = 1, \dots, d$, and to be specific, that we want to measure the smoothness of f in the space $L^p(\mathbb{R}^d)$. So f has smoothness α_i in the $\mathbf{e}_i$ direction, $i = 1, \dots, d$, with $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ being the ith coordinate vector, which has a 1 as its ith element and 0 as all other elements.

For any $\mathbf{h} = (h_1, \dots, h_d)$ in $\mathbb{R}^d$, we define the rth difference of f in the direction $\mathbf{h}$ at the point $\mathbf{x} \in \mathbb{R}^d$ recursively as

$$\Delta^0_{\mathbf{h}} f(\mathbf{x}) := f(\mathbf{x}),$$

$$\Delta^r_{\mathbf{h}} f(\mathbf{x}) := \Delta^{r-1}_{\mathbf{h}} f(\mathbf{x}+\mathbf{h}) - \Delta^{r-1}_{\mathbf{h}} f(\mathbf{x}), \quad \text{for } r > 0. \tag{1.2.1}$$

$\Delta^r_{\mathbf{h}} f(\mathbf{x})$ is defined on the set

$$\Omega(r\mathbf{h}) := \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x}, \mathbf{x}+\mathbf{h}, \dots, \mathbf{x}+r\mathbf{h} \in \Omega\}.$$

We let $\mathbf{t} = (t_1, \dots, t_d)$, $t_i > 0$ for all i, and define the rth modulus of smoothness of f in $L^p(I)$ to be

$$\omega_r(f, \mathbf{t}, I)_p := \omega_r(f, t_1, \dots, t_d, I)_p := \sup_{|h_i| \le t_i} \|\Delta^r_{\mathbf{h}} f\|_p(I(r\mathbf{h})), \tag{1.2.2}$$

where

$$\|g\|_p(I) := \left(\int_I |f(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p}$$

and $L^p(I)$ is the set of all functions $g : I \to \mathbb{R}$ such that $\|g\|_p(I) < \infty$. We shall omit the parameter I when this will not cause any confusion. We require some basic facts about p-norms which are used as substitutes for the triangle inequality when $p < 1$. First of all, for any f, $g \in L^p(I)$, $p < 1$, we have

$$\|f+g\|_p \le 2^{\frac{1}{p}-1} \left[\|f\|_p + \|g\|_p \right] \tag{1.2.3}$$

((1.2.3) is called the *quasi-triangle* inequality). Also, for any sequence of functions $\{f_i\}$,

$$\left\| \sum_i f_i \right\|_p \le \left[\sum_i \|f_i\|_p^\mu \right]^{1/\mu} \tag{1.2.4}$$

where $\mu \le \min(1, p)$.

As a special case of the mixed modulus of smoothness, we present, for $i = 1, \dots, d$ the ith partial modulus of smoothness $\omega^{(i)}_r(f, t, I)_p$ defined by

$$\omega^{(i)}_r(f, t, I)_p := \omega^{(i)}_r(f, t\mathbf{e}_i, I)_p.$$

From now on, the default value of the argument I in the mixed modulus of smoothness and in $\|\cdot\|_p(I)$ will be Ω. The anisotropic Besov space $B^{\boldsymbol{\alpha}}_q(L^p(\Omega))$ for $0 < \alpha_1, \dots, \alpha_d < r$ is defined to be the set of all functions f for which

$$|f|_{B^{\boldsymbol{\alpha}}_q(L^p(\Omega))} := \left(\sum_{k=0}^{\infty} [2^k \omega_r(f, 2^{-k/\alpha_1}, \dots, 2^{-k/\alpha_d})_p]^q \right)^{1/q}$$

is finite, and

$$\|f\|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} := \|f\|_p(\Omega) + |f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))}.$$

Heuristically, a function f in $B_q^{\boldsymbol{\alpha}}(L^p(\Omega))$ has α_i "partial derivatives" in $L^p(\Omega)$ in the ith coordinate direction.

If $\alpha_1 = \cdots = \alpha_d$, then

$$|f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} = \left(\sum_{k=0}^{\infty}[2^k \omega_r(f, 2^{-k/\alpha_1}, \ldots, 2^{-k/\alpha_1})_p]^q\right)^{1/q},$$

which is equivalent to the usual semi-norm

$$|f|_{B_q^{\alpha_1}(L^p(\Omega))} := \left(\sum_{k=0}^{\infty}[2^{\alpha_1 k}\omega_r(f, 2^k)_p]^q\right)^{1/q}$$

of the isotropic Besov space $B_q^{\alpha_1}(L^p(\Omega))$.

1.3 Isotropic Multiresolution Decompositions. We briefly give an overview of previously known multiresolution decompostions, which we shall call isotropic decompositions, to contrast with the nonisotropic decompositions given in the next section. Let us begin with the univariate B-spline $\psi(x)$ of order r with knots at $0, \ldots, r$, defined in Section 3.2. Define a corresponding multivariate B-spline

$$\psi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$$

by $\psi(\mathbf{x}) := \psi(x_1, \ldots, x_d) := \psi(x_1)\cdots\psi(x_d)$. Then $\psi(\mathbf{x})$ satisfies a rewrite rule

$$\psi(\mathbf{x}) = \sum_{\mathbf{j}} a_{\mathbf{j}}\psi(2\mathbf{x} - \mathbf{j}) \tag{1.3.1}$$

for a finite set of nonzero coefficients $a_{\mathbf{j}}$, $\mathbf{j} \in \mathbb{Z}^d$. We let $\mathcal{S}_k$ be the space of all functions

$$\mathcal{S}_k := \left\{\sum_{\mathbf{j}\in\mathbb{Z}^d} c_{\mathbf{j}}\psi(2^k\mathbf{x} - \mathbf{j}) \mid c_{\mathbf{j}} \in \mathbb{R}\right\}$$

and we choose a bounded projection P_k from $L^p(\mathbb{R}^d)$ to $\mathcal{S}_k$. Under certain conditions (see [D-P1]) we can write any $f \in L^p(\mathbb{R}^d)$ as

$$f = \lim_{k\to\infty} P_k = P_0 f + \sum_{k=1}^{\infty}(P_k f - P_{k-1}f),$$

where, because of the rewrite rule (1.3.1), we know that $P_k f - P_{k-1}f$ is in $\mathcal{S}_k$. Thus, since $P_0 \in \mathcal{S}_0$, we can write

$$\begin{aligned} f &= P_0 f + \sum_{k=1}^{\infty}(P_k f - P_{k-1}f) \\ &= \sum_{k=0}^{\infty}\sum_{\mathbf{j}\in\mathbb{Z}^d} d_{\mathbf{j},k}\psi(2^k\cdot - \mathbf{j}). \end{aligned} \tag{1.3.2}$$

1.4 Nonisotropic Multiresolution Decompositions. If a function $f : \Omega \to \mathbb{R}$ is in the nonisotropic Besov space $B_q^{\boldsymbol{\alpha}}(L^p(\Omega))$, then

$$|f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} := \left(\sum_{k=0}^{\infty}[2^k\omega_r(f, 2^{-k/\alpha_1}, \dots, 2^{-k/\alpha_d})_p]^q\right)^{1/q}$$

is finite. However, the quantity on the right is equivalent to

$$\left(\sum_{k=0}^{\infty}[2^{k\underline{\alpha}}\omega_r(f, 2^{-k\underline{\alpha}/\alpha_1}, \dots, 2^{-k\underline{\alpha}/\alpha_d})_p]^q\right)^{1/q}, \tag{1.4.1}$$

where

$$\underline{\alpha} := \min_{1\le i\le d} \alpha_i.$$

Note also that (1.4.1) is equivalent to

$$\left(\sum_{k=0}^{\infty}[2^{k\underline{\alpha}}\omega_r(f, 2^{-\lfloor k\underline{\alpha}/\alpha_1\rfloor}, \dots, 2^{-\lfloor k\underline{\alpha}/\alpha_d\rfloor})_p]^q\right)^{1/q}, \tag{1.4.2}$$

(where $\lfloor x\rfloor$ denotes the largest integer $\le x$) since this increases the size of each argument of the modulus of smoothness by at most a factor of 2.

Motivated by these considerations, we define $\mathcal{S}_k$ to be the linear span of the *anisotropic B-splines*

$$\psi_{\mathbf{j},k}(\mathbf{x}) := \psi(2^{\lfloor k\underline{\alpha}/\alpha_1\rfloor}x_1 - j_1, \dots, 2^{\lfloor k\underline{\alpha}/\alpha_d\rfloor}x_d - j_d)$$

for $\mathbf{j} = (j_1, \dots, j_d) \in \mathbb{Z}^d$. Note that the scaling in each variable is always an integer power of 2, so that $\mathcal{S}_k$ is, indeed, included in S_{k+1}, by the rewrite rule for ψ; furthermore, since $\underline{\alpha}/\alpha_i = 1$ for at least one i, we know that $\mathcal{S}_k$ is *strictly* contained in $\mathcal{S}_{k+1}$, i.e., when moving from $\mathcal{S}_k$ to $\mathcal{S}_{k+1}$, one refines functions by a factor of two in at least one direction. In fact, going from $\mathcal{S}_k$ to $\mathcal{S}_{k+1}$ we refine in all directions $\mathbf{e}_i$ for which

$$\frac{\underline{\alpha}}{\alpha_i}k < m_{k,i} \le \frac{\underline{\alpha}}{\alpha_i}(k+1)$$

for some integer $m_{k,i}$, and in no other directions.

Thus if we define P_k to be the projection onto the new $\mathcal{S}_k$, we again have

$$f = P_0 f + \sum_{k=1}^{\infty}(P_k f - P_{k-1} f),$$

and $P_k f - P_{k-1} f$ is again in $\mathcal{S}_k$, so we can write $P_k f - P_{k-1} f$ as a linear combination of the functions $\psi_{\mathbf{j},k}$, i.e.

$$f = \sum_{k=0}^{\infty}\sum_{\mathbf{j}\in\mathbb{Z}^d} d_{\mathbf{j},k}\psi_{\mathbf{j},k}. \tag{1.4.3}$$

We remark that a similar framework can be applied to all orthogonal and biorthogonal wavelets, wavelet frames, etc.

1.5 Nonlinear Wavelet Approximation. We now describe the process by which we compute approximations to functions in certain anisotropic Besov spaces. First, we define the sets from which we choose these approximations. For $n = 1, 2, \dots$ let Σ_n denote the set of all linear combinations of anisotropic B-splines

$$\sum_{(\mathbf{j},k)\in\Lambda} c_{\mathbf{j},k}\psi_{\mathbf{j},k}$$

such that $\#\Lambda \leq n$. Σ_n is a *nonlinear manifold* because, in general, the sum of two members of Σ_n is not in Σ_n (although this sum is in Σ_{2n}). *Nonlinear approximation* is the process of approximating functions with members of nonlinear manifolds with certain properties (see [D-P2]).

Fix $1 \leq p < \infty$, and let $f \in L^p(\Omega)$. We will choose approximations to f from the sets Σ_n and use the $L^p(\Omega)$ norm to measure the error in this approximation. That is, for $n = 1, 2, \dots$, let $\sigma_n(f)_p = \inf_{S\in\Sigma_n} \|f - S\|_p$. In other words, $\sigma_n(f)_p$ is the error, measured in the $L^p(\Omega)$ norm, of approximating f from Σ_n. We are interested in determining the rate at which $\sigma_n(f)_p$ decreases as n increases. In general, for $f \in L^p(\Omega)$ we cannot say how quickly $\sigma_n(f)_p$ decreases with n. However, if f is a member of a certain subspace of $L^p(\Omega)$, we have very specific information about the rate of decrease of $\sigma_n(f)_p$. This subspace is

$$B^{\boldsymbol{\alpha}} := B^{\boldsymbol{\alpha}}_{\tau}(L^{\tau}(\Omega)),$$

where

$$\frac{1}{\tau} = \frac{H(\boldsymbol{\alpha})}{d} + \frac{1}{p}, \tag{1.5.1}$$

where $H(\boldsymbol{\alpha})$ is defined by

$$\frac{1}{H(\boldsymbol{\alpha})} = \frac{1}{d}\left(\frac{1}{\alpha_1} + \cdots + \frac{1}{\alpha_d}\right).$$

Our main compression result is that, for $1 \leq p < \infty$,

$$f \in B^{\boldsymbol{\alpha}} \Leftrightarrow \left(\sum_{k=1}^{\infty}\left(2^{kH(\boldsymbol{\alpha})/d}\sigma_{2^k}(f)_p\right)^{\tau}\right)^{1/\tau} < \infty. \tag{1.5.2}$$

In the case $\alpha_1 = \cdots = \alpha_d = \alpha$, we have $H(\boldsymbol{\alpha}) = \alpha$, and (1.5.1) reduces to

$$f \in B^{\alpha} \Leftrightarrow \left(\sum_{k=1}^{\infty}\left(2^{k\alpha/d}\sigma_{2^k}(f)_p\right)^{\tau}\right)^{1/\tau} < \infty$$

for

$$\frac{1}{\tau} = \frac{\alpha}{d} + \frac{1}{p},$$

which is the same as the main result in [D-J-P].

1.6 Optimality of the Scaling of $\psi_{\mathbf{j},k}$ for Approximation of $f \in B^{\boldsymbol{\beta}}$. The scaling for the wavelets that we described yields wavelets which are optimal for approximating functions in $B^{\boldsymbol{\alpha}}$. We now explain what we mean by "optimal". Given $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$, $\beta_1, \dots, \beta_d > 0$, define $\psi^{\boldsymbol{\beta}}_{\mathbf{j},k}$ in the same way as $\psi_{\mathbf{j},k}$ and with the same value of r, but with $\boldsymbol{\beta}$ in place of $\boldsymbol{\alpha}$. For $n = 0, 1, \dots$, define $\Sigma^{\boldsymbol{\beta}}_n$ to be the set of all sums $\sum_{(\mathbf{j},k)\in\Lambda} d_{\mathbf{j},k}\psi^{\boldsymbol{\beta}}_{\mathbf{j},k}$ such that $\#\Lambda \leq n$. Then there exists a function $f \in B^{\boldsymbol{\alpha}}$ such that for any $\boldsymbol{\beta}$ which is not a scalar multiple of $\boldsymbol{\alpha}$, we have $\inf_{S\in\Sigma^{\boldsymbol{\beta}}_n} \|f - S\|_p(\Omega) \neq \mathcal{O}(n^{-\gamma})$ for any $\gamma > 0$.

Chapter 2
Anisotropic Besov Spaces

2.1 The Mixed Modulus of Smoothness. Let $\Omega := [0,1]^d$, $f \in L^p(\Omega)$, $0 < p \leq \infty$. From now on, $\mathbf{x}$, $\mathbf{h}$, $\mathbf{t}$ will be vectors in $\mathbb{R}^d$ with $t_i \geq 0$, $i = 1, \dots, d$, and I will represent an interval $I_1 \times \cdots \times I_d$ in $\mathbb{R}^d$, where the I_i are intervals in $\mathbb{R}$. We shall recall some properties of the univariate modulus of smoothness

$$\omega_r(f,t,I)_p := \sup_{|\mathbf{h}|_\infty \leq t} \|\Delta^r_{\mathbf{h}} f\|_p(I(r\mathbf{h})) \tag{2.1.1}$$

and establish analagous properties of $\omega_r(f,\mathbf{t},I)_p$. Let $\mu := r + (1/p-1)_+$, where $(\cdot)_+ := \max(\cdot,0)$. Then for any $\lambda \geq 1$,

$$\omega_r(f,\lambda t,I)_p \leq C\lambda^\mu \omega_r(f,t,I)_p, \tag{2.1.2}$$

where $C = C(r,p)$, while for $\lambda \leq 1$,

$$\omega_r(f,\lambda t,I)_p \geq C\lambda^\mu \omega_r(f,t,I)_p \tag{2.1.3}$$

(see [P-P]). We have the following analogue of (2.1.2):

Theorem 2.1.1. *For $\lambda \geq 1$, we have*

$$\omega_r(f,\lambda\mathbf{t},I)_p \leq C\lambda^\mu \omega_r(f,\mathbf{t},I)_p, \quad C = C(r,p). \tag{2.1.4}$$

Proof. The proof is similiar to the one in [P-P] of (2.1.2). □

Corollary 2.1.2. *For $\lambda_1, \dots, \lambda_d \geq 1$, we have*

$$\omega_r(f,\lambda_1 t_1,\dots,\lambda_d t_d,I)_p \leq C\max(\lambda_1,\dots,\lambda_d)^\mu \omega_r(f,\mathbf{t},I)_p \tag{2.1.8}$$

while for $\lambda_1,\dots,\lambda_d \leq 1$, we have

$$\omega_r(f,\lambda_1 t_1,\dots,\lambda_d t_d,I)_p \geq C\min(\lambda_1,\dots\lambda_d)^\mu \omega_r(f,\mathbf{t},I)_p, \tag{2.1.9}$$

where $C = C(r,p)$.

Proof. Since (2.1.9) follows immediately from (2.1.8), it is enough to prove (2.1.8). Furthermore, it is enough to prove (2.1.8) in the case where $\lambda_1,\dots,\lambda_d$ are positive integers. Let $\overline{\lambda} := \max(\lambda_1,\dots,\lambda_d)$. Clearly,

$$\omega_r(f,\lambda_1 t_1,\lambda_2 t_2,\dots,\lambda_d t_d,I)_p \leq \omega_r(f,\overline{\lambda}\mathbf{t},I)_p \tag{2.1.10}$$

and combining (2.1.10) with Theorem 2.1.1 gives

$$\omega_r(f,\lambda_1 t_1,\dots,\lambda_d t_d,I)_p \leq c\overline{\lambda}^\mu \omega_r(f,\mathbf{t},I)_p$$

as required. □

We will also need the property that for $r' > r$, we have

$$\omega_{r'}(f,\mathbf{t},I)_p \leq C\omega_r(f,\mathbf{t},I)_p, \quad C = C(r,d,p). \tag{2.1.11}$$

This property is well-known for the univariate modulus of smoothness and exactly the same proof applies to the multivariate modulus of smoothness; see [D-L], section 2.7.

There is a useful connection between the multivariate and univariate moduli of smoothness that we shall exploit. For $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$, $\lambda_1, \ldots, \lambda_d > 0$, define the dilation operator

$$(D_{\boldsymbol{\lambda}} f)(\mathbf{x}) := f(\lambda_1 x_1, \ldots, \lambda_d x_d).$$

Then by a change of variable, we have

$$\omega_r(f, \lambda_1 t, \ldots, \lambda_d t, I)_p = \lambda_1^{1/p} \cdots \lambda_d^{1/p} \omega_r(D_{\boldsymbol{\lambda}} f, t, \frac{1}{\lambda_1} I_1 \times \cdots \times \frac{1}{\lambda_d} I_d)_p \tag{2.1.12}$$

Now recall from [D-L] that the univariate modulus of smoothness $\omega_r(f,t,I)_p$ is equivalent to the univariate averaged modulus of smoothness defined by

$$w_r(f,t,I)_p := \left(t^{-d} \int\limits_{\Omega_t} \int\limits_{I(rs)} |\Delta_s^r(f, \mathbf{x})|^p d\mathbf{x}\, d\mathbf{s} \right)^{1/p} \tag{2.1.13}$$

where $\Omega_t := [0,t]^d$. w_r is needed as a substitute for ω_r when we need to add up smoothness estimates over several intervals.

There exists a multivariate version of the equivalence $\omega_r(f,t,I)_p \asymp w_r(f,t,I)_p$ which we now describe. Define

(2.1.14)

$$w_r(f, \mathbf{t}, I)_p := \left(t_1^{-1} \cdots t_d^{-1} \int_0^{t_1} \cdots \int_0^{t_d} \int\limits_{I(r\mathbf{s})} |\Delta_{\mathbf{s}}^r(f, \mathbf{x})|^p \, d\mathbf{x} ds_d \ldots, ds_1 \right)^{1/p}.$$

Lemma 2.1.3. *$\omega_r(f, t_1, \ldots, t_d, I)_p \asymp w_r(f, t_1, \ldots, t_d, I)_p$, with constants of equivalence depending only on r, p, d.*

Proof. We work with the quantities $\omega_r(f, \lambda_1 t, \ldots, \lambda_d t, I)_p$ and $w_r(f, \lambda_1 t, \ldots, \lambda_d t, I)_p$, where $\lambda_1, \ldots, \lambda_d, t > 0$. Now,

$$\begin{aligned} \omega_r(f, \lambda_1 t, \ldots, \lambda_d t, I)_p &= \lambda_1^{1/p} \cdots \lambda_d^{1/p} \omega_r(D_{\boldsymbol{\lambda}} f, t, \frac{1}{\lambda_1} I_1 \times \cdots \times \frac{1}{\lambda_d} I_d)_p \\ &\asymp \lambda_1^{1/p} \cdots \lambda_d^{1/p} w_r(D_{\boldsymbol{\lambda}} f, t, \frac{1}{\lambda_1} I_1 \times \cdots \times \frac{1}{\lambda_d} I_d)_p \\ &= w_r(f, \lambda_1 t, \ldots, \lambda_d t, I)_p \end{aligned}$$

where the last equality is obtained by changing variables. □

2.2 The Anisotropic Smoothness Norm. Let $\alpha_1, \ldots, \alpha_d$ be positive numbers, with $\underline{\alpha} := \min_{1 \le i \le d} \alpha_i$, $\overline{\alpha} := \max_{1 \le i \le d} \alpha_i$, and $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_d)$. Let $r > \overline{\alpha}$. Then the original (and perhaps most natural) definitions (see [L-T]) of the anisotropic smoothness seminorm and norm of a function $f \in L^p(\Omega)$ are

$$|f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} := \sum_{i=1}^{d} \left[\int_0^1 [t^{-\alpha_i} \omega_r^{(i)}(f,t)_p]^q \, \frac{dt}{t} \right]^{1/q} \tag{2.2.1}$$

$$\|f\|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} := \|f\|_p + |f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))}. \tag{2.2.2}$$

We define the (quasi) normed linear spaces $B_q^{\boldsymbol{\alpha}}(L^p(\Omega))$ by

$$B_q^{\boldsymbol{\alpha}}(L^p(\Omega)) := \{f \mid \|f\|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} < \infty\}.$$

This family of spaces is called the *anisotropic Besov spaces* . The astute reader will notice that the parameter r is missing from the definition of the anisotropic Besov spaces. This is because, as will be shown in Chapter 3, when $r, r' > \overline{\alpha}$ we have

$$B_q^{\boldsymbol{\alpha},r}(L^p(\Omega)) = B_q^{\boldsymbol{\alpha},r'}(L^p(\Omega)) \tag{2.2.3}$$

with equivalent norms.

We shall express the seminorm (2.2.1) in the equivalent and more convenient form

$$|f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} \asymp \left[\int_0^1 [t^{-1}\omega_r(f, t^{1/\alpha_1}, \dots, t^{1/\alpha_d})_p]^q \, \frac{dt}{t}\right]^{1/q}. \tag{2.2.4}$$

To prove (2.2.4), first make the change of variable $s = t^{\alpha_i}$ in each of the integrals in the left hand side of (2.2.1) to obtain

$$\begin{aligned} |f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} &\asymp \sum_{i=1}^d \left[\int_0^1 [s^{-1}\omega_r^{(i)}(f, s^{1/\alpha_i})_p]^q \, \frac{ds}{s}\right]^{1/q} \\ &\le C\left[\sum_{i=1}^d \int_0^1 [s^{-1}\omega_r^{(i)}(f, s^{1/\alpha_i})_p]^q \, \frac{ds}{s}\right]^{1/q} \\ &\le C\left[\int_0^1 [s^{-1}\omega_r(f, s^{1/\alpha_1}, \dots, s^{1/\alpha_d})_p]^q \, \frac{ds}{s}\right]^{1/q} \end{aligned} \tag{2.2.5}$$

where the last inequality follows from the fact that

$$\omega_r^{(i)}(f, s^{1/\alpha_i})_p \le \omega_r(f, s^{1/\alpha_1}, \dots, s^{1/\alpha_d})_p$$

for $1 \le i \le d$. To get the other inequality, note that

$$\begin{aligned} \int_0^1 [t^{-1}\omega_r(f, t^{1/\alpha_1}, \dots, t^{1/\alpha_d})_p]^q \, \frac{dt}{t} &\le C \int_0^1 [t^{-1}\sum_{i=1}^d \omega_r^{(i)}(f, t^{1/\alpha_i})_p]^q \, \frac{dt}{t} \\ &\le C \sum_{i=1}^d \int_0^1 [t^{-1}\omega_r^{(i)}(f, t^{1/\alpha_i})_p]^q \, \frac{dt}{t} \end{aligned} \tag{2.2.6}$$

by the triangle inequality when $q \ge 1$, and the quasi-triangle inequality when $q < 1$. Using this new form of the anisotropic smoothness seminorm, it is easy to see that in the special case $\alpha_1 = \cdots = \alpha_d$, $B_q^{\boldsymbol{\alpha}}(L^p(\Omega)) = B_q^{\alpha_1}(L^p(\Omega))$, where $B_q^{\alpha_1}(L^p(\Omega))$ is the traditional isotropic Besov space.

We shall make some further changes to the form of the seminorm (2.2.4) (for reasons which will be given in Chapter 3). Making the change of variable $s = t^{1/\underline{\alpha}}$ in (2.2.4) and defining $\alpha_i' := \underline{\alpha}/\alpha_i$ gives

$$|f|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))} \asymp \left[\int_0^1 [s^{-\underline{\alpha}}\omega_r(f, s^{\alpha_1'}, \dots, s^{\alpha_d'})_p]^q \, \frac{ds}{s}\right]^{1/q}. \tag{2.2.7}$$

As is customary, we discretize the integral in (2.2.7) over the intervals $(2^{-k-1}, 2^{-k}]$, $k = 0, 1, \dots$ to get a still more convenient form of the anisotropic smoothness seminorm as follows:

$$|f|_{B_q^{\alpha}(L^p(\Omega))} \asymp \left[\sum_{k=0}^{\infty}[2^{k\underline{\alpha}}\omega_r(f, 2^{-k\alpha'_1}, \dots, 2^{-k\alpha'_d})_p]^q\right]^{1/q}. \tag{2.2.8}$$

Finally, let us note that

$$|f|_{B_q^{\alpha}(L^p(\Omega))} \asymp \left[\sum_{k=0}^{\infty}\left[2^{k\underline{\alpha}}\omega_r(f, 2^{-\lfloor k\alpha'_1\rfloor}, \dots, 2^{-\lfloor k\alpha'_d\rfloor})_p\right]^q\right]^{1/q} \tag{2.2.9}$$

by Corollary 2.1.2.

Chapter 3
Local Polynomial and Dyadic Spline Approximation

In this chapter we recall several well-known results on polynomial approximation, with the ultimate objective of finding sequence norms equivalent to the anisotropic smoothness norm. The proofs parallel those of analagous results in [D-P].

3.1 Local Polynomial Approximation. Let

$$P = P(x_1, \dots, x_d) = \sum_{l_d=1}^{s_d} \cdots \sum_{l_1=1}^{s_1} c_{l_1,\dots,l_d} x_1^{p_{l_1}} \cdots x_d^{p_{l_d}}$$

be a polynomial defined on $\mathbb{R}^d$. We shall start by defining the *coordinate degree* of P ($\deg P$). Let $L := \{(l_1, \dots, l_d) \mid c_{l_1,\dots,l_d} \neq 0\}$. Then the coordinate degree of P is defined by

$$\deg P := \max_{(l_1,\dots,l_d)\in L} \max\{p_{l_1}, \dots, p_{l_d}\}.$$

Next we define

$$E_r(f, I)_p := \inf_{\deg P < r} \|f - P\|_p(I). \tag{3.1.1}$$

When I is omitted the norm is understood to be taken over Ω.

We shall need the following estimate for the local error of polynomial approximation:

Theorem 3.1.1 (Generalization of Whitney's Theorem). *Let*

$$I = I_1 \times \cdots \times I_d,$$

where I_i is an interval in $\mathbb{R}$ with length l_{I_i}, $i = 1, \dots, d$. Then given $f \in L^p(I)$, $0 < p \le \infty$, we have

$$E_r(f, I)_p \le C\omega_r(f, l_{I_1}, \dots, l_{I_d}, I)_p, \quad C = C(r, d, p). \tag{3.1.2}$$

Proof. Apply a change of variables to the original Whitney's Inequality, or see [S-O]. □

Corollary 3.1.2. *If $f \in L^p(I)$, with $0 < p \le \infty$, and $I = I_1 \times \cdots \times I_d$, then*

$$E_r(f, I)_p \le C w_r(f, l_{I_1}, \dots, l_{I_d}, I)_p, \quad C = C(r, d, p). \tag{3.1.3}$$

Proof. (3.1.3) follows from Lemma 2.1.3 and (3.1.2) □

We say that P is a polynomial of best $L^p(I)$ approximation of coordinate degree less than r if $\|f-P\|_p(I) = E_r(f, I)_p$ (such a polynomial always exists; see [G]). We say that P is a *near best* $L^p(I)$ approximation to f from polynomials of coordinate degree less than r with constant A if

$$\|f - P\|_p(I) \le A E_r(f, I)_p. \tag{3.1.4}$$

We will fix the constant A now.

Lemma 3.1.3. *For $0 < p, q \le \infty$ and a polynomial P of coordinate degree less than r, we have*

(3.1.5)
$$C_1 \left[\frac{1}{|I|} \int_I |P(\mathbf{x})|^q \, d\mathbf{x} \right]^{1/q} \le \left[\frac{1}{|I|} \int_I |P(\mathbf{x})|^p \, d\mathbf{x} \right]^{1/p} \le C_2 \left[\frac{1}{|I|} \int_I |P(\mathbf{x})|^q \, d\mathbf{x} \right]^{1/q}$$
$C_i = C_i(p, q, r, d), \ i = 1, 2.$

Proof. A change of variables shows that we can assume $I = \Omega$. Now the equivalence follows from the fact that any two quasinorms on a finite dimensional space are equivalent. □

3.2 Dyadic Spline Approximation. Let $\alpha_1, \dots, \alpha_d$, $\underline{\alpha}$, $\overline{\alpha}$, $\boldsymbol{\alpha}$, r, $\alpha'_1, \dots, \alpha'_d$ be as described in Section 2.2. For $k = 0, 1, \dots$, let $\mathcal{D}_k := \mathcal{D}_k^{\boldsymbol{\alpha}}$ denote the collection of dyadic intervals $I = I_1 \times \cdots \times I_d$, where $I_i = [j_i 2^{-\lfloor k\alpha'_i \rfloor}, (j_i+1)2^{-\lfloor k\alpha'_i \rfloor}]$, $i = 1, \dots, d$. Let $\mathcal{D} := \mathcal{D}^{\boldsymbol{\alpha}} := \cup_{k=0}^{\infty} \mathcal{D}_k^{\boldsymbol{\alpha}}$. We shall refer to the members of $\mathcal{D}$ as *blocks* . We define $l(I)$ for any block I to be the length of the shortest side of I; then $l(I) = 2^{-k}$ if and only if $I \in \mathcal{D}_k$. Let $\mathcal{D}_k(\Omega)$ be the blocks in $\mathcal{D}_k$ which are contained in Ω. Note that the blocks in $\mathcal{D}_{k+1}$ are *strictly* smaller than those in $\mathcal{D}_k$, because at least one of $\alpha'_1, \dots, \alpha'_d$ equals 1. Define $\Pi_k := \Pi_k(r)$ to be the space of piecewise polynomials of coordinate degree less than r on $\mathcal{D}_k$, and $\Pi_k(\Omega)$ to be the restrictions of functions in Π_k to Ω.

Denote by ψ the univariate B-spline of degree $r-1$ which has knots at $0, \dots, r$, defined by

$$\psi(x) := r[0, 1, \dots, r] \left(((x - \cdot)_+)^{r-1} \right)$$

Next, define

$$\psi(\mathbf{x}) := \psi(x_1) \cdots \psi(x_d). \tag{3.2.1}$$

To obtain splines in Π_k, we let

$$I := [j_1 2^{-\lfloor k\alpha'_1 \rfloor}, (j_1 + 1) 2^{-\lfloor k\alpha'_1 \rfloor}] \times \cdots \times [j_d 2^{-\lfloor k\alpha'_d \rfloor}, (j_d + 1) 2^{-\lfloor k\alpha'_d \rfloor}]$$

for $j_1, \ldots, j_d \in \mathbb{Z}$. We shall say that the vector $(j_1 2^{-\lfloor k\alpha'_1 \rfloor}, \ldots, j_d 2^{-\lfloor k\alpha'_d \rfloor})$ *corresponds* to I and we shall call the multi-index $(j_1, \ldots, j_d)$ the *position vector* of I. We index the splines in Π_k by I as follows:

$$\psi_I(\mathbf{x}) := \psi(2^{\lfloor k\alpha'_1 \rfloor} x_1 - j_1, \ldots, 2^{\lfloor k\alpha'_d \rfloor} x_d - j_d), \quad \mathbf{j} \in \mathbb{Z}^d. \tag{3.2.2}$$

Define $I' := \operatorname{supp} \psi_I$. It is well-known that I' consists of r^d blocks from $\mathcal{D}_k$, where $\mathcal{D}_k \ni I$, and thus $|I'| = r^d |I|$.

A fundamentally important property of ψ is that it can be written as a linear combination of the splines $\psi(2\mathbf{x} - \mathbf{j})$, $\mathbf{j} \in \mathbb{Z}^d$. That is, ψ satisfies a so-called *refinement equation*:

$$\psi(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{Z}^d} d_{\mathbf{j}} \psi(2\mathbf{x} - \mathbf{j})$$

where $d_{\mathbf{j}} \geq 0$ for all $\mathbf{j} \in \mathbb{Z}^d$ and only finitely many of the $d_{\mathbf{j}}$ are nonzero. It follows that, for $I \in \mathcal{D}_k$, we can rewrite ψ_I by

$$\psi_I(\mathbf{x}) = \sum_{I^{(\mathbf{j})}} d_{k,\mathbf{j}} \psi_{I^{(\mathbf{j})}}(\mathbf{x}) \tag{3.2.3}$$

where the $I^{(\mathbf{j})}$ comprise the set of all blocks $J \in \mathcal{D}_{k+1}$ such that $J' \cap I' \neq \emptyset$. Although the sequence of refinement coefficients $(d_{k,\mathbf{j}})_{\mathbf{j} \in \mathbb{Z}^d}$ depends on the dyadic level k of the spline which is rewritten, there are only a finite number of these sequences.

Now define $\Lambda := \Lambda(k)$ to be the set of all blocks $I \in \mathcal{D}_k$ such that ψ_I does not vanish identically on Ω. Let $\mathcal{S}_k := \mathcal{S}_k^{\boldsymbol{\alpha}} := \mathcal{S}_k^{\boldsymbol{\alpha}}(\Omega)$ denote the linear span of the splines ψ_I, $I \in \Lambda(k)$. Then, in fact, $\mathcal{S}_k = C^{r-2}(\Omega) \cap \Pi_k(\Omega)$ for $r > 1$, and $\mathcal{S}_k = \Pi_k(\Omega)$ for $r = 1$ (because we use the coordinate degree). So $\mathcal{S}_k$ is *strictly* contained in $\mathcal{S}_{k+1}$ for $k = 0, 1, \ldots$ since $\Pi_k \subsetneq \Pi_{k+1}$ for $k = 0, 1, \ldots$. Each spline $S \in \mathcal{S}_0$ can be written as

$$S = \sum_{\mathbf{j} \in \mathbb{Z}^d} c_{\mathbf{j}}(S) \psi(\cdot - \mathbf{j}) \tag{3.2.4}$$

with $c_{\mathbf{j}}$ the dual functionals to the basis $\{\psi(\cdot - \mathbf{j}) \mid \mathbf{j} \in \mathbb{Z}^d\}$ of $\mathcal{S}_0$. In fact,

$$c_{\mathbf{j}}(S) = c_0(S(\cdot + \mathbf{j})). \tag{3.2.5}$$

By translation and dilation we obtain for any $S \in \mathcal{S}_k$, $k = 1, 2, \ldots$, that

$$S = \sum_{J \in \mathcal{D}_k} c_J(S) \psi_J \tag{3.2.6}$$

where $c_J(S) := c_0(S(2^{-\lfloor k\alpha'_1 \rfloor} x_1 + j_1, \ldots, 2^{-\lfloor k\alpha'_d \rfloor} x_d + j_d))$ and J corresponds to $(j_1 2^{-\lfloor k\alpha'_1 \rfloor}, \ldots, j_d 2^{-\lfloor k\alpha'_d \rfloor})$. In fact, we can represent the functionals c_I for $I \in \mathcal{D}$ by

$$c_I(S) = \int_{\Omega} S \mu_I \, d\mathbf{x} \tag{3.2.7}$$

where $\mu_I = P_{\tilde{I}} \chi_{\tilde{I}}$, $P_{\tilde{I}}$ is a certain polynomial depending on $\tilde{I}$, and $\tilde{I}$ can be chosen to be any block such that $|\tilde{I}| = |I|$ and $\tilde{I} \subset \Omega \cap I'$.

We call a block J a *support block* of ψ_I if $|J| \leq |I|$ and $J \subset I'$. Note that if J is a support block of ψ_I then so is any subblock of J. We shall need an estimate for the functionals c_I, which is given in Lemma 3.2.1 and proved in [D-J-P], Chapter 4.

Lemma 3.2.1. *There is a constant $0 < \delta < 1$, $\delta = \delta(\psi, d, p)$, such that for any block $I \in \mathcal{D}_k$ and any support block J of ψ_I with $|J| = |I|$ and any set $E \subset J$ with $|E| \le \delta |J|$, we have*

$$|c_I(S)| \le C \left(\frac{1}{|I|} \int\limits_{J \setminus E} |S(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p}, \quad S \in \mathcal{S}_k, \quad C = C(\psi, p). \tag{3.2.8}$$

Lemma 3.2.2. *If $S = \sum_{I \in \Lambda} c_I(S) \psi_I \in \mathcal{S}_k$ then for any $0 < p \le \infty$, we have*

$$C_1 \|S\|_p \le \left(\sum_{I \in \Lambda} |c_I(S)|^p |I| \right)^{1/p} \le C_2 \|S\|_p, \quad C = C(r, d, p), \quad i = 1, 2. \tag{3.2.9}$$

For a proof, see [D-P], Lemma 4.2, but replace the word "cube" with "block" and use Lemma 3.2.1 (with $E = \emptyset$) in place of the corresponding result (Lemma 4.1) in [D-P].

Corollary 3.2.3. *There exists a constant $0 < \delta < 1$, $\delta = \delta(\phi, d, p)$, such that given $S \in \mathcal{S}_k$ and $J \in \mathcal{D}_k$, $k \in \{0, 1, \dots\}$, and any set $E \subset J'$ with $|E| \le \delta |J|$, we have*

$$\left(\int\limits_{J'} |S(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p} \le C \left(\int\limits_{J' \setminus E} |S(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p}, \quad C = C(\phi, p). \tag{3.2.10}$$

Proof. Write $S(\mathbf{x})$ as $\sum_{I \in \Lambda(k)} c_I(S) \phi_I(S)$. By (3.2.8) there exists $0 < \delta < 1$, $\delta = \delta(\phi, d, p)$, such that given $S \in \mathcal{S}_k$, $J \in \mathcal{D}_k$, and any set E with $E \subset J'$ with $|E| \le \delta |J|$, we have

$$|c_I(S)|^p |I| \le C \int\limits_{J' \setminus E} |S(\mathbf{x})|^p \, d\mathbf{x}$$

for any $I \in \mathcal{D}_k$ such that J is a support block of ψ_I. Adding this inequality over all $I \in \mathcal{D}_k$ such that $|I' \cap J'| \ne 0$ and using (3.2.9) gives us

$$\int\limits_{J'} |S(\mathbf{x})|^p \, d\mathbf{x} \le C \int\limits_{J' \setminus E} |S(\mathbf{x})|^p \, d\mathbf{x}. \qquad \square$$

We need to extend the domain of the functionals c_I, $I \in \mathcal{D}_k$, to Π_k, for $k = 0, 1, \dots$. To this end, for $I \in \mathcal{D}_k$, $S \in \Pi_k$, define the quasi-interpolant operator

$$Q_k(S) = \sum_{I \in \Lambda} c_I(S) \psi_I \tag{3.2.11}$$

where c_I is found using (3.2.7). For $k = 0, 1, \dots$, $I \in \mathcal{D}_k$, and $f \in L^p(\Omega)$, let $P_I := P_I(f)$ be a near-best $L^p(I)$ approximation to f with constant A from polynomials of coordinate degree less than r. Define the operator $S_k : L^p(\Omega) \mapsto \Pi_k$ by

$$(S_k f)(\mathbf{x}) := P_I(\mathbf{x}), \text{ for } \mathbf{x} \in \operatorname{int} I.$$

Finally, define

$$T_k := T_k f := Q_k(S_k(f)), \quad k = 0, 1, \ldots$$

It is shown in [D-P1] that $\|T_k(f)\|_p \le C\|f\|_p$.

We shall give an upper bound on the approximation error $\|f - T_k(f)\|_p$ shortly. First we introduce some notation. If $I \in \mathcal{D}_k$, let $\tilde{I}$ be the smallest block containing all the $J \in \mathcal{D}_k(\Omega)$ such that J is a support block of some $\psi_K \in \mathcal{S}_k$ with $|K' \cap I| \ne 0$. Then $|\tilde{I}| \le C|I|$, $C = C(r, d)$, and $\tilde{I} \subseteq \Omega$.

Theorem 3.2.4. $\|f - T_k(f)\|_p \le C\omega_r(f, 2^{-k\alpha'_1}, \ldots, 2^{-k\alpha'_d})_p$, $C = C(r, d, p, A)$.

Proof. The proof is similar to that of Theorem 4.5 of [D-P1]. In [D-P1], it is shown that

$$\|f - T_k\|_p(I) \le CE_r(f, \tilde{I})_p. \tag{3.2.12}$$

Now, each point $\mathbf{x} \in \Omega$ lies in only $C(r, d)$ of the blocks $\tilde{I}$. So we may raise both sides of (3.2.12) to the pth power, add over all $I \in \mathcal{D}_k(\Omega)$, and use (3.1.3) to obtain

$$\begin{aligned} \|f - T_k\|_p^p(\Omega) &\le C \sum_{I \in \mathcal{D}_k(\Omega)} w_r(f, l_{\tilde{I}_1}, \ldots, l_{\tilde{I}_d}, \tilde{I})_p^p \\ &= C \sum_{I \in \mathcal{D}_k(\Omega)} l_{\tilde{I}_1}^{-1} \cdots l_{\tilde{I}_d}^{-1} \int_0^{l_{\tilde{I}_1}} \cdots \int_0^{l_{\tilde{I}_d}} \int_{\tilde{I}(r\mathbf{s})} |\Delta_{\mathbf{s}}^r(f, \mathbf{x})|^p \, d\mathbf{x} \, ds_d \cdots ds_1. \end{aligned} \tag{3.2.13}$$

Let $t_i := \max l_{\tilde{I}_i}$. Then $t_i \asymp 2^{-k\alpha'_i}$, $i = 1, \ldots, d$, with constants of equivalence depending only upon r and d. Thus the right hand side of (3.2.13) is bounded above by

$$\begin{aligned} &C \sum_{I \in \mathcal{D}_k(\Omega)} t_1^{-1} \cdots t_d^{-1} \int_0^{t_1} \cdots \int_0^{t_d} \int_{\tilde{I}(r\mathbf{s})} |\Delta_{\mathbf{s}}^r(f, \mathbf{x})|^p \, d\mathbf{x} \, ds_d \cdots ds_1 \\ =&C t_1^{-1} \cdots t_d^{-1} \int_0^{t_1} \cdots \int_0^{t_d} \sum_{I \in \mathcal{D}_k(\Omega)} \int_{\tilde{I}(r\mathbf{s})} |\Delta_{\mathbf{s}}^r(f, \mathbf{x})|^p \, d\mathbf{x} \, ds_d \cdots ds_1 \\ \le&C t_1^{-1} \cdots t_d^{-1} \int_0^{t_1} \cdots \int_0^{t_d} \int_{\Omega(r\mathbf{s})} |\Delta_{\mathbf{s}}^r(f, \mathbf{x})|^p \, d\mathbf{x} \, ds_d \cdots ds_1 \\ \le&C t_1^{-1} \cdots t_d^{-1} \int_0^{t_1} \cdots \int_0^{t_d} \omega_r(f, t_1, \ldots, t_d)_p^p \, ds_d \cdots ds_1 \\ \le&C\omega_r(f, 2^{-k\alpha'_1}, \ldots, 2^{-k\alpha'_d})_p^p. \qquad \square \end{aligned}$$

Now let $s_k(f)_p := \inf_{S \in \mathcal{S}_k} \|f - S\|_p$, $s_{-1}(f)_p := \|f\|_p$. Then Theorem 3.2.4 yields Corollary 3.2.5.

Corollary 3.2.5. *For $f \in L^p(\Omega)$, we have*

$$s_k(f)_p \le C\omega_r(f, 2^{-k\alpha_1'}, \dots, 2^{-k\alpha_d'})_p. \tag{3.2.14}$$

As a final step to obtaining seminorms equivalent to the anisotropic smoothness norm, we shall need an upper bound on $\omega_r(f, 2^{-k\alpha_1'}, \dots, 2^{-k\alpha_d'})_p$ in terms of the $s_j(f)_p$, $j = -1, \dots, k$. But first we must prove the following lemma.

Lemma 3.2.6. *Let $\nu : [0,1]^d \to \mathbb{R}$, with $\nu \in C^r([0,1]^d)$ and $\left\| \frac{\partial^{\mathbf{s}} \nu}{\partial \mathbf{x}^{\mathbf{s}}} \right\|_\infty \le C$ for all multi-indices $\mathbf{s} = (s_1, \dots, s_d) \in \mathbb{Z}^d$ with $|\mathbf{s}| := s_1 + \dots + s_d \le r$. Let $\nu_{\mathbf{k}} := \nu(2^{k_1}x_1, \dots, 2^{k_d}x_d)$. Then*

$$\|\Delta_{\mathbf{h}}^r \nu_{\mathbf{k}}\|_\infty(\Omega(r\mathbf{h})) \le C(2^{k_1}h_1 + \dots + 2^{k_d}h_d)^r. \tag{3.2.15}$$

Proof. Repeated applications of the mean value theorem show that, for some $\mathbf{x}$,

$$\Delta_{\mathbf{h}}^r \nu_{\mathbf{k}} = \sum_{|\mathbf{s}|=r} \left. \frac{\partial^{\mathbf{s}} \nu}{\partial \mathbf{x}^{\mathbf{s}}} \right|_{\mathbf{x}} h_1^{s_1} \cdots h_d^{s_d}.$$

So,

$$|\Delta_{\mathbf{h}}^r \nu_{\mathbf{k}}| \le C \sum_{|\mathbf{s}|=r} 2^{k_1 s_1} \cdots 2^{k_d s_d} h_1^{s_1} \cdots h_d^{s_d} = C(2^{k_1}h_1 + \dots + 2^{k_d}h_d)^r. \qquad \square$$

We are now ready to prove the following inverse estimate.

Theorem 3.2.7. *For each $k > 0$ and for $\lambda := \min(r, r-1+1/p)$, and $f \in L^p(\Omega)$, $0 < p \le \infty$, we have*

$$\omega_r(f, 2^{-k\alpha_1'}, \dots, 2^{-k\alpha_d'})_p \le C 2^{-k\lambda \underline{\alpha}/\overline{\alpha}} \left[\sum_{j=-1}^{k} [2^{j\lambda\underline{\alpha}/\overline{\alpha}} s_j(f)_p]^\mu \right]^{1/\mu}, \tag{3.2.16}$$

where $\mu \le \min(1, p)$.

Proof. The proof is basically the same as that of Theorem 4.8 in [D-P1]. Let U_k be a best $L^p(\Omega)$ approximation to f from $\mathcal{S}_k$, $k = 0, 1, \dots$ with $U_{-1} := 0$, and let $u_k := U_k - U_{k-1}$, $k = 0, 1, \dots$ Let $\mathbf{h} = (h_1, \dots, h_d)$ with $0 < h_i \le 1/(dr2^{k\alpha_i'})$, $i = 1, \dots, d$. Using (1.2.4), write

$$\|\Delta_{\mathbf{h}}^r(f, \mathbf{x})\|_p(\Omega(r\mathbf{h})) \le C \left[s_k(f)^\mu + \sum_{j=0}^{k} \|\Delta_{\mathbf{h}}^r u_j\|_p^\mu \right]^{1/\mu}. \tag{3.2.17}$$

To estimate $\|\Delta_{\mathbf{h}}^r u_j\|_p(\Omega(r\mathbf{h}))$, write u_j in its B-spline series as follows:

$$u_j = \sum_{I \in \Lambda(j)} c_I(u_j)\psi_I. \tag{3.2.18}$$

For each $\mathbf{x}$, at most $C(r,d)$ splines at a fixed dyadic level are nonzero at $\mathbf{x}$. Hence

$$|\Delta^r_{\mathbf{h}}(u_j,\mathbf{x})|^p \le C \sum_{I\in\Lambda(j)} |c_I|^p |\Delta^r_{\mathbf{h}}(\psi_I,\mathbf{x})|^p. \tag{3.2.19}$$

Now we give two estimates for $|\Delta^r_{\mathbf{h}}(\psi_I,\mathbf{x})|$. The first estimate is for the set $\Gamma :=$ the set of all $\mathbf{x}$ such that $\mathbf{x}$ and $\mathbf{x}+r\mathbf{h}$ are in the same block $J\in\mathcal{D}_j$, and $\psi_I \not\equiv 0$ on J. Since $\psi_I \in C^r(J)$ we have by (3.2.15), for $x\in\Gamma$,

$$|\Delta^r_{\mathbf{h}}(\psi_I,\mathbf{x})| \le C(2^{j\alpha'_1}h_1+\cdots+2^{j\alpha'_d}h_d)^r. \tag{3.2.20}$$

The second estimate of $|\Delta^r_{\mathbf{h}}(\psi_I,\mathbf{x})|$ is for the set $\Gamma' :=$ the set of all $\mathbf{x}$ such that $\mathbf{x}$ and $\mathbf{x}+r\mathbf{h}$ are in different blocks and ψ_I does not vanish identically on both of those blocks. Since $\psi_I\in C^{r-1}(\Omega)$, we have, again by (3.2.15),

(3.2.21)

$$|\Delta^r_{\mathbf{h}}(\psi_I,\mathbf{x})| \le C(|\Delta^{r-1}_{\mathbf{h}}(\psi_I,\mathbf{x})|+|\Delta^{r-1}_{\mathbf{h}}(\psi_I,\mathbf{x}+\mathbf{h})|) \le C(2^{j\alpha'_1}h_1+\cdots+2^{j\alpha'_d}h_d)^{r-1}.$$

Now, $|\Gamma| \le C2^{-j\alpha'_1-\cdots-j\alpha'_d}$ because the support of ψ_I has measure not exceeding $C2^{-j\alpha'_1-\cdots-j\alpha'_d}$. We also claim that

$$|\Gamma'| \le C2^{-j\alpha'_1-\cdots-j\alpha'_d}\left[h_1 2^{j\alpha'_1}+\cdots+h_d 2^{j\alpha'_d}\right]. \tag{3.2.22}$$

To prove this, let $\Gamma'_i := \{\mathbf{x}\in\mathbb{R}^d | (\mathbf{x}$ and $\mathbf{x}-rh_i\mathbf{e}_i)$ or $(\mathbf{x}$ and $\mathbf{x}+rh_i\mathbf{e}_i)$ are in different blocks from $\mathcal{D}_j\}$, for $i=1,\ldots,d$. Then, if $[a_1,b_1]\times\cdots\times[a_d,b_d]$ is an interval from $\mathcal{D}_j$ containing a point from Γ'_1, then

$$\Gamma'_1 \subseteq ([b_1-rh_1,b_1]\cup[a_1,a_1+rh_1])\times[a_2,b_2]\times\cdots\times[a_d,b_d].$$

So

$$|\Gamma'_1| \le Ch_1 2^{-j\alpha'_2}\cdots 2^{-j\alpha'_d} = Ch_1 2^{-j\alpha'_1-\cdots-j\alpha'_d}2^{j\alpha'_1}.$$

Similiarly, for $i=2,\ldots,d$, $|\Gamma'_i| \le Ch_i 2^{-j\alpha'_1-\cdots-j\alpha'_d}2^{j\alpha'_i}$. Thus, since ψ_I vanishes on all but C cubes from $\mathcal{D}_j$, we have

$$|\Gamma'| \le C\sum_{i=1}^d |\Gamma'_i| \le C2^{-j\alpha'_1-\cdots-j\alpha'_d}\left[h_1 2^{j\alpha'_1}+\cdots+h_d 2^{j\alpha'_d}\right]$$

as required.

Now we can make the estimate

$$\int_{\Omega(r\mathbf{h})} |\Delta^r_{\mathbf{h}}\psi_I|^p \le |\Gamma| \|\Delta^r_{\mathbf{h}}\psi_I\|^p_\infty(\Gamma) + |\Gamma'| \|\Delta^r_{\mathbf{h}}\psi_I\|^p_\infty(\Gamma')$$

(3.2.23)

$$\le 2^{-j(\alpha'_1-\cdots-\alpha'_d)}[(2^{j\alpha'_1}h_1+\cdots+2^{j\alpha'_d}h_d)^{rp} + (2^{j\alpha'_1}h_1+\cdots+2^{j\alpha'_d}h_d)^{1+(r-1)p}].$$

But

$$2^{j\alpha'_1}h_1+\cdots+2^{j\alpha'_d}h_d \le 2^{j\alpha'_1}\frac{1}{rd}2^{-k\alpha'_1}+\cdots+2^{j\alpha'_d}\frac{1}{rd}2^{-k\alpha'_d} < 1,$$

so the right hand side of (3.2.23) is bounded above by

$$C2^{-j\alpha_1'-\cdots-j\alpha_d'}(2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d)^{\lambda p}. \tag{3.2.24}$$

Now integrate (3.2.19), then use the estimates (3.2.23) and (3.2.24) to get

$$\begin{aligned}
&\|\Delta_{\mathbf{h}}^r u_j\|_p(\Omega(r\mathbf{h}))\\
&\le C\Bigg(\sum_{I\in\Lambda(j)}|c_I|^p 2^{-j\alpha_1'-\cdots-j\alpha_d'}\Bigg)^{1/p}\Big(2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d\Big)^{\lambda}\\
&\le C\|u_j\|_p(2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d)^{\lambda}\ \text{(by 3.2.9)}\\
&\le C[\|f-U_{j-1}\|_p+\|f-U_j\|_p](2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d)^{\lambda}\\
&\le C(s_j(f)_p+s_{j-1}(f)_p)(2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d)^{\lambda}.
\end{aligned} \tag{3.2.25}$$

Substituting (3.2.25) into (3.2.17) gives

$$\|\Delta_{\mathbf{h}}^r(f,\mathbf{x})\|_p(\Omega(r\mathbf{h}))\le C\left[s_k(f)_p^{\mu}+\sum_{j=-1}^{k}(2^{j\alpha_1'}h_1+\cdots+2^{j\alpha_d'}h_d)^{\lambda\mu}s_j(f)_p^{\mu}\right]^{1/\mu}. \tag{3.2.26}$$

Taking the sup of $\|\Delta_{\mathbf{h}}^r(f,\mathbf{x})\|_p(\Omega(r\mathbf{h}))$ over $0<h_i\le \frac{1}{dr}2^{-k\alpha_i'}$, $i=1,\dots,d$, (3.2.26) gives us

$$\begin{aligned}
\omega_r(f,2^{-k\alpha_1'},\dots,2^{-k\alpha_d'})_p &\le C\omega_r(f,\frac{1}{rd}2^{-k\alpha_1'},\dots,\frac{1}{rd}2^{-k\alpha_d'},)_p\\
&\le C[s_k(f)_p^{\mu}+\sum_{j=-1}^{k}[2^{(j-k)\alpha_1'}+\cdots+2^{(j-k)\alpha_d'}]^{\lambda\mu}s_j(f)_p^{\mu}]^{1/\mu}\\
&\le C\left[s_k(f)_p^{\mu}+\sum_{j=-1}^{k}\left[2^{(j-k)\underline{\alpha}/\overline{\alpha}}\right]^{\lambda\mu}s_j(f)_p^{\mu}\right]^{1\mu}\\
&\le C2^{-k\lambda\underline{\alpha}/\overline{\alpha}}\left[\sum_{j=-1}^{k}\left[2^{j\lambda\underline{\alpha}/\overline{\alpha}}s_j(f)_p\right]^{\mu}\right]^{1/\mu}. \qquad \square
\end{aligned}$$

We can now prove a Marchaud-Type inequality for $\omega_r(f,\mathbf{t})_p$ in essentially the same way as the corresponding result is proved in [D-P1].

Theorem 3.2.8. *For $f\in L^p(\Omega)$, $r'<r$, there exists a constant $C=C(r,p,d)$ such that*

$$\begin{aligned}
&\omega_{r'}(f,2^{-k\alpha_1'},\dots,2^{-k\alpha_d'},)_p\\
&\qquad\le C2^{-kr'\underline{\alpha}/\overline{\alpha}}\left[\|f\|_p^{\mu}+\sum_{j=0}^{k}\left[2^{jr'}\omega_r(f,2^{-j\alpha_1'},\dots,2^{-j\alpha_d'})_p\right]^{\mu}\right]^{1/\mu}.
\end{aligned} \tag{3.2.27}$$

Proof. We first claim that

$$\omega_{r'}(f, 2^{-k\alpha'_1}, \ldots, 2^{-k\alpha'_d})_p \le C 2^{-kr'\underline{\alpha}/\overline{\alpha}} \left[\sum_{j=-1}^{k} \left[2^{jr'} s_j(f)_p \right]^{\mu} \right]^{1/\mu}. \tag{3.2.28}$$

This is proved in the same way as is Theorem 3.2.7, except that we replace (3.2.20) and (3.2.21) with the single inequality

$$|\Delta_{\mathbf{h}}^{r'}(\psi_I, \mathbf{x})| \le (2^{j\alpha'_1} h_1 + \cdots + 2^{j\alpha'_d} h_d)^{r'}$$

and replace λ with r'. Now, (3.2.27) follows immediately from (3.2.28) and (3.2.14). □

With the Marchaud inequality in hand, it can be shown that (2.2.3) holds, using the same proof given in [D-L] of the corresponding result for isotropic Besov spaces.

3.3 Seminorms Equivalent to the Anisotropic Smoothness Norm. We shall use the estimates of the last section to obtain several norms which are equivalent to (2.2.9). We shall make use of the $l_q^{\theta}(\mathbb{Z})$ norm of a sequence $\mathbf{a} = (a_k)_{k=-\infty}^{\infty}$, defined by

$$\|\mathbf{a}\|_{\theta,q} := \begin{cases} \left(\sum_{k \in \mathbb{Z}} \left[2^{k\theta} |a_k| \right]^q \right)^{1/q} & \text{If } 0 < q < \infty \\ \sup_{k \in \mathbb{Z}} 2^{k\theta} |a_k| & \text{If } q = \infty \end{cases}$$

In the special case $\theta = 0$, the $l_q^{\theta}(\mathbb{Z})$ norm of $\mathbf{a}$ is just the $l_q(\mathbb{Z})$ norm of $\mathbf{a}$, denoted $\|\mathbf{a}\|_{l_q(\mathbb{Z})}$. The space of all sequences $\mathbf{a}$ such that $\|\mathbf{a}\|_{l_q(\mathbb{Z})} < \infty$ is denoted $l_q(\mathbb{Z})$. We note the following well-known property of the $l_q(\mathbb{Z})$ norm for future use.

Theorem 3.3.1. *For $q < p$, we have $l_q(\mathbb{Z}) \hookrightarrow l_p(\mathbb{Z})$, and the containment is strict.*

We now need to recall the following inequality.

Theorem 3.3.2 (Hardy's Inequality). *Let $\mathbf{a} = (a_k)$, $\mathbf{b} = (b_k)$, $k \in \mathbb{Z}$ be two positive sequences. Suppose that for some $\mu > 0$, $C > 0$, one of the following holds:*

$$b_k \le C \left(\sum_{j=k}^{\infty} a_j^{\mu} \right)^{1/\mu}, \quad \text{for all } k \in \mathbb{Z}. \tag{3.3.1}$$

For some $\lambda > 0$,

$$b_k \le C 2^{-k\lambda} \left(\sum_{j=-\infty}^{k} (2^{j\lambda} a_j)^{\mu} \right)^{1/\mu}, \quad \text{for all } k \in \mathbb{Z}. \tag{3.3.2}$$

If (3.3.1) holds then

$$\|\mathbf{b}\|_{\theta,q} \le C \|\mathbf{a}\|_{\theta,q}, \quad C = C(\theta, q),$$

for all $\theta > 0$, $q > 0$.

If (3.3.2) *holds, then we have*

$$\|\mathbf{b}\|_{\theta,q} \le C\|\mathbf{a}\|_{\theta,q}, \quad C = C(\theta, q),$$

for all $0 < \theta < \lambda$ *and all* $q > 0$.

For a proof of Theorem 3.3.2, see [D-L].

We are finally ready to present norms equivalent to the mixed smoothness norm. We define T_k and P_I as in section 3.2, and define $t_k := T_k - T_{k-1}$, with $T_{-1} := 0$, and define the coefficients $c_I := c_I(t_k)$ for $I \in \Lambda(k)$. Again, we let

$$\lambda := \min(r, r - 1 + 1/p).$$

Also, for a sequence $(g_k) := (g_k)_{k=0}^{\infty}$ of functions in $L^p(\Omega)$, $0 < p \le \infty$, and for $0 < \theta < \infty$, $0 < q \le \infty$, define

$$\|(g_k)\|_{l_q^\theta(L^p(\Omega))} := \left(\sum_{k=0}^{\infty} [2^{k\theta}\|g_k\|_p]^q\right)^{1/q}.$$

Theorem 3.3.3. *Let* $0 < p, q \le \infty$, $\lambda > \overline{\alpha}$. *Then the following norms are equivalent to* $N(f) := \|f\|_{B_q^{\boldsymbol{\alpha}}(L^p(\Omega))}$ *with constants of equivalence depending only on* d, r, A, *and the constant in* (3.3.2)*:*

(*i*) $N_1(f) := \|(s_k(f))\|_{l_q^{\underline{\alpha}}(L^p(\Omega))} + \|f\|_p$.
(*ii*) $N_2(f) := \|(f - T_k(f))\|_{l_q^{\underline{\alpha}}(L^p(\Omega))} + \|f\|_p$.
(*iii*) $N_3(f) := \|(t_k(f))\|_{l_q^{\underline{\alpha}}(L^p(\Omega))}$.
(*iv*) $N_4(f) := \left[\sum_{k=0}^{\infty} 2^{kq\underline{\alpha}}\left[\sum_{I\in\Lambda(k)} |c_I|^p|I|\right]^{q/p}\right]^{1/q}$.

Proof. We can show that $N_1(f) \le N_2(f) \le N(f)$ in the same way the corresponding result from [D-P1] (Theorem 5.1.1) is proved. On the other hand, from Theorem 3.2.7 and Hardy's Inequality (3.3.2), we have $N(f) \le CN_1(f)$ for $\lambda\underline{\alpha}/\overline{\alpha} > \underline{\alpha}$. Thus $N(f) \asymp N_1(f) \asymp N_2(f)$ for $\lambda > \overline{\alpha}$. The other equivalences are proved in the same way as is Theorem 5.1 and Corollary 5.3 from [D-P1]. □

Chapter 4
The Direct and Inverse Theorems

4.1 The Direct Theorem. At this point we shall restrict our attention to the class of anisotropic Besov spaces $B^{\boldsymbol{\beta}} := B_\tau^{\boldsymbol{\beta}}(L^\tau(\Omega))$, where

$$\frac{1}{\tau} = \frac{H(\boldsymbol{\beta})}{d} + \frac{1}{p}, \quad 0 < p < \infty. \tag{4.1.1}$$

We define $\mathcal{D} := \mathcal{D}^{\boldsymbol{\beta}}$, $\mathcal{D}_k := \mathcal{D}_k^{\boldsymbol{\beta}}$, $\underline{\beta}$, $\overline{\beta}$, β_i', $i = 1, \dots, d$ the same way as the corresponding symbols in Chapter 3, but using $\boldsymbol{\beta}$ in place of $\boldsymbol{\alpha}$.

Define the seminorm

$$|f|_{\boldsymbol{\beta}} := \left[\sum_{I\in\mathcal{D}} |c_I(f)|^\tau |I|^{\tau/p}\right]^{1/\tau} \asymp |f|_{B^{\boldsymbol{\beta}}}, \tag{4.1.2}$$

where the equivalence follows from Theorem 3.3.3(iv). For $n = 1, 2, \ldots$, define Σ_n to be the set of all linear combinations $\sum_{I\in\Lambda} c_I\psi_I$, where Λ is any subset of $\mathcal{D}$ with $\#\Lambda \le n$. Then define $\sigma_n(f)$ to be the error, measured in the $L^p(\Omega)$ norm, of approximating f from Σ_n, i.e. $\sigma_n(f) := \inf_{S\in\Sigma_n} \|f - S\|_p$.
Let

$$F(\mathbf{x}) := \left[\sum_{I\in\mathcal{D}} \left[|I|^{-H(\boldsymbol{\beta})/d}|c_I(f)|\chi_I(\mathbf{x})\right]^\tau\right]^{1/\tau}. \tag{4.1.3}$$

Then, by (4.1.2),

$$\|F\|_\tau = |f|_{\boldsymbol{\beta}} \le C|f|_{B^{\boldsymbol{\beta}}}. \tag{4.1.4}$$

We now present a modified version of Theorem 2.1 from [D-J-P], called a Direct Theorem or Jackson Inequality.

Theorem 4.1.1. *For $C = C(d, \psi, \overline{\beta}/\underline{\beta})$,*

$$\sigma_n(f)_p \le Cn^{-H(\boldsymbol{\beta})/d}|f|_{B^{\boldsymbol{\beta}}}, \tag{4.1.5}$$

Proof. We shall indicate the modifications that must be made to Theorem 2.1 from [D-J-P]. Replace (2.14) with

$$\Sigma_1(\mathbf{x}) \le C_1\epsilon 2^{-(N/p)(\beta_1'+\cdots+\beta_d')}.$$

Then redefine N to be the smallest integer such that $C_1\epsilon 2^{-\frac{N}{p}(\beta_1'+\cdots+\beta_d')} \le \dfrac{t}{2}$. In (2.17), replace all occurences of $|I|^{\beta/d}$ with $|I|^{H(\boldsymbol{\beta})/d}$ to obtain

$$\Sigma_2(\mathbf{x}) = C2^{N\underline{\beta}}F(\mathbf{x}).$$

Using this with the definition of N gives

$$\Sigma_2(\mathbf{x}) \le C\epsilon^{pH(\boldsymbol{\beta})/d}t^{-pH(\boldsymbol{\beta})/d}F(\mathbf{x}),$$

which implies

$$\mu(\tilde{E}, t) \le \mu\left(F(\mathbf{x}), C\epsilon^{-pH(\boldsymbol{\beta})/d}t^{1+pH(\boldsymbol{\beta})/d}\right).$$

Making the change of variable

$$u = C\epsilon^{-pH(\boldsymbol{\beta})/d}t^{1+pH(\boldsymbol{\beta})/d}$$

gives us $\left\|\tilde{E}\right\|_p^p \le C\epsilon^{\tau H(\boldsymbol{\beta})/d}$, from which (4.1.5) follows. □

4.2 The Inverse Theorem. In this section we outline the proof of the Inverse Theorem for approximation by B-splines with our anisotropic scaling. This proof is a slight generalization of Theorem 5.6 from [D-J-P]. Let $S \in \Sigma_n$. By the definition of Σ_n we have that

$$S = \sum_{I \in \Lambda} a_I \psi_I, \ |\Lambda| \le n. \tag{4.2.1}$$

We may assume that all the a_I are nonzero. We shall let m be the maximum dyadic level of all the I appearing in (4.2.1). To prepare for the proof of the Inverse Theorem we must use a suitable representation of S that is different from (4.2.1). Let Λ_j be the set of blocks $I \in \Lambda \cap \mathcal{D}_j$ and define

$$S_j := \sum_{I \in \Lambda_j} a_I \psi_I. \tag{4.2.2}$$

Now, using the definition of "high density" blocks given in [D-J-P], let $\mathcal{B}_j$ denote the collection of high density blocks $I \in \mathcal{D}_j$. We shall say the dyadic block Q has distance $\le k$ from the dyadic block $I = I_1 \times \cdots \times I_d$ with $l(Q) = l(I)$ if $Q_i = \nu_i l(I_i) + I_i$ with $|\nu_i| \le k$, $i = 1, \dots, d$. Let $\tilde{\mathcal{B}}_i$ be the set of blocks at dyadic level i which have distance $\le 100r$ from one of the blocks in $\mathcal{B}_i$.

We can now write

$$S = S'_0 + \cdots + S'_m, \quad S'_j = \sum_{j \in \Lambda'_j} \gamma_I \psi_J,$$

$$\gamma_J := c_J(S'_j) = c_J(S''_j), \quad J \in \Lambda'_j. \tag{4.2.3}$$

where we construct S'_j, Λ'_j as in Chapter 5 of [D-J-P], but substituting our definition of $\tilde{\mathcal{B}}_i$ for that in [D-J-P].

We now estimate $|\Lambda'|$, where $\Lambda' := \cup_{j=0}^m \Lambda'_j$.

Lemma 4.2.1. *We have* $|\Lambda'| \le C(\psi, \delta)n$.

Proof. Make the following change in the proof of Lemma 5.5 from [D-J-P]: In order to bound the inner sum in (5.19) above by $C \sum_{\nu=j+1}^m 2^{jd} 2^{-\nu d} N_\nu$, first bound the inner sum in (5.19) above by $C \sum_{\nu=j+1}^m 2^{j(\beta'_1 + \cdots + \beta'_d)} 2^{-\nu(\beta'_1 + \cdots + \beta'_d)} N_\nu$. $\square$

The last result we need before we can start the proof of the Inverse Theorem is an upper bound on $|f|_{B^{\boldsymbol{\beta}}}$.

Lemma 4.2.2. *If* $f = \sum_{I \in \mathcal{D}} b_I \psi_I$, $0 < \underline{\beta}, \overline{\beta} < r - 1 + 1/p$, *then*

$$|f|_{B^{\boldsymbol{\beta}}} \le C \left(\sum_{I \in \mathcal{D}} |b_I|^\tau |I|^{\tau/p} \right)^{1/\tau}, \quad C = C(\psi, p).$$

Proof. Let $T_k := \sum_{I \in \mathcal{D}_k} b_I \psi_I$. Then by (1.2.4)

$$s_k(f)_\tau \le \left(\sum_{j=k+1}^{\infty} \|T_j\|_\tau^\mu \right)^{1/\mu}, \quad \mu := \min(1, \tau). \tag{4.2.4}$$

Now apply Hardy's inequality to (4.2.4):

$$\sum_{k=1}^{\infty}[2^{k\underline{\beta}}s_k(f)_\tau]^\tau \le C\sum_{k=1}^{\infty}2^{k\tau\underline{\beta}}\|T_k\|_\tau^\tau. \tag{4.2.5}$$

Now, by Theorem 3.3.3 we have that the left hand side of (4.2.5) is equivalent to $|f|^\tau_{B^{\boldsymbol\beta}}$. Also, by Lemma 3.2.2, we have

$$\begin{aligned} 2^{k\tau\underline{\beta}}\|T_k\|_\tau^\tau &\asymp \sum_{I\in\mathcal{D}_k}|b_I|^\tau 2^{k\tau\underline{\beta}}2^{-k(\beta_1'+\cdots+\beta_d')} \\ &= \sum_{I\in\mathcal{D}_k}|b_I|^\tau 2^{-(k\tau/p)(\beta_1'+\cdots+\beta_d')} \\ &\le C\sum_{I\in\mathcal{D}_k}|b_I|^\tau|I|^{\tau/p}. \qquad \square \end{aligned}$$

Finally, we are ready to prove the Inverse Theorem.

Theorem 4.2.3. *Let S have the form (4.2.1). If $1\le p<\infty$ then for $0<\underline{\beta},\overline{\beta}<r-1+1/p$,*

$$|S|_{B^{\boldsymbol\beta}} \le Cn^{H(\boldsymbol\beta)/d}\|S\|_p, \tag{4.2.6}$$

where $C=C(\psi,p,d)$.

Proof. Use the proof of Theorem 5.6 in [D-J-P] with the equality $1/\tau=\beta/d+1/p$ replaced with (1.5.1). □

Chapter 5
Interpolation of Anisotropic Besov Spaces and Applications

5.1 Definition and Properties of Interpolation Spaces. We now require interpolation theorems for anisotropic Besov spaces. The first application of these theorems will be to prove a Sobolev-type imbedding theorem for anisotropic Besov spaces; the second will be to obtain certain approximation spaces.

Recall that if X_0, X_1 is a pair of quasi-normed spaces which are continuously imbedded in a Hausdorff space X, then Peetre's K-functional is defined by

$$K(f,t) := K(f,t,X_0,X_1) := \inf_{\{f_0\in X_0, f_1\in X_1 | f=f_0+f_1\}}\{\|f_0\|_{X_0}+t\|f_1\|_{X_1}\} \tag{5.1.1}$$

for all $f\in X_0+X_1$. Now, if $0<\theta<1$ and $0<q\le\infty$, we define the interpolation space $X_{\theta,q}=(X_0,X_1)_{\theta,q}$ to be the set of all f such that

$$\|f\|_{X_{\theta,q}} := \|f\|_{X_0+X_1} + \left(\int_0^\infty [t^{-\theta}K(f,t)]^q\,\frac{dt}{t}\right)^{1/q} < \infty \tag{5.1.2}$$

where $\|f\|_{X_0+X_1} := K(f,1)$.

From [B-L], we have that $X_0\cap X_1$ is dense in $X_{\theta,q}$ if $q<\infty$. Therefore, if T is a continuous linear mapping defined on both X_0 and X_1, then T is defined on $X_{\theta,q}$. Moreover, from [B-L], if $T: X_i\to Y_i$ for $i=0,1$, then $T: X_{\theta,q}\to Y_{\theta,q}$ continuously. In particular,

$$\begin{gathered} X_i \hookrightarrow Y_i \text{ for } i=0,1 \implies X_{\theta,q}\hookrightarrow Y_{\theta,q} \\ 0<\theta<1,\ 0<q<\infty. \end{gathered} \tag{5.1.3}$$

5.2 A Sobolev-Type Imbedding Theorem. To obtain the Sobolev-type imbedding theorem, we first need to determine the interpolation spaces

$$(B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega)), B_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega)))_{\theta,q}$$

where $0<\gamma<1$. The seminorm for $B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega))$ is

$$|f|_{B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega))} = \left[\sum_{k=0}^{\infty}[2^k\omega_r(f,2^{-k/(\gamma\alpha_1)},\dots,2^{-k/(\gamma\alpha_d)})_{p_0}]^{q_0}\right]^{1/q_0}$$

$$\asymp \left[\sum_{k=0}^{\infty}[2^{k\gamma}\omega_r(f,2^{-k/\alpha_1},\dots,2^{-k/\alpha_d})_{p_0}]^{q_0}\right]^{1/q_0}. \tag{5.2.1}$$

Now, starting with (5.2.1) and using the same proof as that given for Theorem 3.3.3, we can show that

$$N_3'(f) := \|(t_k(f))\|_{l_{q_0}^{\gamma}(L^{p_0}(\Omega))} + \|f\|_{p_0} \tag{5.2.2}$$

is an equivalent norm for $B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega))$, provided $\gamma\overline{\alpha}+1<r$. Thus the interpolation spaces between $B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega))$, $B_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega))$ can be found using known results on the interpolation spaces between $l_{q_0}^{\gamma}(L^{p_0}(\Omega))$ and $l_{q_1}^{1}(L^{p_1}(\Omega))$.

To make use of these known results, we will need the following lemma proved in [D-P1].

(5.2.3)

$$f \in (B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega)), B_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega)))_{\theta,q} \iff (t_k(f)) \in (l_{q_0}^{\gamma}(L^{p_0}(\Omega)), l_{q_1}^{1}(L^{p_1}(\Omega)))_{\theta,q}$$

and

(5.2.4)

$$\|f\|_{(B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega)), B_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega)))_{\theta,q}} \asymp \|(t_k(f))\|_{(l_{q_0}^{\gamma}(L^{p_0}(\Omega)), l_{q_1}^{1}(L^{p_1}(\Omega)))_{\theta,q}}.$$

Now, from [P], we have for $0<q_0,q_1\le\infty$,

$$(l_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega)), l_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega)))_{\theta,q} = l_q^{\boldsymbol{\alpha}'}(L_{p,q}) \tag{5.2.5}$$

with $\boldsymbol{\alpha}' := \theta\gamma\boldsymbol{\alpha}+(1-\theta)\boldsymbol{\alpha}$, $1/q=\theta/q_0+(1-\theta)/q_1$, $1/p=\theta/p_0+(1-\theta)/p_1$. Here $L_{p,q}$ are the Lorentz spaces, whose definition is not needed here, but may be found in [B-B]. When $q=p$ we have $L_{p,q}=L^p$. So by (5.2.3), (5.2.4), (5.2.5), and (5.2.2), we obtain

Theorem 5.2.1. *If $0<p_0,p_1,q_0,q_1\le\infty$, $0<\theta<1$, $1/q:=\theta/q_0+(1-\theta)/q_1$, $1/p:=\theta/p_0+(1-\theta)/p_1$, then*

$$(B_{q_0}^{\gamma\boldsymbol{\alpha}}(L^{p_0}(\Omega)), B_{q_1}^{\boldsymbol{\alpha}}(L^{p_1}(\Omega)))_{\theta,q} = B_q^{\boldsymbol{\alpha}'}(L_p(\Omega))$$

with $\boldsymbol{\alpha}' := \theta\gamma\boldsymbol{\alpha}+(1-\theta)\boldsymbol{\alpha}$, provided $p=q$.

Define $\boldsymbol{\alpha}^0 := \gamma\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^1 := \boldsymbol{\alpha}$. It is important to notice that, in particular, if p is fixed and $1/q_i = H(\boldsymbol{\alpha}^i)/d+1/p$, $i=0,1$, then

$$(B_{q_0}^{\boldsymbol{\alpha}^0}(L^{q_0}(\Omega)), B_{q_1}^{\boldsymbol{\alpha}^1}(L^{q_1}(\Omega)))_{\theta,q} = B_q^{\boldsymbol{\alpha}'}(L^q(\Omega)) \tag{5.2.6}$$

with $1/q = H(\boldsymbol{\alpha}')/d + 1/p$.

Fix $0 < p < \infty$. We want to show that $B^{\boldsymbol{\alpha}}_\tau(L^\tau(\Omega)) \hookrightarrow L^p(\Omega)$, where

$$1/\tau = H(\boldsymbol{\alpha})/d + 1/p.$$

Now, since $\tau < p$, Theorems 3.3.1 and 3.3.3 imply that

$$B^{\boldsymbol{\alpha}}_\tau(L^\tau(\Omega)) \hookrightarrow B^{\boldsymbol{\alpha}}_p(L^\tau(\Omega)).$$

So it is enough to show that

$$B^{\boldsymbol{\alpha}}_p(L^\tau(\Omega)) \hookrightarrow L^p(\Omega).$$

First we need the inequality

$$\|S\|_p \le C 2^{k\underline{\alpha}} \|S\|_\tau \tag{5.2.7}$$

for $S \in L^p(\Omega)$. To prove (5.2.7) let $I \in \mathcal{D}_k$. By Lemma 3.1.3,

$$\begin{aligned} \|S\|^p_p(I) &\le |I|^{1/p-1/\tau} \|S\|_\tau(I) = |I|^{-H(\boldsymbol{\alpha})/d} \|S\|_\tau(I) \\ &\le C 2^{k(\alpha'_1+\cdots+\alpha'_d)H(\boldsymbol{\alpha})/d} \|S\|_\tau(I) = C 2^{k\underline{\alpha}} \|S\|_\tau(I). \end{aligned}$$

Thus,

$$\|S\|^p_p \le C 2^{k\underline{\alpha}p} \sum_{I\in\mathcal{D}_k} \|S\|^p_\tau(I) \le C 2^{k\underline{\alpha}p} \left(\sum_{I\in\mathcal{D}_k} \|S\|^\tau_\tau(I) \right)^{p/\tau} = C 2^{k\underline{\alpha}p} \|S\|^p_\tau, \tag{5.2.8}$$

so (5.2.7) is proved.

Finally, we can establish the Sobolev-type imbedding theorem.

Theorem 5.2.2. *If $1/\tau = H(\boldsymbol{\alpha})/d + 1/p$ then*

$$\|f\|_p \le C \|f\|_{B^{\boldsymbol{\alpha}}_p(L^\tau(\Omega))}, \quad C = C(H(\boldsymbol{\alpha}), r, d) \tag{5.2.9}$$

for all $f \in B^{\boldsymbol{\alpha}}_p(L^\tau(\Omega))$.

Proof. We follow the proof of Theorem 7.1 in [D-P]. Let $r > \overline{\alpha} + 1$, so that $\lambda := \min\left(r, r-1+\dfrac{1}{p}\right) > \overline{\alpha}$ and Theorem 3.3.3 holds. Write $f = \sum_{j=0}^\infty t_j$ as in Theorem 3.3.3. Then by (1.2.6), (5.2.7), (3.1.3) and Theorem 3.3.3(iii), we have for $\mu := \min(1, p)$,

$$\|f\|_p \le \left(\sum_{j=0}^\infty \|t_j\|^\mu_p \right)^{1/\mu} \le C \left(\sum_{j=0}^\infty \left(2^{j\underline{\alpha}} \|t_j\|_\tau\right)^\mu \right)^{1/\mu} \le C \|f\|_{B^{\boldsymbol{\alpha}}_\mu(L^\tau(\Omega))}. \tag{5.2.10}$$

So we are done if $p \le 1$. If $p > 1$ choose $1 \le p_0 < p < p_1 < \infty$ and choose $\gamma_0, \gamma_1 > 0$ such that

$$\frac{1}{\tau} = \frac{H(\gamma_0\boldsymbol{\alpha})}{d} + \frac{1}{p_0} \tag{5.2.11}$$

$$\frac{1}{\tau} = \frac{H(\gamma_1\boldsymbol{\alpha})}{d} + \frac{1}{p_1}. \tag{5.2.12}$$

Let $\boldsymbol{\alpha}^i := \gamma_i \boldsymbol{\alpha}$, $i = 0, 1$. Then by (5.2.10),

$$\|f\|_{p_i} \le C\|f\|_{B_1^{\boldsymbol{\alpha}^i}(L^\tau(\Omega))}, \quad i = 0, 1. \tag{5.2.13}$$

Now choose θ such that $1/p = \theta/p_0 + (1-\theta)/p_1$ and let $q := p$, $\boldsymbol{\alpha}' := \theta\boldsymbol{\alpha}^0 + (1-\theta)\boldsymbol{\alpha}^1$. Then by Theorem 5.2.1 and (5.2.13) we have

$$\|f\|_p \le C\|f\|_{B_p^{\boldsymbol{\alpha}'}(L^\tau(\Omega))}. \tag{5.2.14}$$

But $\boldsymbol{\alpha}' = \boldsymbol{\alpha}$. To see this multiply (5.2.11) by θ and (5.2.12) by $1-\theta$ and add to get

$$\frac{1}{\tau} = \frac{(\gamma_0\theta + \gamma_1(1-\theta))H(\boldsymbol{\alpha})}{d} + \frac{1}{p}. \tag{5.2.15}$$

Using (5.2.15) with the fact that $1/\tau = H(\boldsymbol{\alpha})/d + 1/p$ gives $\gamma_0\theta + \gamma_1(1-\theta) = 1$, and thus $\boldsymbol{\alpha}' = \boldsymbol{\alpha}$. □

5.3 Characterization of Approximation Spaces. Let $A_q^\alpha := A_q^\alpha(L^p(\Omega))$ be the set of all f such that

$$|f|_{A_q^\alpha} := \left[\sum_{n=0}^{\infty} [2^{n\alpha}\sigma_{2^n}(f)_p]^q\right]^{1/q} < \infty. \tag{5.3.1}$$

The norm of A_q^α is defined by $\|f\|_{A_q^\alpha} := \|f\|_p + |f|_{A_q^\alpha}$. Using standard results from nonlinear approximation theory (see [D-P2]), the results in section 5.2, along with (4.1.5) and (4.2.6) imply that

$$B^{\theta\boldsymbol{\alpha}} = A^{H(\theta\boldsymbol{\alpha})/d}(L^p(\Omega)) \tag{5.3.2}$$

with equivalent norms, for $0 < \theta < 1$, $1 \le p < \infty$, $0 < \underline{\alpha}, \overline{\alpha} < r - 1 + 1/p$.

CHAPTER 6
Further results

6.1 Optimal Scaling for $B^{\boldsymbol{\alpha}}$. Let $\boldsymbol{\alpha} \in \mathbb{R}^d$, $\boldsymbol{\beta} \in \mathbb{R}^d$ be vectors of smoothness parameters with $\boldsymbol{\alpha}' \ne \boldsymbol{\beta}'$. Let

$$\Phi^{\boldsymbol{\beta}} := \{\phi(2^{\lfloor k\beta_1' \rfloor}x_1 - j_1, \dots, 2^{\lfloor k\beta_d' \rfloor}x_d d - j_d) \mid k = 0, 1, \dots,\ j_1, \dots, j_d \in \mathbb{Z}\}$$

and define $\Phi^{\boldsymbol{\alpha}}$ similarly. Also define $\mathcal{S}_k^{\boldsymbol{\beta}}$, $k = 0, 1, \dots$ in the same way as $\mathcal{S}_k^{\boldsymbol{\alpha}}$ but with $\boldsymbol{\beta}$ in place of $\boldsymbol{\alpha}$. Given $f \in B^{\boldsymbol{\beta}}$, we wish to show that the wavelets from $\Phi^{\boldsymbol{\beta}}$ have the optimal scaling for decomposing f and compressing this decomposition. First we would like to make a few remarks about approximating members $\phi_J^{\boldsymbol{\beta}}$ of $\Phi^{\boldsymbol{\beta}}$ by members of $\mathcal{S}_k^{\boldsymbol{\alpha}}$.

First of all, let ϕ^1 denote the univariate B-spline defined in Section 3.2 and let

$$\phi_{j,k}^1 := \phi^1(2^k x - j), \quad j \in \mathbb{Z}, \quad k \in \{0, 1, \dots\}.$$

Redefine $\Lambda(k)$ to be the set of all $j \in \mathbb{Z}$ such that $|\operatorname{supp}(\phi_{j,k}^1) \cap [0,1]| \ne 0$, and let

$$\mathcal{S}_k^1 := \operatorname{span}\{\phi_{j,k}^1 \mid j \in \Lambda(k)\}$$

for $k = 0, 1, \dots$. For $f \in L^p([0,1])$, define

$$\operatorname{dist}(\mathcal{S}_k^1, f)_p := \inf_{g \in \mathcal{S}_k^1} \|g - f\|_p([0,1]).$$

Lemma 6.1.1. *There exists a positive constant $C_0 = C_0(r, d, p)$ such that*

$$\mathrm{dist}(\mathcal{S}_k^1, t\phi_{j,k+1}^1)_p = |t|C_0$$

for any $j \in \Lambda(k)$, $k \geq 0$, $t \in \mathbb{R}$.

Proof. It is enough to prove the lemma for $j = 0$, $k = 0$; the general case follows by translation and dilation. Suppose $t = 1$. Let $(S_i)_{i=1}^\infty$ be a sequence of functions in $\mathcal{S}_0^1$ such that $\lim_{i\to\infty} \|S_i - \phi_{0,1}^1\|_p = \mathrm{dist}(S_0^1, \phi_{0,1})_p$. Since $\mathcal{S}_0^1$ is a closed subspace of $L^p([0,1])$, we may assume that $(S_i)_{i=1}^\infty$ converges to a member S of $\mathcal{S}_0^1$, so that $\|S - \phi_{0,1}^1\|_p([0,1]) = \mathrm{dist}(\mathcal{S}_0^1, \phi_{0,1}^1)_p$. But S must differ from $\phi_{0,1}^1$ at at least one point in some neighborhood of a knot x of $\phi_{0,1}^1$, since S must have derivatives of all orders at x while $\phi_{0,1}^1$ only has derivatives up to order $r-1$ at x. Thus S differs from $\phi_{0,1}^1$ on a set of positive measure, because S and $\phi_{0,1}^1$ are continuous. Thus $\|S - \phi_{0,1}^1\|_p([0,1]) =: C_0 > 0$.

Now, for arbitrary t (ignoring the trivial case $t = 0$) we have

$$\mathrm{dist}(\mathcal{S}_k^1, t\phi_{j,k+1}^1)_p = |t| \cdot \inf_{g\in\mathcal{S}_k^1} \|g - \phi_{j,k+1}^1\|_p([0,1]) = |t|C_0. \qquad \square$$

Now define $\mathrm{dist}(\mathcal{S}_k^{\boldsymbol\alpha}, f)_p := \inf_{g\in\mathcal{S}_k^{\boldsymbol\alpha}} \|g - f\|_p$ for $f \in L^p(\Omega)$.

Lemma 6.1.2. *If $d > 1$, $t \in \mathbb{R}$, and k, k' are such that*

$$\mathrm{diam}(\operatorname{supp}\phi_I^{\boldsymbol\alpha}) > \mathrm{diam}(\operatorname{supp}\phi_J^{\boldsymbol\beta}) \tag{6.1.1}$$

for $I \in \mathcal{D}_k^{\boldsymbol\alpha}(\Omega)$, $J \in \mathcal{D}_{k'}^{\boldsymbol\beta}(\Omega)$, then there exists a positive constant

$$C_1 = C_1(r, d, p, \boldsymbol\beta, \boldsymbol\alpha)$$

such that

$$\mathrm{dist}(\mathcal{S}_k^{\boldsymbol\alpha}, t\phi_J^{\boldsymbol\beta})_p > C_1|t|\|\phi_J^{\boldsymbol\beta}\|_p.$$

Proof. By (6.1.1) we have $2^{-\lfloor k\alpha_i'\rfloor} > 2^{-\lfloor k'\beta_i'\rfloor}$ for some i. Without loss of generality we shall assume $2^{-\lfloor k\alpha_1'\rfloor} > 2^{-\lfloor k'\beta_1'\rfloor}$. Then it will be enough to show that there exists a positive constant C_1 such that

$$\mathrm{dist}(\mathcal{S}_{\lfloor k'\beta_1'\rfloor-1}^1 \otimes L^p([0,1]^{d-1}), t\phi_J^{\boldsymbol\beta})_p \geq C_1|t|\|\phi_J^{\boldsymbol\beta}\|_p \tag{6.1.2}$$

because $\mathcal{S}_k^{\boldsymbol\alpha} \subset \mathcal{S}_{\lfloor k'\beta_1'\rfloor-1}^1 \otimes L^p([0,1]^{d-1})$. Furthermore, it is enough to prove (6.1.2) under the assumption $\lfloor k'\beta_1'\rfloor = 1$ since the general result follows by dilation.

Take $g(\mathbf{x}) = S(x_1)h(x_2, \ldots, x_d)$ where $S \in \mathcal{S}_0^1$ and $h \in L^p([0,1]^{d-1})$. Define E to be the set of all $(x_2, \ldots, x_d) \in [0,1]^{d-1}$ such that

$$\phi(2^{\lfloor k'\beta_1'\rfloor}x_1, \ldots, 2^{\lfloor k'\beta_d'\rfloor}x_d) \geq \frac{1}{2}\phi^1(2^{\lfloor k'\beta_1'\rfloor}x_1).$$

Note that $|E| = C\|\phi_J^{\boldsymbol\beta}\|_p^p$. Now, for any fixed $(x_2, \ldots, x_d) \in E$, we have

$$\begin{aligned}
&\|g(x_1, \ldots, x_d) - t\phi_J^{\boldsymbol\beta}\|_p([0,1] \times \{(x_2, \ldots, x_d)\}) \\
&\qquad > C_0\phi^1(2^{\lfloor k'\beta_2'\rfloor}x_2)\cdots\phi^1(2^{\lfloor k'\beta_d'\rfloor}x_d)|t| \text{ (by Lemma 6.1.1)} \\
&\qquad \geq C_1|t|.
\end{aligned}$$

Thus, $\|g(x_1, \ldots, x_d) - t\phi_J^{\boldsymbol\beta}\|_p > C_1|E|^{1/p}|t| = C_1|t|\|\phi_J^{\boldsymbol\beta}\|_p$. $\square$

Lemma 6.1.3. *Given a nonnegative integer k, let k' be the smallest integer such that $2^{-\lfloor k'\alpha'_i\rfloor} \le 2^{-\lfloor k\beta'_i\rfloor}$ for $i = 1, \dots, d$. For $J \in \mathcal{D}^{\boldsymbol{\beta}}_k$, $I \in \mathcal{D}^{\boldsymbol{\alpha}}_{k'}$, define $m_k := |\operatorname{supp}\phi^{\boldsymbol{\beta}}_J|/|\operatorname{supp}\phi^{\boldsymbol{\alpha}}_I|$. Then*

$$\lim_{k\to\infty} m_k = \infty. \tag{6.1.3}$$

Proof. Since $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are linearly independent, we may assume without loss of generality that $\beta'_2/\alpha'_2 < \beta'_1/\alpha'_1$, i.e.

$$\frac{\beta'_1}{\alpha'_1}\alpha'_2 - \beta'_2 > 0. \tag{6.1.4}$$

Also, $2^{-\lfloor k'\alpha'_1\rfloor} \le 2^{-\lfloor k\beta'_1\rfloor}$ implies that

$$k' \ge \frac{\lfloor k\beta'_1\rfloor}{\alpha'_1}. \tag{6.1.5}$$

Now, to prove (6.1.3) it is enough to show that the ratio $2^{-\lfloor k\beta'_2\rfloor}/2^{-\lfloor k'\alpha'_2\rfloor}$ of corresponding side lengths of blocks from $\mathcal{D}^{\boldsymbol{\alpha}}_k$ and $\mathcal{D}^{\boldsymbol{\beta}}_{k'}$ increases without bound as $k \to \infty$. We do so as follows:

$$\begin{aligned}\frac{2^{-\lfloor k\beta'_2\rfloor}}{2^{-\lfloor k'\alpha'_2\rfloor}} &\asymp 2^{k'\alpha'_2 - k\beta'_2}\\ &\ge 2^{(\lfloor k\beta'_1\rfloor/\alpha'_1)\alpha'_2 - k\beta'_2} \quad \text{(by (6.1.5))}\\ &\ge C2^{k((\beta'_1/\alpha'_1)\alpha'_2 - \beta'_2)}\end{aligned}$$

which increases without bound as $k \to \infty$, by (6.1.4). □

Now let $\Sigma^{\boldsymbol{\alpha}}_n$ denote the set of all sums $\sum_{I\in\Lambda} a_I\phi^{\boldsymbol{\alpha}}_I$ where $\#\Lambda \le n$, for $n = 1, 2, \dots$. For $f \in B^{\boldsymbol{\beta}}$, let $\sigma^{\boldsymbol{\alpha}}_n(f)_p := \inf_{S\in\Sigma^{\boldsymbol{\alpha}}_n} \|f - S\|_p$. $\sigma^{\boldsymbol{\alpha}}_n(f)_p$ is the $L^p(\Omega)$ error in approximating f with an n-term linear combination of "incorrectly" scaled wavelets from $\Phi^{\boldsymbol{\alpha}}$ when $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are linearly independent.

Theorem 6.1.4. *Given $B^{\boldsymbol{\beta}}$, there exists $f \in B^{\boldsymbol{\beta}}$ such that for any $\boldsymbol{\alpha}$ which is not a scalar multiple of $\boldsymbol{\beta}$, we have $\sigma^{\boldsymbol{\alpha}}_n(f)_p \ne \mathcal{O}(n^{-\gamma})$ for any $\gamma > 0$.*

Proof. Let $(I^{(i)})_{i=1}^\infty$ be a sequence of blocks with $I^{(i)} \in \mathcal{D}^{\boldsymbol{\beta}}_{k_i}(\Omega)$ such that the members of the set $\{{I^{(i)}}' \mid i = 1, 2, \dots\}$ are pairwise disjoint (recall that ${I^{(i)}}' := \operatorname{supp}\phi^{\boldsymbol{\beta}}_{I^{(i)}}$). Then $|I^{(i)}| \to 0$ as $k \to \infty$. Let us further assume that $(|I^{(i)}|)_{i=1}^\infty$ decreases so quickly that, for $i = 1, 2. \dots,$

$$m_{k_i} > 2^{i^2} \quad \text{and} \tag{6.1.6}$$

$$\frac{m_{k_i}}{m_{k_{i+1}}} < \frac{\delta}{r^d} \tag{6.1.7}$$

where $\delta > 0$ is so small that, given a positive integer k, a function $S \in \mathcal{S}^{\boldsymbol{\beta}}_k$, a block $I \in \mathcal{D}^{\boldsymbol{\beta}}_k$, and a subset E of I with $|E| < \delta|I|$, we have $\|S\|_p(I' \setminus E) > (1/2)\|S\|_p(I')$ (see Corollary 3.2.3). Let

$$f := \sum_{i=1}^\infty 2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol{\beta}}_{I^{(i)}}.$$

Then

$$|f|^{\tau}_{B^{\boldsymbol\beta}} \asymp \sum_{i=1}^{\infty} 2^{-i}|I^{(i)}|^{-\tau/p}|I^{(i)}|^{\tau/p} = 1$$

and

$$\|f\|_p^p \asymp \sum_{i=1}^{\infty} 2^{-ip/\tau}|I^{(i)}|^{-1}|I^{(i)}| = \frac{2^{-p/\tau}}{1-2^{-p/\tau}} =: C_{p,\tau}.$$

Let $S_n = \sum_{j=1}^n a_j \phi^{\boldsymbol\alpha}_{J(j)}$ be an n-term approximation to f. The integer n will be specified later. Clearly

$$\|f - S_n\|_p^p \geq \left\| \sum_{i=1}^{\infty} (f - S_n) \cdot \chi_{I^{(i)\prime}} \right\|_p^p$$

$$= \sum_{i=1}^{\infty} \|2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol\beta}_{I^{(i)}} - S_n \cdot \chi_{I^{(i)\prime}}\|_p^p. \tag{6.1.8}$$

Let us concentrate on the single term

$$\|2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol\beta}_{I^{(i)}} - S_n \cdot \chi_{I^{(i)\prime}}\|_p^p \tag{6.1.9}$$

from (6.1.8), where $i > 1$. Break up the sum

$$S_n \cdot \chi_{I^{(i)\prime}} = \sum_{j=1}^{n} a_k \phi^{\boldsymbol\alpha}_{J(j)} \chi_{I^{(i)\prime}}$$

into two parts as follows:

$$\sum_{j=1}^{n} a_j \phi^{\boldsymbol\alpha}_{J(j)} \chi_{I^{(i)\prime}} = \sum_{j \in E_1} a_j \phi^{\boldsymbol\alpha}_{J(j)} \chi_{I^{(i)\prime}} + \sum_{j \in E_2} a_j \phi^{\boldsymbol\alpha}_{J(j)} \chi_{I^{(i)\prime}} =: \Sigma_1 + \Sigma_2 \tag{6.1.10}$$

where

$$E_1 := \{j \mid |J_k^{(j)}| > |I_k^{(i)}|, k = 1, \ldots, d\}$$
$$E_2 := \{j \mid |J_k^{(j)}| \leq |I_k^{(i)}|, k = 1, \ldots, d\}.$$

Rewrite (6.1.9) as

$$\|2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol\beta}_{I^{(i)}} - \Sigma_1 - \Sigma_2\|_p^p. \tag{6.1.11}$$

Let $n = m_{k_{i-1}}$. Then

$$\begin{aligned} \frac{\left|\cup_{j\in E_2} \operatorname{supp} \phi^{\boldsymbol\alpha}_{J(j)}\right|}{|I^{(i)}|} &< \frac{n \cdot \max_{j \in E_2} |\operatorname{supp} \phi^{\boldsymbol\alpha}_{J(j)}|}{|I^{(i)}|} \\ &= \frac{nr^d \max_{j\in E_2} |J^{(j)}|}{|I^{(i)}|} \\ &\leq \frac{nr^d |I^{(i)}|/m_{k_i}}{|I^{(i)}|} \\ &< \delta \text{ (by (6.1.7))}. \end{aligned} \tag{6.1.12}$$

We now have

$$\begin{aligned}
&\|2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol{\beta}}_{I^{(i)}} - \Sigma_1 - \Sigma_2\|_p^p \\
&\geq \left(\frac{1}{2}\right)^p \|2^{-i/\tau}|I^{(i)}|^{-1/p}\phi^{\boldsymbol{\beta}}_{I^{(i)}} - \Sigma_1\|_p^p \text{ (by (6.1.12))} \\
&\geq \left(\frac{1}{2}\right)^p C_1^p 2^{-ip/\tau}|I^{(i)}|^{-1}\|\phi^{\boldsymbol{\beta}}_{I^{(i)}}\|_p^p \text{ (by Lemma 6.1.2)} \\
&= C2^{-ip/\tau}. \qquad (6.1.13)
\end{aligned}$$

Thus from (6.1.13) and (6.1.8) we have, for any i,

$$\begin{aligned}
\|f - S_n\|_p^p &\geq \sum_{k=i}^{\infty} \|2^{-k/\tau}|I^{(k)}|^{-1/p}\phi^{\boldsymbol{\beta}}_{I^{(k)}} - S_n \cdot \chi_{I^{(k)\prime}}\|_p^p \\
&\geq C\sum_{k=i}^{\infty} 2^{-kp/\tau} =: C_{p,\tau}2^{-ip/\tau}. \qquad (6.1.14)
\end{aligned}$$

Choose any $C > 0$, $\gamma > 0$. Then for i sufficiently large, we have

$$(i-1)^2 > \frac{1}{\gamma p}\left[\log_2 C + \frac{ip}{\tau} - \log_2 C_{p,\tau}\right]. \qquad (6.1.15)$$

But (6.1.15) and (6.1.6) imply

$$Cn^{-\gamma p} < C_{p,\tau}2^{-ip/\tau} \qquad (6.1.16)$$

and thus, using (6.1.14) with (6.1.16), we obtain for $n = m_{k_{i-1}}$, i sufficiently large,

$$\|f - S_n\|_p^p \geq Cn^{-\gamma p}. \qquad \square$$

6.2 An Imbedding Result. We immediately obtain from Theorem 6.1.4 the following imbedding result.

Corollary 6.2.1. *$B^{\boldsymbol{\beta}} \hookrightarrow B^{\boldsymbol{\alpha}}$ for a fixed value of p if and only if $\boldsymbol{\beta} = \lambda\boldsymbol{\alpha}$ for some $\lambda \geq 1$.*

Proof. If $\boldsymbol{\beta} = \lambda\boldsymbol{\alpha}$, $\lambda \geq 1$, then $\boldsymbol{\beta}' = \boldsymbol{\alpha}'$. Letting

$$\begin{aligned}
\frac{1}{\tau(\boldsymbol{\beta})} &:= \frac{H(\boldsymbol{\beta})}{d} + \frac{1}{p}, \\
\frac{1}{\tau(\boldsymbol{\alpha})} &:= \frac{H(\boldsymbol{\alpha})}{d} + \frac{1}{p},
\end{aligned}$$

we get $\tau(\boldsymbol{\beta}) < \tau(\boldsymbol{\alpha})$ and thus

$$\begin{aligned}
|f|_{B^{\boldsymbol{\beta}}} &= \left[\sum_{I\in\mathcal{D}} |c_I(f)|^{\tau(\boldsymbol{\beta})}|I|^{\tau(\boldsymbol{\beta})/p}\right]^{1/\tau(\boldsymbol{\beta})} \\
&\geq \left[\sum_{I\in\mathcal{D}} |c_I(f)|^{\tau(\boldsymbol{\alpha})}|I|^{\tau(\boldsymbol{\alpha})/p}\right]^{1/\tau(\boldsymbol{\alpha})} = |f|_{B^{\boldsymbol{\alpha}}}. \qquad (6.2.1)
\end{aligned}$$

So $B^{\boldsymbol\beta} \hookrightarrow B^{\boldsymbol\alpha}$.

Conversely, suppose that $\boldsymbol\beta = \lambda\boldsymbol\alpha$ for some $0 < \lambda < 1$ or that $\boldsymbol\beta$ and $\boldsymbol\alpha$ are linearly independent. If $\boldsymbol\beta = \lambda\boldsymbol\alpha$ for some $0 < \lambda < 1$ then we obtain $B^{\boldsymbol\beta} \not\subseteq B^{\boldsymbol\alpha}$ using (4.1.2) and Theorem 3.3.1. Now assume that $\boldsymbol\beta$ and $\boldsymbol\alpha$ are linearly independent and $B^{\boldsymbol\beta} \hookrightarrow B^{\boldsymbol\alpha}$. Let $f \in B^{\boldsymbol\beta}$. Then since $B^{\boldsymbol\beta} \hookrightarrow B^{\boldsymbol\alpha}$ we have

$$\sigma_n(f)_p^{\boldsymbol\alpha} \le C|f|_{B^{\boldsymbol\alpha}} n^{-H(\boldsymbol\alpha)/d} \le C|f|_{B^{\boldsymbol\beta}} n^{-H(\boldsymbol\alpha)/d}$$

for $n = 1, 2, \ldots$, contradicting Theorem 6.1.4. □

References

[B-B] P.L. Butzer and H. Berens, *Semi-groups of Operators and Approximation*, Springer-Verlag, New York, 1967.

[B-L] J. Bergh and J. Löfström, *Interpolation Spaces: An Introduction*, Springer-Verlag, 1976.

[D-J-P] Ronald A. DeVore, Björn Jawerth, and Vasil Popov, *Compression of Wavelet Decompositions*, American Journal of Mathematics **114** (1992), 737–785.

[D-L] Ronald A. DeVore and G. Lorentz, *Constructive Approximation*, Heidelberg: Springer-Verlag, 1993.

[D-P1] Ronald A. DeVore and V.A. Popov, *Interpolation of Besov Spaces*, Trans. Amer. Math. Soc. **305** (1988), 397–414.

[D-P2] ———, *Interpolation Spaces and Nonlinear Approximation*, Function Spaces and Applications, (M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, eds.), vol. 1302, Springer Lecture Notes in Math, 1988, pp. 191–207.

[F-S] C. Fefferman and E. Stein, *Some Maximal Inequalities*, Amer. J. Math. **93** (1971), 107–115.

[G] C.W. Groetsch, *Elements of Applicable Functional Analysis*, Marcel Dekker, Inc., New York, 1980.

[L-T] T.D. Long and H. Triebel, *Equivalent Norms and Shauder Bases in Anisotropic Besov Spaces*, Proceedings of the Royal Society of Edinburgh **84A** (1979), 177–183.

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

[P-S] J. Peetre and G. Sparr, *Interpolation of Normed Abelian Groups*, Ann. Mat. Pura. Appl. **92** (1972), 217–262.

[P] J. Peetre, *New Thoughts on Besov Spaces*, Duke Univ. Math. Ser. I, Durham, NC, 1976.

[S-O] E.A. Storozhenko and P. Osval'd, *Jackson's Theorem in the Spaces* $L^p(\mathbb{R}^k), 0 < p < 1$, Siberian Math. J. **19** (1978), 630–639.