JOURNAL OF APPROXIMATION THEORY **88**, 47–79 (1997)
ARTICLE NO. AT963008

On the Computation of Wavelet Coefficients

Bernard Delyon and Anatoli Juditsky

IRISA-INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

Communicated by C. K. Chui

Received September 15, 1994; accepted in revised form January 2, 1996

We consider fast algorithms of wavelet decomposition of a function f when discrete observations of f ($\operatorname{supp} f \subseteq [0, 1]^d$) are available. The properties of the algorithms are studied for three types of observation design which for $d = 1$ can be described as follows: the regular design, when the observations $f(x_i)$ are taken on the regular grid $x_i = i/N$, $i = 1, ..., N$; the case of a jittered regular grid, when it is only known that for all $1 \leqslant i \leqslant N$, $i/N \leqslant x_i < i+1)/N$; and the random design case; in which x_i, $i = 1, ..., N$, are independent and identically distributed random variables on $[0, 1]$. We show that these algorithms are in a certain sense efficient when the accuracy of the approximation is concerned. The proposed algorithms are computationally straightforward: the whole effort to compute the decomposition is order N for the sample size N. © 1997 Academic Press

1. INTRODUCTION

The now classical orthogonal (biorthogonal) wavelet transform of a function $f \in L_2(R)$ can be written as

$$f(x) = \sum_{k \in Z} \alpha_k \phi_k(x) + \sum_{j \geqslant 0} \sum_{k \in Z} \beta_{jk} \psi_{jk}(x),$$

where

$$\phi_k(x) = \phi(x-k) \qquad \text{and} \qquad \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k).$$

Here $\phi(x)$ and $\psi(x)$ are the scale function and the mother wavelet respectively. Wavelet coefficients are given by the integrals

$$\alpha_k = \int f(x)\, \tilde{\phi}_k(x)\, dx, \qquad \beta_{jk} = \int f(x)\, \tilde{\psi}_{jk}(x)\, dx \tag{1}$$

(in the orthogonal case, $\tilde{\phi} = \phi$ and $\tilde{\psi} = \psi$).

If f is C^s and the wavelet $\tilde{\psi}$ is compactly supported and orthogonal to x^l, $l = 0, ..., M$ with $s < M+1$, the sequence β_{jk} decreases as fast as

0021-9045/97 $25.00

$2^{-j(s+1/2)}$ when j increases (this is easily guessed by using Taylor's formula). Thus, large values of β_{jk} are encountered only when f is irregular in the neighborhood of the $2^{-j}k$. These ideas are rigorously formalized in Section 3.2.

Generally, the values of f are not available except for a sequence $f_i = f(X_i)$, $i = 1, ..., N$, of discrete observations; the analyzing wavelets $\tilde{\psi}$ and $\tilde{\phi}$ generally cannot be expressed in a compact analytical form. Thus the integral above cannot be computed exactly, and some approximate method should be used to compute the quadrature in (1). On the other hand, the main interest of using wavelets in data compression and estimation lies in the fact for a smooth function f the coefficients β_{jk} decay rapidly, and consequently, the wavelet projection $P_{j_0} f$,

$$P_{j_0} f(x) = \sum_{k \in Z} \alpha_k \phi_k(x) + \sum_{j=0}^{j_0} \sum_{k \in Z} \beta_{jk} \psi_{jk}(x), \tag{2}$$

converges rapidly to f when j_0 increases. We will seek to preserve this property when computing the empirical wavelet coefficients of f on the basis of observations $f(X_i)$. The objective of this work is to design "fast" numerical algorithms to compute the "estimates" of true wavelet coefficients β_{jk} when discrete observations of f are available. For the sake of simplicity we suppose that f is a compactly supported function.

1.1. *Fast Wavelet Algorithms*

In the whole paper, we will work under the following assumption (cf. Section 2.1):

Assumption 1. The tuple $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ forms a biorthogonal multiresolution analysis such that ϕ and ψ are C^{M+1} for some $M \in \mathbb{N}$. $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ have compact support. $\tilde{\psi}$ is orthogonal to polynomials of degree $\leqslant M$.

Note that by Corollary 5.5.2 of [4], in the orthogonal case, the orthogonality to polynomials required in the assumption is a consequence of the regularity.

The algorithm for computing the wavelet coefficients α_k and β_{jk} is based on a calculation of the coefficients

$$\alpha_{jk} = \int f(x)\, 2^{j/2} \tilde{\phi}(2^j x - k)\, dx, \tag{3}$$

at one fine level j (i.e., large j) which leads to the $\alpha_k = \alpha_{0k}$'s and $\beta_{j'k}$'s, $j' = 0, ..., j$, through the filtering relations (18) and (19) below. This is this last step which requires the multi-resolution structure imposed in Assumption 1. Since $2^j \tilde{\phi}(2^j x - k)$ is a function with integral 1 with a sharp peak on $2^{-j}k$, $2^{-j/2}\alpha_{jk}$ is quite close to $f(2^{-j}k)$. The formulas we propose below are finer approximations of this integral, consistent with the smoothness of f and the observations of this function we have at our disposal.

1.2. *Regularity Classes*

It is evident that for the estimate $\hat{f}_N$ of f, based on observations $f_1, ..., f_N$, to converge to f with a good rate, we have to suppose that f satisfies some regularity constraints. In this paper we consider the Besov constraints. These constraints can be easily expressed in the form of conditions on wavelet coefficients of the function f. As a consequence, in the wavelet decomposition of a function from a Besov class generally most coefficients are close to zero and can be neglected. Finally, we use the Besov classes because of their exceptional expressive power (cf. [17]). Proposition 1 says that functions from the Besov spaces can be "well" (in some sense) approximated by piecewise polynomials.

We will develop fast approximate quadrature formulas to compute wavelet coefficients which are *exact for the polynomials*. Then good approximation of the underlying polynomials will imply optimal rates of convergence for a function f satisfying Besov constraints. The calculation of the wavelet coefficients α_{jk} depends on the nature of points X_i where the observations of f are taken. Let us now turn to the model of observations. We suppose that the available sample (f_i), $i=1, ..., N$, is noiseless, i.e., we consider the model $f_i = f(X_i)$, and we call the vector $X = (X_1, ..., X_N)^T$ the observation design.[1]

1.3. *Observation Design*

We consider two types of observation designs:

- observations f_k, where k is a multi-index taken on a regular grid, i.e., $X_k = (k_1/n, ..., k_d/n)$, $k_i = 1, ..., n$, $i = 1, ..., d$, $n = 2^j$. We refer to this case as a *regular sampling*, or regular design. A variation of the regular design is the situation when n^d observations of means $A^f_1, ..., A^f_{(n, ..., n)}$ are available,

$$A^f_k = \int_I f((x+k)/n)\, dx, \qquad k = (1, ..., 1)\ldots, (n, ..., n),$$

where $I = [-1/2, 1/2]^d$. Following Donoho [6], we refer to this case as a *regular boxcar design*. This type of observation is usual in the problems of computer vision, when the average intensity over the receiving retina cells is measured;

- the case of *irregular design*, when N observations $f_i = f(X_i)$ are available at the points X_i, $i = 1, ..., N$, which do not constitute a regular grid. We are particularly interested in two special cases of irregular design:

[1] From now on we use A^T to denote the transpose of A.

1. *Jittered regular design* is a simple case of irregular sampling. In this case we suppose that the observations $f(X_k)$, where $k=(k_1, ..., k_d)$ is a multi-index ($k_i=1, ..., n$, $n^d=N$), are available. Here nX_k belongs to the cubes $k+[-1, 0]^d$. A particular case of this is a *random jittered grid*, when X_k are independent random variables, and each nX_k is uniformly distributed in $k+[-1, 0]^d$ (cf. [10, Sect. 1.9]);

2. *Random design*, when observations X_i, $i=1, ..., N$, are independent and identically distributed on $[0, 1]^d$ with some density $p(x): 0<p_{\min}\leqslant p(x)\leqslant p_{\max}<\infty$. This situation is common when general scattered data models are concerned.

1.4. *Organization and Notations*

The paper is organized as follows: first in Section 2 we briefly introduced and discuss the algorithms we use to compute wavelet coefficients for different cases of observation design. For the sake of clarity the algorithms are presented in this section in a one-dimensional context. However, they can be easily generalized for the general multi-dimensional case. Then in Section 3 we recall some notions on wavelets and Besov spaces and introduce notations to be used later. Then we study the properties of proposed algorithms in case of regular design in Section 4. Finally, we consider the case of irregular design and provide the estimates of the accuracy of algorithms for the case of jittered grid and random design.

The Fourier transform of a function f is

$$\hat{f}(\omega)=\int f(x)\, e^{-i\omega x}\, dx$$

For sequences α_k and β_{jk} we will denote

$$\|\alpha\|_p=\left(\sum_k |\alpha_k|^p\right)^{1/p} \qquad \|\beta_j\|_p=\left(\sum_k |\beta_{jk}|^p\right)^{1/p}$$

Recall that L_p is the classical Lebesgue space with the norm $\|f\|_p=(\int g^p(x)\, dx)^{1/p}$. We denote by C a generic positive constant.

2. ALGORITHMS DESCRIPTION

2.1. *Regular Point Design*

A simple (or naive) method in this case consists of considering the values $2^{-j/2}f_k$ as the wavelet coefficients α_{jk}. This method was introduced by Mallat [11] and widely spread in engineering practice without much of a rigorous foundations (cf. [1]). A theoretical analysis of this method was recently performed by Donoho [5], who has shown that it produces a sort of multi-resolution analysis, though the very form of the wavelet ψ_{jk} in the decomposition (2) depends on the parameter j; thus, reconstruction formula (2) is inapplicable. However, he shows that this method is efficient for coding (this was Mallat's purpose) since the coefficients β_{jk} are still small if f is smooth.

We propose here a simple quadrature method for computing wavelet coefficients of a function with regularity s, $s < M+1$; this method coincides with preceding "naive" transform if $\tilde{\phi}(x)$ has M vanishing moments, i.e. $\int x^l \tilde{\phi}(x)\, dx = 0$ for $l = 1, ..., M$ (cf. Section 2.3).

ALGORITHM 1. Find a finite sequence (c_i) such that

$$\int x^l \tilde{\phi}(x)\, dx = \sum_i c_i i^l \qquad l = 0, ..., M \tag{4}$$

(where $0^0 = 1$) and put

$$\bar{\alpha}_{jk} = 2^{-j/2} \sum_i c_i f(2^{-j}(i+k)). \tag{5}$$

Remark. In d dimensions, $l = (l_1, ..., l_d)$ and $i = (i_1, ..., i_d)$ are multi-integers, $\sum_{k=1}^{d} l_k \leqslant M$, and $2^{-j/2}$ becomes $2^{-jd/2}$.

Theoretical bases for this procedure are given in Proposition 2. In simpler words, the idea is that this quadrature formula is exact for polynomials of degree $\leqslant M$; thus, using Taylor's formula in Eq. (3), one suggests that the quadrature error would be quite small.

Consider $N = 2^{j_0}$ and suppose that a filter $\{c_i\}$ of order l is used. Then for a data sample of size N the algorithm above uses $2l \times N$ elementary operations in order to compute N coefficients $\bar{\alpha}_{j_0 k}$.

Then the whole effort to compute the coefficients α_{0k} and β_{jk} for $0 \leqslant j < j_0$ from $\bar{\alpha}_{jk}$ is of order N (cf. (18), (19) in Section 3).

2.2. *Regular Boxcar Design*

The algorithm consisting in taking the observations A_k^f for the estimates of the wavelet coefficients α_{jk} $(N = 2^j)$ was first analyzed theoretically by Donoho in [6]. He also considered an alternative approach which consists in using a it biorthogonal wavelet transform [4]. We consider the following

ALGORITHM 2. j is fixed and we observe

$$A_{jk}^f = \int_{-1/2}^{1/2} f(2^{-j}(k+x))\, dx \qquad k = 1, ..., 2^j. \tag{6}$$

Find a finite sequence (c_i) such that

$$\int x^l \tilde{\phi}(x)\, dx = \sum_i c_i \frac{[(i+1/2)^{l+1} - (i-1/2)^{l+1}]}{l+1} \qquad \text{for} \quad l = 0, ..., M \tag{7}$$

and set

$$\tilde{\alpha}_{jk} = 2^{-j/2} \sum_i c_i A^f_{j,\, k+i}. \tag{8}$$

The remark after Algorithm 1 remains valid for this algorithm. It can be easily verified that the approximation (8) is exact for polynomials of degree $\leqslant M$.

2.3. *Computing the* (c_i) *Sequence and the Choice of the Wavelet*

Note first that the choice of this sequence is generally non-unique since the number of unknowns may be larger that the number of equations.

If one uses Coiflets (cf. [4], p. 258), $c_i = \delta_{0i}$ is also convenient for Algorithm 1 since ϕ itself has M vanishing moments, i.e.,

$$\int x^l \phi(x)\, dx = 0, \qquad l = 1, ..., M$$

(see, for instance, Chap. 6 of [4]).

One the other hand, if orthogonal wavelets are used, Lemma 4 below implies that $c_i = \phi(i)$ is a convenient choice of the filter c_i for Algorithm 1. It can be easily verified that in this case the computed coefficients α_k, β_{jk} are

$$\alpha_k = \frac{1}{N} \sum_{i=1}^{N} f(i/N)\, \phi_k(i/N), \qquad \beta_{jk} = \frac{1}{N} \sum_{i=1}^{N} f(i/N)\, \psi_{jk}(i/N).$$

In this case the result of Algorithm 1 is the "empirical wavelet transform" (cf. [7]).

In the case of the boxcar design and orthogonal wavelets, Lemma 4 and Assumption 1 imply that $c_i = \sum_{k>0} \phi'(i - k/2)$ is a convenient filter for Algorithm 2.

The c_i sequence required in Algorithm 2 may be chosen as (δ_{0i} in the case of biorthogonal wavelets with $\tilde{\phi}$ = "Haar function" (cf. [4, p. 272).

2.4. *Irregular Design*

We use the same algorithm in both cases of jittered and random designs, though the parameters of the algorithm are chosen differently.

The algorithms consist of two stages: in the first step we compute the least-squares estimates $\hat{f}_k$ of the values of $f(2^{-j}(K+1/2))$ at the knots of the regular grid (of the averages over cells of the regular grid) for some resolution j (we discuss the choice of j in Section 5 below). Then we use Algorithm 1 (Algorithm 2, respectively) to compute wavelet coefficients.

This leads to the following algorithms:

ALGORITHM 3. Consider 2^j intervals $\Lambda_k = [2^{-j}k, 2^{-j}(k+1)[$, $k = 0, ..., 2^j - 1$, where the resolution j will be given below. We denote by I_k the set of indices $I_k = \{i: X_i \in \Lambda_k\}$, $|I_k| = \mathrm{card}(I_k)$.

Choose j such that

$$\frac{N}{4(M+1)} < 2^j \leqslant \frac{N}{2(M+1)}$$

for the case of regular jittered grid and

$$\frac{N}{2\lambda \ln N} \leqslant 2^j < \frac{N}{\lambda \ln N}$$

for the random design case (the value of λ will be chosen later).

Let $\hat{P}_k(x)$ be the solution of the minimization problem

$$\hat{P}_k = \arg \min_{d \circ P = M} \sum_{i=1}^{N} (P(X_i) - f(X_i))^2 \, 1_{i \in I_k},$$

where the minimum is taken over all polynomials of degree M. Then we put

$$\hat{f}_k = \hat{P}_k(2^{-j}(k + 1/2)). \tag{9}$$

Next we use the estimates $\hat{f}_k$ to compute the wavelet coefficients $\hat{\alpha}_{jk}$ using Algorithm 1 (Eq. (5)):

$$\hat{\alpha}_{jk} = \sum_i c_i 2^{-j/2} \hat{f}_{k+i}.$$

Remarks. In d-dimensions, $k = (k_1, ..., k_d)$ and $i = (i_1, ..., i_d)$ are multi-integers, Λ_k becomes a hypercube, and $2^{-j/2}$ and 2^{-j} become $2^{-jd/2}$ and 2^{-jd} respectively.

Note that $P^{(k)}$ is a least squares polynomial approximation of f on Λ_k. So $\hat{f}_k$ is a kind of least squares estimate of $f(2^{-j}(k + 1/2))$. It can be easily verified that $\hat{f}_k = f(2^{-j}(K + 1/2))$ exactly when the function f is a polynomial of degree $\leqslant M$. The choice of the argument in (9) is not unique. For instance, one can estimate the values of $f(2^{-j}k)$ or $f(2^{-j}(k+1))$, $k = 0, ..., 2^j - 1$, etc. Another interesting use of the estimate $\hat{P}_k$ is provided by the following.

ALGORITHM 4. Consider the same construction as in Algorithm 3. Set

$$\hat{A}_k = \int_{2^{-j}k}^{2^{-j}(k+1)} \hat{P}_k(x)\,dx.$$

Here $\hat{A}_{jk}$ are the estimates of A^f_{jk} of (6). Then we compute the estimates $\hat{\alpha}_{jk}$ of wavelet coefficients using Algorithm 2:

$$\hat{\alpha}_{jk} = \sum_i c_i 2^{j/2} \hat{A}_{j,\,k+i}.$$

3. WAVELET DECOMPOSITION AND BESOV SPACES

This section summarizes basic properties of biorthogonal wavelet and relations between wavelet coefficients norms and norms in Besov spaces (Theorem 2). d is here the dimension of the Euclidean space.

3.1. *Biorthogonal Wavelet Bases*

We recall in this section some basic properties of biorthogonal wavelets.

DEFINITION 1. A biorthogonal basis of a Hilbert space H is a pair of Riesz bases $((a_k)_{k \in N}, (b_k)_{k \in N})$ such that $\langle a_k, b_l \rangle = \delta_{kl}$.[2]

One can prove that any element f of H has the biorthogonal decomposition

$$f = \sum_k \langle f, a_k \rangle\, b_k.$$

DEFINITION 2. A biorthogonal multi-resolution analysis is a pair of functions $(\phi, \tilde{\phi})$ of norm 1 of $L_2(\mathbb{R})$ and spaces

$$V_j = \operatorname{span}(\phi_{jk}, k \in Z), \qquad \phi_{jk}(x) = 2^{j/2}\phi(2^j x - k),$$
$$\tilde{V}_j = \operatorname{span}(\tilde{\phi}_{jk}, k \in Z), \qquad \tilde{\phi}_{jk}(x) = 2^{j/2}\tilde{\phi}(2^j x - k)$$

with the properties

(AMB1) $\cap V_j = \{0\}$ and $\cap \tilde{V}_j = \{0\}$,

(AMB2) $\overline{\cup V_j} = L_2(\mathbb{R})$ and $\overline{\cup \tilde{V}_j} = L_2(\mathbb{R})$,

(AMB3) $V_j \subset V_{j+1}$ and $\tilde{V}_j \subset \tilde{V}_{j+1}$,

(AMB4) $((\phi_{0k})_{k \in Z}, (\tilde{\phi}_{0k})_{k \in Z})$ is a biorthogonal basis.

[2] $\langle \cdot, \cdot \rangle$ denotes the inner product.

These assumptions imply the existence of two square integrable sequences (h_k) and $(\tilde{h}_k)$,

$$\phi(x)=\sqrt{2}\sum h_k\phi(2x-k), \tag{10}$$

$$\tilde{\phi}(x)=\sqrt{2}\sum \tilde{h}_k\tilde{\phi}(2x-k), \tag{11}$$

which are the key of the construction of biorthogonal bases of $L_2(R)$:

THEOREM 1. *Under assumptions* (AMB1–4) *we define* ψ, $\tilde{\psi}$, W_j, $\tilde{W}_j$ *with*

$$\psi(x)=\sqrt{2}\sum g_k\phi(2x-k), \qquad g_k=(-1)^{k+1}\tilde{h}_{1-k}, \tag{12}$$

$$\tilde{\psi}(x)=\sqrt{2}\sum \tilde{g}_k\tilde{\phi}(2x-k), \qquad \tilde{g}_k=(-1)^{k+1}h_{1-k}, \tag{13}$$

$$W_j=\mathrm{span}(\psi_{jk}, k\in Z), \qquad \tilde{W}_j=\mathrm{span}(\tilde{\psi}_{jk}, k\in Z). \tag{14}$$

Then the pair $(\{\phi_k, \psi_{jk}, j\geqslant 0, k\in Z\}, \{\tilde{\phi}, \tilde{\psi}_{jk}, j\geqslant 0, k\in Z\})$ *is a biorthogonal basis of* $L_2(R)$ *and more precisely*

$$V_{j+1}=V_j\oplus W_j, \qquad \tilde{V}_{j+1}=\tilde{V}_j\oplus\tilde{W}_j$$

$$V_j\perp\tilde{W}_j, \qquad \tilde{V}_j\perp W_j, \qquad W_j\perp\tilde{W}_k, \qquad k\neq j.$$

We have also

$$\sqrt{2}\,\tilde{\phi}(2x-m)=\sum_k g_{m-2k}\tilde{\psi}(x-k)+h_{m-2k}\tilde{\phi}(x-k). \tag{15}$$

The proof of this theorem is standard Hilbert space manipulations and will be omitted; more details may be found in [4]. In practice, ϕ, ψ, $\tilde{\phi}$, $\tilde{\psi}$ have compact support and h, g, $\tilde{h}$, $\tilde{g}$ have finite length. Most popular examples of such bases are given in [4].

Equations (11), (13), (15) imply that f can be represented as

$$f(x)=\sum_k \alpha_k\phi_{0k}+\sum_{j\geqslant 0,\, k}\beta_{jk}\psi_{jk} \tag{16}$$

where the coefficients

$$\alpha_{jk}=\langle f, \tilde{\phi}_{jk}\rangle, \qquad \beta_{jk}=\langle f, \tilde{\psi}\rangle \tag{17}$$

satisfy the analysis/synthesis relations:

$$\alpha_{jk} = \sum \tilde{h}_{l-2k} \alpha_{j+1,l} \tag{18}$$

$$\beta_{jk} = \sum \tilde{g}_{l-2k} \alpha_{j+1,l} \tag{19}$$

$$\alpha_{jk} = \sum h_{k-2l} \alpha_{j-1,l} + g_{k-2l} \beta_{j-1,l}. \tag{20}$$

Wavelet Bases in Dimension $d > 1$. We start with a one-dimensional multi-resolution analysis and define

$$\phi_{jk}(x) = \phi_{jk_1}(x_1) \cdots \phi_{jk_d}(x_d) \tag{21}$$

$$\tilde{\phi}_{jk}(x) = \tilde{\phi}_{jk_1}(x_1) \cdots \tilde{\phi}_{jk_d}(x_d) \tag{22}$$

where $k = (k_1, ..., k_d)$ is a multi-index and

$$\begin{aligned} V_j &= \mathrm{span}(\phi_{jk}(x), k \in Z^d) \\ \tilde{V}_j &= \mathrm{span}(\tilde{\phi}_{jk}(x), k \in Z^d). \end{aligned} \tag{23}$$

We use the same apparent notation for the wavelet in dimension one and the d-dimensional wavelet; there is however no ambiguity since the dimension of the argument and of the indices indicates what function is considered. To construct the spaces W_j we introduce the $2^d - 1$ functions $\psi^{(l)}(x)$ made of arbitrary products of $\phi(x_i)$ or $\psi(x_i)$, $i = 1, ..., d$, such that at least one function $\psi(x_i)$ is used. For instance, for $d = 2$ we obtain

$$\begin{aligned} \psi^{(1)}(x) &= \phi(x_1)\, \psi(x_2) \\ \psi^{(2)}(x) &= \psi(x_1)\, \phi(x_2) \\ \psi^{(3)}(x) &= \psi(x_1)\, \psi(x_2). \end{aligned}$$

If we put

$$\begin{aligned} \psi^{(l)}_{jk}(x) &= 2^{j/2} \psi^{(l)}(Z^j x - k) \\ W_j &= \mathrm{span}(\psi^{(l)}_{jk}, k \in Z^d, l = 1, ..., 2^d - 1) \end{aligned}$$

and do the same with the "tilded" variables, we obtain the same relations as given in the theorem (this is straightforward point). Thus, setting

$$\psi_{jk}(x) = (\psi^{(1)}_{jk}(x), ..., \psi^{(2^d-1)}_{jk}(x))^T \tag{24}$$

(the notation is unambiguous, cf. the comment after Eq. (23)) and

$$\alpha_{jk} = \langle f, \tilde{\phi}_{jk} \rangle, \qquad \beta^T_{jk} = \langle f, \tilde{\psi}_{jk} \rangle,$$

(β_{jk} is row vector) we can, as before, represent f as

$$f(x) = \sum_k \alpha_k \phi_{0k}(x) + \sum_{j \geqslant 0,\, k} \beta_{jk} \psi_{jk}(x). \tag{25}$$

3.2. *Besov Spaces*

We follow the presentation in Sections 3.4 and 3.5 of [17].

For any measurable function f, $u > 0$, and integer M we define the functions

$$\operatorname{osc}_u^M f(x, t) = \inf_P \left(\frac{1}{t^d} \int_{|x-y|<t} |f(y) - P(y)|^u \, dy \right)^{1/u}, \tag{26}$$

where the inf is taken over all polynomials P of degree no more than M (usual modification when $u = \infty$). Then we put

$$\theta_{jk}^{(u)} = \operatorname{osc}_u^M f(2^{-j}k, 2^{-j}). \tag{27}$$

The following result is a simple corollary of Theorem 3.5.1 in [17].

Proposition 1. *Let* $0 < p$, $q \leqslant \infty$, $s > d(1/p - 1/\max(u, 1))_+$, *and* $M \geqslant [s]$. *We set*

$$v_{spq}^{(u)}(f) = \|f\|_p + \left(\sum_{j=0}^{\infty} 2^{j(s-d/p)q} \, \|\theta_{j\cdot}^{(u)}\|_p^q \right)^{1/q} \tag{28}$$

(*modification if* $q, p = \infty$). *Then, for* s, p, q *fixed, those norms* (*quasi-norms if* p *or* q *is* <1) *are equivalent* (*when* u *varies in such a way that* $u < (1/p - s/d)^{-1}$ *if* $s \leqslant d/p$, $u > 0$ *if* $s > d/p$).

Proposition 1 can be obtained from Theorem 3.5.1 in [17] if we note that

$$\operatorname{osc}_u^m f(x, 2^{-j}) \geqslant 2^{-d/u} \operatorname{osc}_u^M f(x, 2^{-j-1}),$$

and

$$\operatorname{osc}_u^M f(x, 2^{-j-1}) \leqslant 2^{d/u} \operatorname{osc}_u^M f(2^{-j}k, 2^{-j})$$

for $|x - 2^{-j}k| \leqslant 2^{-j-1}$.

Remark. Note that $\theta_{jk}^{(u)}$ represents the error of approximation of f by a polynomial in the neighborhood of radius 2^{-j} of $2^{-j}k$; thus $\|\theta_{j\cdot}^{(u)}\|_p$ is the L_p error of the approximation with the best piecewise polynomial function.

Now we define the Besov space $B^s_{pq}(\mathbb{R}^d)$:

$$B^s_{pq} = \{ f \in L_{\max(p,\,1)}(\mathbb{R}^d),\ v^{(u)}_{spq}(f) < \infty \} \tag{29}$$

with the norm (or semi-norm) $\|\cdot\|_{spq}$ being one of the equivalent norms (semi-norms) $v^{(u)}_{spq}(\cdot)$. Theorem 3.5.1 from [17] states that this definition of Besov spaces is equivalent to other characterizations (see, for instance, Section 2.5 of [17]).

We recall now some injections between Besov spaces and more standard function spaces; we denote the Hölder and Sobolev spaces

$$C^s = \{ f, \sup_{x,\,h} |h|^{-s}\, |f(x+h) - f(x)| < \infty, \text{ and } \|f\|_\infty < \infty \} \qquad 0 < s \leqslant 1$$

$$W^s_p = \{ f, \|\mathscr{F}((1+|x|^s)\,\hat{f})\|_p < \infty \} \qquad s \geqslant 0$$

(W^s_p is the space of function of L_p such that their derivative of order up to s are in L_p).

- $B^s_{\infty\infty} = C^s$ for $0 < s < 1$
- $B^s_{pq} = C^0$ for $s > d/p$
- $B^s_{pp} \subset W^s_p \subset B^s_{p2}$ for $p \leqslant 2$
- $B^s_{p2} \subset W^s_p \subset B^s_{pp}$ for $p \geqslant 2$

The next result is proved in the Appendix; it relates Besov norms and wavelet coefficients, and it explains the popularity of Besov spaces in the wavelet community [5]–[8]:

THEOREM 2. *Let* $0 < p,\ q \leqslant \infty$ *and* $s > d(1/p - 1)_+$. *Suppose that Assumption* 1 *is satisfied with* $M = [s]$. *Then if* α_k *and* β_{jk} *are the wavelet coefficients of Eq.* (25), *the norm* $\|f\|_{spq}$ *is equivalent to the norm*

$$v_{spq}(f) = \|\alpha\|_p + \left(\sum_{j \geqslant 0} 2^{jq(s + d/2 - d/p)}\, \|\beta_{j\cdot}\|_p^q \right)^{1/q}.$$

An analogous result was shown in [12, Section 4.6] for the orthogonal wavelets and $p, q \geqslant 1$. The case $p \leqslant 1$ was considered in quite general settings in [15] and [8]. In order to obtain the result analogous to that of Theorem 2 the following condition was required in two latter papers:

- $\{\phi(\cdot - k)\colon \operatorname{supp} \phi(\cdot - k) \cap (0,1)^d \neq \varnothing,\ k \in Z^d\}$ form a set of linearly independent functions on $[0,1]^d$.

This condition is verified for the B-spline biorthogonal wavelets. However, it is well known that this condition is not satisfied, for instance, for orthogonal Daubechies' wavelets [13], widely used in estimation and compression [1, 4].

It is interesting that the proof of the theorem does not actually require the multiresolution structure (scaling property) and Assumption 1 can be replaced by

Assumption 2. The tuple $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ is such that $\{\phi(x-k), \psi(2^j x-k), j \geqslant 0, k \in Z^d\}$ and $\{\tilde{\phi}(x-k), \tilde{\psi}(2^j-x-k), j \geqslant 0, k \in Z^d\}$ are a biorthogonal pair of two bases of $L_2(R^d)$, such that ϕ and ψ are C^{M+1} for some $M \in N$. $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ have compact support. $\tilde{\psi}$ is orthogonal to polynomials of degree $\leqslant M$.

Note that Besov spaces over $(0, 1)$ can be defined as in (26)–(29) with the only difference that the oscillation in the definition (26) should be restricted to $(0, 1)$.

Suppose that ϕ and ψ generate an orthonormal basis on R and Assumption 1 holds. It has been shown in [3] and [2] that an orthogonal wavelet basis on $[0, 1]$ can be constructed by retaining those basis elements whose support is included in $[0, 1]$ and adding a finite number of adapted edge wavelets and scaling functions at each scale. These edge elements can be tailored so that

1. the edge-adapted wavelets are orthogonal (on $[0, 1]$) to polynomials of degree M, for some $M \in \mathbb{N}$;
2. edge-adapted wavelets and scale functions are $C^{M+1}(0, 1)$;
3. the total number is exactly 2^j at resolution j.

Then the characterization of Besov spaces on $(0, 1)$ using the coefficients of wavelet decompositions analogous to Theorem 2 can obtained; the proof can be taken over following exactly the lines of that of Theorem 2.

4. REGULAR DESIGN

In this section we consider the general multi-dimensional situation. We recall the convention we use for the notation of multidimensional wavelets (cf. the remark after Eq. (23) and Eq. (24): ψ_{jk}, $k \in Z^d$, is a (2^d-1)-dimensional vector-valued function and β_{jk} is a (2^d-1)-dimensional vector. $P_j f = \sum_k \alpha_{jk} \phi_{jk}(x)$ is the projection of f on V_j.

4.1. *Regular Point Design*

We denote $\bar{P}_j f = \sum_k \bar{\alpha}_{jk} \phi_{jk}(x)$, where $\bar{\alpha}_{jk}$ are computed according to Algorithm 1. We show first that the approximation $\bar{P}_j f$ is closed to f when f satisfies Besov constraints:

PROPOSITION 2. *Let $s > d/p$, $0 < p$, $q \leqslant \infty$, $p' \geqslant p$. Suppose that Algorithm* 1 *is used and that Assumption* 1 *holds with $M = [s]$. Then there exists C such that for any $f \in B^s_{pq}$, $\operatorname{supp}(f) \in [0, 1]^d$, and j, the approximation $\bar{P}_j f$ satisfies*

$$\left(\sum_j \|\bar{P}_j f - P_j f\|^q_{spq} \right)^{1/q} \leqslant C \|f\|_{spq} \tag{30}$$

$$\|\bar{P}_j f - f\|_{p'} \leqslant C \|f\|_{spq} 2^{-js'} \qquad \text{where} \quad s' = s - d/p + d/p'. \tag{31}$$

Proof of the Proposition. We will need the following

LEMMA 1. *Let Assumption* 1 *hold. Then there exists C and j_a such that for any f and $j \geqslant j_a$*

$$\left| 2^{jd} \int f(x)\, \tilde{\phi}(2^j x - k)\, dx - \sum_i c_i f(2^{-j}(k+i)) \right| \leqslant C \theta^{(\infty)}_{j - j_a,\, [2^{-j_a} k]}$$

where the oscillation $\theta^{(\infty)}_{jk}$ is defined by the Eq. (27) ($[\cdot]$ *denotes the integer part*).

Proof. Let us choose j_a such that $\operatorname{supp}(\tilde{\phi}) \cup \{i, c_i \neq 0\} \subset \{|x| < 2^{j_a}\}$ and consider the best polynomial P of degree $M = [s]$, to approximate f at $2^{-j}k$ at the resolution $j - j_a$ (in the sense of the definition (26) with $u = \infty$). Then

$$\begin{aligned}
&\left| 2^{jd} \int f(x)\, \tilde{\phi}(2^j x - k)\, dx - \sum_i c_i f(2^{-j}(k+1)) \right| \\
&\quad \leqslant \left| \int (f - P)(2^{-j}(k+x))\, \tilde{\phi}(x)\, dx \right| \\
&\qquad + \left| \int P(2^{-j}(k+x))\, \tilde{\phi}(x)\, dx - \sum_i c_i P(2^{-j}(k+1)) \right| \\
&\qquad + \left| \sum_i c_i (P - f)(2^{-j}(k+i)) \right| \\
&\quad = \left| \int (f - P)(2^{-j}(k+x))\, \tilde{\phi}(x)\, dx \right| + \left| \sum_i c_i (P - f)(2^{-j}(k+1)) \right| \\
&\quad \leqslant C \theta^{(\infty)}_{j - j_a,\, [2^{-j_a} k]} \quad \blacksquare
\end{aligned}$$

LEMMA 2. *Let Assumption* 1 *hold and $s - d/p + d/p' > 0$. Then the projection $P_{j_0} f$ on V_{j0} satisfies*:

$$\|P_{j_0} f - f\|_{p'} \leqslant C 2^{-j_0(s - d/p + d/p')} \|f\|_{spq}.$$

Proof. Let $p' \geqslant 1$. We have by the Minkowski inequality:

$$\|P_{j_0}f - f\|_{p'} = \left\| \sum_{j=j_0}^{\infty} \sum_k \beta_{jk} \psi_{jk}(x) \right\|_{p'} \leqslant \sum_{j=j_0}^{\infty} \left\| \sum_k \beta_{jk} \psi_{jk}(x) \right\|_{p'}. \tag{32}$$

If l denotes the volume of the wavelet support, we get (recall that $k \in Z^d$ is a multi-index):

$$\begin{aligned} \|P_{j_0}f - f\|_{p'} &\leqslant (l \, \|\psi\|_{\infty})^{1/p'} \sum_{j=j_0}^{\infty} \left\| \sum_k \beta_{jk} 2^{jd/2} 1_{\{2^j x \leqslant k < 2^j x + 1\}} \right\|_{p'} \\ &\leqslant C \sum_{j=j_0}^{\infty} 2^{jd/2 - jd/p'} \, \|\beta_{j\cdot}\|_{p'}. \end{aligned} \tag{33}$$

Note that $\|\beta_{j\cdot}\|_{j'} \leqslant \|\beta_{j\cdot}\|_p$ for $p' > p$. So Theorem 2 implies that

$$\|\beta_{j\cdot}\|_{p'} \leqslant C \, \|f\|_{spq} \, 2^{-j(s + d/2 - d/p)}.$$

When substituting this bound into (33) we obtain the lemma for $p' \geqslant 1$. The proof for $p' < 1$ can be carried out in an analogous way if one uses the inequality (51) instead of the Minkowski inequality in (32). ∎

We can continue the proof of the proposition

$$\begin{aligned} \|\bar{P}_j f - P_j f\|_{p'} &= \left\| \sum_k (\bar{\alpha}_{jk} - \alpha_{jk}) \, \phi_{jk} \right\|_{p'} \\ &\leqslant C 2^{j(d/2 - d/p')} \left\| \sum_k (\bar{\alpha}_{jk} - \alpha_{jk}) \, \phi_{0k} \right\|_{p'}. \end{aligned} \tag{34}$$

Since ϕ is compactly supported, we have (cf. (54) in the Appendix)

$$\left\| \sum_k (\bar{\alpha}_{jk} - \alpha_{jk}) \, \phi_{0k} \right\|_{p'} \leqslant C \, \|\phi\|_{p'} \, \|\bar{\alpha}_{j\cdot} - \alpha_{j\cdot}\|_{p'}.$$

When combining with (34) we obtain from Lemma 1

$$\|\bar{P}_j f - P_j f\|_{p'} = C 2^{j(d/2 - d/p')} \, \|\theta^{(\infty)}_{j - j_a, \cdot}\|_{p'} \leqslant C \, 2^{-j(s - d/p + d/p')} \, \|f\|_{spq},$$

which gives the second statement of the proposition. To show the first one we note that $\bar{P}_j f \in B^s_{pq}$. Furthermore, since

$$\|f(\lambda \cdot)\|_{spq} \leqslant C \lambda^{s - d/p} \, \|f(\cdot)\|_{spq} \tag{35}$$

(cf. [17, 2.3.3, comments after Eq. (13)], we have, from Theorem 2 and Assumption 1:

$$\begin{aligned}\|P_j f - \bar{P}_j f\|_{spq} &= \left\| \sum_k (\alpha_{jk} - \bar{\alpha}_{jk})\, 2^{jd/2} \phi(2^j x - k) \right\|_{spq} \\ &\leqslant C 2^{j(s-d/p+d/2)} \left\| \sum_k (\alpha_{jk} - \bar{\alpha}_{jk})\, \phi(x-k) \right\|_{spq} \\ &= C 2^{j(s-d/p+d/2)} \|\alpha_{j\cdot} - \bar{\alpha}_{j\cdot}\|_p \\ &\leqslant C 2^{j(s-d/p)} (1 + \|\tilde{\phi}\|_\infty)\, \|\theta^{(\infty)}_{j-j_a,\cdot}\|_p.\end{aligned}$$

Note that for $0 \leqslant j < j_a$, $\theta^{(\infty)}_{j-j_a,k} = \theta^{(\infty)}_{j,k}(f(2^{-j_a}\cdot))$, and for these j's, $\|\theta^{(\infty)}_{j-j_a,\cdot}\|_p \leqslant C\, \|f(2^{-j_a}\cdot)\|_{spq} \leqslant C'\, \|f(2^{-j_a}\cdot)\|_{spq}$. Along with the definition of the norm $\|f\|_{spq}$ (Proposition 1) this implies

$$\left(\sum_{j \geqslant 0} \|\bar{P}_j f - P_j f\|^q_{spq} \right)^{1/q} \leqslant C\, \|f\|_{spq}. \quad \blacksquare$$

4.2. *Regular Boxcar Design*

Set

$$\tilde{P}_j f = \sum_k \tilde{\alpha}_{jk} \phi_{jk}(x), \tag{36}$$

where $\tilde{\alpha}_{jk}$ are obtained using Algorithm 2.

PROPOSITION 3. *Let* $s > d(1/p - 1)_+$, $0 < p$, $q \leqslant \infty$, $p' \geqslant p$, $j_a = \log_2 a + 1$. *Suppose that Algorithm* 2 *is used and that Assumption* 1 *holds with* $M = [s]$. *Then there exists a constant* C *such that for any* f, $\mathrm{supp}(f) \in [0, 1]^d$ *and* j *the approximation* $\tilde{P}_j f$ *satisfies*

$$\left(\sum_j \|\tilde{P}_j f - P_j f\|^q_{spq} \right)^{1/q} \leqslant C\, \|f\|_{spq} \tag{37}$$

$$\|\tilde{P}_j f - Pf\|_{p'} \leqslant C\, \|f\|_{spq}\, 2^{-js'} \qquad \text{where} \quad s' = s - d/p + d/p'. \tag{38}$$

The proof of the proposition can be carried out in the same way as that of Proposition 2 if we substitute Lemma 1 with the one below:

LEMMA 3. *Suppose that Assumption* 1 *holds and Algorithm* 2 *is used. Then there exists* C *and* j_a *such that for any* f *and* $j \geqslant j_a$,

$$\left| 2^{jd} \int f(x)\, \phi(2^j x - k)\, dx - \sum_i c_i A^f_{j,k+i} \right| \leqslant C \theta^{(1)}_{j-j_a,[2^{-j_a}k]}.$$

Proof. We choose j_a as in Lemma 1 and the best polynomial P of degree M to approximate f at $2^{-j}k$ at the resolution $j-j_a$. It follows from Eq. (7) that for any polynomial P of degree $l \leqslant M$

$$\int P(x)\,\phi(x)\,dx = \sum_i c_i A^P_{0i}.$$

Then

$$\begin{aligned}
&\left|2^{jd/2}\int f(x)\,\phi_{jk}(x)\,dx - \sum_i c_i A^f_{j,\,k+i}\right| \\
&\quad\leqslant \left|\int (f(2^{-j}(k+x)) - P(2^{-j}(k+x)))\,\phi(x)\,dx\right| \\
&\qquad + \left|\int P(2^{-j}(k+x))\,\phi(x)\,dx - \sum_i c_i A^P_{j,\,k+i}\right| \\
&\qquad + \left|\sum_i c_i A^P_{j,\,k+i} - \sum_i c_i A^f_{j,\,k+i}\right| \\
&\quad= \|\phi\|_\infty \int_{|x|<2^{ja}} |f(2^{-j}(k+x)) - P(2^{-j}(k+x))|\,dx \\
&\qquad + \|c\|_\infty \sum_i \int_{|x|<1/2} |P(2^{-j}(k+i+x)) - f(2^{-j}(k+i+x))|\,dx \\
&\quad\leqslant C\theta^{(1)}_{j-j_a,\,2^{-j_a}k}.
\end{aligned}$$

4.3. *Filter* (c_i)

As we have mentioned in the Introduction there are many different ways to define the filter sequences (c_i) in Algorithms 1 and 2. Since $\tilde{\phi}$ is a continuous function with compact support which satisfies Eq. (11) above, it is not difficult to check that the numbers $M_i = \int x^i \tilde{\phi}(x)\,dx$ satisfy $M_0 = 1$ and

$$(2^i - 1)\,M_i = \frac{1}{\sqrt{2}} \sum_{l=0}^{i-1} \left(\sum_k h_k k^{-il}\right)\binom{i}{l} M_l \tag{39}$$

(the sequence h_k here is defined in (10)). There exists quite a simple way to obtain the sequence (c_i) for the orthogonal wavelet basis (ϕ_k, ψ_{jk}). The idea of the following lemma is borrowed in [16]:

LEMMA 4. *Let* (ϕ) *be an orthogonal multi-resolution analysis satisfying Assumption* 1. *Then for any* $k = 0, ..., M$,

$$\sum_j j^k \phi(j) = \int y^k \phi(y)\,dy.$$

Lemma 4 suggests that a possible way to choose c_i in the Algorithm 1 when orthogonal wavelets are used is simply to put

$$c_i = \phi(i).$$

We note that the length of the filter (c_i) in this case is at least $2M$ (cf. Chap. 6 of [4]), though the minimal length solution of the system (39) contains only $M+1$ coefficients.

Proof. The condition of the compact support along with vanishing moment assumption imply (see the Proof of Corollary 5.5.4, p. 155 of [4]) that ϕ and $\hat{\phi}$ are rapidly decreasing, and $\hat{\phi}(2\pi j) = \hat{\phi}'(2\pi j) = \cdots = \hat{\phi}^{(l)}(2\pi j) = 0$ for $j \neq 0$. We use the Poisson formula (cf. [18, formula 13.4 on p. 68])

$$\sum_j f(j) = \sum_j \hat{f}(2\pi j)$$

with $f(t) = t^k \phi(t-x)$:

$$\begin{aligned}
\sum_j j^k \phi(j-x) &= \sum_j i^k \partial^k (e^{-i\omega x} \hat{\phi}(\omega)) \Big|_{\omega = 2\pi j} \\
&= i^k \partial^k (e^{-i\omega x} \hat{\phi}(\omega)) \,|_{\omega = 0} \\
&= \sum_{r \leqslant k} \binom{k}{r} i^{k-r} x^r \partial^{k-r} \hat{\phi}(0) \\
&= \sum_{r \leqslant k} \binom{k}{r} i^{k-r} x^r \int (-iy)^{k-r} \, \phi(y) \, dy \\
&= \int (x+y)^k \, \phi(y) \, dy. \qquad (40)
\end{aligned}$$

This implies the lemma. ∎

Using the same method, we could prove the following result: denote by $\phi_0(x)$ the function defined by the equation

$$\phi(x) = \int_{-1/2}^{1/2} \phi_0(x+u) \, dy, \qquad (41)$$

then for any $k = 0, ..., l-2$,

$$\frac{1}{k+1} \sum_j \left[\left(j + \frac{1}{2} \right)^{k+1} - \left(j - \frac{1}{2} \right)^{k+1} \right] \phi_0(j) = \int y^k \phi(y) \, dy.$$

4.4. *Lower Bounds for the Approximation Rate*

Let us consider the case $d=1$ for the sake of simplicity. We claim that the approximation bounds obtained in Propositions 2 and 3 are tight. Indeed, the following lower bound for the approximation rate can be easily obtained (compare to the Kolmogorov–Tikhomirov diameters [9]):

PROPOSITION 4. *Let $s>1/p$, $0<p, q\leqslant\infty$. Then there exists a function f, $\|f\|_{spq}\leqslant 1$, and $c>0$ such that for any approximation f_N based on the discrete observations $f(i/N)$ $i=1, ..., N$*

$$\|f_N-f\|_{p'}\geqslant c\,\|f\|_{spq}\,N^{-s'} \qquad \text{where} \quad s'=s-1/p+1/p'.$$

The same bound holds true for the averaged scheme.

Proof. Take j such that $N\leqslant 2^j<2N$. Suppose that the wavelet $\psi(x)$ is compactly supported, $\operatorname{supp}(\psi)\subseteq[-a, a]$, and $\psi(x)\in C^r$, $r>s$. If $j_a\leqslant \log_2 a+1$ then for all $l=2^{j_a}(i+1/2)$, $=0, ..., N-1$, we have $\phi_{j+j_a, l}(i/N)=0$. Consider two functions

$$f_0(x)=0 \qquad \text{and} \qquad f_1(x)=\beta\phi_{j+j_a, 2^{j_a}-1}(x), \tag{42}$$

with $\beta=2^{-(j+j_a)(s+1/2-1/p)}$. Obviously, using the norm of Theorem 2, $\|f_1\|_{spq}=1$. On the other hand, we cannot distinguish f_0 and f_1 using only the observations $f(i/N)$, $i=1, ..., N$. Thus

$$\begin{aligned}\max_{f=f_1, f_0}\|f-f_N\|_{p'}&\geqslant 1/2\beta\,\|\psi_{j+j_a, j^{j_a}-1}\|_{p'}\\ &=1/2\beta\,2^{(j+j_a)(1/2-1/p')}\|\psi\|_{p'}=c2^{-js'}.\end{aligned}$$

The second statement of the theorem can be obtained in the same way if we note that for f defined in (42) all $A^f_{jk}=0$ because $\int\psi(x)\,dx=0$. ∎

5. IRREGULAR DESIGN

In order to simplify the presentation we consider only the case $d=1$ in this section. The generalization of the results stated in this section to $d>1$ is rather straightforward, though tedious. We suppose that the observations $f_i=f(X_i)$ are available at the points X_i, $i=1, ..., N$.

First we present an explicit solution of the least squares problem in Algorithm 3; a change of variables is introduced for the easiness of the proof. We put

$$\begin{aligned}Z_k(x)&=(1, 2^jx-k, ..., (2^jx-k)^M)^T\\ Z_{k,i}&=Z_k(X_i).\end{aligned}$$

Then the polynomial $\hat{P}_k$ may be written

$$\hat{P}_k(x) = \hat{\theta}_k^T Z_k(x)$$

where the vector $\hat{\theta}_k$ is the solution of the minimization problem

$$\hat{\theta}_k = \arg\min_{\theta} \sum_{i \in I_k} (\theta^T Z_{k,i} - f_i)^2.$$

Thus

$$\hat{\theta}_k = \frac{V^{-1}}{|I_k|} \sum_{i \in I_k} f_i Z_{k,i} \tag{43}$$

where

$$V_k = \frac{1}{|I_k|} \sum_{i \in I_k} Z_{k,i} Z_{k,i}^T.$$

Hence $\hat{f}_k = \hat{P}_k(2^{-j}(k+1/2))$ (cf. Eq. (9)) and $\hat{A}_{jk} = 2^{j/2} \int_{\lambda_k} \hat{P}_k(x)\, dx$ in Algorithm 4 can be computed as

$$\hat{f}_k = u^T \hat{\theta}_k, \qquad u = (1, 1/2, ..., 2^{-M})^T$$

$$\hat{A}_{jk} = 2^{-j/2} \hat{\theta}_k^T u, \qquad u = \left(1, 1/2, ..., \frac{1}{M+1}\right)^T.$$

The following simple proposition provides us with the estimate of the rate of approximation obtained using Algorithms 3 and 4 when f belongs to a Besov class.

PROPOSITION 5. *Let $s > 1/p$, $0 < p, q \leqslant \infty$, $p' \geqslant p$. Suppose that Assumption 1 holds with $M = [s]$ and that j is such that for any $k = 0, ..., 2^j - 1$ the gain matrices V_k^{-1} are bounded. Then for any f,* $\operatorname{supp}(f) \in [0, 1]$, *the approximation*

$$\hat{P}_j f = \sum_k \hat{\alpha}_{jk} \phi_{jk}$$

satisfies

$$\left(\sum_j \|\hat{P}_j f - f\|_{spq}^q \right)^{1/q} \leqslant C \|f\|_{spq} \tag{44}$$

$$\|\hat{P}_j f - f\|_{p'} \leqslant C \|f\|_{spq} 2^{-js'} \qquad \textit{where} \quad s' = s - 1/p + 1/p'. \tag{45}$$

Remark. Proposition 5 gives an upper bound of the rate of convergence of the projection $\hat{P}_j f$ if the norm of "gain" matrices V_k^{-1} can be controlled. We show in the next subsections how this condition can be checked in two particular cases considered in the Introduction.

Proof. We provide the proof for Algorithm 3. The same estimate for Algorithm 4 can be shown in a completely analogous way. Recall that the algorithm creates new "observations" on the regular grid $2^{-j}k$ using the least squares method, and then uses Algorithm 1 to compute the estimates $\hat{\alpha}_{jk}$ of wavelet coefficients. Thus the approximation error $\|\hat{P}_j f - f\|_{p'}$ can be separated into two parts:

$$\|\hat{P}_j f - f\|_{p'} \leqslant C(p)(\|\hat{P}_j f = \bar{P}_j f\|_{p'} + \|\bar{P}_j f - f\|_{p'}) = C(p)(\delta_1 + \delta_2).$$

Here δ_1 is the error of estimating the projection $\bar{P}_j f$ with $\hat{P}_j f$, and δ_2 is the error of approximation of $P_j f$ using the observations on the regular grid. From Proposition 2 we have the estimate for δ_2:

$$\delta_2 \leqslant C\,\|f\|_{spq}\, 2^{-js'}.$$

As in (34) we have for δ_1

$$\delta_1 = \|\hat{P}_j f - \bar{P}_j f\|_{p'} = \leqslant C\, 2^{j(1/2 - 1/p')}\, \|\hat{\alpha}_{j\cdot} - \bar{\alpha}_{j\cdot}\|_{p'}. \tag{46}$$

On the other hand,

$$\begin{aligned} |\hat{\alpha}_{jk} - \bar{\alpha}_{jk}| &\leqslant 2^{-j/2} \left| \sum_l c_l (\hat{f}_{k+l} - f(2^{-j}(k+l+1/2))) \right| \\ &\leqslant C\, 2^{-j/2} \sum_l |\hat{f}_{k+l} - f(2^{-j}(k+l+1/2))|. \end{aligned} \tag{47}$$

Let P be the optimal polynomial in the definition (26) of $\theta_{jk}^{(\infty)}$. If $|V_k^{-1}| \leqslant C_v$ then

$$\begin{aligned} &|\hat{f}_{k+l} - f(2^{-j}(k+l+1/2))| \\ &\quad\leqslant \left| \frac{1}{|I_k|} \sum_{i \in I_k} r^T V_k^{-1} Z_{k,i} (f(X_i) - P(X_i)) \right| \\ &\qquad + \left| \frac{1}{|I_k|} \sum_{i \in I_k} u^T V_k^{-1} Z_{k,i} P(X_i) - P(2^{-j}(k+l+1/2)) \right| \\ &\qquad + |P(2^{-j}(k+l+1/2)) - f(2^{-j}(k+l+1/2))| \\ &\quad\leqslant c_v \theta^{(\infty)}_{j-j_a, [2^{-ja}k]} + 0 + \theta^{(\infty)}_{j-j_a, [2^{-ja}k]} \end{aligned}$$

(cf. the comment after Algorithm 3). Along with (46) and (47) we have by (28)

$$\delta_1 \leqslant C 2^{-j/p'} \|\theta^{(\infty)}_{j-j_a,\,[2^{-j_a}k]}\|_{p'} \leqslant C 2^{-j(s-1/p+1/p')} \|f\|_{spq} \quad \blacksquare$$

5.1. *Jittered Grid*

Let X_i, $i=1, ..., N$, be the points of the jittered grid. We take

$$\frac{N}{4(M+1)} < 2^j \leqslant \frac{N}{2(M+1)}. \tag{48}$$

Then $|I_k| \geqslant 2(M+1)$ and we get the following

LEMMA 5. *There is $C<\infty$ such that $|V_k^{-1}| \leqslant C$.*

Proof. For the sake of simplicity we suppose that $2^j = N/2(M+1)$ exactly, and $|I_k| = 2(M+1)$. Let λ_0 be the smallest eigenvalue of V_k, and μ, $|\mu|=1$, the corresponding eigenvector. Consider the polynomial

$$B(z) = \sum_{l=0}^{M} \mu_l z^l \qquad \text{for} \quad z \in [0, 1[.$$

Note that among $2(M+1)$ points $2^j X_i - k$, $i \in I_k$, there are at least $M+1$ "well distanced" points $z_1, ..., z_{M+1}$ such that $\inf_{m,\,l \leqslant M+1} |z_m - z_l| \geqslant 2^{-M-1}$. Then

$$\lambda_0 = \frac{1}{2(M+1)} \sum_{i \in I_k} B^2(z_i) \geqslant \frac{1}{2(M+1)} \sum_{m=1}^{M+1} B^2(z_m) \geqslant c_0 > 0,$$

because the polynomial $B^2(z)$ has only M different roots and the sequence (z_m) is taken from a compact set where the latter functional is strictly positive. $\blacksquare$

When combining the results of Proposition 5 with the lemma above we obtain the following:

PROPOSITION 6. *Let $s > 1/p$, $0 < p$, $q \leqslant \infty$, $p' \geqslant p$. Suppose that Assumption 1 holds with $M = [s]$ and that X_i, $i = 1, ..., N$, are the points of the jittered grid. Let j be such that*

$$\frac{N}{4(M+1)} < 2^j \leqslant \frac{N}{2(M+1)}.$$

Then the approximation $\hat{P}_j f$ satisfies

$$\left(\sum_j \|\hat{P}_j f - f\|_{spq}^q\right)^{1/q} \leqslant C \|f\|_{spq}$$

$$\|\hat{P}_j f - f\|_{p'} \leqslant C \|f\|_{spq} N^{-s'} \qquad \textit{where} \quad s' = s - 1/p + 1/p'.$$

Remark. It can be easily verified that one needs $O(M^3)$ elementary operations to compute $\hat{\theta}_k$ from (43) on a segment Λ_k. The bound (48) for 2^j (for 2^{dj} in the d-dimensional case) then implies that the total effort to compute the coefficients $\hat{\alpha}_{jk}$ will be of order $N \times M$.

5.2. *Random Grid*

Let X_i, $i = 1, ..., N$, be independent and identically distributed random variables on $[0, 1]$ with density $p(x)$, $0 < p_{\min} \leqslant p(x) \leqslant p_{\max} < \infty$. In order to obtain the bound for the rate of approximation it suffices to find the largest j such that all $|V_k^{-1}|$ are bounded with overwhelming probability. Consider the following algorithm: Take j such that

$$\frac{N}{2\lambda \ln N} \leqslant 2^j < \frac{N}{\lambda \ln N}. \tag{49}$$

Compute

$$\hat{f}_k = \frac{1}{|I_k|} \sum_{i \in I_k} u^T V_k^{-1} Z_{k,i} f(X_i).$$

Then form

$$\hat{\alpha}_{jk} = \sum_i 2^{-j/2} c_i \hat{f}_k$$

and

$$\hat{P}_j f(x) = \sum_k \hat{\alpha}_{jk} \phi_{jk}(x).$$

Note. Note that $O(M^2 \times |I_k|)$ elementary operations are required to compute $\hat{f}_k$. Since we have the overwhelming probability that $|I_k| = O(\ln N)$, we conclude from (49) that the average effort to compute the coefficients $\hat{\alpha}_{jk}$ will be of order $N \times M$.

PROPOSITION 7. *Let $s > 1/p$, $0 < p$, $q \leqslant \infty$, and $p' \geqslant p$. Then for any $\gamma < \infty$ there are $\lambda_0(\gamma)$ such that if $\lambda_0 \leqslant \lambda < \infty$ then*

$$P\left(\|\hat{P}_j f - P_j f\|_{p'} \geqslant C_1(\lambda) \|f\|_{spq} \left(\frac{\ln N}{N}\right)^{s'}\right) \leqslant C_2(\lambda) N^{-\gamma},$$

where $s' = s - 1/p + 1/p'$. (*Here* $P(\cdot)$ *is the probability by the joint measure generated by* (X_i), $i = 1, ..., N$.)

Proof. The proof of the proposition is based on the following simple

LEMMA 6. *There is* $\alpha > 0$,

$$E[V_k \, |I_k|] \geqslant \alpha N 2^{-j} p_{\min} I$$

(*here* $E[\cdot]$ *stands for expectation by the joint measure generated by* (X_i), $i = 1, ..., N$). *Moreover, for any* $\varepsilon > 0$ *there are* C *and* $C' < \infty$ *such that*

$$P(|\,|I_k| - NP(X_1 \in \Lambda_k)| \geqslant \varepsilon NP(X_1 \in \Lambda_k)) \leqslant CN^{-\gamma},$$

$$P(|V_k \, |I_k| - E[V_k \, |I_k|]| \geqslant \varepsilon E[V_k \, |I_k|]) \leqslant CN^{-\gamma}$$

for any $\lambda \geqslant \lambda_0(\gamma, \varepsilon)$ *in the statement of the proposition.*

Proof. To prove the first assertion we compute for an $a \in \mathbb{R}^{M+1}$, $|a| = 1$,

$$a^T E[V_k |I_k|] \, a = N a^T E[Z_{k,1} Z_{k,1}^T 1_{\{X_1 \in \Lambda_k\}}] \, a = NE[B(Z^j X_1 - k) \, 1_{\{X_1 \in \Lambda_k\}}],$$

where $B(z) = \sum_{l=0}^{M} a_l z^l$ we get

$$a^T E[V_k \, |I_k|] \, a = N \int_{\Lambda_k} B(2^j x - k) \, p(x) \, dx \geqslant N 2^{-j} p_{\min} \int_0^1 B^2(z) \, dz.$$

The latter integral is positive uniformly in the compact set $|a| = 1$:

$$\int_0^1 B^2(z) \, dz \geqslant \alpha(M) \, |a| = \alpha(M) > 0.$$

This gives the first inequality of the lemma.

Note that $N^{-1} E[|I_k|] = P(X_1 \in \Lambda_k) \stackrel{\text{def}}{=} p_k$, and

$$E[|\,|I_k| - Np_k|]^2 = E\left[\left|\sum_{i=1}^{N} (1_{\{i \in I_k\}} - p_k)\right|^2\right] = N(p_k - p_k^2) \leqslant Np_k.$$

When using the Bernstein inequality, we obtain for $\varepsilon < 1$

$$P(|\,|I_k| - Np_k| \geqslant \varepsilon Np_k) \leqslant 2 \exp\left(-\frac{\varepsilon^2 N^2 p_k^2}{2Np_k + (2/3)\varepsilon Np_k}\right) \leqslant 2 \exp\left(-\frac{\varepsilon^2 Np_k}{3}\right).$$

Since

$$p_k \geqslant p_{\min} 2^{-j} \geqslant \frac{p_{\min} \lambda \ln N}{N},$$

we have

$$P(|\,|I_k| - Np_k|) \geqslant \varepsilon N p_k \leqslant 2\exp\left(-\frac{p_{\min}\lambda\varepsilon^2 \ln N}{3}\right) \leqslant 2N^{-\gamma}$$

for any $\lambda \geqslant 3\gamma/\varepsilon^2 p_{\min}$. The last inequality of the lemma can be shown in an analogous way. ▮

We continue now the proof of the proposition. We denote $\lambda_{\min}(V_k)$ be the minimum eigenvalue of V_k. Note that

$$\begin{aligned}
\lambda_{\min}(V_k) &= |I_k|^{-1}\,\lambda_{\min}(V_k\,|I_k|)\\
&\geqslant |I_k|^{-1}\,\lambda_{\min}(E[V_k\,|I_k|])\\
&\quad + |I_k|^{-1}\,\lambda_{\min}(V_k\,|I_k| - E[V_k\,|I_k|])\\
&\geqslant \alpha_+^{-1} 1_{\{|I_k| \leqslant \alpha_+\}}\,\lambda_{\min}(E[V_k\,|I_k|])\\
&\quad - \alpha_-^{-1} 1_{\{|I_k| \geqslant \alpha_-\}} |V_k\,|I_k| - E[V_k\,|I_k|]|\\
&\geqslant \alpha_+^{-1} 1_{\{|I_k| \leqslant \alpha_+\}}\,\lambda_{\min}(E[V_k\,|I_k|])\\
&\quad - \alpha_-^{-1} 1_{\{|I_k| \geqslant \alpha_-\}} \beta 1_{\{|V_k\,|I_k| - E[V_k\,|I_k|]| \leqslant \beta\}}.
\end{aligned}$$

We put

$$\alpha_+ = (1+\varepsilon)\,Np_k, \qquad \alpha_- = (1-\varepsilon)\,Np_k, \qquad p_k = P(X_1 \in \Lambda_k)$$

and

$$\beta = \varepsilon E[V_k\,|I_k|];$$

then using the result of Lemma 6, we obtain

$$\begin{aligned}
&P\left(\lambda_{\min}(V_k) \leqslant \frac{1-2\varepsilon-\varepsilon^2}{1-\varepsilon^2} N^{-1} p_k^{-1} E[V_k\,|I_k|]\right)\\
&\quad\leqslant P(|\,|I_k| - Np_k| \geqslant \varepsilon N p_k) + P(|V_k\,|I_k| - E[V_k\,|I_k|]| \geqslant \varepsilon E[V_k\,|I_k|])\\
&\quad\leqslant CN^{-\gamma}.
\end{aligned} \tag{50}$$

Since for $\varepsilon < 1/3$

$$\frac{1-2\varepsilon-\varepsilon^2}{1-\varepsilon^2} N^{-1} p_k^{-1} E[V_k\,|I_k|] \geqslant c\,2^{-j} p_{\min} p_k^{-1} \geqslant c',$$

the bound (50) implies

$$P(|V_k^{-1}| \leqslant (c')^{-1}) \leqslant CN^{-\gamma}.$$

Along with the result of Proposition 5 this gives the proof. ∎

Remark. Note that the bounds shown in Proposition 6 and 7 are tight. It is demonstrated by the following

PROPOSITION 8. *Let $s > 1/p$, $0 < p, q \leqslant \infty$. Then there exists a function f, $\|f\|_{spq} \leqslant 1$, $c > 0$, and $\alpha > 0$ such that for any approximation method $\hat{f}(f(X_1), ..., f(X_N))$ based on the discrete observations $f(X_i)$, $i = 1, ..., N$,*

1. *if $X_1, ..., X_N$ are the points of a regular jittered grid, then $\|f_N - f\|_{p'} \geqslant c\, \|f\|_{spq}\, N^{-s'}$;*
2. *if $X_1, ..., X_N$ are the points of a random grid, then $P(\|f_N - f\|_{p'} \geqslant c\, \|f\|_{spq}\, N^{-s'}) > \alpha$ where $s' = s - 1/p + 1/p'$.*

Proof. The proof of this statement is analogous to that of Proposition 4. For instance, in the case of the random grid it suffices to note that there exists k such that $P(X_1 \in [k/N, (k+1)/2)/N]) \leqslant 1/2N$, thus for some $\alpha > 0$,

$$P\left(\operatorname{card}\left\{i \colon X_i \in \left[\frac{k}{N}, \frac{(k+1/2)}{N}\right]\right\} = 0\right) > \alpha. \quad ∎$$

APPENDIX

When compared with other proofs [8, 15] of the characterization of Besov spaces by wavelet coefficients, the Proof of Theorem 2 below is very direct: in order to prove the lower characterization, a polynomial approximation of functions ϕ_k and ψ_{jk} is used rather than the condition of independence of translations $\phi(\cdot - k)$ and $\psi(\cdot - k)$ on an interval.

We shall use the following inequality for L_p quasi-norm: for $p < 1$, if f and g are measurable functions, and if (x_n) and (y_n) are real sequences one has

$$\|x + y\|_p^p \leqslant \|x\|_p^p + \|y\|_p^p \qquad \text{and} \qquad \|f + g\|_p^p \leqslant \|f\|_p^p + \|g\|_p^p. \qquad (51)$$

We recall the convention we use for the notation of multidimensional wavelets (cf. the remark after Eq. (23) and Eq. (24)): ψ_{jk}, $k \in Z^d$, is a

(2^d-1)-dimensional vector-valued function and β_{jk} is a (2^d-1)-dimensional vector.

We need the following lemma.

Lemma 7. *For any $a>1$ and $q>0$, there exists C_{aq} such that for any non-negative sequence $(x_j)_{j\geqslant 0}$,*

$$\sum_{j\geqslant 0} a^{jq}\left(\sum_{l\geqslant j} x_l\right)^q \leqslant C_{aq}\sum_{l\geqslant 0} a^{lq}x_l^q. \tag{52}$$

For any $a<1$ and $q>0$, there exist C_{aq} such that for any non-negative sequence $(x_j)_{j\geqslant 0}$,

$$\sum_{j\geqslant 0} a^{jq}\left(\sum_{l\leqslant j} x_l\right)^q \leqslant C_{aq}\sum_{l\geqslant 0} a^{lq}x_l^q. \tag{53}$$

Proof. We consider first the case $a>1$. If $q<1$:

$$\sum_{j\geqslant 0} a^{jq}\left(\sum_{l\geqslant j} x_l\right)^q \leqslant \sum_{j\geqslant 0} a^{jq}\sum_{l\geqslant j} x_l^q \leqslant C_a\sum_{l\geqslant 0} a^{lq}x_l^q,$$

and if $q\geqslant 1$, setting $b=(1+a)/2$:

$$\begin{aligned}
\sum_{j\geqslant 0} a^{jq}\left(\sum_{l\geqslant j} x_l\right)^q &\leqslant \sum_{j\geqslant 0} a^{jq}\left(\sum_{l\geqslant j} x_l b^l b^{-l}\right)^q \\
&\leqslant \sum_{j\geqslant 0} a^{jq}\left(\sum_{l\geqslant j} x_l^q b^{lq}\right)\left(\sum_{l\geqslant j} b^{-lq/(q-1)}\right)^{q-1} && \text{(Hölder)}\\
&\leqslant C_a\sum_{j\geqslant 0} (a/b)^{jb}\left(\sum_{l\geqslant j} x_l^q b^{lq}\right) && \text{since } b>1\\
&\leqslant C'_a\sum_{l\geqslant 0} (a/b)^{lq}x_l^q b^{lq} && \text{since } b<a\\
&= C'_a\sum_{l\geqslant 0} x_l^q a^{lq}.
\end{aligned}$$

If now $a<1$ and $q<1$,

$$\sum_{j\geqslant 0} a^{jq}\left(\sum_{l\leqslant j} x_l\right)^q \leqslant \sum_{j\geqslant 0} a^{jq}\sum_{l\leqslant j} x_l^q \leqslant C_a\sum_{l\geqslant 0} a^{lq}x_l^q.$$

If $q \geqslant 1$, setting $b = (a+1)/2$, we have as before:

$$\sum_{j \geqslant 0} a^{jq} \left(\sum_{l \leqslant j} x_l \right)^q \leqslant \sum_{j \geqslant 0} a^{jq} \left(\sum_{l \leqslant j} x_l b^l b^{-l} \right)^q$$
$$\leqslant \sum_{j \geqslant 0} a^{jq} \left(\sum_{l \leqslant j} x_l^q b^{lq} \right) \left(\sum_{l \leqslant j} b^{-lq/(q-1)} \right)^{q-1}$$
$$\leqslant C_a \sum_{l \geqslant 0} x_l^q a^{lq}.$$

Proof of Theorem 2. The proof is made in eight steps. In Steps 1–5 we show that $v_{spq}(f) \leqslant \|f\|_{spq}$: in Steps 1–2 we bound the β's using the polynomial approximation of f in Proposition 1 and (26); in Step 3 we bound the α's in the case $p \geqslant 1$; a particular difficulty arises when $p < 1$: this case is dealt with in Steps 4–5. The proof of $\|f\|_{spq} \leqslant v_{spq}(f)$ is made in Steps 6–7, but essentially in Step 7, where the polynomial approximation for f is obtained using the polynomial approximation of ϕ and ψ.

Step 1. There exist j_0 such that for any $j \geqslant 0$, $2^{jd/2} \|\beta_{j\cdot}\|_p \leqslant C \|\theta^{(1)}_{j-j_0,\,\cdot}\|_p$.

Indeed,

$$2^{jd/2} \beta_{jk} = \int f(2^{-j}(x+k))\, \tilde{\psi}(x)\, dx = 2^{jd} \int (f(x) - P(x))\, \tilde{\psi}(2^j x - k)\, dx$$

for any polynomial P of degree smaller than s; thus

$$2^{jd/2} |\beta_{jk}| \leqslant \theta^{(1)}_{j-j_0,\,[2^{-j_0}k]} \|\tilde{\psi}\|_\infty,$$

where j_0 is such that $|x| < 2^{j0}$ for any x of the support of $\tilde{\psi}$.

Step 2. For $j < j_0$, $\|\theta^{(1)}_{j-j_0,\,\cdot}\|_p \leqslant C\|f\|_{spq}$; $2^{jd/2} \|\beta_{j\cdot}\|_p \leqslant C\|f\|_{spq}$.

The first inequality is a consequence of two relations: first, if we set $g(x) = f(2^{j0}x)$ we have

$$\theta^{(1)}_{j-j_0,\,k}(f) = \theta^{(1)}_{j,\,2^{j0}k}(g)$$

and second

$$\|g\|_{spq} \leqslant 2^{j_0(s-d/p)} \|f\|_{spq}$$

(cf. the inequality (35)). This bound along with Step 1 (for $j < j_0$) and (28) (for $j \geqslant j_0$) gives us the second inequality of the step.

We obtain from Steps 1, 2 and (28):

$$\sum_j 2^{qj(s+d/2-d/p)} \|\beta_{j\cdot}\|_p^q \leqslant C \|f\|_{spq}^q.$$

Step 3. If $p \geqslant 1$, $\|\alpha\|_p \leqslant C\|f\|_p$.

Since $\alpha_k = \langle f, \tilde{\phi}_k \rangle$, we have

$$\frac{|\alpha_k|}{\|\tilde{\phi}\|_1} \leqslant \left(\int |f(x)|^p \frac{|\tilde{\phi}(x-k)|}{\|\tilde{\phi}\|_1} \, dx \right)^{1/p} \qquad \text{(Hölder)}.$$

Thus

$$\begin{aligned} \sum_k |\alpha_k|^p &\leqslant \|\tilde{\phi}\|_1^{p-1} \int |f(x)|^p \sum_k |\tilde{\phi}(x-k)| \, dx. \\ &\leqslant \|\tilde{\phi}\|_1^{p-1} \left\| \sum_k |\tilde{\phi}(x-k)| \right\|_\infty \int |f(x)|^p \, dx. \end{aligned}$$

Step 4. If $p < 1$, $\|\alpha\|_p \leqslant C \, \|\sum_k \alpha_k \phi_k\|_p$.

Let I and J denote the support of functions $\tilde{\phi}(x)$ and $\phi(x)$ respectively. Put $g = \sum_k \alpha_k \phi_k$. Then we get for α_k

$$|\alpha_k|^p = \left| \int g(x) \, \tilde{\phi}(x-k) \, dx \right|^p \leqslant C \left(\int_I |g(x+k)|^p \, dx \right)^p \max_{x \in I} |g(x+k)|^{p(1-p)}.$$

The inequality ($p \leqslant 1$, $q = 1 - p$)

$$xy \leqslant p\varepsilon^{-1/p} x^{1/p} + q\varepsilon^{1/q} y^{1/q}$$

implies

$$\begin{aligned} |\alpha_k|^p &\leqslant Cp\varepsilon^{-1/p} \int_I |g(x+k)|^p \, dx + Cq\varepsilon^{1/q} \max_{x \in I} |g(x+k)|^p \\ &\leqslant Cp\varepsilon^{-1/p} \int_{I+k} |g(x)|^p \, dx + C \, \|\phi\|_\infty \, q\varepsilon^{1/q} \left(\sum_{l \in J-I+k} |\alpha_l| \right)^p \\ &\leqslant C' \int_{I+k} |g(x)|^p \, dx + \varepsilon' \sum_{l \in J-I+k} |\alpha_l|^p. \end{aligned}$$

Choosing ε small enough, we can have ε' smaller than half the number of multiintegers of $J - I$, and summing both sides over k, we obtain

$$\|\alpha\|_p^p \leqslant C'' \, \|g(x)\|_p^p + \|\alpha\|_p^p / 2$$

which is the result.

Step 5. If $p<1$, $\|\sum_k \alpha_k \phi_k\|_p^p \leqslant C\|f\|_{spq}$.

In this case, we have

$$\left\|\sum_k \alpha_k \phi_k\right\|_p^p \leqslant C\,\|f\|_p^p + C\sum_{jk} |\beta_{jk}|^p\, \|\psi_{jk}\|_p^p,$$

and, using Steps 1 and 2,

$$\begin{aligned}\sum_{jk} |\beta_{jk}|^p\, \|\psi_{jk}\|_p^p &\leqslant C\sum_{jk} |\beta_{jk}|^p\, 2^{jd(p/2-1)} = C\sum_j \|\beta_{j\cdot}\|_p^p\, 2^{jd(p/2-1)}\\ &\leqslant C'\, \|f\|_{spq}^p \sum_j 2^{-jd} \leqslant C''\, \|f\|_{spq}^q.\end{aligned}$$

Step 6. $\|f\|_p \leqslant C v_{spq}(f)$.

If $p>1$,

$$\|f\|_p \leqslant \left\|\sum_k \alpha_k \phi_k\right\|_p + \sum_j \left\|\sum_k \beta_{jk}\psi_{jk}\right\|_p.$$

Since ϕ has compact support, we can find a partition $I_1, ..., I_l$ of $\mathbb{Z}^d$ such that ϕ_k and $\phi_{k'}$ do not overlap if k and k' are in the same I_i, and

$$\left\|\sum_k \alpha_k \phi_k\right\|_p \leqslant \sum_{i=1}^{l} \left\|\sum_{k\in I_i} \alpha_k \phi_k\right\|_p = \sum_{i=1}^{l} \left(\sum_{k\in I_i} |\alpha_k|^p\, \|\phi_k\|_p^p\right)^{1/p} \leqslant l\, \|\phi\|_p\, \|\alpha\|_p \quad (54)$$

(the equality is due to the non-overlapping property). In the same way

$$\left\|\sum_k \beta_{jk}\psi_{jk}\right\|_p \leqslant C\, \|\beta_{j\cdot}\|_p\, \|\psi_{jk}\|_p = C\, \|\psi\|_p\, \|\beta_{j\cdot}\|_p\, 2^{j(d/2-d/p)}.$$

And since

$$\|\beta_{j\cdot}\|_p \leqslant C v_{spq}(f)\, 2^{-j(s+d/2-d/p)}.$$

we obtain immediately the result (s is >0).

The case $p<1$ is treated exactly the same way where Eq. (51) is used instead of the Minkowski inequality.

Step 7. $\sum_j 2^{j(s-d/p)q}\, \|\theta_{j\cdot}^{(p)}\|_p^q \leqslant C v_{spq}^q(f)$.

Denote

$$Q_j f = \sum_k \beta_{jk}\psi_{jk}$$

$$R_j f = \sum_{l\geqslant j} Q_l f$$

$$P_j f = \sum_k \alpha_k \phi_k + \sum_{l\leqslant j} Q_l f.$$

$P_j f$ is the projection of f on V_j. Since ϕ and ψ are C^{M+1}, we can find polynomials $\Phi_{k'}$, $\Psi_{j'k'}$, and C independent of j or k, such that for $|2^{-j}k - x| < 2^{-j}$, $j' \leqslant j$, and any k' we have

$$|\phi_{k'}(x) - \Phi_{k'}(x)| < C 2^{-j(M+1)}, \qquad |\psi_{j'k'}(x) - {}_{jk'}(x)| < C 2^{-(j-j')(M+1)} 2^{j'd/2}.$$

On each interval of length 2^{-j} we will approximate f with the polynomial

$$\Pi_{jk}(x) = \sum_{k'} \alpha_{k'} \Phi_{k'}(x) + \sum_{0 \leqslant j' \leqslant j} \sum_{k'} \beta_{j'k'} \Psi_{j'k'}(x)$$

where each sum is restricted to those k' for which the function $\phi_{k'}$ or $\psi_{j'k'}$ is non-zero on the interval; each sum on k' contains at most K terms where K depends on the support of ϕ and ψ; in other words, there exist K such that the first sum is restricted to $|k' - 2^{-j}k| < K$ and, for each j', the sum in the second term is restricted to $|k' - 1^{-j+j'}k| < K$.

$$\Pi_{jk}(x) = \sum_{|k' - 2^{-j}k| < K} \alpha_{k'} \Phi_{k'}(x) + \sum_{0 \leqslant j' \leqslant j} \sum_{|k' - 2^{-j+j'}k| < K} \beta_{j'k'} \Psi_{j'k'}(x).$$

This implies that

$$\sup_{|2^{-j}k - x| < 2^{-j}} |P_j f(x) - \Pi_{jk(x)}| \leqslant C \sum_{|k' - 2^{-j}k| < K} 2^{-j(M+1)} |\alpha_{k'}| + C \sum_{l \leqslant j} \sum_{|k' - 2^{-j+l}k| < K} |\beta_{lk'}| \, 2^{-(j-l)(M+1) + ld/2}$$

and

$$\sum_k \sup_{|2^{-j}k - x| < 2^{-j}} |P_j f(x) - \Pi_{jk}(x)|^p \leqslant C' \, 2^{jd} 2^{-j(M+1)\,p} \, \|\alpha\|_p^p + C' \sum_{l \leqslant j} 2^{(j-l)\,d} \, \|\beta_{l\cdot}\|_p^p \, 2^{-(j-l)\,p(M+1) + lpd/2},$$

where C depends here only on p and K.

Then

$$(\theta_{jk}^{(p)})^p \leqslant 2^{jd} \int_{|2^{-j}k - y| < 2^{-j}} |f(x) - \Pi_{jk}(x)|^p \, dx$$

$$\leqslant C \sup_{|2^{-j}k - x| < 2^{-j}} |P_j f(x) - \Pi_{jk}(x)|^p + C \, 2^{jd} \int_{|2^{-j}k - x| < 2^{-j}} |R_j(x)|^p \, dx$$

$$\sum_k (\theta_{jk}^{(p)})^p \leqslant C'\, 2^{jd} 2^{-j(M+1)\,p}\, \|\alpha\|_p^p$$
$$+ C' \sum_{l<j} 2^{d(j-l)}\, \|\beta_{l\cdot}\|_p^p\, 2^{-(j-l)\,p(M+1)+lpd/2}$$
$$+ C\, 2^{jd} \int |R_j(x)|^p\, dx$$

$$2^{j(sp-d)}\, \|\theta_{j\cdot}^{(p)}\|_p^p \leqslant C'\, 2^{-j(M+1-s)\,p}\, \|\alpha\|_p^p$$
$$+ C'\, 2^{-jp(M+1-s)} \sum_{l<j} 2^{lp(M+1+d/2-d/p)}\, \|\beta_{l\cdot}\|_p^p$$
$$+ C 2^{jsp} \int |R_j(x)|^p\, dx = A_j + B_j + C_j.$$

We have to show that $\sum_{j \geqslant 0} (A_j + B_j + C_j)^{q/p} \leqslant C v_{spq}^q(f)$.

Note that obviously, since $s < M+1$,

$$\sum_{j \geqslant 0} A_j^{q/p} \leqslant C\, \|\alpha\|_p^q \leqslant C v_{spq}(f)^q.$$

Now using Eq. (53)

$$\sum_{j \geqslant 0} B_j^{q/p} \leqslant C \sum_{j \geqslant 0} 2^{j(s+d/2-d/p)\,q}\, \|\beta_{j\cdot}\|_p^{q/p} \leqslant C v_{spq}(f)^q.$$

If $p \geqslant 1$ we obtain

$$\sum_{j \geqslant 0} c_j^{q/p} = \sum_{j \geqslant 0} 2^{jsq}\, \|R\|_p^q \leqslant \sum_{j \geqslant 0} 2^{jsq} \left(\sum_{l \geqslant j} \|Q_l\|_p \right)^q \leqslant C \sum_{j \geqslant 0} 2^{jsq}\, \|Q_j\|_p^q$$

(thanks to Eq. (52)), and if $p < 1$

$$\sum_{j \geqslant 0} C_j^{q/p} = \sum_{j \geqslant 0} 2^{jsq} (\|R_j\|_p^p)^{q/p} \leqslant \sum_{j \geqslant 0} 2^{jsq} \left(\sum_{l \geqslant j} \|Q_l\|_p^p \right)^{q/p} \leqslant \sum_{j \geqslant 0} 2^{jsq}\, \|Q_j\|_p^q.$$

On the other hand, because of the compactness of the support of ψ:

$$\|Q_j\|_p^p \leqslant C \sum_k |\beta_{jk}|^p\, \|\psi_{jk}\|_p^p \leqslant C\, 2^{j(pd/2-d)}\, \|\beta_{j\cdot}\|_p^p,$$

and finally, for any $p > 0$,

$$\sum_{j \geqslant 0} C_j^{q/p} \leqslant C \sum_{j \geqslant 0} 2^{jsq} 2^{j(d/2-d/p)\,q}\, \|\beta_{j\cdot}\|_p^q \leqslant C' v_{spq}(f)^q.$$

Step 8. End of the proof.

Putting together the results of Steps 2, 3, 5, 6, and 7, we obtain the theorem. ▮

REFERENCES

1. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using Wavelet transforms, *IEEE Proc. Acoustics, Speech, Signal Process.* (1991), to appear.
2. A. Cohen, I. Daubechies, B. Jawerth, and P. Vial, Multiresolution analysis, Wavelets, and fast algorithms on interval, *Comp. Rend. Acad. Sci. Paris (A)* **316** (1993), 417–421.
3. A. Cohen, B. Daubechies, and P. Vial, Wavelets on the interval and fast Wavelet transforms, *Appl. Comput. Harmonic Anal. (A)* **1** (1993), 54–81.
4. I. Daubechies, "Ten Lectures on Wavelets," CBMS–NSF Regional Conference Series in Applied Mathematics, SIAM, 1992.
5. D. Donoho, Interpolating wavelet transforms. Dept. of Statistics, Stanford University, 1993; ftp:plaifair.stanford.edu.
6. D. Donoho, Smooth wavelet decompositions with blocky coefficient kernels, Dept. of Statistics, Stanford University, 1993; ftp:plaifair.stanford.edu.
7. D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, Density estimation by wavelet thresholding, submitted, 1993.
8. B. Jawerth, R. A. DeVore, and V. Popov, Compression of wavelet decompositions, *Amer. J. Math.* **114** (1993), 737–785.
9. A. N. Kolmogorov and V. M. Tikhomirov, ε-entropy and ε-capacity of sets in functional spaces, *Usp. Mat. Nauk.* **14**, No. 2 (1995), 3–86.
10. A. Korostelev and A. Tsybakov, "Minimax Theory of Image Reconstruction," Springer-Verlag, New York/Berlin, 1993.
11. S. Mallat and W. L. Hwang, Singularity detection and processing with wavelets, *IEEE Trans. Info Theory* **38**, No. 2 (1993), 617–643.
12. Y. Meyer, "Ondelettes et opérateurs," Hermann, Paris, 1990.
13. Y. Meyer, Ondelettes sur l'intervalle, *Rev. Mat. Iberoamericana* **7** (1991), 115–134.
14. D. Pollard, "Convergence of Stochastic Processes," Springer-Verlag, New York/Berlin, 1984.
15. W. Stickel, Spline representations of functions in Besov-Triebel-Lizorkin spaces on R^n, *Forum Math.* **2** (1990), 451–476.
16. G. Strang and G. Fix, A Fourier analysis of the finite element variational method, *in* "Constructive Aspects of Functional Analysis" (Geyminat, Ed.), pp. 793–840, C.I.M.E., 1978.
17. H. Triebel, "Theory of Function Spaces II," Birkhauser, Basel, 1983.
18. A. Zygmund, "Trigonometric Series," Vols. I and II combined, Cambridge Univ. Press, Cambridge, 1993.