On the Computation of Wavelet Coefficients

Bernard Delyon, Anatoli Juditsky*

March 9, 1999

*IRISA-INRIA, Campus de Beaulieu, 35042 RENNES Cedex, FRANCE

Abbreviated title : Wavelet coefficients

Anatoli Juditsky
IRISA-INRIA, Campus de Beaulieu
35042 RENNES Cedex
FRANCE
e-mail: iouditsk@irisa.fr.

Abstract

We consider fast algorithms of wavelet decomposition of a function f when discrete observations of f ($\text{supp} f \subseteq [0,1]$) are available. The properties of the algorithms are studied for three types of observation design: 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 jittered regular grid, when it is only known that for all $1 \leq i \leq N$ $i/N \leq x_i < (i+1)/N$; the random design case: x_i $i = 1, ..., N$ are independent and identically distributed random variables on $[0,1]$. We show that these algorithms are in certain sense efficient when the accuracy of approximation is concerned.

The proposed algorithms are computationally straightforward: the whole effort to compute the decomposition is order N for the sample size N.

1 Introduction

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

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

where

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

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

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

(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 $2^{-j(s+1/2)}$ 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 $2^{-j}k$. These ideas are rigorously formalized in section 2.2.

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 that for a smooth function f the coefficients β_{jk} decay rapidly, and consequently, the wavelet projection $P_{j_0}f$,

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

converges rapidly to f when j_0 increases. We would like 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 multi-resolution analysis such that ϕ and ψ are C^{M+1} for some $M \in \mathbf{N}$. $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ have compact support. $\tilde{\psi}$ is orthogonal to polynomials of degree $\leq M$.

Note that by corollary 5.5.2 of [2], 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 lead to the $\alpha_k = \alpha_{0k}$'s and $\beta_{j'k}$'s, $j' = 0, \ldots j$, 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 $f_1, \ldots f_N$, 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. [15]). Theorem 1 says that functions from the Besov spaces can be "well" (in some sense) approximated by piecewise polynomials.

We are to 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 $f_i = f(X_i)$, we call the vector $X = (X_1, \ldots, X_N)^T$ the observation design.[1] We consider three types of observation designs below.

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

1.3 Regular point design

We suppose that the observations f_k are taken on a regular grid, i.e. $X_k = k/N, \quad k = 1, ..., N$, $N = 2^j$. 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 S. Mallat [9] and widely spread in engineering practice without much rigorous foundation (cf. [1]). A theoretical analysis of this method was recently performed by D. Donoho [3], 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 1.5).

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, $\sum_{k=1}^d l_k \leq M$ and $2^{-j/2}$ becomes $2^{-jd/2}$.

Theoretical bases for this procedure are given in proposition 1. In simpler words, the idea is that this quadrature formula is exact for polynomials of degree $\leq M$; thus, using Taylors's formula in equation (3), one guesses easily that the quadrature error will be quite small.

1.4 Regular boxcar design

In this case we suppose that N observations of interval means $A_1^f, ..., A_N^f$ are available,

$$A_k^f = \int_{-1/2}^{1/2} f((x+k)/N)dx, \qquad k = 1..., N.$$

This type of observations is usual in the problems of computer vision, when the average intensity over receipting retinas cells is measured. 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 D. Donoho in [4]. He also considered an alternative approach which consists in using *biorthogonal* wavelet transform [2]. We consider the following

Algorithm 2 :

j is fixed and we observe

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

Find a finite sequence (c_i) such that

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

and set

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

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

1.5 Calculation of the (c_i) sequence and 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 [2] p.258), $c_i = \delta_{0i}$ is also convenient for algorithm 1 since ϕ itself has M vanishing moments, i.e.

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

(see, for instance, ch. 6 of [2]).

On the other hand, if orthogonal wavelets are used, lemma 3 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:

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

In this case the result of algorithm 1 is the "empirical wavelet transform" (cf. [5]).

In the case of the boxcar design and orthogonal wavelets, lemma 3 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 [2] p.272).

1.6 Irregular design

We suppose that N observations $f_i = f(X_i)$ are available at the points X_i, $i = 1, ..., N$. We are particularly interested in two special cases of irregular design:

- *jittered regular design* is a simple case of irregular sampling. In this case we suppose that the observations $f(X_i)$, $i = 1, ..., N$, where X_i belongs to the interval $[(i-1)/N, i/N]$. A particular case of this is a *random jittered grid*, when X_i are independent random variables, each X_i is uniformly distributed in $[(i-1)/N, i/N]$ (cf. [8], sect.1.9);
- *random design,* when observations X_i, $i = 1, ..., N$ are independent identically distributed on $[0,1]^d$ with some density $p(x)$: $0 < p_{min} \le p(x) \le p_{max} < \infty$. This situation is common when general scattered data models are concerned.

We use the same algorithm for these two cases, 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 4 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 $\Lambda_k = [2^{-j}k, 2^{-j}(k+1)[$, $k = 0, \dots, 2^j - 1$, 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

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

for the case of regular jittered grid and

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

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

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

where the minimum is taken over all polynomials of degree $\leq M$. 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 $\widehat{\alpha}_{jk}$ using algorithm 1 ((5)) :

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

Commentary : Note that $P_\theta^{(k)}$ 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 $\leq M$. 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

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

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

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

1.7 Organization and notations

The paper is organized as follows : in section 2 we recall some notions on wavelets and Besov spaces and introduce notations to be used later. Then we study the properties of algorithms 1 and 2 in section 3. Finally, we consider the case of irregular design and provide the estimates of the accuracy of algorithms 3 and 4 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

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

We denote by C a generic positive constant.

2 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.

2.1 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 $((a_k)_{k\in\mathbf{N}}, (b_k)_{k\in\mathbf{N}})$ 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 $L_2(\mathbf{R})$ and spaces*

$$\begin{aligned} V_j &= span(\phi_{jk}, k \in \mathbf{Z}) & \phi_{jk}(x) = 2^{j/2}\phi(2^j x - k) \\ \tilde{V}_j &= span(\tilde{\phi}_{jk}, k \in \mathbf{Z}) & \tilde{\phi}_{jk}(x) = 2^{j/2}\tilde{\phi}(2^j x - k) \end{aligned}$$

with the properties

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

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

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

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

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 $L_2(\mathbf{R})$:

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

Theorem 1 *Under assumptions (AMB1-4) we define* $\psi, \tilde{\psi}, W_j, \tilde{W}_j$ *with*

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

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

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

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

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

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

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 [2]. In practice, $\phi, \psi, \tilde{\phi}, \tilde{\psi}$ have compact support and $h, g, \tilde{h}, \tilde{g}$ have finite length. Most popular examples of such bases are given in [2].

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

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

where the coefficients

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

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 1-dimensional multi-resolution analysis and define

$$\boldsymbol{\phi}_{jk}(x) = \phi_{jk_1}(x_1) ... \phi_{jk_d}(x_d)$$

$$\tilde{\boldsymbol{\phi}}_{jk}(x) = \tilde{\phi}_{jk_1}(x_1) ... \tilde{\phi}_{jk_d}(x_d)$$

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

$$V_j = span(\boldsymbol{\phi}_{jk}(x), k \in \mathbf{Z}^d)$$

$$\tilde{V}_j = span(\tilde{\boldsymbol{\phi}}_{jk}(x), k \in \mathbf{Z}^d).$$

To construct the spaces W_j we introduce the $2^d - 1$ functions $\boldsymbol{\psi}^{(l)}(x)$ made of arbitrary products of $\phi(x_i)$ or $\psi(x_i)$, $i = 1 \ldots d$, such that at least one function $\psi(x_i)$ is used. For instance, for $d = 2$ we obtain

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

If we put

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

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

$$\boldsymbol{\psi}_{jk}(x) = (\boldsymbol{\psi}^{(1)}_{jk}(x), \ldots \boldsymbol{\psi}^{(2^d-1)}_{jk}(x))^T$$

and

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

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

$$f(x) = \sum_k \alpha_k \boldsymbol{\phi}_{0k} + \sum_{j \geq 0, k} \beta_{jk}\boldsymbol{\psi}_{jk}. \tag{21}$$

2.2 Besov spaces

We follow the presentation in section 3.4 and 3.5 of [15].

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

$$\mathrm{osc}^M_u f(x,t) \;=\; \inf_P \left(\frac{1}{t^d}\int_{|x-y|<t} |f(y) - P(y)|^u dy\right)^{1/u}, \tag{22}$$

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

$$\theta^{(u)}_{jk} = \mathrm{osc}^M_u f(2^{-j}k, 2^{-j}). \tag{23}$$

The following result is a simple corollary of Theorem 3.5.1 in [15]

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

$$\nu^{(u)}_{spq}(f) = \|f\|_p + \left(\sum_{j=0}^{\infty} 2^{j(s-d/p)q} \|\theta^{(u)}_{j.}\|_p^q \right)^{1/q} \tag{24}$$

(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 $s \leq d/p$, $u > 0$ if $s > d/p$) and

$$B^s_{pq} = \{f \in L_{\max(p,1)}(\mathbf{R}^d), \ \ \nu^{(u)}_{spq}(f) < \infty\}. \tag{25}$$

Theorem 1 can be obtain from theorem 3.5.1 in [15] if we note that

$$\begin{aligned} &\mathrm{osc}^M_u f(x, 2^{-j}) \geq 2^{-d/u} \mathrm{osc}^M_u f(x, 2^{-j-1}), \quad \text{and} \\ &\mathrm{osc}^M_u f(x, 2^{-j-1}) \leq 2^{d/u} \mathrm{osc}^M_u f(2^{-j}k, 2^{-j}). \end{aligned}$$

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

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

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

$$\begin{aligned} &C^s = \{f, \sup_{x,h} |h|^{-s}|f(x+h) - f(x)| < \infty, \ \ \textit{and} \ \ \|f\|_\infty < \infty\} \quad s \geq 0 \\ &W^s_p = \{f, \|\mathcal{F}((1 + |x|^s)\hat{f})\|_p < \infty\} \quad s \geq 0 \end{aligned}$$

(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 $p \leq 2$
- $B^s_{p2} \subset W^s_p \subset B^s_{pp}$ for $p \geq 2$

We see that B^s_{pq} is quite close to W^s_p.

Next result is proved in appendix; it relates Besov norms and wavelet coefficients, and explains the popularity of Besov spaces in wavelet community [3] – [6]:

Theorem 2 *Let $0 < p, q \leq \infty$ and $s > d(1/p-1)_+$. Assumption 1 is satisfied with $M=[s]$. Then if α_k and β_{jk} are the wavelet coefficients of equation (21), the norm $\|f\|_{spq}$ is equivalent to the norm*

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

An analogous result was shown in [10] section 4.6 for the orthogonal wavelets and $p, q \geq 1$. The case $p \leq 1$ was considered in quite general settings in [13] and [6]. In order to obtain the result analogous to that of theorem 2 the following condition was required:

- $\{\phi(\cdot - k) : \mathrm{supp}\phi(\cdot - k) \cap (0,1)^d \neq \emptyset, \quad k \in \mathbf{Z}^d\}$ form a set of linearly independent functions on $[0,1]^d$.

This condition is verified for the B-spline biorthogonal wavelets. It is well known, however, that this condition is not satisfied when the general orthogonal wavelets are concerned [11].

3 Regular design

Recall that $P_j f = \sum_k \alpha_{jk}\phi_{jk}(x)$ is the projection of f on V_j.

3.1 Regular point design

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

Proposition 1 *Let $s > d/p$, $0 < p, q \leq \infty$, $p' \geq p$. Suppose that algorithm 1 is used and that assumption 1 holds with $M = [s]$. Then there exists C such that for any $f \in B^s_{pq}$ and j, the approximation $\bar{P}_j f$ satisfies*

$$\left(\sum_j \|\bar{P}_j f - P_j f\|_{spq}^q \right)^{1/q} \leq C\|f\|_{spq} \tag{26}$$

$$\|\bar{P}_j f - f\|_{p'} \leq C\|f\|_{spq} 2^{-js'} \qquad \textit{where } s' = s - d/p + d/p'. \tag{27}$$

Notes: The optimality of the rate of convergence of equation (27) is proved in proposition 3. Equation (26) indicates that $\|\bar{P}_j f - P_j f\|$ is generically smaller than $\|f - P_j f\|$.

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 $j \geq j_a$*

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

where the oscillation $\theta^{(\infty)}_{jk}$ is defined by the equation (23) ($[\cdot]$ denotes the integer part).

Proof: Let us choose j_a such that $\mathrm{supp}(\tilde{\phi})\cup\{i, c_i\neq 0\}\subset\{|x|<2^{j_a}\}$ 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 (22) with $u=\infty$). Then

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

∎

We can continue the proof of the proposition.

$$\|\bar{P}_j f - P_j f\|_{p'} = \|\sum_k(\bar{\alpha}_{jk}-\alpha_{jk})\phi_{jk}\|_{p'} \leq C2^{j(d/2-d/p')}\|\sum_k(\bar{\alpha}_{jk}-\alpha_{jk})\phi_{0k}\|_{p'}. \tag{28}$$

Since ϕ is compactly supported, we have (cf. (45) in the appendix)

$$\|\sum_k(\bar{\alpha}_{jk}-\alpha_{jk})\phi_{0k}\|_{p'} \leq C\|\phi\|_{p'}\|\bar{\alpha}_{j.}-\alpha_{j.}\|_{p'}.$$

When combining with (28) we obtain from lemma 1

$$\|\bar{P}_j f - P_j f\|_{p'} = C2^{j(d/2-d/p')}\|\theta^{(\infty)}_{j-j_a,\cdot}\|_{p'} \leq C2^{-j(s-d/p+d/p')}\|f\|_{spq}.$$

Note that for the real projection $P_j f$ on V_j we have

$$\|P_j f - f\|_{p'} \leq C2^{-j(s-d/p+d/p')}\|f\|_{spq},$$

$\boxed{\text{preuve?}}$ 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

$$\|f(\lambda\cdot)\|_{spq} \leq C\lambda^{s-d/p}\|f(\cdot)\|_{spq} \tag{29}$$

(cf [15], 2.3.3, comments after equation (13)), we have, using theorem 2,

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

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

$$\left(\sum_{j\geq 0} \|\bar{P}_j f - P_j f\|^q_{spq}\right)^{1/q} \leq C\|f\|_{spq}.$$

∎

3.2 Regular boxcar design

Set

$$\tilde{P}_j f = \sum_k \tilde{\alpha}_{jk}\phi_{jk}(x), \tag{30}$$

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

Proposition 2 *Let $s > d(1/p-1)_+$, $0 < p, q \leq \infty$, $p' \geq p$, $j_a = \log_2 a + 1$. 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 and j the approximation $\tilde{P}_j f$ satisfies*

$$\left(\sum_j \|\tilde{P}_j f - P_j f\|^q_{spq}\right)^{1/q} \leq C\|f\|_{spq} \tag{31}$$

$$\|\tilde{P}_j f - Pf\|_{p'} \leq C\|f\|_{spq} 2^{-js'} \qquad \text{where } s' = s - d/p + d/p'. \tag{32}$$

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

Lemma 2 *Suppose that assumption 1 holds and algorithm 2 is used. Then there exists C and j_a such that for any f and $j \geq j_a$*

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

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 equation (7) that for any polynomial P of degree $l \leq M$

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

Then

$$\begin{aligned}
|2^{jd/2}\int f(x)\phi_{jk}(x)dx - \sum_i c_i A^f_{j,k+i}| \quad \leq \quad & |\int (f(2^{-j}(k+x)) - P(2^{-j}(k+x)))\phi(x)dx| \\
& +|\int P(2^{-j}(k+x))\phi(x)dx - \sum_i c_i A^P_{j,k+i}| \\
& +|\sum_i c_i A^P_{j,k+i} - \sum_i c_i A^f_{j,k+i}|
\end{aligned}$$

$$\begin{aligned}
&= \|\phi\|_\infty \int_{|x|<2^{j_a}} |f(2^{-j}(k+x)) - P(2^{-j}(k+x))|dx \\
&\quad +\|c\|_\infty \sum_i \int_{|x|<1/2} |P(2^{-j}(k+i+x)) - f(2^{-j}(k+i+x))|dx \\
&\leq \ C\theta^{(1)}_{j-j_a,2^{-j_a}k}
\end{aligned}$$

∎

3.3 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 equation (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

$$(2^i - 1)M_i = \frac{1}{\sqrt{2}} \sum_{l=0}^{i-1} \left(\sum_k h_k k^{i-l} \right) \left(\begin{array}{c} i \\ l \end{array} \right) M_l \tag{33}$$

(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 [14]:

Lemma 3 *Let (ϕ) be an orthogonal multi-resolution analysis satisfying assumption 1. Then for any $k = 0, \ldots, M$*

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

Lemma 3 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. ch.6. of [2]), though the minimal length solution of the system (33) 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 [2]) that ϕ and $\widehat{\phi}$ are rapidly decreasing, and $\widehat{\phi}(2\pi j) = \widehat{\phi}'(2\pi j) = \ldots = \widehat{\phi}^{(l)}(2\pi j) = 0$ for $j \neq 0$. We use the Poisson formula (cf. [16] formula 13.4 on p. 68)

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

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

$$\begin{aligned}
\sum_j j^k \phi(j-x) &= \sum_j i^k \partial^k (e^{-i\omega x} \widehat{\phi}(\omega))\Big|_{\omega=2\pi j} \\
&= i^k \partial^k (e^{-i\omega x} \widehat{\phi}(\omega))\Big|_{\omega=0} \\
&= \sum_{r \leq k} C_k^r i^{k-r} x^r \partial^{k-r} \widehat{\phi}(0) \\
&= \sum_{r \leq k} C_k^r i^{k-r} x^r \int (-iy)^{k-r} \phi(y) dy \\
&= \int (x+y)^k \phi(y) dy. \qquad (34)
\end{aligned}$$

This implies the lemma. ∎

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

$$\phi(x) = \int_{-1/2}^{1/2} \phi_0(x+u) du. \qquad (35)$$

then for any $k = 0, \ldots, l-2$

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

3.4 Lower bounds for the approximation rate

Note that the approximation bounds obtained in propositions 1 and 2 are tight. Indeed, the following lower bound for the approximation rate can be easily obtained (compare to the Kolmogorov-Tikhomirov diameters, [7]) :

Proposition 3 *Let $s > 1/p$, $0 < p, q \le \infty$. Then there exist a function f, $\|f\|_{spq} \le 1$ and $c > 0$ such that for any approximation f_N based on the discrete observations $f(i/N)$ $i = 1, \ldots, N$*

$$\|f_N - f\|_{p'} \ge c\|f\|_{spq} N^{-s'} \qquad \textit{where } s' = s - 1/p + 1/p'.$$

The same bound holds true for the averaged scheme.

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

$$f_0(x) = 0, \quad \text{and} \quad f_1(x) = \beta\phi_{j+j_a,2^{j_a-1}}(x), \tag{36}$$

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 $f(i/N)$ $i = 1, \ldots, N$. Thus

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

The second statement of the theorem can be obtained in the same way if we note that for f defined in (36) all $A^f_{jk} = 0$ because $\int \psi(x)dx = 0$. ∎

4 Irregular design

In what follows only the case $d = 1$ is considered. 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

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

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

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

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

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

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 4 *Let $s > 1/p$, $0 < p, q \leq \infty$, $p' \geq p$. 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 the approximation*

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

satisfies

$$\left(\sum_j \|\widehat{P}_j f - f\|_{spq}^q\right)^{1/q} \leq C\|f\|_{spq} \tag{37}$$
$$\|\widehat{P}_j f - f\|_{p'} \leq C\|f\|_{spq}2^{-js'} \qquad \textit{where } s' = s - 1/p + 1/p'. \tag{38}$$

Remark: proposition 4 gives an upper bound of the rate of convergence of the projection $\widehat{P}_j f$ if the norm of "gain" matrices V_k^{-1} can be controlled. We show in 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 the algorithm 4 can be shown in a completely analogous way. Recall that the algorithm consists to create new "observations" on the regular grid $2^{-j}k$ using the least squares method, and then to use algorithm 1 to compute the estimates $\widehat{\alpha}_{jk}$ of wavelet coefficients. Thus the approximation error $\|\widehat{P}_j f - f\|_{p'}$ can be separated into two parts :

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

Here δ_1 is the error of estimating the projection $\bar{P}_j f$ with $\widehat{P}_j f$, and δ_2 is the error of approximation of $P_j f$ using the observations on the regular grid. From proposition 1 we have the estimate for δ_2 :

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

As in (28) we have for δ_1 :

$$\delta_1 = \|\widehat{P}_j f - \bar{P}_j f\|_{p'} = \leq C 2^{j(1/2-1/p')} \|\widehat{\alpha}_{j\cdot} - \bar{\alpha}_{j\cdot}\|_{p'}. \tag{39}$$

On the other hand,

$$\begin{aligned} |\widehat{\alpha}_{jk} - \bar{\alpha}_{jk}| &\leq 2^{-j/2} |\sum_l c_l (\widehat{f}_{k+l} - f(2^{-j}(k+l+1/2))| \\ &\leq C 2^{-j/2} \sum_l |\widehat{f}_{k+l} - f(2^{-j}(k+l+1/2))|. \end{aligned} \tag{40}$$

Let P be the optimal polynomial in the definition (22) of $\theta_{jk}^{(\infty)}$. If $|V_k^{-1}| \leq C_v$ then

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

(cf. the comment after algorithm 3). Along with (39) and (40) we have by (24)

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

∎

4.1 Jittered grid

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

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

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

Lemma 4 *There is $C < \infty$ such that $|V_k^{-1}| \leq C$.*

Proof: For the sake of simplicity we suppose that $2^j = \frac{N}{2(M+1)}$ 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

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

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 $\inf_{m,l \leq M+1} |z_m - z_l| \geq 2^{-M-1}$. Then

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

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. ∎

When combining the results of proposition 4 with lemma above we obtain the following :

Proposition 5 *Let $s > 1/p$, $0 < p, q \leq \infty$, $p' \geq p$. Suppose that assumption 1 holds with $M = [s]$ and that X_i, $i = 1, \ldots, N$ are the points of jittered grid. Let j be such that*

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

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

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

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

4.2 Random grid

Let X_i, $i = 1, ..., N$ be independent and identically distributed random variables on $[0,1]$ with density $p(x)$, $0 < p_{min} \leq p(x) \leq p_{max} < \infty$. 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

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

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

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

and

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

Proposition 6 *Let $s > 1/p$, $0 < p, q \leq \infty$ and $p' \geq p$. Then for any $\gamma < \infty$ there are $\lambda_0(\gamma)$ such that if $\lambda_0 \leq \lambda < \infty$ then*

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

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 5 *There is $\alpha > 0$,*

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

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

$$P\left(|\,|I_k| - NP(X_1 \in \Lambda_k)| \geq \epsilon NP(X_1 \in \Lambda_k)\right) \leq CN^{-\gamma},$$
$$P\left(|V_k|I_k| - E[V_k|I_k|]| \geq \epsilon E[V_k|I_k|]\right) \leq CN^{-\gamma}$$

for any $\lambda \geq \lambda_0(\gamma, \epsilon)$ in the statement of the proposition.

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

$$a^T E[V_k|I_k|]a = N a^T E[Z_{k,1} Z_{k,1}^T 1_{\{X_1 \in \Lambda_k\}}]a = NE[B(2^j X_1 - k)1_{\{X_1 \in \Lambda_k\}}],$$

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

$$a^T E[V_k|I_k|]a = N \int_{\Lambda_k} B(2^j x - k)p(x)dx \geq N2^{-j} p_{min} \int_0^1 B^2(z)dz.$$

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

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

This gives the first inequality of the lemma.

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

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

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

$$P(|\,|I_k| - Np_k| \geq \epsilon Np_k) \quad \leq \quad 2\exp\left(-\frac{\epsilon^2 N^2 p_k^2}{2Np_k + \frac{2}{3}\epsilon Np_k}\right) \leq 2\exp\left(-\frac{\epsilon^2 Np_k}{3}\right).$$

Since

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

we have

$$P(|\,|I_k| - Np_k| \geq \epsilon N p_k \leq\leq 2\exp(-\frac{p_{min}\lambda\epsilon^2 \ln N}{3}) \leq 2N^{-\gamma}$$

for any $\lambda \geq \frac{3\gamma}{\epsilon^2 p_{min}}$. 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

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

We put

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

and

$$\beta = \epsilon E[V_k|I_k|];$$

then using the result of lemma 5, we obtain

$$\begin{aligned}
&P\left(\lambda_{\min}(V_k) \leq \frac{1-2\epsilon-\epsilon^2}{1-\epsilon^2} N^{-1} p_k^{-1} E[V_k|I_k|]\right) \\
&\leq P(|\,|I_k| - Np_k| \geq \epsilon N p_k) + P(|V_k|I_k| - E[V_k|I_k|]| \geq \epsilon E[V_k|I_k|]) \\
&\leq CN^{-\gamma}. \qquad (41)
\end{aligned}$$

Since for $\epsilon < 1/3$

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

the bound (41) implies

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

Along with the result of Proposition 4 this gives the proof. ∎

Remark : Note that the bounds shown in proposition 5 and 6 are tight. It is demonstrated by the following

Proposition 7 *Let $s > 1/p$, $0 < p, q \le \infty$. Then there exists a function f, $\|f\|_{spq} \le 1$, $c > 0$ and $\alpha > 0$ such that for any approximation method $\hat{f}(f(X_1), ..., f(X_N))$ based on the discrete observations $f(X_i)$ $i = 1, \ldots, N$*

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

where $s' = s - 1/p + 1/p'$.

Proof: The proof of this statement is analogous to that of proposition 3. For instance, in the case of random grid it suffices to note that there exists k such that $P(X_1 \in [\frac{k}{N}, \frac{(k+1/2)}{N}]) \le \frac{1}{2N}$, thus for some $\alpha > 0$

$$P(\text{card}\{i :\ X_i \in [\frac{k}{N}, \frac{(k+1/2)}{N}]\} = 0) > \alpha.$$

∎

5 Appendix

The proof of theorem 2 consists in tedious variations on Minkowski and Hölder inequalities. 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

$$\|x + y\|_p^p \le \|x\|_p^p + \|y\|_p^p \qquad \text{and} \qquad \|f + g\|_p^p \le \|f\|_p^p + \|g\|_p^p. \tag{42}$$

We need the following lemma.

Lemma 6 *For any $a > 1$ and $q > 0$, there exist C_{aq} such that for any non-negative sequence $(x_j)_{j\ge 0}$*

$$\sum_{j\ge 0} a^{jq} \left(\sum_{l \ge j} x_l \right)^q \le C_{aq} \sum_{l \ge 0} a^{lq} x_l^q. \tag{43}$$

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

$$\sum_{j\ge 0} a^{jq} \left(\sum_{l \le j} x_l \right)^q \le C_{aq} \sum_{l \ge 0} a^{lq} x_l^q. \tag{44}$$

Proof:

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

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

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

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

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

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

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

$$\begin{aligned}
\sum_{j\geq 0} a^{jq}\left(\sum_{l\leq j} x_l\right)^q &\leq \sum_{j\geq 0} a^{jq}\left(\sum_{l\leq j} x_l b^l b^{-l}\right)^q \\
&\leq \sum_{j\geq 0} a^{jq}\left(\sum_{l\leq j} x_l^q b^{lq}\right)\left(\sum_{l\leq j} b^{-lq/(q-1)}\right)^{q-1} \\
&\leq C_a \sum_{l\geq 0} x_l^q a^{lq}.
\end{aligned}$$

Proof of theorem 2 **Step 1:** There exist j_0 such that for any $j \geq 0$, $2^{jd/2}\|\beta_{j\cdot}\|_p \leq C\|\theta^{(1)}_{j-j_0,\cdot}\|_p$.

Indeed,

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

for any polynomial P of degree smaller than s; thus

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

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

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

The first inequality is a consequence of two relations: first, if we set $g(x) = f(2^{j_0}x)$ we have

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

and second

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

(cf. the inequality (29)). This bound along with Step 1 (for $j < j_0$) and (24) (for $j \geq j_0$) give us the second inequality of the step.

We obtain from Step 1, 2 and (24) :

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

Step 3 : If $p \geq 1$, $\|\alpha\|_p \leq C\|f\|_p$.

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

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

Thus

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

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

Denote $I = \{\tilde{\phi}(x) \neq 0\}$, $J = \{\phi(x) \neq 0\}$ and $g = \sum_k \alpha_k \phi_k$.

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

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

$$xy \leq p\epsilon^{-1/p}x^{1/p} + q\epsilon^{1/q}y^{1/q}$$

implies

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

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

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

which is the result.

Step 5 : If $p<1$, $\|\sum_k \alpha_k \boldsymbol{\phi}_k\|_p^p \leq C\|f\|_{spq}$.

In this case, we have

$$\|\sum_k \alpha_k \boldsymbol{\phi}_k\|_p^p \leq C\|f\|_p^p + C\sum_{jk} |\beta_{jk}|^p \|\boldsymbol{\psi}_{jk}\|_p^p,$$

and, using step 1 and 2,

$$\begin{aligned}\sum_{jk} |\beta_{jk}|^p \|\boldsymbol{\psi}_{jk}\|_p^p &\leq C\sum_{jk} |\beta_{jk}|^p 2^{jd(p/2-1)} = C\sum_j \|\beta_{j.}\|_p^p 2^{jd(p/2-1)} \\ &\leq C'\|f\|_{spq}^p \sum_j 2^{-jd} \leq C''\|f\|_{spq}^p.\end{aligned}$$

Step 6 : $\|f\|_p \leq C\nu_{spq}(f)$.

If $p>1$:

$$\|f\|_p \leq \|\sum_k \alpha_k \boldsymbol{\phi}_k\|_p + \sum_j \|\sum_k \beta_{jk}\boldsymbol{\psi}_{jk}\|_p.$$

Since $\boldsymbol{\phi}$ has compact support, we can find a partition $I_1, ..I_l$ of $\mathbf{Z}^d$ such that $\boldsymbol{\phi}_k$ and $\boldsymbol{\phi}_{k'}$ do not overlap if k and k' are in the same I_i, and

$$\|\sum_k \alpha_k \boldsymbol{\phi}_k\|_p \leq \sum_{i=1}^{l} \|\sum_{k\in I_i} \alpha_k \boldsymbol{\phi}_k\|_p = \sum_{i=1}^{l} \left(\sum_{k\in I_i} |\alpha_k|^p \|\boldsymbol{\phi}_k\|_p^p\right)^{1/p} \leq l\|\boldsymbol{\phi}\|_p\|\alpha\|_p \tag{45}$$

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

$$\|\sum_k \beta_{jk}\boldsymbol{\psi}_{jk}\|_p \leq C\|\beta_{j.}\|_p\|\boldsymbol{\psi}_{jk}\|_p = C\|\boldsymbol{\psi}\|_p\|\beta_{j.}\|_p 2^{j(d/2-d/p)}.$$

And since

$$\|\beta_{j.}\|_p \leq C\nu_{spq}(f)2^{-j(s+d/2-d/p)}.$$

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

The case $p<1$ is treated exactly the same way where equation (42) is used instead of the Minkowski inequality.

Step 7 : $\sum_j 2^{j(s-d/p)q}\|\boldsymbol{\theta}_{j.}^{(p)}\|_p^q \leq C\nu_{spq}^q(f)$.

Denote

$$Q_j f = \sum_k \beta_{jk} \boldsymbol{\psi}_{jk}$$
$$R_j f = \sum_{l \geq j} Q_l f$$
$$P_j f = \sum_k \alpha_k \boldsymbol{\phi}_k + \sum_{l \leq j} Q_l f$$

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

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

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

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

where each sum is restricted to those k' for which the function $\boldsymbol{\phi}_{k'}$ or $\boldsymbol{\psi}_{j'k'}$ is non-zero on the interval; each sum on k' contains at most K term, where K depends on the support of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$; 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 $|k' - 2^{-j+j'}k| < K$.

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

This implies that

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

and

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

where C depends here only on p and K.

Then

$$\begin{aligned} \left(\theta_{jk}^{(p)}\right)^p \quad & \leq \quad 2^{jd} \int_{|2^{-j}k - y| < 2^{-j}} |f(x) - \Pi_{jk}(x)|^p dx \\ & \leq \quad C \sup_{|2^{-j}k - x| < 2^{-j}} |P_j f(x) - \Pi_{jk}(x)|^p + C 2^{jd} \int_{|2^{-j}k - x| < 2^{-j}} |R_j(x)|^p dx \end{aligned}$$

$$
\begin{aligned}
\sum_k \left(\theta_{jk}^{(p)}\right)^p &\leq C' 2^{jd} 2^{-j(M+1)p} \|\alpha\|_p^p + C' \sum_{l<j} 2^{d(j-l)} \|\beta_{l.}\|_p^p 2^{-(j-l)p(M+1)+lpd/2} \\
&\quad + C 2^{jd} \int |R_j(x)|^p dx \\
2^{j(sp-d)} \|\theta_{j.}^{(p)}\|_p^p &\leq C' 2^{-j(M+1-s)p} \|\alpha\|_p^p + C' 2^{-jp(M+1-s)} \sum_{l<j} 2^{lp(M+1+d/2-d/p)} \|\beta_{l.}\|_p^p \\
&\quad + C 2^{jsp} \int |R_j(x)|^p dx = A_j + B_j + C_j
\end{aligned}
$$

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

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

$$
\sum_{j \geq 0} A_j^{q/p} \leq C \|\alpha\|_p^q \leq C \nu_{spq}(f)^q.
$$

Now using equation (44)

$$
\sum_{j \geq 0} B_j^{q/p} \leq C \sum_{j \geq 0} 2^{j(s+d/2-d/p)q} \|\beta_{j.}\|_p^{q/p} \leq C \nu_{spq}(f)^q.
$$

If $p \geq 1$ we obtain

$$
\sum_{j \geq 0} C_j^{q/p} = \sum_{j \geq 0} 2^{jsq} \|R_j\|_p^q \leq \sum_{j \geq 0} 2^{jsq} \left(\sum_{l \geq j} \|Q_l\|_p \right)^q \leq C \sum_{j \geq 0} 2^{jsq} \|Q_j\|_p^q
$$

(thanks to equation (43)), and if $p < 1$

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

On the other hand, because of the compactness of the support of $\boldsymbol{\psi}$:

$$
\|Q_j\|_p^p \leq C \sum_k |\beta_{jk}|^p \|\boldsymbol{\psi}_{jk}\|_p^p \leq C 2^{j(pd/2-d)} \|\beta_{j.}\|_p^p,
$$

and finally, for any $p > 0$

$$
\sum_{j \geq 0} C_j^{q/p} \leq C \sum_{j \geq 0} 2^{jsq} 2^{j(d/2-d/p)q} \|\beta_{j.}\|_p^q \leq C' \nu_{spq}(f)^q.
$$

Step 8 : End of the proof.

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

References

[1] ANTONINI, M., BARLAUD, M., MATHIEU, P. AND DAUBECHIES, I. Image coding using wavelet transforms, *IEEE Proc. Acoustics, Speech, Signal Processing*, 1991, to appear.

[2] I.DAUBECHIES, *Ten Lectures on Wavelets,* CBMS-NSF regional conference series in applied mathematics.

[3] D. DONOHO *Interpolating Wavelet Transforms* Dept. of Statistics, `ftp plaifair.stanford.edu`, 1993.

[4] D. DONOHO *Smooth wavelet decompositions with blocky coefficient kernels*, Dept. of Statistics, `ftp plaifair.stanford.edu`, 1993.

[5] D. DONOHO, I. JOHNSTONE, G. KERKYACHARIAN, D. PICARD, Density estimation by wavelet thresholding, submitted, 1993.

[6] B. JAWERTH, R.A. DEVORE, V. POPOV, Compression of wavelet decompositions. *Amer.J.Math.* **114**, 737-785.

[7] A.N. KOLMOGOROV, V.M. TIKHOMIROV, ϵ-entropy and ϵ-capacity of sets in functional spaces, *Uspekhi Mat. Nauk*, **14**, N. 2, 1959, 3-86.

[8] A. KOROSTELEV, A. TSYBAKOV *Minimax theory of image reconstruction*, Springer-Verlag, 1993.

[9] S. MALLAT, HWANG, W.L. Singularity detection and processing with wavelets. *IEEE Trans. Info Theory.* **38**, 2, 1993, 617-643.

[10] Y.MEYER , *Ondelettes et opérateurs,* Hermann, 1990

[11] Y.MEYER ,Ondelettes sur l'intervalle *Revista Matematica Iberoamericana*, **7**, 1991, 115-134.

[12] D.POLLARD, *Convergence of Stochastic Processes*, Springer-Verlag, 1984.

[13] W. SICKEL, Spline representations of functions in Besov-Triebel-Lizorkin spaces on $\mathbf{R}^n$. *Forum Math.*, 2, 451-476, 1990.

[14] G.STRANG, G.FIX, *A Fourier analysis of the finite element variational method*, Constructive aspects of functional analysis, Geymonat ed., C.I.M.E., 1978, pp 793-840.

[15] H.TRIEBEL, *Theory of Function Spaces II,* Birkhauser 1983.

[16] ANTONI ZYGMUND *Trigonometric series : volumes I and II combined* Cambridge : Cambridge University Press, 1993.