Regression in random design and warped wavelets.

Gérard Kerkyacharian[1,2], Dominique Picard[2]

January 23, 2003

Abstract

We consider the problem of estimating an unknown function f in a regression setting with random design. Instead of expanding the function on a regular wavelet basis, we expand it on the basis $\{\psi_{jk}(G), j,\ k\}$ warped with the design. This allows to perform a very stable and computable thresholding algorithm. We investigate the properties of this new basis. In particular, we prove that if the design has a property of Muckenhoupt type, this new basis has a behavior quite similar to a regular wavelet basis. This enables us to prove that the associated thresholding procedure achieves rates of convergence which have been proved to be minimax in the uniform design case.

Key words and phrases: nonparametric regression, random design, wavelet thresholding, warped wavelets, maxisets, Muckenhoupt weights.

1- Département de Mathématiques, Université de Paris X, 92001 Nanterre, France
2- Laboratoire de Probabilités et Modèles Aléatoires, Universités de Paris VI et Paris VII, 75252 Paris, France

1 Introduction

In this paper we consider the problem of estimating an unknown function f in a regression setting with random design. We will consider the problem in the framework of wavelet thresholding. Of course, if the design is regular, the procedures are now standard (see Donoho, Johnstone (1995) [16], Donoho, Johnstone, Kerkyacharian, Picard (1994) [18]). In the case of irregular design, various attempts to solve this problem have been studied : see for instance the interpolation methods of Hall and Turlach (1997) [25], Kovac and Silverman (2000) [30], the binning method of Antoniadis, Gregoire and Vial (1997) [2], the transformation method of Cai and Brown (1998) [6], the weighted wavelet transform of Foster [22] , the isometric method of Sardy et al (1999) [34], the penalisation method of Antoniadis and Fan (2001) [3], the specific construction of ad hoc wavelets of Delouille et al (2001) [11]...

Our aim here is to stay as close as possible to a standard WaveShrink. Doing this, means that we accept to consider instead of the wavelet expansion of the function f its expansion on the basis $\{\psi_{jk}(G), j,\ k\}$, where G is the distribution function of the design (or its estimation, when it is not known). This obviously creates some difficulties since $\{\psi_{jk}(G), j,\ k\}$ is no longer an orthonormal basis, but has also clear advantages : among them, let us emphasize the fact that our procedure is computationaly very simple. Compared for instance, to the transformation method of Cai Brown, which considers this point of view for the finest scale, but then project on the regular wavelet basis, the calculation are more direct, but overall, doing this, we stabilize the variance of the estimated coefficients which avoids using a thresholding rule that needs to be re-calculated (or estimated) for each coefficient.

Adopting such a point of view obviously pushes the difficulty toward the analytic part since we need to study the behavior of the family of 'warped wavelet basis ' $\{\psi_{jk}(G), j,\ k\}$. Of course the properties of this basis truely depends on the warping factor G. Obvioulsy, if G is uniform, then, $\{\psi_{jk}(G), j,\ k\}$ is a regular wavelet basis. We will prove that under a condition on G, it is 'almost' the case e.g. for statistical purposes, we can use the warped basis as a standard one. As expected, this condition is quantifying in a way the departure from the uniform distribution and is associated to Muckenhoupt weights.

The Muckenhoupt weights have been introduced in [31] (see also [23] and[8]) and widely used afterwards in the context of Calderon-Zygmund theory.

Our results will prove that under conditions mixing the regularity of f and the fact that G is not degenerating, we find the rate of convergence of the procedures. For instance, in the case where the density g of G is bounded above and below, we found exactly the same behavior as in the regular design except that here the conditions of regularity are formulated on the function $f \circ G^{-1}$. In a way, this is strongly linked with the results about the equivalences of experiments (see, for instance Brown, Cai, Low and Zhang (2002) [5]). The assumption of boundedness above and below for g will not be required in full generality. In this case, the regularity of f will be expressed in terms of 'warped' Besov spaces.

2 Model, Warped bases, Estimation procedures

2.1 Regression with random design

Let us consider the following model: We observe $Y_1, \dots, Y_n$ n independent variables with

$$Y_i = f(X_i) + \varepsilon_i \tag{1}$$

where X_i and ε_i are independent random variables, ε_i has a known distribution with density g_0. The X_i's are observed, ε_i's are not. Our aim is to estimate the function f. To simplify, in the sequel, we will assume that the ε_i's are normal variables with zero mean and variance σ^2. σ^2 will be assumed to be known or replaced by an estimator. For sake of simplicity, in the sequel, we assume $\sigma^2 = 1$. The X_i's have a density g which may be known or unknown. g is assumed to be compactly supported on the interval $\mathcal{I} = [a, b]$, as well as f.

2.2 Wavelet shrinkage

Wavelet shrinkage is now a well established statistical procedure used for nonparametric estimation. A generic wavelet estimator of an unknown function f is written as

$$\hat{f} = \sum_{\{I=(j,k), -1 \leq j \leq J(n)\}} \hat{\beta}_I \, \psi_I \, \mathbb{I}_{\{|\hat{\beta}_I| \geq \lambda\}} \tag{2}$$

where $\{\psi_{j,k}, j \geq -1, \ k \in \mathbb{Z}\}$ is a compactly supported wavelet basis (we recall that : $\psi_{-1,k} = \phi_{0k}$) $\hat{\beta}_I$ is an estimator of the true wavelet coefficients

$$\beta_I = \int f \, \psi_I \tag{3}$$

Note that the procedure (2) is non-linear since only statistically significant coefficients (e.g. $|\hat{\beta}_I| \geq \lambda$) are kept. Here λ, is a threshold parameter which depends on the problem at hand. This procedure has been investigated in many cases. See for the regression with equispaced design Donoho Johnstone (1996) [17], where this estimator has been proposed with the following estimators of the wavelet coefficients

$$\hat{\beta}_I = \sum_{i=1}^{n} Y_i \, \psi_I(i/n) \tag{4}$$

In the case, with non equispaced but still fixed design, many adaptations of this first estimator have been provided. Let us only mention here Cai and Brown (1998) [6] and Hall and Turlach (1997) [25], which are the closest to the forthcoming discussion.

2.3 Warping the basis

The main idea developed in this paper is that instead of expanding the function on a wavelet basis and obtaining as a consequence an estimator which is adapted to the basis but no so well adapted to the statistical problem, we are going to adopt a different strategy : We will warp the wavelet in such a way that in this new basis, the estimates of the coefficients will be more natural.

Let us devote the following lines to explain this idea : In the case where the design is fixed and equispaced, the estimators of the coefficients β_I appear to be in a *natural* way $\hat{\beta}_I$ given in (4). If we follow this idea in the random design case and suppose for a while that

$$G(x) = \int_a^x g(u)du$$

is a known function, continuous and strictly monotone from $[a, b]$ to $[0, 1]$, then

$$\hat{\beta_I}^* = \frac{1}{n} \sum_{i=1}^{n} \psi_I(G(X_i))Y_i \tag{5}$$

is a natural extension of (4).

We have :

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_I^*) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(\psi_{j,k}(G(X_i))(f(X_i) + \epsilon_i)) = \mathbb{E}(\psi_I(G(X))f(X)) \\
&= \int_a^b \psi_I(G(x))f(x)g(x)dx = \int_a^b \psi_I(y)f(G^{-1}(y))dy := \beta_I
\end{aligned}$$

where β_I is now the coefficient of the new function $f(G^{-1}(y))$ in the wavelet basis $\{\psi_I, j \geq -1, \ k \in \mathbb{Z}\}$. This can be rewritten as

$$f(G^{-1}(y)) = \sum_I \beta_I \psi_I(y)$$

or,

$$f(x) = \sum_I \beta_I \psi_I(G(x)) \tag{6}$$

and we can associate to this decomposition the following estimate :

$$\hat{f}^*(x) = \sum_{j=-1}^{J} \sum_{k \in \mathbb{Z}} \hat{\beta}_I^* I\{|\hat{\beta}_I^*| \geq \kappa t_n\} \psi_I(G(x)) \tag{7}$$

with,

$$t_n = (\frac{\log n}{n})^{1/2},\ 2^J \sim t_n^{-1}. \tag{8}$$

Obviously, (6) considers an expansion of f in a new basis :

$$\{\psi_{j,k}(G), j \geq -1,\ k \in \mathbb{Z}\}.$$

Notice here, that the formula (5) is also a key tool in Cai and Brown [6], since it is the estimator which is used at the highest level J' ($2^{J'} = n$)), and then the wavelet coefficients are deduced by considering the projection of $\phi_{J',k}(G)$ over each ψ_{jk}. In the present case, the calculation is very simple since starting from the same first step, we just need to use the classical pyramidal algorithm, and as a consequence any wavelet software can be used to perform the estimation.

Then one might ask what is to be done if G is not known, which is the most frequent case. The answer is simple : Let

$$\hat{G}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\}$$

be the empirical distribution function of the X_i's. Let us define the new empirical wavelet coefficients :

$$\hat{\beta}'_{jk} = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(\hat{G}_n(X_i))Y_i.$$

And let us now consider the estimator :

$$\hat{f}'' = \sum_{j=-1}^{J} \sum_{k \in \mathbb{Z}} \hat{\beta}'_{jk} I\{|\hat{\beta}'_{jk}| \geq \kappa t_n\} \psi_{jk}(\hat{G}_n(x)) \tag{9}$$

With again:

$$t_n = (\frac{\log n}{n})^{1/2},\ 2^J \sim t_n^{-1}$$

The difference between the two estimators is the substitution of the empirical distribution function. Notice however that this substitution makes the computation even easier.

The only calculation steps are :

1. Sort the X_i's,
2. Change the numbering in such a way that X_i has rank i,
3. Calculate the highest level $alpha$-coefficients using the formula :

$$\hat{\alpha}_{J'k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{J'k}(i/n)Y_i, \quad (2^{J'} = n)$$

4. Calculate the wavelet coefficients using the classical pyramidal algorithm
5. Perform a thresholding algorithm giving rise to $\tilde{\beta}_{jk}$ coefficients,

6. Reconstruct the estimator, using again the standard backward pyramidal algorithm, and obtain

$$\hat{f}'' = \sum_{j=-1}^{J} \sum_{k \in \mathbb{Z}} \tilde{\beta}_{jk} \psi_{jk}(\hat{G}_n(x))$$

which is a function especially easy to draw.

The aim of this paper will be to study the performances of the procedures $\hat{f}^*$ and $\hat{f}$" under conditions of regularity which will take into account the regularity of the function f as well as the concentration properties of the underlined design. It is interesting, at this step to notice that there is a slight difference here with the standard setting in the fact that we set $2^J \sim t_n^{-1}$ whereas, usually, we set $2^L \sim t_n^{-2}$, for the greatest level. This will be commented below. It is also worthwhile to notice that for technical reasons, the results will be proved not exactly for $\hat{f}$", but for a procedure which is a bit less direct from the computation point of view (but still very simple): Instead of estimating G over the whole sample, we divide the sample into 2 (independent) parts and use the first part (for i in $\{1, \ldots, [n/2]\}$) giving rise to $\hat{G}_{[n/2]}(x)$. Then we estimate the wavelet coefficients using the other part of the data :

$$\hat{\beta}_{jk}^{@} = \frac{2}{n} \sum_{i=[n/2]+1}^{n} \psi_{jk}(\hat{G}_{[n/2]}(X_i))Y_i.$$

And let us now consider the estimator :

$$\hat{f}^{@} = \sum_{j=-1}^{J} \sum_{k \in \mathbb{Z}} \hat{\beta}_{jk}^{@} I\{|\hat{\beta}_{jk}^{@}| \geq \kappa t_n\} \psi_{jk}(\hat{G}_{[n/2]}(x)) \tag{10}$$

3 Muckenhoupt weight and warped bases

3.1 Muckenhoupt Weight

Let us first recall the following notion :

Definition 1. (Muckenhoupt weights) *For $1 < p < \infty,\ 1/p + 1/q = 1$, a measurable function $\omega \geq 0$ belongs to the Muckenhoupt class A_p if there exists $0 < C < \infty$ such that for any interval I included in $\mathbb{R}$,*

$$\left(\frac{1}{|I|}\int_I \omega(x)dx\right)^{1/p} \left(\frac{1}{|I|}\int_I \omega(x)^{-\frac{q}{p}} dx\right)^{1/q} \leq C$$

For $p = 1$, the definition is modified in the following way : $\omega \geq 0$ belongs to the Muckenhoupt A_1 class if there exists $0 < C < \infty$ such that,

$$\omega^*(x) \leq C\omega(x) \ \ a.e.$$

where $\omega^*(x)$ is the Hardy-Littlewood maximal function.
For $p = \infty$, we set

$$A_\infty = \cup_{p \geq 1} A_p.$$

Definition 2. (Maximal function)*If $\mathcal{B}$ is the set of all the intervals of $\mathbb{R}$ and if f is a measurable function, then the Hardy-Littlewood maximal function associated to f is*

$$f^*(x) = \sup_{I \in \mathcal{B},\ x \in I} (\frac{1}{|I|}\int_I |f(x)|dx.)$$

The concept of Muckenhoupt weight has been introduced in [31] (see also [23] and[8]) and widely used afterwards in the context of Calderon-Zygmund theory. It is easy to observe that the Muckenhoupt spaces are increasing. Many functions belong to one of these classes. Of course, if ω is bounded above and below, it belongs to A_∞, but ω can also approach zero. For instance $w(x) = |x|^a$ belongs to A_p for $-1 < a < p - 1$.

We see on the definition that this property somehow quantifies the way how ω is closed to a uniform weight, where the function and its inverse are evenly charging each interval. Some of the important properties of these functional classes will be recalled in Appendix I.

In the sequel we will assume the following condition :

$(\mathcal{H}_p) \quad y \mapsto \omega(y) = \frac{1}{g(G^{-1}(y))}$ *is a Muckenhoupt weight belonging to* $A_p([a, b])$

This will be proved to be equivalent (see Proposition 9, Appendix I) to :

There exists C, such that for all interval $I \subset [a, b]$,

$$(\frac{1}{|I|}\int_I g(x)^q dx)^{1/q} \leq C\frac{1}{|I|}\int_I g(x)dx \quad 1/p + 1/q = 1.$$

Again, these conditions are obviously true when the design g is uniform or uniformly bounded above and below. More generally they obviously quantify the usual assumption that the design gives enough mass to any interval.

3.2 Properties of the warped wavelet basis

As is shown in formula (6), our construction builds on the new 'basis' $\{\psi_{jk}(G(.)),\ j \geq -1,\ k \in \mathbb{Z}\}$.

Let us consider, the following $\mathbb{L}_p$ risk :

$$\mathbb{E}\|\hat{f} - f\|_p^p = \mathbb{E}\int_{[a,b]} |\hat{f}(x) - f(x)|^p dx.$$

As is proved in [27], generally speaking (but with a mathematical sense that will be detailed later) thresholding methods are working especially well in $\mathbb{L}_p$-risk, not only if one thresholds the wavelet coefficients but also if one thresholds the coefficients associated to a 'well adapted basis' $\{e_i,\ i \in \mathbb{N}\}$. In this context, being well adapted precisely means the two following properties :

Shrinkage (or unconditional) property: There exists an absolute constant K such that if $|\theta_i| \leq |\theta_i'|$ for all i, then

$$\|\sum_i \theta_i e_i\|_p \leq K\|\sum_i \theta_i' e_i\|_p. \tag{11}$$

Temlyakov property: There exist c_p and C_p such that for any finite set of integers F we have :

$$c_p \int \sum_{i\in F} |e_i|^p \leq \int (\sum_{i\in F} |e_i|^2)^{\frac{p}{2}} \leq C_p \int \sum_{i\in F} |e_i|^p. \tag{12}$$

Let us now state the following theorem, borrowed from [29]:

Let $1 < p < \infty,\ \ \omega \in A_p,$ and $\psi_{j,k}$ be a compactly supported wavelet. Let T and S be two real measurable functions defined on $\mathbb{R}$ such that

$$S(T(x)) = x,\ \ a.e.;\quad T(S(x)) = x,\ \ a.e. \tag{13}$$

$$\forall h \geq 0,\ \ \text{measurable function,} \quad \int_{\mathbb{R}} h(T(x))dx = \int_{\mathbb{R}} h(y)\omega(y)dy \tag{14}$$

Theorem 1. *Under the conditions (13) and (14), the family* $\{\psi_{jk}(T(.)),\ j \geq -1,\ k \in \mathbb{Z}\}$ *satisfies the properties of shrinkage and* p*-Temlyakov.*

Typically, these conditions are realized if we take $T(x) = G(x)$, defined on $]a, b[$ and if $S = G^{-1}$ is a locally lipschitz function on $]0, 1[$. It is well known then that if S is almost everywhere differentiable, then the following change of variable formula is true (cf [24]) :

$$\forall h \geq 0,\ \ \text{measurable function,} \quad \int_{\mathbb{R}} h(x)\omega(x)dx = \int_{\mathbb{R}} h(T(y))dy$$

where ω is the Jacobian of S. i.e. $\omega(y) = \frac{1}{g(G^{-1}(y))}$. Then, we see that our assumption $(\mathcal{H}_p)$ precisely states that (14) is realized, with $\omega \in\ A_p$.

3.3 Weighted Besov spaces

It is natural in this context, if we want to obtain a global rate of convergence in terms of $\mathbb{L}_p$ risk, to impose regularity conditions taking into account the fact that the design is non equispaced. This is what we are going to express in this section. Let us define, for every measurable function

$$\Delta_h f(x) = f(x+h) - f(x)$$

Then, recursively, $\Delta_h^2 f(x) = \Delta_h(\Delta_h f)(x)$ and identically, for $N \in \mathbb{N}_*$, $\Delta_h^N f(x)$.

Let

$$\rho^N(t, f, \omega, p) = \sup_{|h|\leq t} (\int |\Delta_h^N f(u)|^p \omega(u) du)^{1/p}$$

with the usual modification for $p = \infty$. and let us define the following modified Besov space :

$$B_{s,p,q}(\omega) = \{f: \ (\int_0^1 (\frac{(\rho^N(t, f, \omega, p)}{t^s})^q \frac{dt}{t})^{1/q} < \infty\}.$$

The only difference with the usual Besov spaces is the fact that the modulus of continuity ρ^N is calculated with the weight ω that possibly leads to a space inhomogeneity.

One of the major advantages of Besov spaces is that they can be expressed in terms of wavelet coefficients. In fact, if ω is a reasonable weight, one can show that it is still the case. The following proposition proves the direct sense, which is the useful one in the context of this paper.

Proposition 1. *For $1 \leq p \leq \infty$, let us suppose that ω is in A_p, and let us put for every interval $I \subset \mathbb{R}$*

$$\omega(I) = \int_I \omega(x) dx$$

Then, if ψ is a real compactly supported wavelet, such that

$$\int \psi(x) x^k dx = 0, \ k = 0, \dots, N-1$$

then for

$$f = \sum_{j,k} \beta_{jk} \psi_{j,k}, \quad I_{j,k} = [\frac{k}{2^j}, \frac{k+1}{2^j}]$$

$$(\int_0^1 (\frac{(\rho^N(t, f, \omega, p)}{t^s})^q \frac{dt}{t})^{1/q} < \infty \Longrightarrow [\sum_j (2^{js} 2^{j/2} (\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p \omega(I_{j,k}))^{1/p})^q]^{1/q} < \infty$$

with the usual modification if $q = \infty$.

This proposition is proved in Appendix I (see Theorem 6). We will use the following corollary, which will be necessary since we are not expanding the function in the wavelet basis but in the warped basis. Let us define :

$$\Delta_h(G) f(x) = f(G^{-1}[G(x) + h]) - f(x).$$

As above, recursively, $\Delta_h^2(G) f(x) = \Delta_h(G)(\Delta_h(G) f)(x)$ and identically, for $N \in \mathbb{N}_*$, $\Delta_h^N(G) f(x)$, and again,

$$\tilde{\rho}^N(t, f, G, p) = \sup_{|h|\leq t} (\int |\Delta_h^N(G) f(u)|^p du)^{1/p}.$$

Notice that $\tilde{\rho}^N$ is defined with the standard uniform weight, the 'spatial inhomogeneity' now lies in the definition of $\Delta(G)$. Let us define the following spaces :

$$B_{s,p,q}^G = \{f: \ (\int_0^1 (\frac{(\tilde{\rho}^N(t, f, G, p)}{t^s})^q \frac{dt}{t})^{1/q} < \infty.\}$$

Notice that in the particular case $p = q = \infty$, it is easy to prove that :

$$f \in B_{s,\infty,\infty}^G \iff f \circ G^{-1} \in B_{s,\infty,\infty}.$$

The following corollary concerns the representation of spaces $B_{s,p,q}^G$ in term of coefficients in the expansion using the warped basis.

Corollary 1. *Under the conditions of Proposition 1, for*

$$f = \sum_{j,k} \beta_{jk}\psi_{j,k}(G),\ (i.e.\ \sum_{j,k} \beta_{jk} = \int [f \circ G]\psi_{j,k})$$

we have

$$(\int_0^1 (\frac{(\tilde{\rho}^N(t,f,G,p)}{t^s})^q \frac{dt}{t})^{1/q} < \infty \Longrightarrow [\sum_j (2^{js}2^{j/2}(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p \omega(I_{j,k}))^{1/p})^q]^{1/q} < \infty$$

with the usual modification if $q = \infty$.

The corollary is an obvious consequence of the previous proposition applied to $f \circ G^{-1}$ just observing that $\tilde{\rho}^N(t,f,G,p) = \rho^N(t, f\circ G^{-1}, \omega, p)$.

4 Performances of the estimation procedures

4.1 Maxisets

The properties of the procedures $\hat{f}^*$ or $\hat{f}^@$ will be expressed in two different ways. The first one is commonly used: It consists in proving that we obtain minimax rates of convergence for a large variety of loss functions and a wide class of regularity spaces (Theorem (2) and Proposition (2)).

The second way (Theorem (4) and Proposition (5)) consists in determining the maxiset of the procedures. Let us quickly recall this notion. For a sequence of models $\mathcal{E}_n = \{P^n_\theta, \theta \in \Theta\}$, where the P^n_θ's are probability distributions, Θ is the set of parameters, we consider a sequence of estimates $\hat{q}_n$ of a quantity $q(\theta)$, a loss function $\rho(\hat{q}_n, q(\theta))$ and a rate of convergence α_n tending to 0.

Definition 3. *We define the* **maxiset** *associated with the sequence* $\hat{q}_n$*, the loss function* ρ*, the rate* α_n *and the constant* T *as the following set:*

$$Max(\hat{q}_n, \rho, \alpha_n)(T) = \{\theta \in \Theta,\ \sup_n \mathbb{E}^n_\theta \rho(\hat{q}_n, q(\theta))(\alpha_n)^{-1} \leq T\}$$

This way of measuring the performances of procedures has been particularly sucessful in the nonparametric framework (see for instance [9], [27], [33]). It has the advantage of giving less arbitrary and pessimistic comparisons of procedures. It also has the advantage of being very powerful at giving as subproducts the comparisons of traditional types as quoted above; Here we will obtain the maxisets for the procedure $\hat{f}^*$ (theorem 4). From this result, we deduce the rates of convergence of $\hat{f}^*$ over a large amount of regularity classes by proving their inclusions into the maxiset. We will then deduce the results for the more general procedure $\hat{f}^@$ taking advantage of the proximity of $\hat{f}^*$ and $\hat{f}^@$, when n is large (theorem 2).

4.2 Properties of the procedures $\hat{f}^*$ and $\hat{f}^@$:

Theorem 2. *Assume that we observe the model (1), with the unknown function* g *satisfying the conditions* $(\mathcal{H}_p)$*, where* $p > 1$*,* $\pi \geq p$ *are given real numbers. Let us suppose that* f *is bounded and let us take* $0 < r \leq \frac{p}{2s+1}$*, the two estimators* $\hat{f}^*$ *and* $\hat{f}^@$ *defined in (7), (10), have the following rates of convergence :*

$$\mathbb{E}\|\hat{f}^* - f\|_p^p \leq C[n/\log n]^{-\alpha} \quad if\ f \in B^G_{s,\pi,r}, \quad s \geq 1/2 \tag{15}$$

$$\mathbb{E}\|\hat{f}^@ - f\|_p^p \leq C[n/\log n]^{-\alpha} if\ f \in B^G_{s,\infty,\infty}, \quad s > 1/2. \tag{16}$$

where

$$\alpha = \frac{sp}{1+2s} \tag{17}$$

Remarks and comments:
The rates of convergence obtained here for $\hat{f}^*$ corresponds to the rates which were proved to be minimax in a uniform design, up to logarithmic factors. Notice however, that we don't observe the elbow, and the division between a sparse and a dense zone as was the case for the uniform design. This is essentially due to the fact that the Sobolev embeddings which are true with regular Besov spaces, no longer occur in the context of weighted spaces.
The results on $\hat{f}^{@}$ are almost the same as for $\hat{f}^*$, except that we need uniform conditions on the wavelet coefficients.
The limitation $s \geq 1/2$ is standard in the regression setting. Let us observe that this restriction appear in our choice of J. In standard thresholding (standard denoising or density estimation for instance) one usually set the highest level L so that $2^L \sim n/\log n$. Here we have to stop much sooner ($2^J \sim (n/\log n)^{1/2}$). This is necessary to obtain, in particular the exponential inequalities of Proposition 3.
We can also want to express the results in terms of 'regular' Besov spaces. This can be done if we are ready to more restrictive assumptions on the underlined design (e.g. its density is bounded above and below). We have the following proposition :

Proposition 2. *Assume that we observe the model (1), with the unknown function g satisfying $0 < m \leq g \leq M < \infty$, for $p > 1, \pi \geq 1$ given real numbers, $0 < r \leq \frac{p}{2s+1}$, the two estimators $\hat{f}^*$ and $\hat{f}^{@}$ defined in (7), (10), have the following rates of convergence :*

$$\mathbb{E}\|\hat{f}^* - f\|_p^p \leq C[n/\log n]^{-\alpha(s)} \quad if\ f \circ G^{-1} \in B_{s,\pi,r}, \quad s \geq 1/2 \tag{18}$$

$$\mathbb{E}\|\hat{f}^{@} - f\|_p^p \leq C[n/\log n]^{-\alpha} \quad if\ f \circ G^{-1} \in B_{s,\infty,\infty}, \quad s > 1/2. \tag{19}$$

where

$$\alpha(s) = \alpha = \frac{sp}{1+2s},\ for\ s \geq \frac{p-\pi}{2\pi} \tag{20}$$

$$= \frac{(s - 1/\pi + 1/p)p}{1 + 2(s - 1/\pi)},\ otherwise \tag{21}$$

This proposition proves that, under the condition that g is bounded above and below (case already investigated in Stone 1982 [36]), we observe exactly the same behavior as in the regular setting with the only exception that the regularity is stated with the function $f \circ G^{-1}$ instead of f.

The proof of theorem 2 will be given in the next section. It will be decomposed into the following items : First we investigate the behavior of $\hat{f}^*$. The first step takes advantage of the following theorem 3 borrowed from [27].
The aim of theorem 3 is to determine the 'maxiset' of the thresholding method for a completely general basis. We refer to [27] for its proof. This theorem will be applied to obtain Theorem 4, which is determining the maxiset of the particular procedure $\hat{f}^*$. The proof of Theorem 4 is given in the first part of the next section.
The second step (proving (15)), consists then in proving that the space $B^G_{s,\pi,r}$ is included into the maxiset $Max(q)$ with q properly chosen to obtain the prescribed rate of convergence ($\alpha = (p-q)/2$). This is done in Appendix II.

When the result is established for $\hat{f}^*$, we just need to transfer it to $\hat{f}^{@}$ by proving that the two estimators are reasonably close. This is done in the following section, part 3.

We need now to introduce the following notations : Let $\{\ e_{jk}, j \geq -1, k \in \mathbb{N}\}$ be a set of functions in $L^p(\mathbb{R})$, ν will denote the measure such that for $j \in \mathbb{N},\ k \in \mathbb{Z}$,

$$\nu\{(j,k)\} = \|e_{jk}\|_p^p,$$

and we define the following functions spaces :

$$l_{q,\infty}(\nu) = \left\{ f = \sum \beta_{jk} e_{jk},\ \sup_{\lambda > 0} \lambda^q \nu\{(j,k)/\ |\beta_{jk}| > \lambda\} < \infty \right\}.$$

Theorem 3. *Let $p > 1, 0 < q < p$. Suppose that $\{ e_{jk}, j \geq -1, k \in \mathbb{N}\}$ satisfies the shrinkage and p-Temlyakov's properties (11) and (12). Suppose that $c(n)$ is a sequence of real numbers tending to zero and Λ_n is a set of pairs (j,k) such that :*

$$\sup_n \nu\{\Lambda_n\} c(n)^p < \infty. \tag{22}$$

We suppose in addition that, for any pair (j,k) in Λ_n we have an estimator $\hat{\beta_{jk}}$, such that, the two following inequalities hold :

$$\mathbb{E}|\hat{\beta_{jk}} - \beta_{jk}|^{2p} \leq Cc(n)^{2p} \tag{23}$$

$$P\Big(|\hat{\beta_{jk}} - \beta_{jk}| \geq \kappa c(n)/2\Big) \leq Cc(n)^{2p} \wedge c(n)^4. \tag{24}$$

Then, the thresholding estimator (25)

$$\hat{f} = \sum_{(j,k)\in\Lambda_n} \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq \kappa c(n)\} e_{jk} \tag{25}$$

is such that there exists $C > 0$

$$\forall n \in \mathbb{N}^*, \qquad \mathbb{E}_f^n \|\hat{f}_n - f\|_p^p \leq Cc(n)^{p-q}.$$

if and only if,

$$\begin{aligned} f &\in l_{q,\infty}(\nu), \quad and, \\ \sup_n c(n)^{q-p} \quad &\| \quad f - \sum_{(j,k)\in\Lambda_n} \beta_{jk} e_{jk}\|_p^p < \infty. \end{aligned}$$

Remarks and comments:

1. Rephrasing the theorem is saying that the maxiset of the procedure $\hat{f}$,

$$\begin{aligned} Max(q) &= \{f,\ \mathbb{E}\|\hat{f}^* - f\|_p^p \, (c(n))^{(q-p)} < \infty\} \\ &= l_{q,\infty}(\nu) \cap \{f = \sum \beta_{jk} e_{jk}\ \sup_n c(n)^{q-p} \|f - \sum_{(j,k)\in\Lambda_n} \beta_{jk} e_{jk}\|_p^p < \infty\}. \end{aligned}$$

2. This theorem will be applied to obtain the following theorem 4 with

$$e_{jk} = \psi_{jk} \circ G, \quad \hat{f} = \hat{f}^*, \quad \Lambda_n = \{(j,k);\ |k| \leq D2^j,\ -1 \leq j \leq J\}$$

   The basis satisfies the shrinkage and Temlyakov properties because of condition $(\mathcal{H}_p)$ and the Theorem (1).

3. The estimators of the coefficients will be taken to be $\hat{\beta}^*_{jk}$ It will be proved in Proposition 3 (see section 5.1) that inequalities (23), (24) hold with

$$c(n) = t_n = \left(\frac{\log n}{n}\right)^{1/2} \text{ and } 2^J \sim c(n)^{-1} \tag{26}$$

4. It will be proved in Appendix I (see Theorem 5) that condition $(\mathcal{H}_p)$ implies that $\nu\{(j,k)\} = \|e_{jk}\|_p^p = \|\psi_{jk}\|^p_{\mathbb{L}_p(\omega)} \sim 2^{jp/2}\omega(I_{jk})$. Then the condition (22) is verified as soon as we have :

$$\sum_{j=-1}^{J} 2^{j(p/2)} \sum_k \omega(I_{jk}) c(n)^p < \infty \tag{27}$$

   This is obviously true as soon as ω belongs to $\mathbb{L}_1$ and $2^J c(n)^2$ is bounded, which is the case under our assumptions.

Hence, as a consequence, we obtain the following theorem.

Theorem 4. *Let $p > 1$, $0 < q < p$. Under the condition $(\mathcal{H}_p)$, the maxiset of the estimator $\hat{f}^*$*

$$Max(q) = \{f,\ \mathbb{E}\|\hat{f}^* - f\|_p^p \left(\frac{\log n}{n}\right)^{(q-p)/2} < \infty\} \tag{28}$$

can be expressed in the following form if $\nu\{(j,k)\} = 2^{jp/2}\omega(I_{jk})$,

$$Max(q) = \left\{ f = \sum_I \beta_I \psi_I \circ G,\ \sup_{\lambda>0} \lambda^q \nu\{(j,k)/\ |\beta_{jk}| > \lambda\} < \infty,\ \sup_{l\geq 0} \| \sum_{j\geq l,\ k} \beta_{jk}\psi_{jk} \circ G\|_p^p 2^{l(p-q)} < \infty \right\} \tag{29}$$

5 Proofs

In this section we will first prove Theorem 4. As a consequence of Theorem 3 and the remarks above, we only need to prove inequalities (23) and (24). This is done in subsection 5.1. Then, we prove inequalities (15) and (18) by inclusion, in subsection 5.2, which is very short since we rejected all technicalities about Besov classes in the Appendix I. Subsection (5.3), is devoted to prove that $\hat{f}^@$ and $\hat{f}^*$ are close enough, at least under regularity conditions on the unknown function. This will be done in two steps reflecting the fact that the difference between $\hat{f}^@$ and $\hat{f}^*$ is decomposable into two parts with different level of difficulties : 1-replacing $\hat{\beta}^*$ by $\hat{\beta}^@$. 2- replacing G by $\hat{G}_{n/2}$ in $\psi_I(G)$.

5.1 Proof of Theorem 4

We begin with the following proposition :

Proposition 3. *if f is bounded, there exist constants $C_p,\ C'_p,$ and for any $\gamma > 0$ there exists a constant κ_0, with :*

$$\mathbb{E}(|\hat{\beta_{jk}} - \beta_{jk}|^p) \leq C_p \frac{(1 + \|f\|_\infty^p)}{n^{p/2}},\ for\ 2^j \leq n \tag{30}$$

$$P(|\hat{\beta_{jk}} - \beta_{jk}| > \kappa\sqrt{\frac{\log n}{n}}) \leq C'_p n^{-\gamma p}\ for\ \kappa \geq \kappa_0, 2^j \leq (\frac{n}{\log n})^{1/2} \tag{31}$$

Remark : Let us observe that (35) implies (23), and choosing γ large enough will obviously ensure (24).

5.1.1 Proof of the Proposition

1. Using Rosenthal inequality (see [26] p. 241), for $p \geq 2$,

$$\mathbb{E}(|\hat{\beta_{jk}} - \beta_{jk}|^p) \leq C(\frac{\mathbb{E}|\psi_{j,k}(G(X))Y|^p}{n^{p-1}} + \frac{(\mathbb{E}|\psi_{j,k}(G(X))Y|^2)^p/2}{n^{p/2}})$$

$$\mathbb{E}|\psi_{j,k}(G(X))Y|^p = \mathbb{E}|\psi_{j,k}(G(X))(f(X)+\epsilon)|^p \leq 2^{p-1}(\mathbb{E}|\psi_{j,k}(G(X))f(X)|^p + \mathbb{E}|\psi_{j,k}(G(X))\epsilon|^p)$$

But,

$$\begin{aligned}\mathbb{E}|\psi_{j,k}(G(X))f(X)|^p &= \int |\psi_{j,k}(G(x))f(x)|^p g(x)dx \leq \|f\|_\infty^p \int |\psi_{j,k}(G(x))|^p g(x)dx \\ &= \|f\|_\infty^p \int |\psi_{j,k}(u)|^p du \leq \|f\|_\infty^p 2^{j\frac{p-2}{2}} \int |\psi_{j,k}(u)|^2 du = \|f\|_\infty^p 2^{j(\frac{p}{2}-1)}\end{aligned}$$

Furthermore,

$$\mathbb{E}|\psi_{j,k}(G(X))\epsilon|^p = \mathbb{E}|\psi_{j,k}(G(X))|^p\mathbb{E}|\epsilon|^p = \mathbb{E}|\epsilon|^p \int |\psi_{j,k}(G(x))|^p g(x)dx \leq C_p 2^{j(\frac{p}{2}-1)}$$

So :

$$\mathbb{E}|\psi_{j,k}(G(X))Y|^p \leq C_p(1+\|f\|_\infty^p)2^{j(\frac{p}{2}-1)}$$

So : if $p \geq 2$,

$$\mathbb{E}(|\hat{\beta_{jk}} - \beta_{jk}|^p) \leq C_p(\frac{(1+\|f\|_\infty^p)2^{j(\frac{p}{2}-1)}}{n^{p-1}} + \frac{(1+\|f\|_\infty^p)}{n^{p/2}})$$

So obviously if $2^j \leq n$ (and a fortiori if $2^j \leq \sqrt{\frac{n}{\log n}}$) we have

$$\mathbb{E}(|\hat{\beta_{jk}} - \beta_{jk}|^p) \leq C_p\frac{(1+\|f\|_\infty^p)}{n^{p/2}})$$

The same is obviously true for $0 < p < 2$, using Jensen inequality :

$$\mathbb{E}(|\hat{\beta_{jk}} - \beta_{jk}|^p) \leq \frac{(\mathbb{E}|\psi_{j,k}(G(X))Y|^2)^p/2}{n^{p/2}}$$

2.

$$\begin{aligned}
\frac{1}{n} \quad \textstyle\sum_{i=1}^n \quad & \psi_{j,k}(G(X_i))(f(X_i)+\epsilon_i) - \beta_{jk} \\
= \quad & (\frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))(f(X_i) - \mathbb{E}(\psi_{j,k}(G(X))f(X))) + \frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))\epsilon_i
\end{aligned}$$

Hence

$$\begin{aligned}
P(|\hat{\beta_{jk}} - \beta_{jk}| \quad &> \quad \kappa\sqrt{\frac{\log n}{n}}) \\
&\leq \quad P(|\frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))(f(X_i) - \mathbb{E}(\psi_{j,k}(G(X))f(X))| > \kappa/2\sqrt{\frac{\log n}{n}}) \\
&+ \quad P(|\frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))\epsilon_i| > \kappa/2\sqrt{\frac{\log n}{n}})
\end{aligned}$$

- Let us observe that conditionally to $(X_1, ..., X_n) = (x_1, ..., x_n)$ we have

$$\frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))\epsilon_i \sim N(0, \frac{1}{n^2}\sum_{i=1}^n \psi_{j,k}^2(G(x_i)))$$

So

$$\begin{aligned}
P(|\frac{1}{n}\sum_{i=1}^n \psi_{j,k}(G(X_i))\epsilon_i| \quad &> \quad \kappa/2\sqrt{\frac{\log n}{n}}) \\
&\leq \quad \int ... \int \exp - \frac{\kappa \log n}{\frac{8}{n}\sum_{i=1}^n \psi_{j,k}^2(G(x_i))} g(x_1)..g(x_n)dx_1..dx_n \\
&\leq \quad P(|\frac{1}{n}\sum_{i=1}^n \psi_{j,k}^2(G(X_i)) - 1| > \alpha) + \exp - \frac{\kappa \log n}{8(1+\alpha)}
\end{aligned}$$

Using Hoeffding inequality (see [26] p. 241), we have, using the fact that $\psi_{j,k}^2(G(X_i))$ are i.i.d. variables bounded by $2^j\|\psi\|_\infty^2$, and such that $\mathbb{E}\psi_{j,k}^2(G(X_i)) = 1$:

$$P(|\frac{1}{n}\sum_{i=1}^n \psi_{j,k}^2(G(X_i)) - 1| > \alpha) \leq 2\exp - \frac{2n^2\alpha^2}{n\|\psi\|_\infty^2 2^{2j}} \leq 2n^{-2\alpha^2/\|\psi\|_\infty^2} \qquad (32)$$

if $2^j \leq \sqrt{\frac{n}{\log n}}$.
Hence, we can easily fix α and then κ large enough in such a way that

$$P(|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i| > \kappa/2\sqrt{\frac{\log n}{n}}) \leq Cn^{-\gamma},$$

if $2^j \leq \sqrt{\frac{n}{\log n}}$.

- Using Bernstein inequality :

$$\begin{aligned} P(|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))(f(X_i) &- \mathbb{E}(\psi_{j,k}(G(X))f(X))| > \kappa/2\sqrt{\frac{\log n}{n}}) \\ &\leq 2\exp - \frac{n^2(\kappa/2\sqrt{\frac{\log n}{n}})^2}{2/3(3\sigma^2 + M\kappa/2\sqrt{\frac{\log n}{n}})} \end{aligned}$$

where

$$\begin{aligned} M &= \|\psi_{j,k}(G(X))(f(X) - \mathbb{E}(\psi_{j,k}(G(X))f(X))\|_\infty \leq 22^{j/2}\|\psi\|_\infty\|f\|_\infty \\ \sigma^2 &= \mathbb{E}|\psi_{j,k}(G(X))(f(X) - \mathbb{E}(\psi_{j,k}(G(X))f(X))|^2 \leq \mathbb{E}|\psi_{j,k}(G(X))(f(X)|^2 \leq \|f\|_\infty. \end{aligned}$$

Furthermore,

$$\begin{aligned} 2\exp - \frac{n^2(\kappa/2\sqrt{\frac{\log n}{n}})^2}{2/3(3\sigma^2 + M\kappa/2\sqrt{\frac{\log n}{n}})} &\leq 2\exp - \frac{3\kappa^2\log n}{4\|f\|_\infty(3 + 22^{j/2}\kappa/2\sqrt{\frac{\log n}{n}})} \\ &\leq 2\exp - \frac{3\kappa^2\log n}{4\|f\|_\infty(3 + \kappa(\frac{\log n}{n})^{1/4})} \end{aligned}$$

if $2^j \leq \sqrt{\frac{n}{\log n}}$.
Hence, we find that for any γ, there exists κ large enough such that

$$P(|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))(f(X_i) - \mathbb{E}(\psi_{j,k}(G(X))f(X))| > \kappa/2\sqrt{\frac{\log n}{n}}) \leq C'n^{-\gamma}.$$

$\square$

5.2 Proof of inequalities (15) and (18)

To prove these inequalities we only need the following proposition :

Proposition 4. *For all $p > 1,\ \pi \geq p,\ s \geq 1/2,\ r \geq p/(1+2s)$, for q such that : $p - q = 2sp/(1+2s)$, we have,*

$$B^G_{s,\pi,r} \subset Max(q).$$

Furthermore, if $0 < m \leq g \leq M < \infty$,

$$B_{s,\pi,r} \subset Max(q).$$

This proposition is proved in Appendix II.

5.3 Behavior of the estimator $\hat{f}^{@}$

5.3.1 Maxiset for an intermediate estimate

Let us consider an intermediate estimate (which will only be used for the convenience of the proof).

$$\hat{f}'(x) = \sum_{j=-1}^{J} \sum_{k \in \mathbb{Z}} \hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geq K t_n\} \psi_{jk}(G(x))$$

$\hat{f}'$ is intermediate between $\hat{f}^{@}$ and $\hat{f}^*$. The difference between $\hat{f}^{@}$ and $\hat{f}'$ only lies into the basis system which is (as for $\hat{f}^*$) $\psi_{jk}(G(x))$ for $\hat{f}'$ whereas it is a random system for $\hat{f}^{@}$.

Our first concern consists in investigating the behaviour of $\hat{f}'$ by proving the following proposition, using a technology similar to the one used for $\hat{f}^*$.

Proposition 5. *Let $p > 1$, $0 < q < p$. Under the condition $(\mathcal{H}_p)$, the maxiset of the estimator $\hat{f}'$*

$$Max'(q) = \{f,\ \mathbb{E}\|\hat{f}' - f\|_p^p \left(\frac{\log n}{n}\right)^{(q-p)/2} < \infty\} \tag{33}$$

is such that,

$$Max'(q) = Max(q) \tag{34}$$

The proof of this result completely mimic the proof of the result concerning $\hat{f}^*$. The only involving problem consists in showing that we have a result similar to Proposition 3, if we replace the estimates $\hat{\beta}^*$ by $\hat{\beta}^{@}$.

Proposition 6. *if f is bounded, there exist constants $C_p,\ C'_p,$ such that for any $\gamma > 0$ there exists κ_0 :*

$$\mathbb{E}(|\hat{\beta_{jk}}^{@} - \beta_{jk}|^p) \leq C_p \frac{(1 + \|f\|_\infty^p)}{n^{p/2}},\ for\ 2^j \leq n \tag{35}$$

$$P(|\hat{\beta_{jk}}^{@} - \beta_{jk}| > \kappa\sqrt{\frac{\log n}{n}}) \leq C'_p n^{-\gamma p}\ for\ \kappa \geq \kappa_0, 2^j \leq (\frac{n}{\log n})^{1/2} \tag{36}$$

Here will appear the fact that dividing the sample in 2 sets obviously simplifies a great deal the proof. We will mimic the proof of Proposition 3, just arguing conditionally to the first part of the sample (i.e. conditionally to $\hat{G}_{[n/2]}$).

We easily obtain :

$$\mathbb{E}[(|\hat{\beta_{jk}}^{@} - \beta_{jk}|^p) \quad | \quad \hat{G}_{[n/2]}] \leq C(\frac{\mathbb{E}[|\psi_{j,k}(\hat{G}_{[n/2]}(X))Y|^p\ |\hat{G}_{[n/2]}]}{n^{p-1}} + \frac{(\mathbb{E}[|\psi_{j,k}(\hat{G}_{[n/2]}(X))Y|^2\ |\hat{G}_{[n/2]}])^p/2}{n^{p/2}})$$

$$\begin{aligned}
P(|\frac{1}{n} \quad \textstyle\sum_{i=[n/2]+1}^{n} \quad & \psi_{j,k}(\hat{G}_{[n/2]}(X_i))\epsilon_i| > \kappa/2\sqrt{\frac{\log n}{n}}\ |\hat{G}_{[n/2]}) \\
& \leq \int ... \int \exp - \frac{\kappa}{8}\frac{\log n}{n} \frac{1}{\frac{1}{n^2}\sum_{i=1}^{n} \psi_{j,k}^2(\hat{G}_{[n/2]}(x_i))} g(x_1)..g(x_n) dx_1..dx_n \\
P(|\frac{1}{n} \quad \textstyle\sum_{i=[n/2]+1}^{n} \quad & \psi_{j,k}(\hat{G}_{[n/2]}(X_i))(f(X_i) - \mathbb{E}(\psi_{j,k}(\hat{G}_{[n/2]}(X))f(X)\ |\hat{G}_{[n/2]})| > \kappa/2\sqrt{\frac{\log n}{n}}\ |\hat{G}_{[n/2]}) \\
& \leq 2\exp - \frac{3\kappa^2 \log n}{4\|f\|_\infty(3 + 22^{j/2}\kappa/2\sqrt{\frac{\log n}{n}})}
\end{aligned}$$

To finish the proof as above, we just need the following lemma: Let us define

$$\varepsilon_{jk}(l) = \mathbb{E}[|\psi_{j,k}(\hat{G}_{[n/2]}(X))|^p h(X)\ |\hat{G}_{[n/2]}] - \int |\psi_{j,k}(G(x))|^l |h(x) g(x) dx$$

Lemma 1. *For $l \leq 1$, h uniformy bounded, if $2^j \leq (n/\log n)^{1/2}$, we have*

$$\mathbb{E}|\varepsilon_{jk}(l)| \leq C2^{j(l/2-1)} \tag{37}$$

$$\forall\, \gamma > 0,\ \exists \lambda,\ P(|\varepsilon_{jk}(2)| \geq \lambda) \leq Cn^{-\gamma} \tag{38}$$

Proof of the lemma : Let us recall the following inequalities (see for instance the review on the subject in Devroye Lugosi [**?**] section 12.) : For any $r > 0,\ \lambda > 0$, there exist constants $C_1,\ C_2$, such that:

$$\mathbb{E}\|\hat{G}_{[n/2]} - G\|_\infty^r \leq C_1 n^{-r/2} \tag{39}$$

$$P(\|\hat{G}_{[n/2]} - G\|_\infty \geq \lambda) \leq C_2 4n \exp -n\lambda^2/32 \tag{40}$$

The proof of the lemma consists in using these inequalities and the following bound : Let us put

$$R_n = \|\hat{G}_{[n/2]} - G\|_\infty$$

If N is the size of the wavelet support, we have

$$\begin{aligned}
|\varepsilon_{jk}(p)| &\leq p\int (2^{j/2}\|\psi\|_\infty)^{p-1}(2^{3j/2}\|\psi'\|_\infty)R_n I\{G(x) \in [k/2^j - R_n, (k+N)/2^j + R_n]\}g(x)dx \\
&\leq C2^{j(p+2)/2}R_n\left[\int I\{G(x) \in [(k-K)/2^j, (k+N+K)/2^j]\}g(x)dx + I\{R_n \geq K2^{-j}\}\right] \\
&\leq C2^{j(p+2)/2}R_n\left[\int I\{y \in [(k-K)/2^j, (k+N+K)/2^j]\}dy + I\{R_n \geq K2^{-j}\}\right] \\
&\leq C'2^{jp/2}R_n + C2^{j(p+2)/2}R_n I\{R_n \geq K2^{-j}\}
\end{aligned}$$

Then, obviously, we get :

$$\begin{aligned}
\mathbb{E}|\varepsilon_{jk}(p)| &\leq C'[2^{jp/2}n^{-1/2} + 2^{j(p+2)/2}n^{-1/2-\gamma/2}] \\
P(|\varepsilon_{jk}(2)| \geq \lambda) &\leq P(R_n \geq C'\lambda 2^{-(j+1)}) + P(R_n \geq K2^{-(j+1)})] \\
&\leq C[n^{-C'^2\lambda/64+1} + n^{-K^2/64+1}]
\end{aligned}$$

To finish the proof of Proposition 6, we just have to replace (32), by

$$\begin{aligned}
P(\frac{1}{n} \quad &|\textstyle\sum_{i=[n/2]+1}^n \quad \psi_{j,k}^2(\hat{G}_{[n/2]}(X_i)) - 1| \geq \alpha) \\
&\leq P(\tfrac{1}{n} \qquad |\sum_{i=[n/2]+1}^n \psi_{j,k}^2(\hat{G}_{[n/2]}(X_i)) - \int \psi_{j,k}^2(\hat{G}_{[n/2]}(x))g(x)dx| \geq \alpha/2) + P(|\varepsilon_{jk}(2)| \geq \alpha/2) \\
&\leq Cn^{-\gamma}
\end{aligned}$$

using Hoeffding inequality for the first part, as in (32) and the lemma for the second one.

5.3.2 Evaluating the difference $\hat{f}^@ - \hat{f}'$

The second part of the proof consists in evaluating the difference

$$\hat{f}^@ - \hat{f}'.$$

Proposition 7. *Under the conditions of Theorem 2, if $f \in B^G_{s,\infty,\infty},\ s > 1/2$,*

$$\mathbb{E}\|\hat{f}^@ - \hat{f}'\|_p^p \leq C((\log n/n)^{sp/2}(\log n)^{p-1} \wedge n^{-p/2})$$

The end of this subsection will be devoted to proving the proposition. Let us observe that the Proposition proves inequality (16) since it is not difficult to verify that under the condition $s > 1/2$, the quantities of order $(\log n/n)^{sp/2}(\log n)^{p-1}$ will always be negligeable compared to the rates that we are expecting, in theorem 2. Notice also that we will need the condition $B^G_{s,\infty,\infty}$ as is clear in the following lemma where we will need a uniform condition on the wavelet coefficients.

Lemma 2. *Since g is compactly supported, if there exists $s > 0$, such that $|\beta_{jk}| \leq C2^{-j(s+1/2)}$, for all $j \geq 0,\ k \in \mathbb{Z}$, for*

$$\Delta_{jk}(x) := \psi_{jk}(\hat{G}_{[n/2]}(x)) - \psi_{jk}(G(x))$$

we have the following bound as soon as $l_{jk} \leq 1$ is random or fixed,

$$\mathbb{E}\|\sum_{j=-1}^{J}\sum_{k\in\mathbb{Z}}\beta_{jk}l_{jk}\Delta_{jk}\|_p^p \leq (b-a)^p C^p C_1((\log n/n)^{sp/2}(\log n)^{p-1} \wedge n^{-p/2}) \tag{41}$$

Proof of the lemma:

It will be useful to remark than if

$$R_n = \|\hat{G}_{[n/2]} - G\|_\infty$$

is smaller than $1/2^{j+1}$, we can affirm that if $G(x)$ belongs to the support of ψ_{jk}, then it belongs to only a finite number of supports of other $\psi_{jk'}$'s and the same is true for $\hat{G}_{[n/2]}(x)$. More generally, if $G(x)$ belongs to the support of ψ_{jk}, then $\psi_{jk}(\hat{G}_{[n/2]}(x))$ may not disappear only for a number of k which is proportional to $2^j R_n$. Hence, we can write :

$$\|\sum_{k\in\mathbb{Z}}\beta_{jk}l_{jk}\Delta_{jk}\|_\infty \leq K\sup_k|\beta_{jk}|2^{j/2}\|\psi\|_\infty 2^j R_n \tag{42}$$

where K only depends on the length of support of ψ. Hence, using (42), and (39)

$$\begin{aligned}
\mathbb{E}\|\sum_{j\leq J}\sum_{k\in\mathbb{Z}}\beta_{jk}l_{jk}\Delta_{jk}\|_p^p &\leq (b-a)\mathbb{E}\|\sum_{j\leq J}\sum_{k\in\mathbb{Z}}\beta_{jk}l_{jk}\Delta_{jk}\|_\infty^p \\
&\leq (b-a)J^{p-1}\sum_{j\leq J}\mathbb{E}\|\sum_{j\leq J}\sum_{k\in\mathbb{Z}}\beta_{jk}l_{jk}\Delta_{jk}\|_\infty^p \\
&\leq C_1(b-a)J^{p-1}\sum_{j\leq J}\sup_k|\beta_{jk}|^p 2^{3jp/2}\|\psi\|_\infty^p\mathbb{E}R_n^p \\
&\leq C^pC_1(b-a)J^{p-1}\sum_{j\leq J}2^{jp(1-s)}n^{-p/2} \\
&\leq (b-a)^pC^pC_1((\log n/n)^{sp/2}(\log n)^{p-1}\wedge n^{-p/2})
\end{aligned}$$

We write for

$$H = \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}}\beta_{jk}I\{|\beta_{jk}|\geq Kt_n/2\}\Delta_{jk}$$

$$
\begin{aligned}
\hat{f}^{@} - \hat{f}' &= \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} \hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\}\Delta_{jk} \\
&= \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} \hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\}\Delta_{jk}(I\{|\beta_{jk}| \geq Kt_n/2\} + I\{|\beta_{jk}| < Kt_n/2\}) \\
&- \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} \beta_{jk} I\{|\beta_{jk}| \geq Kt_n/2\}\Delta_{jk}(I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} + I\{|\hat{\beta}^{@}_{jk}| < Kt_n\}) + H \\
&= \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} (\hat{\beta}^{@}_{jk} - \beta_{jk}) I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| \geq Kt_n/2\}\Delta_{jk} \\
&+ \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} (\hat{\beta}^{@}_{jk} - \beta_{jk}) I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| < Kt_n/2\}\Delta_{jk} \\
&+ \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} \beta_{jk} I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| < Kt_n/2\}\Delta_{jk} \\
&- \sum_{J\geq j\geq -1}\sum_{k\in\mathbb{Z}} \beta_{jk} I\{|\beta_{jk}| \geq Kt_n/2\} I\{|\hat{\beta}^{@}_{jk}| < Kt_n\})\Delta_{jk} + H \\
&= a_1 + a_2 + a_3 + a_4 + H
\end{aligned}
$$

We have :

$$
\mathbb{E}\|\hat{f}^{@} - \hat{f}'\|_p^p \leq 5^{p-1}[\mathbb{E}\|H\|_p^p + \mathbb{E}\|a_1\|_p^p + \mathbb{E}\|a_2\|_p^p + \mathbb{E}\|a_3\|_p^p + \mathbb{E}\|a_4\|_p^p] \tag{43}
$$

Using lemma 2, we get

$$
\mathbb{E}\|H\|_p^p + \mathbb{E}\|a_3\|_p^p + \mathbb{E}\|a_4\|_p^p \leq C((\log n/n)^{sp/2}(\log n)^{p-1} \wedge n^{-p/2}). \tag{44}
$$

Let us now investigate the terms a_1.

$$
\mathbb{E}\|a_1\|_p^p \leq J^{p-1}(b-a) \sum_{J\geq j\geq -1} \mathbb{E}\|\sum_{k\in\mathbb{Z}} (\hat{\beta}^{@}_{jk} - \beta_{jk}) I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| \geq Kt_n/2\}\Delta_{jk}\|_\infty^p
$$

Notice that if j_s is such that $2^{j_s} \sim n^{\frac{1}{1+2s}}$, we observe that for $j \geq j_s$ the terms in the right hand side disappear since there is no k's with $|\beta_{jk}| \geq Kt_n/2$. Hence, using (42) and (39) and for $\frac{1}{r_1} + \frac{1}{r_2} = 1$,

$$
\begin{aligned}
\mathbb{E}\|a_1\|_p^p &\leq J^{p-1}(b-a) \sum_{j_s\geq j\geq -1} \mathbb{E}\|\sum_{k\in\mathbb{Z}} (\hat{\beta}^{@}_{jk} - \beta_{jk}) I\{|\hat{\beta}^{@}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| \geq Kt_n/2\}\Delta_{jk}\|_\infty^p \\
&\leq KJ^{p-1}(b-a) \sum_{j_s\geq j\geq -1} \mathbb{E}\sup_{k\in\mathbb{Z}} |\hat{\beta}^{@}_{jk} - \beta_{jk}|^p 2^{3jp/2}\|\psi\|_\infty^p R_n^p \\
&\leq KJ^{p-1}(b-a) \sum_{j_s\geq j\geq -1} 2^{3jp/2}\mathbb{E}[\sup_{k\in\mathbb{Z}} |\hat{\beta}^{@}_{jk} - \beta_{jk}|^{pr_1}]^{1/r_1}\mathbb{E}[R_n^{pr_2}]^{1/r_2} \\
&\leq KJ^{p-1}(b-a) \sum_{j_s\geq j\geq -1} 2^{3jp/2}[\frac{2^j}{n^{pr_1/2}}]^{\frac{1}{r_1}}[\frac{1}{n^{pr_2/2}}]^{\frac{1}{r_2}} \\
&\leq Kn^{\frac{3p/2+1/r_1}{1+2s}-p}
\end{aligned}
$$

As $s > 1/2$ we can choose r_1 such that this term is not significant either.
The term a_2 is even simpler :

$$
\begin{aligned}
\mathbb{E}\|a_2\|_p^p &= \mathbb{E}\|\sum_{j\geq -1}\sum_{k\in\mathbb{Z}}(\hat\beta^{@}_{jk}-\beta_{jk})I\{|\hat\beta^{@}_{jk}|\geq Kt_n\}I\{|\beta_{jk}|<Kt_n/2\}\Delta_{jk}\|_p^p \\
&\leq CJ^{p-1}\sum_{J\geq j\geq -1}2^{3jp/2}[\mathbb{E}\sup_{k\in\mathbb{Z}}|\hat\beta^{@}_{jk}-\beta_{jk}|^{pr_1}]^{\frac{1}{r_1}}[P\{\sup_{k\in\mathbb{Z}}|\hat\beta^{@}_{jk}-\beta_{jk}|\geq Kt_n/2\}]^{\frac{1}{r_2}} \\
&\leq J^{p-1}C\sum_{J\geq j\geq -1}2^{3jp/2}[\frac{2^j}{n^{pr_1/2}}]^{\frac{1}{r_1}}[\frac{2^j}{n^{\gamma}}]^{\frac{1}{r_2}} \\
&\leq CJ^{p-1}n^{\frac{p}{4}+\frac{1}{2}-\frac{\gamma}{r_2}}\leq Cn^{-p/2}
\end{aligned}
$$

if we choose γ large enough. We have used (36).

6 Appendix I: Muckhenhoupt weights, Besov spaces

6.1 Definitions

The definition of a Muckhenhoupt weight has been given in subsection 2.4. There are several equivalent definitions which are well known (see [35]). We give here another important one together with the very helpful 'doubling property'.

Proposition 8. *If I denote a bounded interval of $\mathbb{R}$, and $|I|$ its Lebesgue measure, for $1\leq p<\infty$ and q such that $1/p+1/q=1$, ω a non-negative locally integrable function, the following statement are equivalent :*

1. $\omega\in A_p$ $i.E.$

$$\forall I,\ \ (\frac{1}{|I|}\int_I\omega)^{1/p}(\frac{1}{|I|}\int_I\omega^{-q/p})^{1/q}\leq C<\infty, \tag{45}$$

(with the obvious modification if $q=\infty,\ p=1$.)

2.

$$\text{For any measurable function } f,\quad (\frac{1}{|I|}\int_I|f|)\leq C(\frac{1}{\omega(I)}\int_I|f|^p\omega)^{1/p} \tag{46}$$

(where $\omega(I)=\int_I\omega$.)

Moreover, the measure $\omega(A)=\int_A\omega(x)dx$ satisfies the following 'doubling' property :
If $I=[a-h,a+h]$ and $2I=[a-2h,a+2h]$ then

$$\omega(2I)\leq(2C)^p\omega(I) \tag{47}$$

Proof: (46) implies easily (45) taking $f=\omega^{-q/p}$. To prove that (45) implies (46), we apply Hölder inequality to $|f|=(|f|\omega^{1/p})(\omega^{-1/p})$:

$$(\frac{1}{|I|}\int_I|f|)\leq(\frac{1}{|I|}\int_I|f|^p\omega)^{1/p}(\frac{1}{|I|}\int_I\omega^{-q/p})^{1/q}\leq C(\frac{1}{|I|}\int_I|f|^p\omega)^{1/p}(\frac{1}{|I|}\int_I\omega)^{-1/p}$$

Applying now (46) with $2I$ instead of I and $f=1_I$ we get (47). $\square$

6.2 Muckhenhoupt weight and densities.

here we prove the following proposition :

Proposition 9. *Let $1 \leq p < \infty$, let g be a density on $[a,b]$ and $G(x) = \int_a^x g(s)ds$ be the associated repartition function. Let us suppose that G is strictly increasing from $[a,b]$ to $[0,1]$. The following statements are equivalent*

1. $\frac{1}{g(G^{-1}(t))} \in A_p([0,1])$ *i.e. for any I subinterval of* $[0,1]$
$$(\frac{1}{|I|}\int_I \frac{1}{g(G^{-1}(t))}dt)^{1/p}(\frac{1}{|I|}\int_I (\frac{1}{g(G^{-1}(t))})^{q/p}dt)^{/q} \leq C$$
2. *For q such that $1/p + 1/q = 1$, for any J subinterval of $[a,b]$, we have*
$$(\frac{1}{|J|}\int_J g(s)^q ds)^{1/q} \leq C(\frac{1}{|J|}\int_J g(s)ds$$

Proof:

Since G is strictly increasing from $[a,b]$ to $[0,1]$, we have a bijection between the intervals of $[a,b]$ and those of $[0,1]$. So if $I = [\alpha,\beta] \subset [0,1]$ then $[\alpha,\beta] = [G(u), G(v)]$. And if then $J = [u,v]$, we have $1_I(G(s)) = 1_J(s)$, or $1_I(t) = 1_J(G^{-1}(t))$. In addition, as for any non negative measurable function Φ :

$$\int_{[0,1]} \Phi(G^{-1}(t))dt = \int_{[a,b]} \Phi(s)g(s)ds$$
$$|I| = G(v) - G(u) = \int_{[u,v]} g(s)ds = \int_J g(s)ds$$

So

$$(\frac{1}{|I|}\int 1_I(t)\frac{1}{g(G^{-1}(t))}dt)^{1/p}(\frac{1}{|I|}\int 1_I(t)g(G^{-1}(t))^{q/p}dt)^{1/q} \leq C$$

is obviously equivalent to

$$\frac{1}{|I|}(\int 1_I(G(s))ds)^{1/p}(\int 1_I(G(s))g(s)^{q/p+1}ds)^{1/q} \leq C$$

or

$$(\frac{1}{|J|}\int_J g(s)^q ds)^{1/q} \leq C\frac{|I|}{|J|} = C\frac{1}{|J|}\int_J g(s)ds$$

□

6.3 Weighted spaces, wavelets and approximation

In this section ϕ is a compactly supported scaling function of a multiresolution analysis and ψ an associated compactly supported wavelet. We fix the notations in the following way :

$$supp(\phi) \subset [0,L]; \;\; supp(\psi) \subset [0,L].$$

$$\hat{\phi}(\xi) = m_0(\xi/2)\mathcal{F}(\phi)(\xi/2) \quad (48)$$
$$\hat{\psi}(\xi) = m_1(\xi/2)\mathcal{F}(\phi)(\xi/2) \quad (49)$$

where $\hat{g}$ denotes here the Fourier transform of g and $m_0(\xi)$ and $m_1(\xi)$ are trigonometric polynomials.

As usual for k,j in $\mathbb{Z}$ and any function g, we put $g_{j,k}(x) = 2^{j/2}g(2^jx - k)$. We put

$$I_{j,k} = [\frac{k}{2^j}, \frac{k+1}{2^j}]), \; \tilde{I}_{j,k} = [\frac{k}{2^j}, \frac{k+L}{2^j}],$$

So that $supp(\phi_{j,k}) \subset \tilde{I}_{j,k}, \; supp(\psi_{j,k}) \subset \tilde{I}_{j,k}$. For a measurable function f we define:

$$\alpha_{j,k} = \int f(x)\overline{\phi_{j,k}}(x)dx; \qquad \beta_{j,k} = \int f(x)\overline{\psi_{j,k}}(x)dx$$
$$P_jf = \sum_k \alpha_{j,k}\phi_{j,k} = P_{V_j}f; \qquad P_{j+1}f - P_jf = P_{W_j}f = \sum_k \beta_{j,k}\psi_{j,k}$$

6.3.1 Linear approximation in $\mathbb{L}_p(\omega)$

The following theorem express the equivalence of the $\mathbb{L}_p(\omega)$ norms of functions in V_j or W_j in terms of wavelet coefficients. Notice however that here the weight ω is appearing in the sum.

Theorem 5. *Let $1 \leq p < \infty$, and suppose ω belongs to $A_p(\mathbb{R})$, then*

1. *There exists C only depending on ϕ, ψ and ω, such that :*

$$\frac{1}{C}\sum_k |\alpha_{j,k}|^p \omega(I_{j,k}) \leq \quad 2^{-jp/2}\|\textstyle\sum_k \alpha_{j,k}\phi_{j,k}\|^p_{\mathbb{L}_p(\omega)} \quad \leq C\displaystyle\sum_k |\alpha_{j,k}|^p\omega(I_{j,k}) \tag{50}$$

$$\frac{1}{C}\sum_k |\beta_{j,k}|^p \omega(I_{j,k}) \leq \quad 2^{-jp/2}\|\textstyle\sum_k \beta_{j,k}\psi_{j,k}\|^p_{\mathbb{L}_p(\omega)} \quad \leq C\displaystyle\sum_k |\beta_{j,k}|^p\omega(I_{j,k}) \tag{51}$$

2.

$$\forall j \in \mathbb{Z},\ \|P_j f\|_{\mathbb{L}_p(\omega)} \quad \leq \quad C^2\|f\|_{\mathbb{L}_p(\omega)} \tag{52}$$

$$\lim_{j\to\infty} \|P_j f - f\|_{\mathbb{L}_p(\omega)} \quad = \quad 0 \tag{53}$$

3. *Let $0 < q \leq \infty$, and $f \in \mathbb{L}_p(\omega)$,*

$$[\sum_j (2^{js}\|P_j f - f\|_{\mathbb{L}_p(\omega)})^q]^{1/q} < \infty \Longleftrightarrow [\sum_j (2^{js}2^{j/2}(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p\omega(I_{j,k}))^{1/p})^q]^{1/q} < \infty \tag{54}$$

   with the usual modification if $q = \infty$.

This theorem is the consequence of the following lemmas.

Lemma 3. *Let ω be in $A_\infty(\mathbb{R})$. Let θ a be bounded function, whith support in $[0, L]$ and $\theta_{j,k}(x) = 2^{j/2}\theta(2^j x - k)$. Then for $0 < p < \infty$,*

$$\|\sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x)\|_{\mathbb{L}_p(\omega)} \leq C' 2^{j/2}(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p\omega(I_{j,k}))^{1/p}$$

and $p = \infty$,

$$\|\sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x)\|_{\mathbb{L}_\infty(\omega)} \leq C' 2^{j/2}(\sup_{k\in\mathbb{Z}} |\lambda_{j,k}|)$$

Proof of the lemma: The main tool of this proof is the doubling property (47) of the measure $\omega(x)dx$.

1. $p = \infty$ is obvious.
2. $1 < p < \infty$.

   As θ is a bounded function, with support in $[0, L]$, $\theta_{j,k}$ in supported in $\tilde{I}_{j,k}$. Hence there exists $C < \infty$ such that $\sum_k |\theta(x - k)| \leq C$. Hence,

$$\begin{aligned}
|\sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x)|^p &\leq 2^{jp/2}(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p|\theta(2^j x - k)|)(\sum_{k\in\mathbb{Z}} |\theta(2^j x - k)|)^{p/q} \\
&\leq C^{p/q}2^{jp/2}(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p|\theta(2^j x - k)|) \\
\int |\sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x)|^p\omega(x)dx &\leq C^{p/q}2^{jp/2}(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p \int_{\tilde{I}_{j,k}} |\theta(2^j x - k)|)\omega(x)dx \\
&\leq C^{p/q}\|\theta\|_\infty 2^{jp/2}(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p\omega(\tilde{I}_{j,k})
\end{aligned}$$

   We finish the proof using the doubling property (47), since it implies :

$$\omega(\tilde{I}_{j,k}) \leq c\,\omega(I_{j,k}).$$

3. $0 < p \leq 1$.

$$\begin{aligned}\int |\sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x)|^p \omega(x)dx &\leq \sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p \int_{\tilde{I}_{j,k}} |\theta_{j,k}(x)|^p \omega(x)dx \\ &\leq \sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p \|\theta\|_\infty^p 2^{jp/2}\omega(\tilde{I}_{j,k}) \\ &\leq c\|\theta\|_\infty^p 2^{jp/2} \sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p \omega(I_{j,k})\end{aligned}$$

□

Lemma 4. *For* $1 \leq p \leq \infty, \omega \in A_p$,

$$2^{j/2}(\sum_k |\int f\bar{\phi}_{j,k}dx|^p \omega(I_{j,k}))^{1/p} \leq C\|f\|_{\mathbb{L}_p(\omega)}$$

(with the obvious modification if $p = \infty$*.)*

The same inequality is true if we replace ϕ *by* ψ*.*

Proof: The main tool is here property (46).

$$\begin{aligned}2^{jp/2}\sum_k |\int f\bar{\phi}_{j,k}dx|^p \omega(I_{j,k}) &\leq 2^{jp/2}\sum_k (\int_{\tilde{I}_{j,k}} |f||\phi_{j,k}|dx)^p \omega(I_{j,k}) \\ &\leq C2^{jp/2}\sum_k \frac{1}{\omega(\tilde{I}_{j,k})}\int |f\phi_{j,k}|^p\omega(x)dx \;\; \omega(I_{j,k}) \;\; |\tilde{I}_{j,k}|^p \\ &\leq C'2^{-jp/2}\int |f(x)|^p \sum_k 2^{jp/2}|\phi(2^jx-k)|^p\omega(x)dx \\ &\leq C" \int |f(x)|^p\omega(x)dx\end{aligned}$$

Using $|\tilde{I}_{j,k}| \sim 2^{-j}$ and the doubling property (47). Of course ϕ and ψ can be exchanged.

□

Remarks :

1. From the two previous lemma we deduce (50) and (51).
2. Using these lemmas we deduce (52) :

$$\begin{aligned}\|P_jf\|_{\mathbb{L}_p(\omega)} &= \|\sum_k \int f(y)\phi_{j,k}(y)dy\phi_{j,k}\|_{\mathbb{L}_p(\omega)} \\ &\leq C2^{j/2}(\sum_k |\int f\phi_{j,k}dx|^p\omega(I_{j,k}))^{1/p} \leq C^2\|f\|_{\mathbb{L}_p(\omega)}\end{aligned}$$

3. Now, to prove (53), it is enough to prove that the family $\{\phi_k,\ \psi_{j,k}\}$ is total in $\mathbb{L}_p(\omega)$. But this is obvious since if $g \in \mathbb{L}_p(\omega)^* = \mathbb{L}_q(\omega)$ and $\int g\phi_k\omega = \int g\psi_{j,k'}\omega = 0$ for all k, k', j then $g\omega = 0$ $a.e.$ so $g = 0$ $\omega.$ $a.e.$. (It is clear that if $g \in \mathbb{L}_q(\omega)$ then $g\omega$ is locally Lebesgue integrable.)

   ◇

It remains to prove (54). But for $f \in \mathbb{L}_p(\omega)$,

$$\|P_{W_j} f\|_{\mathbb{L}_p(\omega)} \leq \|P_{j+1} f - f\|_{\mathbb{L}_p(\omega)} + \|P_j f - f\|_{\mathbb{L}_p(\omega)}$$

and

$$\|P_j f - f\|_{\mathbb{L}_p(\omega)} \leq \sum_{l=j}^{\infty} \|P_{W_l} f\|_{\mathbb{L}_p(\omega)}$$

Hence :

$$[\sum_j (2^{js} \|P_j f - f\|_{\mathbb{L}_p(\omega)})^q]^{1/q} < \infty \Longleftrightarrow [\sum_j (2^{js} \|P_{W_j} f\|_{\mathbb{L}_p(\omega)})^q]^{1/q} < \infty$$

We have used the following well known convolution lemma:

Lemma 5. *Let $(a_j)_{j\in\mathbb{Z}}$ and $(b_j)_{j\in\mathbb{Z}}$ two sequence and*

$$a \star b_k = \sum_j a_{k-j} b_j.$$

then

$$\|a \star b\|_{l_q(\mathbb{Z})} \leq \|a\|_{l_{q\wedge 1}(\mathbb{Z})} \|b\|_{l_q(\mathbb{Z})} \tag{55}$$

Moreover, using (4) we get :

$$[\sum_j (2^{js} \|P_{W_j} f\|_{\mathbb{L}_p(\omega)})^q]^{1/q} < \infty \Longleftrightarrow [\sum_j (2^{js} 2^{j/2} (\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p \omega(I_{j,k}))^{1/p})^q]^{1/q} < \infty$$

6.4 Weighted Besov spaces and wavelet expansions

Using the notations of section 3.3, we shall prove the following theorem :

Theorem 6. *Let $\omega \in A_p, \quad 1 \leq p < \infty$*

Let ϕ and ψ be defined as above and let us suppose in addition that :

$$\int x^k \psi(x) dx = 0, \;\; k = 0, 1, .., N-1$$

Let

$$\beta_{j,k} = \int_{\mathbb{R}} f(x) \overline{2^{j/2} \psi(2^j x - k)} dx.$$

Then,

$$(\int_0^1 (\frac{(\rho^N(t, f, \omega, p)}{t^s})^q \frac{dt}{t})^{1/q} < \infty \Longrightarrow [\sum_j (2^{js} 2^{j/2} (\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p \omega(I_{j,k}))^{1/p})^q]^{1/q} < \infty$$

with the usual modification if $q = \infty$.

We will use the standard following lemma.

Lemma 6. *The following statements are equivalent :*

1. *There exists $\theta \in \mathbb{L}_1(\mathbb{R})$ such that*
$$\psi(x) = (-1)^N \Delta^N_{-1/2} \theta(x)$$
2. *There exists $\gamma \in \mathbb{L}_1(\mathbb{R})$ such that*
$$\psi(x) = (D^N \gamma)(x)$$
3. $$\int x^k \psi(x) dx = 0, \;\; k = 0, 1, .., N-1$$

4.

$$m_1(\xi) = \mathcal{O}(|\xi|^N)$$

5. There exists a trigonometric polynomial $\tilde{m}$ such that

$$m_1(\xi) = (1 - \exp -i\xi)^N \tilde{m}(\xi)$$

Moreover, $supp(\theta) \subset [0, L], \quad supp(\gamma) \subset [0, L]$.

For the reader's convenience we give a very short proof of this lemma.

Proof:

1. $1 \Longrightarrow 2$ The hypothesis is equivalent to

$$\hat{\psi}(\xi) = (1 - \exp -i\xi/2)^N \hat{\theta}(\xi)$$

So

$$\hat{\psi}(\xi) = (1 - \exp -i\xi/2)^N \hat{\theta}(\xi) = (i\xi)^N \exp -iN\xi/4 \ \frac{1}{2^N}(\frac{\sin \xi/4}{\xi/4})^N \hat{\theta}(\xi)$$

And obviously $\exp -iN\xi/4 \ \frac{1}{2^N}(\frac{\sin \xi/4}{\xi/4})^N \hat{\theta}(\xi)$ is the Fourier transform of an integrable function.

2. $2 \Longleftrightarrow 3$ This is standard using Taylor formula.

3. $2 \Longrightarrow 4$

$$(i\xi)^N \hat{\gamma}(\xi) \ = \hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2)$$

implies, as $|\hat{\phi}(0)| = 1$,

$$m_1(\xi) = \mathcal{O}(|\xi|^N)$$

4. $4 \Longleftrightarrow 5$ This is due to the following lemma .

5. $5 \Longrightarrow 1$ we have

$$\hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2) = (1 - \exp -i\xi/2)^N \tilde{m}(\xi/2)\hat{\phi}(\xi/2)$$

Lemma 7. *Let $m(\omega)$ be a trigonometric polynomial . The following statements are equivalent.*

1.

$$m(\omega) = (1 - \exp -i\omega)^N \tilde{m}(\omega)$$

with $\tilde{m}$ a trigonometric polynomial.

2.

$$m(\omega) = \mathcal{O}(|\omega|^N) \ \ .$$

Proof:

$1 \Longrightarrow 2$ is obvious.

$2 \Longrightarrow 1$: Let us put

$$m(\omega) = \sum_{k=0}^{M} a_k \exp ik\omega$$

If $N = 1$, we have to find a a trigonometric polynomial $\sum_k b_k \exp ik\omega$ such that

$$\sum_{k=0}^{M} a_k \exp ik\omega = (1 - \exp i\omega) \sum_k b_k \exp ik\omega$$

So

$$\sum_{k=0}^{M} a_k \exp ik\omega = \sum_{k \in \mathbb{Z}} (b_k - b_{k+1}) \exp ik\omega$$

Let us put $\Delta b_k = (b_k - b_{k+1}) = a_k$, so that $b_k = \sum_{j\geq k} a_j$. But, by hypothesis

$$m(0) = 0 = \sum_{l=0}^{M} a_l$$

So $b_k = 0$ for $k < 0$ and $k > M$. We can now finish the proof using a recurrence on N.

□

The following corollary of lemma 6 is now clear :

Corollary 2. *let ψ a compactly supported wavelet satisfying one of the previous equivalent properties of lemma 6. Let f a locally integrable function, and :*

$$\beta_{j,k} = \int f(x)\psi_{j,k}(x)dx = 2^{j/2}\int f(x)\psi(2^j x - k)dx.$$

then

$$\beta_{j,k} = (-1)^N 2^{j/2}\int \Delta^N_{2^{-(j+1)}} f(u)\theta(2^j u - k)du \tag{56}$$

and if $D^N f$ exists

$$\beta_{j,k} = (-1)^N 2^{-jN} 2^{j/2}\int D^N f(u)\gamma(2^j u - k)du \tag{57}$$

Proof :

$$\begin{aligned}
\beta_{j,k} &= 2^{j/2}\int f(x)\psi(2^j x - k)dx = 2^{j/2}\int f(x)\sum_{l=0}^{N} C_N^l(-1)^l\theta(2^j x - l/2 - k)\\
&= 2^{j/2}\int \sum_{l=0}^{N} C_N^l(-1)^l f(u - l2^{-j-1})\theta(2^j u - k)\\
&= (-1)^N 2^{j/2}\int \Delta^N_{2^{-(j+1)}} f(u)\theta(2^j u - k)du
\end{aligned}$$

One can prove simply (57) using integration by part. □

6.4.1 Proof of theorem 6

For $\omega \in A_p$, using (56), (46) and(47), we have :

$$\begin{aligned}
|\beta_{j,k}|^p &\leq 2^{jp/2}(\int_{\tilde{I}_{j,k}} |\Delta^N_{2^{-(j+1)}} f(u)||\theta(2^j u - k)|du)^p\\
&\leq C2^{jp/2}\,\frac{|\tilde{I}_{j,k}|^p}{\omega(\tilde{I}_{j,k})}\int_{\tilde{I}_{j,k}} |\Delta^N_{2^{-(j+1)}} f(u)|^p|\theta(2^j u - k)|^p\omega(u)du
\end{aligned}$$

So

$$2^{jp/2}|\beta_{j,k}|^p\omega(I_{j,k}) \leq C'\int_{\tilde{I}_{j,k}} |\Delta^N_{2^{-(j+1)}} f(u)|^p|\theta(2^j u - k)|^p\omega(u)du$$

and

$$2^{j/2}(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p\omega(I_{j,k}))^{1/p} \leq C"\int_{\mathbb{R}} |\Delta^N_{2^{-(j+1)}} f(u)|^p\omega(u)du \leq C"\rho^N(2^{-(j+1)}, f, \omega, p) \tag{58}$$

7 Appendix II: proof of Proposition 4

We will only prove the first part. For the case where g is bounded above and below, it is enough to notice that we are reduced to the general case with $\omega(I_{jk}) \sim 2^{-j}$.

Let us recall that we are going to consider the following spaces :

$$B^G_{s,\pi,r} = \{f : \ (\int_0^1 (\frac{(\tilde{\rho}^N(t,f,G,\pi)}{t^s})^r \frac{dt}{t})^{1/r} < \infty\}.$$

Let us recall that corollary 1 proves that under the condition $(\mathcal{H}_\pi)$, then, for $I_{j,k} = [\frac{k}{2^j}, \frac{k+1}{2^j}])$, and $f = \sum_{j,k} \beta_{jk}\psi_{j,k}(G)$, we have

$$f \in B^G_{s,\pi,r} \Longrightarrow [\sum_j (2^{js}2^{j/2}(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^\pi \omega(I_{j,k}))^{1/\pi})^r]^{1/r} < \infty$$

with the usual modification if $r = \infty$.

As $Max(q)$ is the intersection of 2 conditions, we will have to prove the inclusions of $B^G_{s,\pi,r}$ into the two following sets :

$$L_1 \quad = \quad \left\{ f = \sum_I \beta_I \psi_I \circ G, \ \sup_{\lambda>0} \lambda^q \nu\{(j,k)/ \ |\beta_{jk}| > \lambda\} < \infty \right\} \tag{59}$$

$$L_2 \quad = \quad \left\{ f = \sum_I \beta_I \psi_I \circ G, \ \sup_{l\geq 0} \| \sum_{j\geq l, \ k} \beta_{jk}\psi_{jk} \circ G\|_p^p 2^{l(p-q)} < \infty \right\} \tag{60}$$

Let us remind that we will concentrate on the case where

$$\nu(I) = \|\psi_I \circ G\|_p^p \sim 2^{jp/2}\omega(I_{jk})$$

and let us introduce the following besov bodies :

$$b^G_{s,\pi,r} = \left\{ f = \sum_I \beta_I \psi_I \circ G, \ [\sum_{j\geq -1} 2^{jsr}2^{jr/2}(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^\pi \omega(I_{j,k}))^{r/\pi}]^{1/r} < \infty \right\} \tag{61}$$

with the usual modification if $r = \infty$. Our aim is to reduce the proof of Proposition 4, to the embeddings of Besov bodies which are quite simple as is shown just below.

7.0.2 Embeddings of the Besov bodies

Because of the fact that ω is a finite weight, the following inclusions are obvious.

$$b^G_{s,\pi,r} \hookrightarrow b^G_{z,\rho,r}, \quad if\ 0 < \rho \leq \pi, \ z \leq s. \tag{62}$$

7.0.3 Condition (59)

Now, let us turn to the problem of embedding a particular body $b^G_{s,\pi,r}$ into

$$l_{q,\infty}(\nu) = \left\{ f = \sum_I \beta_I \psi_I \circ G, \ \sup_{\lambda>0} \lambda^q \nu\{(j,k)/ \ |\beta_{jk}| > \lambda\} < \infty \right\}.$$

We will simplify the problem by considering the embedding into

$$l_q(\nu) = \{f = \sum_I \beta_I \psi_I \circ G, \ \sum_{jk} |\beta_{j,k}|^q 2^{jp/2}\omega(I_{jk}) < \infty\}$$

since, using Markov inequality, obviously $l_q(\nu) \subset l_{q,\infty}(\nu)$. Let us remark that choosing $s = p/2q - 1/2$, we have

$$l_q(\nu) = b^G_{s,q,q}.$$

Then we have the following proposition.

Proposition 10. *Let us define* q *by the relation (63)*

$$s = (\frac{p}{2q} - \frac{1}{2}) \tag{63}$$

then

$$\textit{if } 0 < r \leq q, \quad b^G_{s,\pi,r} \hookrightarrow b^G_{s,q,q}$$

Proof:

We will use the embeddings (62), taking $\rho = q$. As we have $q \leq \pi$, (since $p > q \iff s > 0$), using (62), we get, if moreover $r \leq q$, :

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s,q,q}$$

7.0.4 Condition (60)

Using Theorem 5, we have:

$$\begin{aligned} \| \sum_{j \geq l,k} \beta_{j,k} \psi_{j,k} \circ G \|_p 2^{l \frac{p-q}{p}} &\leq \sum_{j \geq l} \| \sum_k \beta_{j,k} \psi_{j,k} \circ G \|_p 2^{l \frac{p-q}{p}} \\ &\leq C \sum_{j \geq l} 2^{j/2} (\sum_k |\beta_{j,k}|^p \omega(I_{j,k}))^{1/p} 2^{l \frac{p-q}{p}} \end{aligned}$$

Hence, if $f \in b^G_{(1-q/p),p,\infty}$, condition (60) obviously holds. Hence the problem remaining to us is to check whether $b^G_{s,\pi,r}$ is included into $b^G_{(1-q/p),p,\infty}$. Now, if we use the embeddings (62), with $\rho = p$, and we only need to check that $s \geq 2s/(1+2s) = 1 - q/p$, with q chosen as in (63), which is always true for $s \geq 1/2$. Hence, (60) will always hold if $s \geq 1/2$, for $p \leq \pi$.

References

[1] Antoniadis, A., Grégoire, G. and McKeague, I. (1994), "Wavelet methods for curve estimation", *J. Am. Statist. Ass.*, **89**, 1340–1353.

[2] Antoniadis, A., Grégoire, G. and Vial, P. (1997), "Random design wavelet curve smoothing", *Statistics & Probability Letters*, **35**, pp. 225–232.

[3] Antoniadis, A. and J. Fan, (2001) "Regularization of Wavelets Approximations" with discussion, *J. Amer. Statist. Assoc.*, **96**, No 455, pp. 939-963.

[4] Antoniadis, A. and P. Dinh-Tuan, (1998)"Wavelet regression for random or irregular design", *Computational Statistics & data analysis*, **28**, 353–369.

[5] Brown, L. D, Cai, T. T., Low, M. G., Zhang, C.H. (2002) Asymptotic equivalence theory for non parametric regression with random design. *Annals of Statistics*, **30**, No. 3, 688-707.

[6] Cai, T. T. and Brown, L. D. (1998), "Wavelet Shrinkage for nonequispaced samples", *The Annals of Statistics*, **26**, 1783–1799.

[7] Cai, T. and Brown, L. D. (1999) Wavelet regression for random uniform design. Statistics and Probability Letters, 42, 313–321.

[8] Coifman, R., and Fefferman, C. (1974) Weighted norm inequality for maximal functions and singular integrals. Studia Math. 109 (3) pp 241-250.

[9] Cohen,A., DeVore, R. , Kerkyacharian, G., Picard,D.(1999) : Saturation spaces associated with non linear Statistical methods Submitted to Applied and Computational Harmonic Analysis

[10] V. Delouille, J. Simoens and R. Von Sachs (2001), Smooth design-adapted wavelets for nonparametric stochastic regression. Discussion Paper 01/17 de l'Institut de Statistique, UCL, Louvain-la-Neuve

[11] V. Delouille, J. Franke and R. Von Sachs (2001), Nonparametric stochastic regression with design-adapted wavelets. Sankhya, Series A (2001), Vol 63 (3), pp. 328-366.

[12] Donoho, D.L. (1993) Unconditional bases are optimal bases for data compression and statistical estimation *Applied and Computational Harmonic Analysis* **1**, 100-115.

[13] Donoho, D.L. (1996a) Non linear solutions of linear inverse problems by wavelet-vaguelette decomposition *Applied and Computational Harmonic Analysis* **2**, 388-392.

[14] Donoho, D.L. (1996b) Unconditional bases and bit-level compression *Applied and Computational Harmonic Analysis* **3**, 388-392.

[15] Donoho, D.L. (1997) CART and best-ortho-basis: a connection. *Annals of Statistics***25** , no. 5, 1870–1911.

[16] Donoho, D.L., Johnstone I.M. (1995) Ideal spatial adaptation via Wavelet shrinkage *Biometrika* **81**, 425-455.

[17] Donoho, D.L., Johnstone.I.M. (1996) Neoclassical minimax problems, thresholding and adaptive function estimation *Bernoulli* (2), 39-62.

[18] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994) Wavelet Shrinkage: Asymptopia? . *Journal of the Royal Statistical Society* Series B. **57**, 301-369, with discussion.

[19] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Annals of Statistics*, **24**, 508-539.

[20] Donoho, D.L;, Johnstone, I.M., Kerkyacharian, G.,and Picard, D., (1997) Universal near minimaxity of wavelet shrinkage by optimal recovery. *Lucien LeCam Festschrift* D. Pollard and G. Yang (eds.), 183-215, Springer Verlag.

[21] Devroye, L., Lugosi, G. (1996) *A probabilist theory of pattern recognition.*, Springer-Verlag, New-York.

[22] Foster, G. (1996) Wavelet for period analysis of unequally sampled time series. Astronomy Journal, **112**, 4, 1709-1729.

[23] Garcia-Cuerva,J. and Rubio de Francia J.L. (1985)*Weighted norm inequalities and related topics* North-Holland math. stud. 116.

[24] Gol'dshtein,V.M. and Reshetnyak,Y. (1983) *Quasiconformal Mappings and Sobolev Spaces.* Kluwer Academic Publishers.

[25] Hall, P. and Turlach, B.A. (1997), "Interpolation methods for nonlinear wavelet regression with irregularly spaced design", *The Annals of Statistics*, **25**, 1912–1925.

[26] Härdle W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998) *Wavelet, Approximation and Statistical Applications* Lecture Notes in Statistics, 129 Springer Verlag, New York.

[27] Kerkyacharian, G. and Picard, D. (2000) Thresholding algorithms, maxisets and well-concentrated bases, with discussion, Test, vol 9, No 2, P 283-345.

[28] Kerkyacharian, G. and Picard, D. (2001) Entropy, Universal coding, Approximation and bases properties. Technical report.

[29] Kerkyacharian, G. and Picard, D. (2002) Non-linear Approximation and Muckenhoupt weights. Technical report.

[30] Kovac, A. and Silverman, B. W. (2000), "Extending the scope of wavelet regression methods by coefficient-dependent thresholding", *J. Am. Statist. Ass.*, **95**, 172–183.

[31] Muckenhoupt, B.(1972) Weighted norm inequalitiesfor the hardy maximal function, Trans. Amer. Math. Hencec. 165, 207-226.

[32] Pensky, M. and Vidakovic, B. (2000) On non-equally spaced wavelet regression. To appear in Annals of Institute of Statistical Mathematics (2001).

[33] Rivoirard, V. (2002) Etimation bayesienne non parametrique. Thèse de l'université Paris VII.

[34] Sardy, S., Percival, D. B., Bruce A., G., Gao, H.-Y. and Stuelzle, W. (1999) "Wavelet shrinkage for unequally spaced data", *Statistics and Computing*, **9**, 65–75.

[35] Stein, E. (1993) *Harmonic Analysis : Real Variable Methods, orthogonalite, and Oscillatory Integrals.* Princeton University Press.

[36] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. Ann. Statist.**10**, 1040-1053.

[37] Vanraes, E., Jansen, M. and Bultheel, A. (2001) Stabilized wavelet transforms for nonequispaced data smoothing. Technical report, Dept. of Computer Science, K.U.Leuven, Belgium.

[38] Voichitsa, M. (2002) Personal communication.