Bernoulli **10**(6), 2004, 1053–1105

Regression in random design and warped wavelets

GÉRARD KERKYACHARIAN[1,2] and DOMINIQUE PICARD[2]

[1]*Département de Mathématiques, Université de Paris X, 92001 Nanterre, France. E-mail: kerk@math.jussieu.fr*
[2]*CNRS, Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VII, 75252 Paris, France. E-mail: picard@math.jussieu.fr*

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 us to employ 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 behaves quite similarly 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.

Keywords: maxisets; Muckenhoupt weights; nonparametric regression; random design; warped wavelets; wavelet thresholding

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 and Johnstone (1994) and Donoho *et al*. (1995). In the case of irregular design, various attempts to solve this problem have been made: see, for instance, the interpolation methods of Hall and Turlach (1997) and Kovac and Silverman (2000); the binning method of Antoniadis *et al*. (1997); the transformation method of Cai and Brown (1998), or its recent refinement by Maxim (2002) for a random design; the weighted wavelet transform of Foster (1996); the isometric method of Sardy *et al*. (1999); the penalization method of Antoniadis and Fan (2001); and the specific construction of wavelets adapted to the design of Delouille *et al.* (2001; 2004). See also Penski and Vidakovic (2001).

Our aim here will be to stay as close as possible to the standard thresholding. For a signal observed at some design points (for denoising or other purposes), $Y(t_i)$, $i \in \{1, \ldots, 2^J\}$, if the design is regular ($t_k = k/2^J$), the standard wavelet decomposition algorithm starts with $\alpha_{Jk} = 2^{J/2} Y(k/2^J)$ which approximates $\int Y(x)\phi_{Jk}(x)\mathrm{d}x$. Then the cascade algorithm is employed to obtain the β_{jk} for $j \leqslant J$, which in turn are thresholded. If the design is not regular, and we still employ the *same* algorithm, then for a function H such that $H(k/2^J) = t_k$, we have $\alpha_{Jk} = 2^{J/2} Y(H(k/2^J))$. Roughly, then, what we are

1350–7265

focusing on is in fact the standard expansion of the function $Y(H(x))$, or, if $G \circ H(x) \equiv x$, the decomposition of Y on the 'warped' basis, $Y(x) = \sum_{I=(j,k)} \beta_I \psi_I(G(x))$. In the regression setting, this means replacing the standard wavelet expansion of the function f by its expansion on a new basis $\{\psi_{jk}(G), j, k\}$, where G is adapting to the design: it may be the distribution function of the design, or its estimation, when it is unknown. This obviously creates some new difficulties since $\{\psi_{jk}(G), j, k\}$ is no longer an orthonormal basis. Still, this also has clear advantages: among them, let us emphasize the fact that our procedure is computationally very simple. Compared, for instance, to the transformation method of Cai and Brown (1998), which considers this approach for the finer scale, but then projects on the regular wavelet basis, the calculations are more direct. But overall, by doing this, we stabilize the variance of the estimated coefficients and avoid a necessary but heavy systematic calculation of the threshold of each coefficient.

An appealing feature of this method is that it does not provide a new algorithm. As far as algorithms are concerned, this one uses the standard thresholding procedure and as a consequence the two estimators do coincide at their respective design points. From this perpective, our paper answers the interesting theoretical question: for which class of design densities is it optimal to run the classical wavelet algorithm, ignoring the irregularity of the grid? In this sense this paper can be considered as a generalization of Cai and Brown (1999).

However, considered as functional objects, the two estimators (since they project on very different atoms) are really different, and this will be revealed when we study the behaviour of the warped estimator in various $\mathbb{L}_p$ norms.

Adopting such a point of view shifts the difficulty towards the study of the bias, that is, the approximation properties of the warped wavelet bases $\{\psi_{jk}(G), j, k\}$.

Of course the properties of this basis truly depend on the warping factor G. Obviously, if G is uniform, then $\{\psi_{jk}(G), j, k\}$ is a regular wavelet basis. We will prove that, under a condition on G, we can expect behaviour of the warped basis which is almost as good for statistical purposes as a standard one. As expected, this condition properly quantifies the departure from the uniform distribution and happens to be associated with a notion introduced some years ago in Muckenhoupt (1972) (see also García-Cuerva and Rubio de Francia 1985; Coifman and Fefferman 1974) and widely used afterwards in the context of Calderón–Zygmund theory: the Muckenhoupt weights.

Beyond the study of the regression problem with random design, we wish to emphasize that we investigate here the properties of a new kind of basis, which although definitively not wavelet bases, turn out to behave, at least for statistical concerns, as well as ordinary wavelet bases.

Our results will provide the rate of convergence of the procedures for various $\mathbb{L}_p$ norms under conditions associating the regularity of f with the design G. For instance, in the case where the density g of G is bounded from above and below, we found exactly the same behaviour as in the regular design, except that here the conditions of regularity are formulated on the function $f \circ G^{-1}$.

The assumption of boundedness from above and below for g will not be required in full generality. Moreover, it will be proved that there is a deep connection between Besov 'bodies' constructed on the warped basis and Besov spaces which are modelled on a more general measure than the Lebesgue measure.

This paper is organized as follows. In Section 2 we introduce the notions of warped bases, weighted spaces and their relations. In Section 3 we introduce the model and the different procedures. In Section 4 the notion of Muckenhoupt weight is recalled and its relation with the warping problem is detailed. Section 5 describes the performance of our procedures in terms of minimax or maxiset properties. Section 6 is devoted to the proofs, although some of these are postponed to Sections 7–9. Each of these three sections is of interest in its own right and specifically illustrates one aspect of the paper: Section 7 is devoted to Muckenhoupt weights and associated Besov spaces, Section 8 details the embeddings of Besov bodies in the presence of weight, and Section 9 details the probability bounds, some of which are a little delicate and require the use of very precise concentration inequalities for the empirical process.

2. Warped bases, weighted spaces and properties

2.1. Warped bases and weighted spaces

Consider a compactly supported wavelet basis $\{\psi_{j,k},\, j \geqslant -1,\, k \in \mathbb{Z}\}$ (note that $\psi_{-1,k} = \phi_{0k}$ denotes the scaling function). The usual expansion of a function f in $\mathbb{L}_2(\mathbb{R})$ is

$$f(x) = \sum_{I=(j,k)} \beta_I \psi_I(x), \tag{1}$$

with

$$\beta_I = \int f(x)\psi_I(x)\mathrm{d}x. \tag{2}$$

Suppose now that instead of being in a homogeneous space, we are sitting in a medium $G(x)$ describing the fact that some zones in the space are dense and some are sparse. In such a situation we may find some advantage in replacing (1) by the atomic decomposition

$$f(x) = \sum_{I=(j,k)} \beta_I \psi_I(G(x)). \tag{3}$$

Such 'warped' bases have been considered, for instance, in order to catch certain geometric features or to handle local stationarity; see Clerc and Mallat (2003) and Le Pennec and Mallat (2003). The new family $\{\psi_{j,k}(G),\, j \geqslant -1,\, k \in \mathbb{Z}\}$ is no longer an orthonormal basis. For instance, the coefficients β_I can be calculated (under mild conditions on G) using the following formula, which obviously differs from the standard case (2):

$$\beta_I = \int \psi_I(G(x))f(x)g(x)\mathrm{d}x,$$

where g is the derivative of G.

A closely connected problem is the following. Regular (Sobolev or Besov) spaces are generally defined with respect to the Lebesgue measure. However (again when an inhomogeneity appears in the spaces), it may be more natural to consider other measures, especially measures of the form $\omega(x)\mathrm{d}x$ where ω is a weighting of the space.

In this case, we are interested in considering the standard expansion (1), with a view to measuring its approximation behaviour in the spaces $\mathbb{L}_p(\omega(x)\mathrm{d}x)$. The connection between the two approaches obviously lies in the following formula, where we have used a simple change of variables (also valid under mild conditions on G):

$$\int\left|f(G^{-1}(x)) - \sum_{I\in\Lambda}\lambda_I\psi_I(x)\right|^p \frac{1}{g(G^{-1}(x))}\mathrm{d}x = \int\left|f(u) - \sum_{I\in\Lambda}\lambda_I\psi_I(G(u))\right|^p \mathrm{d}u. \tag{4}$$

We see in this equation that the approximation properties of f in terms of the warped atoms correspond to the approximation properties of $f(G^{-1})$ in terms of the standard wavelet bases, but measured with the weight $\omega(x) = (g(G^{-1}(x)))^{-1}$.

2.2. Properties of atoms

Many properties of the possible atoms that are shared by wavelet bases can be explored. However, we shall concentrate here on two special properties which are sufficient, in the treatment of most statistical applications, to restrict the complexity of the problem to the level of a Hilbertian framework.

Property 1 *Shrinkage (or unconditional) property. There exists an absolute constant K such that if $|\theta_i| \leqslant |\theta_i'|$ for all i, then*

$$\left\|\sum_i \theta_i e_i\right\|_p \leqslant K\left\|\sum_i \theta_i' e_i\right\|_p. \tag{5}$$

Property 1 means, in particular, that by thresholding or shrinking the coefficients we do not risk exploding the $\mathbb{L}_p$-norm, and has many more important properties of its own, as underlined, for instance, in Donoho (1993).

Property 2 *p-Temlyakov property. There exist c_p and C_p such that, for any finite set of integers F, we have*

$$c_p\int\sum_{i\in F}|e_i|^p \leqslant \int\left(\sum_{i\in F}|e_i|^2\right)^{p/2} \leqslant C_p\int\sum_{i\in F}|e_i|^p. \tag{6}$$

This pair of inequalities was introduced in DeVore (1998) and Temlyakov (1998). They are also referred to as 'democratic'. They are obviously true for $p = 2$ when e_i is an othonormal basis and provide a key tool to allow extension from quadratic losses to $\mathbb{L}_p$ ones in most statistical applications (see Kerkyacharian and Picard 2000).

Properties 1 and 2 are true for compactly supported wavelets. They become false in general for warped bases, even for $p = 2$ (warped bases generally are not orthonormal bases). However, we will see in Section 4 that they remain true if we assume good properties on G.

3. Model and estimation procedures

3.1. Wavelet shrinkage

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

$$\hat{f} = \sum_{\{I=(j,k), -1\leqslant j\leqslant J(n)\}} \hat{\beta}_I \psi_I I\{|\hat{\beta}_I| \geqslant \lambda\} \tag{7}$$

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

$$\beta_I = \int f\,\psi_I.$$

Note that procedure (7) is nonlinear since only statistically significant coefficients (e.g. $|\hat{\beta}_I| \geqslant \lambda$) are kept. Here λ is a threshold parameter which depends on the problem at hand. This procedure has been investigated in many cases. For the case of regression with equispaced design,

$$Y_i = f\left(\frac{i}{n}\right) + \varepsilon_i,$$

where the ε_i are independent standard Gaussian random variables; see Donoho and Johnstone (1996), where this estimator has been proposed with the following estimators of the wavelet coefficients:

$$\hat{\beta}_I = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_I\left(\frac{i}{n}\right). \tag{8}$$

For the case with non-equispaced but still fixed design,

$$Y_i = f(U_i) + \varepsilon_i, \tag{9}$$

where U_i is a fixed sequence, non-decreasing with i, many adaptations of this first estimator have been provided. At this stage we mention only Cai and Brown (1998) and Hall and Turlach (1997), which are the closest to the forthcoming discussion.

A common assumption in this context states that the U_i are of the form $G^{-1}(i/n)$, where

G is a known and regular distribution function. In this case, expression (8) can also be written as

$$\hat{\beta}_I = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_I\left(\frac{i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_I(G(U_i)). \tag{10}$$

This formula is important since it allows very simple computations. Our aim here is to stay as close as possible to this formula. It is also used in Cai and Brown (1998), but at the very first step of the calculation to obtain the α_{Jk} coefficients at the finest scale. A comparison between the Cai and Brown procedures and the ones detailed in this paper is given in Section 5.2.

3.2. Regression with random design

Let us now consider the following model. We observe independent variables $Y_1, \ldots, Y_n$, with

$$Y_i = f(X_i) + \varepsilon_i, \tag{11}$$

where X_i and ε_i are independent random variables, and ε_i has a known distribution with density g_0. The X_i are observed, the ε_i are not. Our aim is to estimate the function f. To simplify, we assume that the ε_i are normal variables with mean zero and variance σ^2; σ^2 is assumed to be known or replaced by an estimator. For the sake of simplicity, we take $\sigma^2 = 1$. The X_i 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 is f.

3.3. 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 not so well adapted to the statistical problem, we adopt a different strategy. We warp the wavelet in such a way that in this new basis the estimates of the coefficients are more natural.

If we follow the idea developed above in the non-equispaced but fixed design case and suppose for a while that

$$G(x) = \int_a^x g(u)\mathrm{d}u$$

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

$$\hat{\beta}_I^* = \frac{1}{n}\sum_{i=1}^{n} \psi_I(G(X_i))Y_i \tag{12}$$

is a natural extension of (8).

We have

$$\mathbb{E}(\hat{\beta}_I^*) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\psi_{j,k}(G(X_i))(f(X_i)+\epsilon_i)) = \mathbb{E}(\psi_I(G(X))f(X))$$

$$= \int_a^b \psi_I(G(x))f(x)g(x)\mathrm{d}x = \int_a^b \psi_I(y)f(G^{-1}(y))\mathrm{d}y := \beta_I,$$

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

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

or

$$f(x) = \sum_I \beta_I \psi_I(G(x)), \tag{13}$$

and we can associate with this decomposition the estimate

$$\hat{f}^*(x) = \sum_{j=-1}^J \sum_{k\in\mathbb{Z}} \hat{\beta}_I^* I\{|\hat{\beta}_I^*| \geqslant \kappa t_n\}\psi_I(G(x)), \tag{14}$$

with

$$t_n = \left(\frac{\log n}{n}\right)^{1/2}, \qquad 2^J \sim t_n^{-1}. \tag{15}$$

Obviously, (13) describes an expansion of f in the new basis $\{\psi_{j,k}(G), j \geqslant -1, k \in \mathbb{Z}\}$.

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

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

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

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

Let us also consider the estimator

$$\hat{f}'' = \sum_{j=-1}^J \sum_{k\in\mathbb{Z}} \hat{\beta}'_{jk} I\{|\hat{\beta}'_{jk}| \geqslant \kappa t_n\}\psi_{jk}(\hat{G}_n(x)), \tag{16}$$

again with

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

The thresholding constant κ is, as usual, a tuning constant of the method. The theory will be stated assuming only that κ is 'large enough' to ensure good concentration inequalities. However, it is known in practice that a universal threshold will suffer from oversmoothing especially if it chosen very large. We do not investigate in this paper the influence of the design on the 'optimal' choice of this constant, postponing such practical investigations to a future paper.

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 as follows:

1. Sort the X_i.
2. Change the numbering in such a way that X_i has rank i.
3. Calculate the highest level *alpha* coefficients using the formula
$$\hat{\alpha}_{J'k} = \frac{1}{n}\sum_{i=1}^{n} \phi_{J'k}\left(\frac{i}{n}\right) Y_i, \qquad 2^{J'} = n.$$
4. Calculate the wavelet coefficients using the classical pyramidal algorithm.
5. Employ a thresholding algorithm giving rise to $\tilde{\beta}_{jk}$ coefficients.
6. Reconstruct the estimator, again using 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.

It is worthwhile to notice that $\hat{f}''$ can be considered as *the same* estimator as the standard thresholding estimator (7) using (8) since the algorithms are exactly the same and the two estimators obviously coincide at the design points. So, in this sense, investigating the properties of this procedure is in some sense measuring the robustness of the standard thresholding estimator with respect to the design. Considered as functions, however, they are quite different because they are decomposed into atoms of very different kinds. The aim of this paper will also be to partly investigate the functional aspects, and specifically to study the behaviour, 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 underlying design. It is interesting, at this stage, to notice that there is a slight difference here from the standard setting in the fact that we set $2^J \sim t_n^{-1}$, whereas we usually set $2^L \sim t_n^{-2}$ for the finest level. This will be commented on below. It is also worthwhile to notice that, for technical reasons, the results will be proved not quite for $\hat{f}''$, but for a procedure which is a little less direct from the computation point of view (but still very simple): instead of estimating G over the whole sample, we assume that our number of observation is $2n$, and divide the sample (before sorting the X_i of course!) into two (independent) parts $(X_1, Y_1, \ldots, X_n, Y_n, X_1', Y_1', \ldots, X_n', Y_n')$. This splitting allows us to simplify, in the proofs, the necessary control of terms of the form $\psi_{j,k}(\hat{G}_n) - \psi_{j,k}(G)$ which is already complicated enough. Note that in practice it makes sense to split the data into 'even and odd observations', especially if one suspects for instance a possible change in the design, although in theory the splitting may be done arbitrarily. We use the first part of the data to estimate $G(x)$ by

$$\hat{G}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}(X_i). \tag{17}$$

Then we estimate the wavelet coefficients using the other part of the data,

$$\hat{\beta}^{@}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k}(\hat{G}_n(X'_i))Y'_i, \tag{18}$$

and then consider the estimator

$$\hat{f}^{@} = \sum_{j=-1}^{J}\sum_{k\in\mathbb{Z}} \hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geqslant \kappa t_n\}\psi_{jk}(\hat{G}_n(x)). \tag{19}$$

4. Muckenhoupt weight and warped bases

4.1. Muckenhoupt weight

Let us first recall the following notions:

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

$$\left(\frac{1}{|I|}\int_I \omega(x)\mathrm{d}x\right)^{1/p}\left(\frac{1}{|I|}\int_I \omega(x)^{-q/p}\,\mathrm{d}x\right)^{1/q} \leqslant C.$$

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

$$\omega^*(x) \leqslant C\omega(x) \qquad \textit{almost everywhere},$$

where $\omega^*(x)$ is the Hardy–Littlewood maximal function.

For $p = \infty$, we set

$$A_\infty = \bigcup_{p\geqslant 1} A_p.$$

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 with f is*

$$f^*(x) = \sup_{I\in\mathcal{B},x\in I}\left(\frac{1}{|I|}\int_I |f(u)|\mathrm{d}u\right).$$

The concept of Muckenhoupt weight was introduced in Muckenhoupt (1972) (see also García-Cuerva and Rubio de Francia 1985; Coifman and Fefferman 1974) and widely used subsequently in the context of Calderón–Zygmund theory. It is easy to observe that the Muckenhoupt spaces form an increasing family when p varies. Many functions belong to one of these classes. Of course, if ω is bounded from above and below, it belongs to A_1 (and so to any A_p), but ω can also approach zero. For instance $w(x) = |x|^a$ belongs to A_p for $-1 < a < p-1$.

We see in the definition that this property in some sense quantifies how close ω is to a uniform weight, where the function and its inverse evenly weight each interval. Some of the important properties of these classes of functions will be recalled in Section 7. In what follows we will assume the following condition:

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

(We recall that $G(x) = \int_a^x g(u)\mathrm{d}u$.) This will be proved to be equivalent (see Proposition 9) to the condition that there exists C such that, for all intervals $I \subset [a, b]$,

$$\left(\frac{1}{|I|}\int_I g(x)^q \,\mathrm{d}x\right)^{1/q} \leqslant C\frac{1}{|I|}\int_I g(x)\mathrm{d}x, \qquad 1/p + 1/q = 1.$$

Again, these conditions are obviously true when the design g is uniform or uniformly bounded from above and below. More generally, they obviously quantify the usual assumption that the design gives enough mass to any interval. To fix the ideas, if we take the example of the density $g(x) = (\alpha+1)x^\alpha$ on the interval [0, 1], it satisfies $(\mathcal{H}_p)$ for $\alpha > -1 + 1/p$. We see in this example that this condition does not require the density to be bounded from above or from below.

4.2. Properties of the warped wavelet basis

As shown in formula (13), our construction builds on the new ‘basis’ $\{\psi_{jk}(G(\cdot)), j \geqslant -1, k \in \mathbb{Z}\}$.

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

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

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

$$S(T(x)) = x, \text{ a.e.}, \qquad T(S(x)) = x, \text{ a.e.};$$

$$\text{for any non-negative measurable function } h, \qquad \int_{\mathbb{R}} h(T(x))\mathrm{d}x = \int_{\mathbb{R}} h(y)\omega(y)\mathrm{d}y. \tag{20}$$

We now state the following theorem from Kerkyacharian and Picard (2002):

Theorem 1. *Under conditions* (20)*, the family* $\{\psi_{jk}(T(\cdot)), j \geqslant -1, k \in \mathbb{Z}\}$ *satisfies Properties 1 and 2.*

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 holds (cf. Gol'dshtein and Reshetnyak 1983):

$$\text{for any non-negative measurable function,} \qquad \int_{\mathbb{R}} h(x)\omega(x)\mathrm{d}x = \int_{\mathbb{R}} h(T(y))\mathrm{d}y,$$

where ω is the Jacobian of S, that is, $\omega(y) = (g(G^{-1}(y)))^{-1}$. Then we see that our assumption $(\mathcal{H}_p)$ states precisely that (20) is realized, with $\omega \in A_p$.

4.3. Weighted Besov spaces

It is natural in this context, if we wish 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, and it is to this that we now turn. 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 so on, for $\Delta_h^N f(x)$, $N \in \mathbb{N}_*$.

Let

$$\rho^N(t, f, \omega, p) = \sup_{|h| \leqslant t} \left(\int |\Delta_h^N f(u)|^p \omega(u)\mathrm{d}u \right)^{1/p},$$

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

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

The only difference from the usual Besov spaces is the fact that the modulus of continuity ρ^N is calculated weighting the space with the measure $\omega(x)\mathrm{d}x$ instead of the Lebesgue measure corresponding to the function $\omega = 1$. One of the major advantages of regular Besov spaces is that they can be expressed in terms of wavelet coefficients: under standard oscillating conditions on the wavelet ψ (see, for instance, Meyer 1990), we have, for $f = \sum_{j \geqslant -1, k} \beta_{jk} \psi_{j,k}$,

$$\left(\int_0^1 \left(\frac{\rho^N(t, f, 1, p)}{t^s} \right)^q \frac{\mathrm{d}t}{t} \right)^{1/q} < \infty \Leftrightarrow \left[\sum_{j \geqslant -1} \left\{ 2^{js} 2^{j/2} \left(\sum_{k \in \mathbb{Z}} |\beta_{j,k}|^p 2^{-j} \right)^{1/p} \right\}^q \right]^{1/q} < \infty.$$

We show in the following proposition that, under conditions on ω, the direct sense of the implication is still true if, in the sum over k, 2^{-j} is replaced by $\omega([k/2^j, (k+1)/2^j])$. It is worthwhile to notice that, in fact, in most statistical applications, only the direct sense is used.

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

$$\omega(I) = \int_I \omega(x)\mathrm{d}x.$$

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

$$\int \psi(x)x^k \,\mathrm{d}x = 0, \qquad k = 0, \ldots, N-1,$$

then for

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

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

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

This proposition is proved in Section 7. A corollary 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 so on for $\Delta_h^N(G)f(x)$, $N \in \mathbb{N}_*$, and again,

$$\tilde{\rho}^N(t, f, G, p) = \sup_{|h|\leqslant t}\left(\int |\Delta_h^N(G)f(u)|^p \,\mathrm{d}u\right)^{1/p}.$$

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 spaces

$$B^G_{s,p,q} = \left\{f : \left(\int_0^1 \left(\frac{\tilde{\rho}^N(t, f, G, p)}{t^s}\right)^q \frac{\mathrm{d}t}{t}\right)^{1/q} < \infty\right\}.$$

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

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

The following corollary concerns the representation of spaces $B^G_{s,p,q}$ in terms of coefficients in the expansion using the warped basis.

Corollary 1. *Under the conditions of Proposition 1 and condition $(\mathcal{H}_p)$, for*

$$f = \sum_{j,k} \beta_{jk}\psi_{j,k}(G)$$

(*i.e.* $\beta_{jk} = \int [f \circ G^{-1}] \psi_{j,k}$), *we have*

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

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

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

5. Performance of the estimation procedures

The properties of the procedures $\hat{f}^*$ (corresponding to the case where G is assumed to be known) and $\hat{f}^@$ (where G is not known) will be expressed in two different ways. The first, developed in Section 5.1, is commonly used. It consists of 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 (see Theorem 3 and Proposition 5 below) consists of determining the maxiset of the procedures.

5.1. Minimax properties of the procedures $\hat{f}^*$ and $\hat{f}^@$

Theorem 2. *Assume that we observe model* (11), *with* f *bounded and* g *satisfying conditions* ($\mathcal{H}_p$), *where* $p > 1$, $\pi \geqslant p$ *are given real numbers, and that the two estimators* $\hat{f}^*$ *and* $\hat{f}^@$ *are those defined in* (14) *and* (19). *Then*

$$\mathbb{E}\|\hat{f}^* - f\|_p^p \leqslant C \left[\frac{n}{\log n} \right]^{-\alpha} \qquad \text{if } f \in B^G_{s,\pi,\infty}, \quad s \geqslant \tfrac{1}{2}. \tag{21}$$

If, in addition, $f \circ G^{-1}$ *belongs to the space* $lip_{1/2}$ *of Hölderian functions of coefficient* $\frac{1}{2}$,

$$\mathbb{E}\|\hat{f}^@ - f\|_p^p \leqslant C \left[\frac{n}{\log n} \right]^{-\alpha} \qquad \text{if } f \in B^G_{s,\pi,\infty}, \quad s \geqslant \tfrac{1}{2}, \tag{22}$$

where

$$\alpha = \frac{sp}{1 + 2s}. \tag{23}$$

A number of remarks and comments are in order. The rates of convergence obtained here for $\hat{f}^*$ correspond to the rates which were proved to be minimax in a uniform design, up to logarithmic factors. Notice, however, that we do not observe the so-called 'elbow' here, that is, the division of the set of parameters into a sparse and a dense region with different rates of convergence, as occurs in the case of regular design (see, for instance, Donoho *et al.*

1995). 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 the same as for $\hat{f}^{*}$, except that we need an additional Lipschitz condition on $f \circ G^{-1}$ in the first case. We do not know if this is necessary.

The limitation $s \geqslant \frac{1}{2}$ is standard in the regression setting. Let us observe that this restriction appears in our choice of J. In standard thresholding (standard denoising or density estimation, for instance) one usually sets 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 especially necessary to obtain the exponential inequalities of Proposition 3.

If we need to be more explicit, we can also express the results in terms of 'regular' Besov spaces. This can be done if we are ready to impose more restrictive assumptions on the underlying design (e.g. its density is bounded from above and below). In this case, we have the following proposition:

Proposition 2. *Assume that we observe model* (11)*, with f bounded and g satisfying $0 < m \leqslant g \leqslant M < \infty$. If $p > 1$, $\pi \geqslant 1$ are given real numbers and if the two estimators $\hat{f}^{*}$ and $\hat{f}^{@}$ are those defined in* (14) *and* (19)*, then*

$$\mathbb{E}\|\hat{f}^{*} - f\|_p^p \leqslant C\left[\frac{n}{\log n}\right]^{-\alpha(s,r)} \qquad \text{if } f \circ G^{-1} \in B_{s,\pi,r}, \quad s \geqslant \tfrac{1}{2}. \tag{24}$$

If, in addition, $f \circ G^{-1}$ belongs to the space $lip_{1/2}$ of Hölderian functions of coefficient $\frac{1}{2}$, then

$$\mathbb{E}\|\hat{f}^{@} - f\|_p^p \leqslant C\left[\frac{n}{\log n}\right]^{-\alpha(s,r)} \qquad \text{if } f \circ G^{-1} \in B_{s,\pi,r}, \quad s \geqslant \tfrac{1}{2}, \tag{25}$$

where

$$\alpha(s, r) = \begin{cases} \alpha = \dfrac{sp}{1+2s}, & \text{for } s > \dfrac{p-\pi}{2\pi},\ r \in [1, \infty], \\[2ex] \dfrac{(s - 1/\pi + 1/p)p}{1 + 2(s - 1/\pi)}, & \text{for } \dfrac{1}{2} + \dfrac{1}{\pi} \leqslant s \leqslant \dfrac{p-\pi}{2\pi},\ r \leqslant \dfrac{p-2}{2(s-1/\pi)+1}. \end{cases} \tag{26}$$

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

The proofs of Theorem 2 and Proposition 2 will be given in the next section.

5.2. Illustration of the minimax results

We now illustrate the above results by a comparison of our procedure with the more sophisticated procedure provided in Cai and Brown (1998). Let us first explain why the Cai–Brown procedure is more difficult to implement: it consists of starting at the finest level with the same estimate – using formula (12) at the highest level J' ($2^{J'} = n$) – and then projecting this high-resolution estimate onto the lower j levels. But as this projection is performed on the true wavelet basis instead of the warped one, the cascade algorithm is no longer usable. This creates two difficulties: first, the algorithm is much more involved; and second, the variance of the estimated wavelet coefficients is no longer constant and not so easy to calculate. As this variance enters into the threshold, it has to be calculated for each coefficient, and this creates another source of difficulty in the procedure.

Among the advantages of our results let us emphasize that we obtain the rates of convergence for various $\mathbb{L}_p$ losses and a wider class of regularity spaces, as well as much lighter conditions on the design. This allows us, for instance, in the case where g is bounded from above and below, to find the elbow again as in the uniform design case, as described in Section 5.1.

Among the drawbacks let us just mention that the results are expressed in terms of $B^G_{s,p,q}$-spaces, which are spaces mixing the regularity of f with the design G. For instance, (22), which considers the case where G is unknown, states the rate $(n/\log n)^{-sp/(1+2s)}$ if $f(G^{-1})$ belongs to the space $B_{s,\infty,\infty}$, when comparatively close (for $p=2$) results in Cai and Brown (1998) are expressed in terms of f belonging to $B_{s,\infty,\infty}$. If we again take the example where

$$g(x) = (\alpha+1)x^{\alpha} I\{[0, 1]\}(x),$$

condition $\mathcal{H}(p)$ is satisfied for $\alpha > -1 + 1/p$ and $f(G^{-1})(x) = f(x^{1/(\alpha+1)})$. Obviously, if, for instance, $s \leqslant k$ and $\alpha \leqslant -1 + 1/k$, $f \in B_{s,\infty,\infty} \Rightarrow f(G^{-1}) \in B_{s,\infty,\infty}$. But in general this may not be true. In fact, this is precisely the type of cases where the Cai–Brown procedure may be more accurate, since estimating $f(G^{-1})$ may lead to loss of regularity. As a consequence, a procedure using this estimation only at the higher scales (as in Cai and Brown, only using the formula (12) at the highest level J' ($2^{J'} = n$)) where the bias is less important, obviously leads to more computations but might exhibit more precise behaviour at this point.

5.3. Maxiset properties 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 are probability distributions and 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:*

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

This way of measuring the performance of procedures has been particularly sucessful in the nonparametric framework (see, for example, Cohen *et al.* 2001; Kerkyacharian and Picard 2000; Rivoirard 2002). 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 upper bound inequalities in the minimax comparisons. And, for instance, the following Theorem 3 providing the maxisets for the procedure $\hat{f}^*$ will be a key tool for the minimax results. Indeed, using Theorem 3, we are able to deduce the rates of convergence of $\hat{f}^*$ over a large number of regularity classes just by proving their inclusion in the maxiset. Then we deduce the results for the more general procedure $\hat{f}^{@}$ by taking advantage of the proximity of $\hat{f}^*$ and $\hat{f}^{@}$ when n is large (Theorem 2).

We now need to introduce the following notation. Let us suppose that ν is the following measure for $j \in \mathbb{N}$, $k \in \mathbb{Z}$:

$$\nu\{(j, k)\} = 2^{jp/2}\omega(I_{jk}).$$

We define the following function spaces:

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

Theorem 3. *Let $p > 1$, $0 < q < p$. Under condition $(\mathcal{H}_p)$, the maxiset of the estimator $\hat{f}^*$,*

$$\text{Max}(q) = \left\{ f, \mathbb{E}\|\hat{f}^* - f\|_p^p \left(\frac{\log n}{n}\right)^{(q-p)/2} < \infty \right\}, \tag{27}$$

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

$$\text{Max}(q) = l_{q,\infty}(\nu) \cap \left\{ f = \sum_I \beta_I \psi_I \circ G, \sup_{l \geqslant 0} \left\| \sum_{j \geqslant l, k} \beta_{jk} \psi_{jk} \circ G \right\|_p^p 2^{l(p-q)} < \infty \right\}. \tag{28}$$

5.4. Illustration of the maxiset properties

Determining the maxiset of a procedure may have various applications. First, it can help in the comparison of the procedure with others, by proving for instance that its maxiset is systematically larger. This presupposes the calculation of the maxiset of the other procedures. It might have been interesting in our context to also calculate the maxiset of the Cai–Brown procedure, but this is beyond the scope of this paper. The second purpose is to use the maxiset as a very powerful tool to prove rates of convergence on specified functional spaces (such as Besov spaces) just by proving their inclusions in the maxiset. This aspect will be extensively used in the following sections. The third aspect is more descriptive, and helps to understand the particular shape of functions which are well estimated (or, on the contrary, badly estimated) by the procedure. In our case it is natural to compare our results with the maxiset of the standard wavelet thresholding (see Kerkyacharian and Picard 2000) since, as explained earlier, this will help us to understand how far we can go with the standard wavelet thresholding instead of turning to more sophisticated procedures when the design is irregular. The previous result shows that in fact the properties of the function are clearly expressed in terms not of f itself, but of $f \circ G^{-1}$. So the design appears once in this aspect, and as we have precisely calculated the maxiset we see that the appearance of $f \circ G^{-1}$ is in some sense unavoidable. But the design also appears in the following way. As in the case of regular design, the functions which are well estimated are characterized by the fact that their 'number' of large wavelet coefficients is small. The way this number is calculated is reflected in the definition of the space $l_{q,\infty}(\nu)$. We see that the design again makes an appearance here through $\omega(I_{jk})$, which might greatly influence the counting. We also see that the closer we are to the uniform design ($\omega(I_{jk}) = 2^{-j}$), the better interpretable this counting is.

6. Proofs

6.1. Proof of Theorem 3

This proof takes advantage of the following Theorem 4 taken from and proved in Kerkyacharian and Picard (2000). The aim of the theorem is to determine the 'maxisets' of the thresholding methods for a completely general basis. It will be applied to obtain Theorem 3.

Let us introduce the following notation. Let $\{e_{jk},\ j \geqslant -1,\ k \in \mathbb{N}\}$ be a set of functions in $L^p(\mathbb{R})$. ν will denote the measure such that, for $j \in \mathbb{N}$, $k \in \mathbb{Z}$,

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

We also define the function 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 4. *Let $p > 1$, $0 < q < p$. Suppose that $\{e_{jk},\ j \geqslant -1,\ k \in \mathbb{N}\}$ satisfies Properties 1 and 2. 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*

$$\sup_n \nu\{\Lambda_n\} c(n)^p < \infty. \tag{29}$$

We suppose in addition that, for any pair (j, k) in Λ_n, there exists an estimator $\hat{\beta}_{jk}$ such that

$$\mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leqslant Cc(n)^{2p} \tag{30}$$

and

$$P\Big(|\hat{\beta}_{jk} - \beta_{jk}| \geqslant \kappa c(n)/2 \Big) \leqslant Cc(n)^{2p} \wedge c(n)^4. \tag{31}$$

Then the thresholding estimator

$$\hat{f} = \sum_{(j,k) \in \Lambda_n} \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geqslant \kappa c(n)\} e_{jk} \tag{32}$$

is such that there exists $C > 0$ such that,

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

if and only if

$$f \in l_{q,\infty}(\nu)$$

and

$$\sup_n c(n)^{q-p} \left\| f - \sum_{(j,k)\in\Lambda_n} \beta_{jk} e_{jk} \right\|_p^p < \infty.$$

In using Theorem 4 to prove Theorem 3, we make the following observations:

1. The principal result of Theorem 4 says that the maxiset of the procedure $\hat{f}$,

$$\mathrm{Max}(q) = \{f, \mathbb{E}\|\hat{f}^* - f\|_p^p (c(n))^{q-p} < \infty\}$$

$$= l_{q,\infty}(\nu) \cap \left\{ f = \sum \beta_{jk} e_{jk} \sup_n c(n)^{q-p} \left\| f - \sum_{(j,k)\in\Lambda_n} \beta_{jk} e_{jk} \right\|_p^p < \infty \right\}.$$

2. This theorem will be applied to obtain Theorem 3 with

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

   The basis satisfies Properties 1 and 2 because of condition $(\mathcal{H}_p)$ and Theorem 1.
3. The estimators of the coefficients will be taken to be $\hat{\beta}^*_{jk}$. It will be proved in the following Proposition 3 that inequalities (30) and (31) hold with

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

4. It will be proved in Section 7 (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 condition (29) is satisfied if

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

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

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

$$\mathbb{E}(|\hat{\beta}_{jk} - \beta_{jk}|^p) \leqslant C_p \frac{1 + \|f\|_\infty^p}{n^{p/2}}, \qquad \text{for } 2^j \leqslant n, \tag{33}$$

$$P\left(|\hat{\beta}_{jk} - \beta_{jk}| > \kappa\sqrt{\frac{\log n}{n}}\right) \leqslant C'_p n^{-\gamma p} \qquad \text{for } \kappa \geqslant \kappa_\gamma,\ 2^j \leqslant \sqrt{\frac{n}{\log n}}. \tag{34}$$

Proposition 3 is proved in Section 9.

6.2. Proof of Theorem 2 and Proposition 2

To prove Theorem 2 and Proposition 2, we begin by investigating the behaviour of $\hat{f}^*$. The first step (proving (21) or (24)), consists of proving that the space $B^G_{s,\pi,r}$ is included in the maxiset Max(q) with q properly chosen to obtain the prescribed rate of convergence (i.e. $\alpha(s) = (p-q)/2$) and the corresponding behaviour for (24).

6.2.1. *Proof of inequalities* (21) *and* (24)

Proposition 4. *For* $p > 1, s \geqslant \frac{1}{2}, s > (p-\pi)/2\pi,\ r \geqslant 1,$ *for* $q = p/(1+2s),$ *we have, if* g *satisfies condition* $(\mathcal{H}_p)$,

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

For $p > 1,\ s \geqslant \frac{1}{2},\ s \leqslant (p-\pi)/2\pi,$ *if furthermore* $0 < m \leqslant g \leqslant M < \infty$, *for* $q = (p-2)/(2(s-1/\pi)+1), r \leqslant q,$ *we have*

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

This proposition is proved in Section 8.

6.2.2. *Proof of inequalities* (22) *and* (25)

Now that our 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 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 levels of difficulty: first replacing $\hat{\beta}^*$ by $\hat{\beta}^@$, then replacing G by $\hat{G}_{n/2}$ in $\psi_I(G)$. It will be seen that the first step is far less difficult than the second, which deals with random atoms and will require fine concentration inequalities.

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

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

$\hat{f}'$ is intermediate between $\hat{f}^@$ and $\hat{f}^*$. The difference between $\hat{f}^@$ and $\hat{f}'$ only lies in 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 is to investigate the behaviour of $\hat{f}'$ by proving the following proposition, using a technique similar to that used for $\hat{f}^*$.

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

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

is such that

$$\mathrm{Max}'(q) = \mathrm{Max}(q).$$

The proof of this result exactly mimics the proof of the result concerning $\hat{f}^*$. The only problem is showing that we have a result similar to Proposition 3 if we replace $\hat{\beta}^*$ by $\hat{\beta}^@$.

Proposition 6. *Let us suppose* $\|f\|_\infty \leqslant D, \|f \circ G^{-1}\|_{lip_{1/2}} \leqslant D$. *There exist constants* C_p, C'_p *such that for any* $\gamma > 0$ *there exists* $\kappa(\gamma, D)$ *such that*

$$\mathbb{E}(|\hat{\beta}^@_{j,k} - \beta_{j,k}|^p) \leqslant C_p \frac{1 + D^p}{n^{p/2}}, \qquad \text{for } 2^j \leqslant \sqrt{\frac{n}{\log n}}, \tag{35}$$

$$P\left(|\hat{\beta}^@_{j,k} - \beta_{j,k}| > \kappa\sqrt{\frac{\log n}{n}} \right) \leqslant C'_p n^{-\gamma p}, \qquad \text{for } \kappa \geqslant \kappa(\gamma, D),\ 2^j \leqslant \sqrt{\frac{n}{\log n}}. \tag{36}$$

Proposition 6 is proved in Section 9.

Evaluating the difference $\hat{f}^@ - \hat{f}'$ The second part of the proof involves evaluating the difference

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

Proposition 7. *Under condition* (20)*, if* f *is bounded and such that* $f(G^{-1})$ *is in the space* $lip_{1/2}$*, then, for* $s \geqslant \frac{1}{2}$,

$$\mathbb{E}(\|\hat{f}^@ - \hat{f}'\|^p_{\mathbb{L}_p(\mathbb{R})}) \leqslant C\|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)},$$

where C *only depends on universal constants and the sup and* $lip_{1/2}$ *norms of* $f(G^{-1})$.

This proposition is proved in Section 9.

7. Muckenhoupt weights and Besov spaces

7.1. Definitions

The definition of a Muckenhoupt weight was given in Section 4.1. There are several equivalent definitions which are well known (see Stein 1993). We give here another important one, together with the very helpful 'doubling property'.

Proposition 8. *If* I *denotes a bounded interval of* $\mathbb{R}$*, and* $|I|$ *its Lebesgue measure, for*

$1 \leqslant p < \infty$ and q such that $1/p + 1/q = 1$, ω a non-negative locally integrable function, the following statements are equivalent:

(i) *$\omega \in A_p$, that is,*

$$\forall I, \qquad \left(\frac{1}{|I|}\int_I \omega\right)^{1/p}\left(\frac{1}{|I|}\int_I \omega^{-q/p}\right)^{1/q} \leqslant C < \infty \tag{37}$$

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

(ii) *For any measurable function f,*

$$\left(\frac{1}{|I|}\int_I |f|\right) \leqslant C\left(\frac{1}{\omega(I)}\int_I |f|^p \omega\right)^{1/p} \tag{38}$$

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

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

$$\omega(2I) \leqslant (2C)^p \omega(I). \tag{39}$$

Proof. Inequality (38) easily implies (37), taking $f = \omega^{-q/p}$. To prove that (37) implies (38), we apply the Hölder inequality to $|f| = (|f|\omega^{1/p})(\omega^{-1/p})$:

$$\left(\frac{1}{|I|}\int_I |f|\right) \leqslant \left(\frac{1}{|I|}\int_I |f|^p\omega\right)^{1/p}\left(\frac{1}{|I|}\int_I \omega^{-q/p}\right)^{1/q} \leqslant C\left(\frac{1}{|I|}\int_I |f|^p\omega\right)^{1/p}\left(\frac{1}{|I|}\int_I \omega\right)^{-1/p}.$$

Applying (38) with $2I$ instead of I and $f = 1_I$, we obtain (39). □

7.2. Muckenhoupt weight and densities

We prove the following proposition:

Proposition 9. *Let $1 \leqslant p < \infty$. Let g be a density on $[a, b]$ and $G(x) = \int_a^x g(s)\mathrm{d}s$ be the associated partition function. Suppose that G is strictly increasing from $[a, b]$ to $[0, 1]$. The following statements are equivalent:*

(i) $(g(G^{-1}(t)))^{-1} \in A_p([0, 1])$.

(ii) *For q such that $1/p + 1/q = 1$, for any J subinterval of $[a, b]$, we have*

$$\left(\frac{1}{|J|}\int_J g(s)^q \,\mathrm{d}s\right)^{1/q} \leqslant C\left(\frac{1}{|J|}\int_J g(s)\mathrm{d}s\right).$$

Proof. Since G is strictly increasing and continuous from $[a, b]$ to $[0, 1]$, we have a natural one-to-one correspondence between the intervals of $[a, b]$ and those of $[0, 1]$. The result is an obvious consequence of the following lemma (with $Y = [0, 1]$, $X = [a, b]$, $\mathrm{d}m$ and $\mathrm{d}\nu$ the Lebesgue measure). □

Lemma 1. *Let (X, m) and (Y, ν) be two measure spaces, $G : X \mapsto Y$ a measurable function. Let us suppose that,*

$$\text{for all non-negative measurable functions } F, \qquad \int_X F(G(x))\mathrm{d}m(x) = \int_Y F(y)\omega(y)\mathrm{d}\nu(y).$$

Let us put $g(x) = \omega(G(x))^{-1}$ and let $\mathcal{Q}$ be a class of measurable subsets of Y and $\tilde{\mathcal{Q}} = G^{-1}(\mathcal{Q})$. Then we have an equivalence between the two following statements:

$$\forall Q \in \mathcal{Q}, \qquad \left(\frac{1}{\nu(Q)}\int_Q \omega(y)\mathrm{d}\nu(y)\right)^{1/p}\left(\frac{1}{\nu(Q)}\int_Q \omega^{-q/p}(y)\mathrm{d}\nu(y)\right)^{1/q} \leqslant C, \tag{40}$$

$$\forall \tilde{Q} \in \tilde{\mathcal{Q}}, \qquad \left(\frac{1}{m(\tilde{Q})}\int_{\tilde{Q}} g(x)^q \mathrm{d}m(x)\right)^{1/q} \leqslant C\left(\frac{1}{m(\tilde{Q})}\int_{\tilde{Q}} g(x)\mathrm{d}m(x)\right). \tag{41}$$

Proof. Inequality (40) is equivalent to,

$$\forall Q \in \mathcal{Q}, \qquad \left(\int_Q \omega(y)\mathrm{d}\nu(y)\right)^{1/p}\left(\int_Q \omega^{-q/p}(y)\mathrm{d}\nu y)\right)^{1/q} \leqslant C\nu(Q)$$

and (41) to,

$$\forall \tilde{Q} \in \tilde{\mathcal{Q}}, \qquad \left(\int_{\tilde{Q}} g(x)^q \mathrm{d}\mu(x)\right)^{1/q} \leqslant Cm(\tilde{Q})^{-1/p}\left(\int_{\tilde{Q}} g(x)\mathrm{d}m(x)\right).$$

But, as $1_Q(G(x)) = 1_{\tilde{Q}}(x)$,

$$\int_Q \omega(y)\mathrm{d}\nu(y) = \int_Y \omega(y)1_Q(y)\mathrm{d}\nu(y) = \int 1_Q(G(x))\mathrm{d}m(x) = m(\tilde{\mathcal{Q}}),$$

$$\int_Q \omega^{-q/p}(y)\mathrm{d}\nu(y) = \int_Y \omega^{-q}(y)\omega(y)1_Q(y)\mathrm{d}\nu(y) = \int_X \omega^{-q}(G(x))1_Q(G(x))\mathrm{d}m(x)$$
$$= \int_{\tilde{Q}} g(x)^q \mathrm{d}m(x)$$

and

$$\nu(Q) = \int_Y \omega(y)\omega(y)^{-1}1_Q(y)\mathrm{d}\nu(y) = \int_X \omega(G(x))^{-1}1_Q(G(x))\mathrm{d}m(x) = \int_{\tilde{Q}} g(x)\mathrm{d}m(x).$$

□

7.3. Weighted spaces, wavelets and approximation

The aim of the rest of this section is to prove Proposition 1. We necessarily begin by expressing the $\mathbb{L}_p$ norm of linear combinations of wavelets at the same resolution level.

In this subsection ϕ is a compactly supported scaling function of a multiresolution analysis and ψ an associated compactly supported wavelet. We fix the notation as follows:

$$\mathrm{supp}(\phi) \subset [0, L], \qquad \mathrm{supp}(\psi) \subset [0, L];$$
$$\hat{\phi}(\xi) = m_0(\xi/2)\mathcal{F}(\phi)(\xi/2), \qquad \hat{\psi}(\xi) = m_1(\xi/2)\mathcal{F}(\phi)(\xi/2), \tag{42}$$

where $\hat{g}$ denotes 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

$$g_{j,k}(x) = 2^{j/2} g(2^j x - k), \quad I_{j,k} = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right], \quad \tilde{I}_{j,k} = \left[\frac{k}{2^j}, \frac{k+L}{2^j}\right],$$

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

$$\alpha_{j,k} = \int f(x)\phi_{j,k}(x)\mathrm{d}x, \qquad \beta_{j,k} = \int f(x)\psi_{j,k}(x)\mathrm{d}x;$$

$$P_j f = \sum_k \alpha_{j,k}\phi_{j,k} = P_{V_j} f, \qquad P_{j+1} f - P_j f = P_{W_j} f = \sum_k \beta_{j,k}\psi_{j,k}.$$

The following theorem states 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 appears in the sum.

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

(i) *There exists C, depending only on ϕ, ψ and ω, such that*

$$\frac{1}{C}\sum_k |\alpha_{j,k}|^p \omega(I_{j,k}) \leqslant 2^{-jp/2} \left\| \sum_k \alpha_{j,k}\phi_{j,k} \right\|^p_{\mathbb{L}_p(\omega)} \leqslant C \sum_k |\alpha_{j,k}|^p \omega(I_{j,k}), \tag{43}$$

$$\frac{1}{C}\sum_k |\beta_{j,k}|^p \omega(I_{j,k}) \leqslant 2^{-jp/2} \left\| \sum_k \beta_{j,k}\psi_{j,k} \right\|^p_{\mathbb{L}_p(\omega)} \leqslant C \sum_k |\beta_{j,k}|^p \omega(I_{j,k}). \tag{44}$$

(ii) *We have*

$$\forall j \in \mathbb{Z}, \qquad \|P_j f\|_{\mathbb{L}_p(\omega)} \leqslant C^2 \|f\|_{\mathbb{L}_p(\omega)}, \tag{45}$$

$$\lim_{j\to\infty} \|P_j f - f\|_{\mathbb{L}_p(\omega)} = 0. \tag{46}$$

(iii) *Let $0 < q \leqslant \infty$ and $f \in \mathbb{L}_p(\omega)$. Then*

$$\left[\sum_j (2^{js}\|P_j f - f\|_{\mathbb{L}_p(\omega)})^q\right]^{1/q} < \infty \Longleftrightarrow \left[\sum_j \left[2^{js}2^{j/2}\left(\sum_{k\in\mathbb{Z}} |\beta_{j,k}|^p \omega(I_{j,k})\right)^{1/p}\right]^q\right]^{1/q} < \infty, \tag{47}$$

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

This theorem is the consequence of the following lemmas.

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

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

and for $p = \infty$,

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

Proof. The main tool of this proof is the doubling property (39) of the measure $\omega(x)\mathrm{d}x$. For $p = \infty$ the result is obvious. We give separate proofs for $1 < p < \infty$ and $0 < p \leqslant 1$.

Let $1 < p < \infty$. As θ is a bounded function, with support in $[0, L]$, $\theta_{j,k}$ is supported in $\tilde{I}_{j,k}$ and there exists $C < \infty$ such that $\sum_k |\theta(x-k)| \leqslant C$. Hence,

$$\begin{aligned} \left| \sum_{k\in\mathbb{Z}} \lambda_{j,k}\theta_{j,k}(x) \right|^p &\leqslant 2^{jp/2} \left(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p |\theta(2^j x - k)| \right) \left(\sum_{k\in\mathbb{Z}} |\theta(2^j x - k)| \right)^{p/q} \\ &\leqslant C^{p/q} 2^{jp/2} \left(\sum_{k\in\mathbb{Z}} |\lambda_{j,k}|^p |\theta(2^j x - k)| \right) \end{aligned}$$

and

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

We finish the proof using the doubling property (39), which implies

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

Finally let $0 < p \leqslant 1$. Then

$$\int\Big|\sum_{k\in\mathbb{Z}}\lambda_{j,k}\theta_{j,k}(x)\Big|^p\omega(x)\mathrm{d}x\leqslant\sum_{k\in\mathbb{Z}}|\lambda_{j,k}|^p\int_{\tilde I_{j,k}}|\theta_{j,k}(x)|^p\omega(x)\mathrm{d}x$$

$$\leqslant\sum_{k\in\mathbb{Z}}|\lambda_{j,k}|^p\|\theta\|_\infty^p 2^{jp/2}\omega(\tilde I_{j,k})$$

$$\leqslant c\|\theta\|_\infty^p 2^{jp/2}\sum_{k\in\mathbb{Z}}|\lambda_{j,k}|^p\omega(I_{j,k}).$$

□

Lemma 3. *For* $1\leqslant p\leqslant\infty$, $\omega\in A_p$, *we have*

$$2^{j/2}\left(\sum_k\left|\int f\bar\phi_{j,k}\mathrm{d}x\right|^p\omega(I_{j,k})\right)^{1/p}\leqslant C\|f\|_{\mathbb{L}_p(\omega)},$$

with the obvious modification if $p=\infty$. *The same inequality holds if we replace* ϕ *by* ψ.

Proof. The main tool here is property (38). We have

$$2^{jp/2}\sum_k\left|\int f\bar\phi_{j,k}\mathrm{d}x\right|^p\omega(I_{j,k})\leqslant 2^{jp/2}\sum_k\left(\int_{\tilde I_{j,k}}|f||\phi_{j,k}|\mathrm{d}x\right)^p\omega(I_{j,k})$$

$$\leqslant C2^{jp/2}\sum_k\frac{1}{\omega(\tilde I_{j,k})}\int|f\phi_{j,k}|^p\omega(x)\mathrm{d}x\,\omega(I_{j,k})|\tilde I_{j,k}|^p$$

$$\leqslant C'2^{-jp/2}\int|f(x)|^p\sum_k 2^{jp/2}|\phi(2^jx-k)|^p\omega(x)\mathrm{d}x$$

$$\leqslant C''\int|f(x)|^p\omega(x)\mathrm{d}x,$$

using $|\tilde I_{j,k}|\sim 2^{-j}$ and the doubling property (39). Of course ϕ and ψ can be exchanged. □

Proof of Theorem 5. From Lemmas 2 and 3 we deduce (43) and (44).

Using these lemmas we also deduce (45):

$$\|P_jf\|_{\mathbb{L}_p(\omega)}=\left\|\sum_k\int f(y)\phi_{j,k}(y)\mathrm{d}y\,\phi_{j,k}\right\|_{\mathbb{L}_p(\omega)}$$

$$\leqslant C2^{j/2}\left(\sum_k\left|\int f\phi_{j,k}\mathrm{d}x\right|^p\omega(I_{j,k})\right)^{1/p}\leqslant C^2\|f\|_{\mathbb{L}_p(\omega)}.$$

Now, to prove (46), 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 $g\omega = 0$ ω-a.e. so $g = 0$ ω-a.e. (It is clear that if $g \in \mathbb{L}_q(\omega)$ then $g\omega$ is locally Lebesgue integrable.)

It remains to prove (47). But for $f \in \mathbb{L}_p(\omega)$,

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

and

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

Hence

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

We have used the following well-known convolution lemma:

Lemma 4. *Let $(a_j)_{j\in\mathbb{Z}}$ and $(b_j)_{j\in\mathbb{Z}}$ be two sequences and $a \star b_k = \sum_j a_{k-j} b_j$. Then*

$$\|a \star b\|_{l_q(\mathbb{Z})} \leqslant \|a\|_{l_{q\wedge 1}(\mathbb{Z})} \|b\|_{l_q(\mathbb{Z})}. \tag{48}$$

Moreover, using (44), we obtain

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

□

7.4. Weighted Besov spaces and wavelet expansions

Using the notation of Section 4.3, we prove Proposition 1. We begin with the following standard lemma.

Lemma 5. *The following statements are equivalent:*

(i) *There exists $\theta \in \mathbb{L}_1(\mathbb{R})$ such that $\psi(x) = (-1)^N \Delta^N_{-1/2} \theta(x)$.*
(ii) *There exists $\gamma \in \mathbb{L}_1(\mathbb{R})$ such that $\psi(x) = (D^N \gamma)(x)$.*
(iii) $\int x^k \psi(x) \mathrm{d}x = 0,\ k = 0, 1, \ldots, N-1.$
(iv) $m_1(\xi) = \mathcal{O}(|\xi|^N)$.
(v) *There exists a trigonometric polynomial $\tilde{m}$ such that*

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

.

Moreover,

$$\mathrm{supp}(\theta) \subset [0, L], \qquad \mathrm{supp}(\gamma) \subset [0, L].$$

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

Proof. (i) $\Rightarrow$ (ii). The hypothesis is equivalent to

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

So

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

And obviously

$$\exp(-\mathrm{i}N\xi/4) \frac{1}{2^N} \left(\frac{\sin \xi/4}{\xi/4}\right)^N \hat{\theta}(\xi)$$

is the Fourier transform of an integrable function.

(ii) $\Leftrightarrow$ (iii). This is standard using Taylor's formula.

(ii) $\Rightarrow$ (iv) $(\mathrm{i}\xi)^N \hat{\gamma}(\xi) = \hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2)$ implies, as $|\hat{\phi}(0)| = 1$, that $m_1(\xi) = \mathcal{O}(|\xi|^N)$.

(v) $\Rightarrow$ (i). We have $\hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2) = (1 - \exp(-\mathrm{i}\xi/2))^N \tilde{m}(\xi/2)\hat{\phi}(\xi/2)$.

(iv) $\Leftrightarrow$ (v). This follows from Lemma 6 below. □

Lemma 6. *Let $m(\omega)$ be a trigonometric polynomial. The following statements are equivalent.*

(i) $m(\omega) = (1 - \exp(-\mathrm{i}\omega))^N \tilde{m}(\omega)$, with $\tilde{m}$ *a trigonometric polynomial.*
(ii) $m(\omega) = \mathcal{O}(|\omega|^N)$.

Proof. (i) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (i). Let us put $m(\omega) = \sum_{k=0}^{M} a_k \exp(\mathrm{i}k\omega)$. If $N = 1$, we have to find a trigonometric polynomial $\sum_k b_k \exp ik\omega$ such that

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

So

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

Let us put $\Delta b_k = (b_k - b_{k+1}) = a_k$, so that $b_k = \sum_{j\geqslant k} a_j$. 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 5 is now clear:

Corollary 2. *Let ψ be a compactly supported wavelet satisfying one of the equivalent properties of Lemma 5. Let f a locally integrable function, with*

$$\beta_{j,k} = \int f(x)\psi_{j,k}(x)\mathrm{d}x.$$

Then

$$\beta_{j,k} = (-1)^N 2^{j/2} \int \Delta^N_{2^{-(j+1)}} f(u)\theta(2^j u - k)\mathrm{d}u \tag{49}$$

and, if $D^N f$ exists,

$$\beta_{j,k} = (-1)^N 2^{-jN} 2^{j/2} \int D^N f(u)\gamma(2^j u - k)\mathrm{d}u. \tag{50}$$

Proof. We have

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

One can prove (50) using integration by parts. □

Proof of Proposition 1. For $\omega \in A_p$, using (49), (38) and (39), we have

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

So

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

and

$$2^{j/2}\left(\sum_{k\in\mathbb{Z}}|\beta_{j,k}|^p\omega(I_{j,k})\right)^{1/p} \leqslant C''\int_{\mathbb{R}}|\Delta^N_{2^{-(j+1)}}f(u)|^p\omega(u)\mathrm{d}u \leqslant C''\rho^N(2^{-(j+1)}, f, \omega, p). \quad (51)$$

□

8. Embeddings of Besov bodies with weight

This section is devoted to the proof of Proposition 4. Recall that we consider the following spaces:

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

Recall also that Corollary 1 proves that under condition $(\mathcal{H}_\pi)$, for $I_{j,k} = [k/2^j, (k+1)/2^j]$ and $f = \sum_{j,k}\beta_{jk}\psi_{j,k}(G)$, we have

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

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

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

$$L_1 = \left\{f = \sum_I\beta_I\psi_I \circ G, \sup_{\lambda>0}\lambda^q\nu\{(j, k)\|\beta_{jk}| > \lambda\} < \infty\right\}, \quad (52)$$

$$L_2 = \left\{f = \sum_I\beta_I\psi_I \circ G, \sup_{l\geqslant 0}\left\|\sum_{j\geqslant l,k}\beta_{jk}\psi_{jk} \circ G\right\|_p^p 2^{l(p-q)} < \infty\right\}. \quad (53)$$

We remind the reader that we will concentrate on the case where

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

Let us introduce the following Besov bodies:

$$b^G_{s,\pi,r} = \left\{f = \sum_I\beta_I\psi_I \circ G, \left[\sum_{j\geqslant -1}2^{jsr}2^{jr/2}\left(\sum_{k\in\mathbb{Z}}|\beta_{j,k}|^\pi\omega(I_{j,k})\right)^{r/\pi}\right]^{1/r} < \infty\right\}, \quad (54)$$

with the usual modification if $r = \infty$. Hence, we reduce the proof of Proposition 4 to embeddings of Besov bodies which are quite simple, as shown below. Some embedding properties are obvious:

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s,\pi,\infty}, \qquad \text{if } 0 < r < r'. \quad (55)$$

Because of the fact that ω is a finite weight, the following embeddings are also obvious:

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s,\rho,\infty}, \qquad \text{if } 0 < \rho \leqslant \pi. \tag{56}$$

Notice, however, that other embeddings which are standard in the regular case ($\omega(I_{jk}) = 2^{-j}$, which occurs for instance if $0 < m \leqslant g \leqslant M < \infty$) may not be satisfied here (see condition (58) in the following proposition).

Consider condition (52).

Proposition 10. (i) *For $q < \pi$, for s related to q by*

$$s = \frac{p}{2q} - \frac{1}{2}, \tag{57}$$

we have

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s,\pi,\infty} \hookrightarrow l_{q,\infty}(\nu).$$

(ii) *For $q \geqslant \pi$, if ω is such that*

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s',\pi',r}, \qquad \forall \pi \leqslant \pi',\ s' - \frac{1}{\pi'} = s - \frac{1}{\pi} \tag{58}$$

and if s is now related to q by

$$s - \frac{1}{\pi} = s' - \frac{1}{q}, \qquad s' = \frac{p}{2q} - \frac{1}{2}, \tag{59}$$

then

$$b^G_{s,\pi,r} \hookrightarrow b^G_{s',q,r} \hookrightarrow l_{q,\infty}(\nu), \qquad \textit{for } r \leqslant q.$$

Proof. Let us put, for simplicity, $\nu_p\{(j, k)\} = 2^{jp/2}\omega(I_{j,k})$, and consider that the support of f and ψ is [0, 1]. Let us consider $f \in b^G_{s,\pi,\infty}$, such that

$$f(G^{-1}(x)) = \sum_{j=-1}^{\infty} \sum_{0 \leqslant k < 2^j} \beta_{j,k}\psi_{j,k}(x); \qquad \sum_{0 \leqslant k < 2^j} |\beta_{j,k}|^\pi \omega(I_{j,k}) \leqslant C^\pi 2^{-j(s+1/2)\pi}.$$

We observe that for all $j \geqslant -1$, $\sum_{0 \leqslant k < 2^j} \nu_p\{(j, k)\} = \sum_{0 \leqslant k < 2^j} 2^{jp/2}\omega(I_{jk}) = 2^{jp/2}\omega([0, 1])$. To simplify, let us suppose that $\omega([0, 1]) = 1$.

We will use the following decomposition, and then consider every j level separetely:

$$\nu_p\{(j, k), |\beta_{j,k}| > \lambda\} = \sum_{j=-1}^{\infty} \nu_p\{(j, k) \in \{j\} \times \mathbb{N}, |\beta_{j,k}| > \lambda\}.$$

Now, for fixed $j \in \mathbb{N}$,

$$\nu_p\{(j,k)\in\{j\}\times\mathbb{N}, |\beta_{j,k}|>\lambda\} \leqslant 2^{jp/2}\wedge \nu_p\{(j,k)\in\{j\}\times\mathbb{N}, |\beta_{j,k}|>\lambda\}$$
$$\leqslant 2^{jp/2}\wedge \frac{\sum_{0\leqslant k<2^j}|\beta_{j,k}|^\pi 2^{jp/2}\omega(I_{j,k})}{\lambda^\pi}$$
$$\leqslant 2^{jp/2}\left\{1\wedge\left(\frac{C2^{-j(s+1/2)}}{\lambda}\right)^\pi\right\}.$$

Let J be such that $C2^{-J(s+1/2)}\sim\lambda$. As $p/2<(s+1/2)\pi$, we have

$$\nu_p\{(j,k)\in\{j\}\times\mathbb{N}, |\beta_{j,k}|>\lambda\} \leqslant \sum_{j=0}^{J}2^{jp/2}+\left(\frac{C}{\lambda}\right)^\pi\sum_{j=J+1}^{\infty}2^{jp/2}2^{-j(s+1/2)\pi}$$
$$\sim 2^{Jp/2}+\left(\frac{C}{\lambda}\right)^\pi 2^{-J((s+1/2)\pi-p/2)}.$$

But as $2^{Jp/2}\sim\lambda^{-q}$, and $2^{-J((s+1/2)\pi-p/2)}=2^{-J(p/2q)\pi}2^{Jp/2}\sim\lambda^\pi\lambda^{-q}$, we obtain the first inclusion. The second one is obtained simply using Markov's inequality, observing that $b^G_{s',q,q}=\{\sum_{jk}|\beta_{jk}|^q<\infty\}$. □

Now consider condition (53). First, if $\pi\geqslant p$, then $b^G_{s,\pi,r}\subset b^G_{s/1+2s,p,\infty}$. Using Theorem 5, we have:

$$\left\|\sum_{j\geqslant l,k}\beta_{j,k}\psi_{j,k}\circ G\right\|_p 2^{l(p-q)/p}\leqslant\sum_{j\geqslant l}\left\|\sum_k\beta_{j,k}\psi_{j,k}\circ G\right\|_p 2^{l(p-q)/p}$$
$$\leqslant C\sum_{j\geqslant l}2^{j/2}\left(\sum_k|\beta_{j,k}|^p\omega(I_{j,k})\right)^{1/p}2^{l(p-q)/p}.$$

Hence, if $f\in b^G_{(1-q/p),p,\infty}$, condition (53) obviously holds. Hence the problem remaining is to check whether $b^G_{s,\pi,r}$ is included in $b^G_{(1-q/p),p,\infty}$. Now, if we use the embeddings (56), with $\rho=p$, we only need to check that $s\geqslant 2s/(1+2s)=1-q/p$, when q is chosen as in (57), which is always true for $s\geqslant\frac{1}{2}$, or $s\geqslant 2(s-1/\pi+1/p)/(1+2(s-1/\pi))=1-q/p$, when q is chosen as in (59), which is always true for $p\geqslant 2\pi$, but observe that this condition is necessary for $1/2\leqslant s\leqslant(p-\pi)/2\pi$. Hence, (53) will always hold if $s\geqslant\frac{1}{2}$, for $p\leqslant\pi$.

For $\pi<p$, g bounded from above and below, we have, using (58), $b^G_{s,\pi,r}\subset b^G_{s-1/\pi+1/p,p,\infty}$. Now, $b^G_{s-1/\pi+1/p,p,\infty}\subset b^G_{(1-q/p),p,\infty}$ for $s-1/\pi+1/p\geqslant 2(s-1/\pi+1/p)/(1+2(s-1/\pi))$ (q is necessarily in this case chosen as in (59)). The last inequality is true for $s\geqslant 1/\pi+1/2$.

9. The probability bounds

In this section we summarize all the proofs for the probability bounds used for the results above. We recall the following elementary facts.

If, for all $x \in \mathbb{R}$, $G(x) = P(X \leqslant x)$ denotes the distribution function of the variable X, then we set

$$\forall y \in [0, 1], \qquad G^{-1}(y) = \sup\{x,\ G(x) \leqslant y\}.$$

It is well known that

$$\forall y \in [0, 1], \qquad G(G^{-1}(y)) \geqslant y, \quad \text{and} \quad G(G^{-1}(y)) \equiv y \iff G \text{ continuous.}$$

$$\forall x \in \mathbb{R}, \qquad G^{-1}(G(x)) \geqslant x, \quad \text{and} \quad G^{-1}(G(x)) \equiv x \iff G \text{ strictly increasing.}$$

Moreover, if U a is random variable with uniform law on [0, 1], then X has the same distribution as $G^{-1}(U)$, and

$$\forall \Phi, \qquad \mathbb{E}\Phi(X) = \mathbb{E}\Phi(G^{-1}(U)) = \int_0^1 \Phi(G^{-1}(y))\mathrm{d}y.$$

If G is a continuous function, $G(X)$ has the uniform law on [0, 1]. The following facts are also equivalent:

- G is absolutely continuous and $G'(x) = g(x)$ a.e.
- For all $F \geqslant 0$,

$$\int F(G(x))g(x)\mathrm{d}x = \int_0^1 F(u)\mathrm{d}u.$$

- For all $F \geqslant 0$,

$$\int F(G(x))\mathrm{d}x = \int_0^1 F(u)\omega(u)\mathrm{d}u.$$

Moreover,

$$g(x) = \frac{1}{\omega(G(x))} \quad \text{and} \quad \omega(u) = \frac{1}{g(G^{-1}(u))}.$$

We recall that the definitions of the estimators of the empirical partition function, and the different estimates of the wavelet coefficients, considered below, are given respectively in (17), (12) and (18). Obviously,

$$\mathbb{E}(\hat{\beta}^*_{j,k}) = \int \psi_{j,k}(G(x))f(x)\mathrm{d}G(x) = \beta_{j,k} = \int_0^1 \psi_{j,k}(G(G^{-1}(y)))f(G^{-1}(y))\mathrm{d}y,$$

$$\mathbb{E}(\hat{\beta}^{@}_{j,k}) = \int \mathbb{E}_X[\psi_{j,k}(\hat{G}_n(x))]f(x)\mathrm{d}G(x) = \int_0^1 \mathbb{E}_X[\psi_{j,k}(\hat{G}_n(G^{-1}(y)))]f(G^{-1}(y))\mathrm{d}y,$$

where $\mathbb{E}_X$ denotes expectation with the respect to $X_1, \ldots, X_n$. If G is a continuous function, we also have

$$\beta_{j,k} = \int_0^1 \psi_{j,k}(y)f(G^{-1}(y))\mathrm{d}y.$$

Moreover,

$$\hat{G}_n(G^{-1}(y)) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty, G^{-1}(y)]}(X_i) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty, y]}(G(X_i)).$$

Let us put $U_1 = G(X_1)$, $U_2 = G(X_2)$, $\ldots$, $U_n = G(X_n)$; these variables are independently and identically uniform on [0, 1]. So if we put

$$\hat{U}_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty, y]}(U_i), \tag{60}$$

then

$$\mathbb{E}[\hat{\beta}^{@}_{j,k}] = \int_0^1 \mathbb{E}_u[\psi_{j,k}(\hat{U}_n(y))] f(G^{-1}(y))\mathrm{d}y, \tag{61}$$

where $\mathbb{E}_u$ denotes expectation with respect to the law of $U_1, \ldots, U_n$.

Now if we put

$$A(y, \alpha) = \{|\hat{U}_n(y) - y| \geqslant \alpha t_n\}, \qquad t_n = \sqrt{\frac{\log n}{n}},$$

and if $\sqrt{\log n/n} \leqslant 2^{-j}$, then there exists $L_{jk} = L_{jk}(\alpha)$, an interval homothetical to the support of ψ_{jk}, with a ratio bounded independently of jk, such that

$$\psi_{j,k}(\hat{U}_n(y)) = I\{A(y, \alpha)\}\psi_{j,k}(\hat{U}_n(y)) + I\{A^c(y, \alpha)\}\psi_{j,k}(\hat{U}_n(y)) I\{L_{jk}\}(y).$$

So we can prove the following lemma which will later be a key tool:

Lemma 7. *Let $2^j \leqslant \sqrt{n/\log n}$. With the previous notation, there exists A such that*

$$\int_0^1 \mathbb{E}|\psi_{j,k}(\hat{U}_n(y))|^p \,\mathrm{d}y \leqslant A\{1 + 2^{j(p/2-1)}\}. \tag{62}$$

Proof. Using Hoeffding's inequality,

$$P\{|\hat{U}_n(y) - y| \geqslant \lambda\} \leqslant 2\exp(-2n\lambda^2) \tag{63}$$

and

$$\begin{aligned}\mathbb{E}|\psi_{j,k}(\hat{U}_n(y))|^p &= \mathbb{E}(I\{A(y, \alpha)\}|\psi_{j,k}(\hat{U}_n(y))|^p) + \mathbb{E}(I\{A^c(y, \alpha)\}|\psi_{j,k}(\hat{U}_n(y))|^p) I\{L_{jk}\}(y)\\ &\leqslant C\left\{\frac{2^{jp/2}}{n^{2\alpha^2}} + 2^{jp/2} I\{L_{jk}\}(y)\right\}.\end{aligned}$$

So

$$\int_0^1 \mathbb{E}|\psi_{j,k}(\hat{U}_n(y))|^p \,\mathrm{d}y \leqslant C\left\{\frac{2^{jp/2}}{n^{2\alpha^2}} + 2^{j(p/2-1)}\right\} \leqslant C\{1 + 2^{j(p/2-1)}\}$$

if α is chosen large enough. □

9.1. Bounds for $\hat{\beta}^*_{jk} - \beta_{jk}$: Proof of Proposition 3

We need to prove the following inequalities:

$$\mathbb{E}(|\hat{\beta}^*_{jk} - \beta_{jk}|^p) \leqslant C_p \frac{1 + \|f\|_\infty^p}{n^{p/2}}, \qquad \text{for } 2^j \leqslant n, \tag{64}$$

$$P\left(|\hat{\beta}^*_{jk} - \beta_{jk}| > \kappa\sqrt{\frac{\log n}{n}}\right) \leqslant C'_p n^{-\gamma p}, \qquad \text{for } \kappa \geqslant \kappa_\gamma,\ 2^j \leqslant \sqrt{\frac{n}{\log n}}. \tag{65}$$

First, we observe that for $p < 2$, using Jensen's inequality,

$$\mathbb{E}(|\hat{\beta}^*_{jk} - \beta_{j,k}|^p) \leqslant (\mathbb{E}(|\hat{\beta}^*_{jk} - \beta_{j,k}|^2))^{p/2}.$$

Hence, it is enough to prove (64) for $2 \leqslant p < \infty$. Using Rosenthal's inequality (see Härdle *et al*. 1998, p. 241), for $p \geqslant 2$,

$$\mathbb{E}(|\hat{\beta}^*_{jk} - \beta_{jk}|^p) \leqslant C\left(\frac{\mathbb{E}|\psi_{j,k}(G(X'))Y'|^p}{n^{p-1}} + \frac{(\mathbb{E}|\psi_{j,k}(G(X'))Y'|^2)^{p/2}}{n^{p/2}}\right),$$

$$\begin{aligned}\mathbb{E}|\psi_{j,k}(G(X'))Y'|^p &= \mathbb{E}|\psi_{j,k}(G(X'))(f(X') + \epsilon)|^p \\ &\leqslant 2^{p-1}(\mathbb{E}|\psi_{j,k}(G(X'))f(X')|^p + \mathbb{E}|\psi_{j,k}(G(X'))\epsilon|^p).\end{aligned}$$

But

$$\begin{aligned}\mathbb{E}|\psi_{j,k}(G(X'))f(X')|^p &= \int |\psi_{j,k}(G(x))f(x)|^p g(x)\mathrm{d}x \leqslant \|f\|_\infty^p \int |\psi_{j,k}(G(x))|^p g(x)\mathrm{d}x \\ &= \|f\|_\infty^p \int |\psi_{j,k}(u)|^p \mathrm{d}u \leqslant \|f\|_\infty^p 2^{j(p-2)/2}\|\psi\|_\infty^{p-2}\int |\psi_{j,k}(u)|^2\mathrm{d}u \\ &= \|f\|_\infty^p \|\psi\|_\infty^{p-2} 2^{j(p/2-1)}.\end{aligned}$$

Furthermore,

$$\mathbb{E}|\psi_{j,k}(G(X'))\epsilon|^p = \mathbb{E}|\psi_{j,k}(G(X'))|^p\mathbb{E}|\epsilon|^p = \mathbb{E}|\epsilon|^p\int |\psi_{j,k}(G(x))|^p g(x)\mathrm{d}x \leqslant C_p 2^{j(p/2-1)}.$$

So

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

This proves (64) if $2^j \leqslant n$ (and a fortiori if $2^j \leqslant \sqrt{n/\log n}$).

Now let us prove (65). We have

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

Hence

$$P\left(|\hat{\beta_{jk}} - \beta_{jk}| > \kappa\sqrt{\frac{\log n}{n}}\right)$$

$$\leqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))(f(X'_i) - \mathbb{E}(\psi_{j,k}(G(X))f(X)))\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right)$$

$$+ \; P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))\epsilon_i\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right).$$

Let us observe that, conditionally on $(X'_1, \ldots, X'_n)$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))\epsilon_i \sim N\left(0, \frac{1}{n^2}\sum_{i=1}^{n}\psi^2_{j,k}(G(X'_i))\right).$$

So

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))\epsilon_i\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right) \leqslant \mathbb{E}_{(X'_1,\ldots,X'_n)}\exp\left(-\frac{\kappa^2\log n}{8n^{-1}\sum_{i=1}^{n}\psi^2_{j,k}(G(X'_i))}\right)$$

$$\leqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi^2_{j,k}(G(X_i)) - 1\right| > \alpha\right)$$

$$+ \; \exp\left(-\frac{\kappa^2\log n}{8(1+\alpha)}\right).$$

Using Hoeffding's inequality (see Härdle *et al*. 1998, p. 241), we have, using the fact that $\psi^2_{j,k}(G(X'_i))$ are i.i.d. variables bounded by $2^j\|\psi\|^2_\infty$ and such that $\mathbb{E}\psi^2_{j,k}(G(X'_i)) = 1$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi^2_{j,k}(G(X'_i)) - 1\right| > \alpha\right) \leqslant 2\exp\left(-\frac{2n^2\alpha^2}{n\|\psi\|^4_\infty 2^{2j}}\right) \leqslant 2n^{-2\alpha^2/\|\psi\|^4_\infty} \qquad (66)$$

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

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))\epsilon_i\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right) \leqslant Cn^{-\gamma},$$

if $2^j \leqslant \sqrt{n/\log n}$.

Using Bernstein's inequality,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(G(X'_i))(f(X'_i)-\mathbb{E}(\psi_{j,k}(G(X'))f(X')))\right|>\frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right)$$

$$\leqslant 2\exp\left(-\frac{n((\kappa/2)\sqrt{\log n/n})^2}{\frac{2}{3}(3\sigma^2+M(\kappa/2)\sqrt{\log n/n})}\right),$$

where

$$M=\|\psi_{j,k}(G(X'))(f(X')-\mathbb{E}(\psi_{j,k}(G(X'))f(X')))\|_\infty\leqslant 2.2^{j/2}\|\psi\|_\infty\|f\|_\infty,$$

$$\sigma^2=\mathbb{E}|\psi_{j,k}(G(X'))(f(X')-\mathbb{E}(\psi_{j,k}(G(X'))f(X')))|^2\leqslant\mathbb{E}|\psi_{j,k}(G(X'))f(X')|^2\leqslant\|f\|_\infty^2,$$

as

$$\mathbb{E}|\psi_{j,k}(G(X'))|^2=\int|\psi_{j,k}(G(x))|^2\mathrm{d}\mu(x)=\int_0^1|\psi_{j,k}(GG^{-1}(y))|^2\,\mathrm{d}y=\int_0^1|\psi_{j,k}(y)|^2\,\mathrm{d}y=1.$$

Furthermore,

$$2\exp\left(-\frac{n((\kappa/2)\sqrt{\log n/n})^2}{\frac{2}{3}(3\sigma^2+M(\kappa/2)\sqrt{\log n/n})}\right)\leqslant 2\exp\left(-\frac{3\kappa^2\log n}{4\|f\|_\infty(3+2.2^{j/2}(\kappa/2)\sqrt{\log n/n})}\right)$$

$$\leqslant 2\exp-\frac{3\kappa^2\log n}{4\|f\|_\infty(3+\kappa)}$$

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

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

□

9.2. Bounds for $\hat{\beta}^{@}_{j,k}-\beta_{j,k}$: Proof of Proposition 6

We now need to prove the same inequalities as above but for $\hat{\beta}^{@}_{j,k}$, namely,

$$\mathbb{E}|\hat{\beta}^{@}_{j,k}-\beta_{j,k}|^p\leqslant C_p\frac{1+D^p}{n^{p/2}},\qquad\text{for } 2^j\leqslant\sqrt{\frac{n}{\log n}},\tag{67}$$

$$P\left(|\hat{\beta}^{@}_{j,k}-\beta_{j,k}|>\kappa\sqrt{\frac{\log n}{n}}\right)\leqslant C'_pn^{-\gamma p},\qquad\text{for }\kappa\geqslant\kappa(\gamma,D),\ 2^j\leqslant\sqrt{\frac{n}{\log n}}.\tag{68}$$

As above, it is enough to prove (67) for $2\leqslant p<\infty$. Let us observe that conditioning on $(X_1,\ldots,X_n)$, we have

$$\mathbb{E}[\hat{\beta}^{@}_{j,k}|(X_1, \ldots, X_n)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{1}{n}\sum_{i=1}^{n}1_{]-\infty,X'_i]}(X_i)\right)(f(X'_i)+\epsilon_i)\middle|(X_1, \ldots, X_n)\right].$$

$$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}\left(\frac{1}{n}\sum_{i=1}^{n}1_{]-\infty,X'_i]}(X_i)\right)(f(X'_i))\middle|(X_1, \ldots, X_n)\right].$$

$$= \int_0^1 \psi_{j,k}(\hat{U}_n(y))f(G^{-1}(y))\mathrm{d}y,$$

with $\hat{U}_n(y)$ defined as in (60). It is natural to introduce

$$\tilde{\beta}_{jk} = \tilde{\beta}_{jk}(X_1, \ldots, X_n) = \int_0^1 \psi_{j,k}(\hat{U}_n(y))f(G^{-1}(y))\mathrm{d}y. \tag{69}$$

Moreover, using (61),

$$\mathbb{E}[\tilde{\beta}_{jk}] = \mathbb{E}[\hat{\beta}^{@}_{j,k}] = \int \mathbb{E}_u[\psi_{j,k}(\hat{U}_n(y))]f(G^{-1}(y))\mathrm{d}y,$$

where $\mathbb{E}_u$ denotes the expectation with respect to the law of $U_1, \ldots, U_n$. We have,

$$\mathbb{E}|\hat{\beta}^{@}_{j,k} - \beta_{jk}|^p \leqslant 2^{p-1}(\mathbb{E}|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}|^p + \mathbb{E}|\tilde{\beta}_{jk} - \beta_{jk}|^p), \qquad 2 \leqslant p < \infty, \tag{70}$$

$$P\left(|\hat{\beta}^{@}_{j,k} - \beta_{j,k}| > \kappa\sqrt{\frac{\log n}{n}}\right) \leqslant P\left(|\hat{\beta}^{@}_{jk} - \tilde{\beta}_{jk}| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right) + P\left(|\tilde{\beta}_{jk} - \beta_{j,k}| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right). \tag{71}$$

9.2.1. Bounds for $\hat{\beta}^{@}_{jk} - \tilde{\beta}_{jk}$

We consider two cases: $p \geqslant 2$ and $p < 2$.

For $p \geqslant 2$, using Rosenthal's inequality, conditionally on $(X_1, \ldots, X_n)$,

$$\mathbb{E}(|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}|^p|(X_1, \ldots, X_n)) \leqslant C_p\left(\frac{\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))(f(X')+\epsilon)|^p|(X_1, \ldots, X_n)}{n^{p-1}}.\right.$$

$$\left. + \left[\frac{\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))(f(X')+\epsilon)|^2|(X_1, \ldots, X_n)}{n}\right]^{p/2}\right).$$

But

$$2^{-(p-1)}\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))(f(X')+\epsilon)|^p|(X_1, \ldots, X_n)$$
$$\leqslant \mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))f(X')|^p|(X_1, \ldots, X_n) + \mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))|^p|\epsilon|^p|(X_1, \ldots, X_n)$$

and

$$
\begin{aligned}
\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))f(X')|^p|(X_1, \ldots, X_n) &= \int |\psi_{j,k}(\hat{G}_n(x'))f(x')|^p \mathrm{d}\mu(x') \\
&= \int |\psi_{j,k}(\hat{G}_n(G^{-1}(y)))f(G^{-1}(y))|^p \,\mathrm{d}y = \int |\psi_{j,k}(\hat{U}_n(y))f(G^{-1}(y))|^p \,\mathrm{d}y \\
&\leqslant \|f\|_\infty^p \int |\psi_{j,k}(\hat{U}_n(y))|^p \,\mathrm{d}y
\end{aligned}
$$

and

$$
\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'))|^p|\epsilon|^p|(X_1, \ldots, X_n) = \mathbb{E}(|\epsilon|^p)\int_0^1 |\psi_{j,k}(\hat{U}_n(y))|^p \,\mathrm{d}y.
$$

So integrating with respect to X_i and using Lemma 7,

$$
\begin{aligned}
\mathbb{E}|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}|^p &= \mathbb{E}[\mathbb{E}|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}|^p|(X_1, \ldots, X_n)] \\
&\leqslant 2^{p-1} C_p \left(\frac{A(1 + 2^{j(p/2-1)})(\mathbb{E}|\epsilon|^p + \|f\|_\infty^p)}{n^{p-1}} + \frac{[2A(1 + \|f\|_\infty^2)]^{p/2}}{n^{p/2}} \right) \\
&\leqslant C'_p \frac{1 + \|f\|_\infty^p}{n^{p/2}}.
\end{aligned}
$$

For $p < 2$, we have

$$
\mathbb{P}\left(|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}| \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}} \right) = \mathbb{E}_{(X_1,\ldots,X_n)}\left[\mathbb{P}\left(|\hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk}| \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}} \middle| (X_1, \ldots, X_n) \right) \right].
$$

But

$$
\begin{aligned}
&\mathbb{P}\left(\left| \hat{\beta}^{@}_{j,k} - \tilde{\beta}_{jk} \right| \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}} \middle| (X_1, \ldots, X_n) \right) \\
&\quad \leqslant \mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^n \psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i) - \tilde{\beta}_{jk}) \right| \geqslant \frac{\kappa}{4}\sqrt{\frac{\log n}{n}} \middle| (X_1, \ldots, X_n) \right) \\
&\qquad + \mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^n \psi_{j,k}(\hat{G}_n(X'_i))\epsilon_i \right| \geqslant \frac{\kappa}{4}\sqrt{\frac{\log n}{n}} \middle| (X_1, \ldots, X_n) \right).
\end{aligned}
$$

The two right-hand-side terms will now be investigated separately. We have

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i)-\tilde{\beta}_{jk})\right|\geqslant\frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\middle|(X_1,\ldots,X_n)\right).$$

$$\leqslant 2\exp\left(-\frac{n\Big((\kappa/4)\sqrt{\log n/n}\Big)^2}{2\sigma^2+M(\kappa/6)\sqrt{\log n/n}}\right)$$

where

$$M\leqslant 2\|\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i))\||(X_1,\ldots,X_n)\|_\infty\leqslant 2.2^{j/2}\|\psi\|_\infty\|f\|_\infty$$

$$\sigma^2\leqslant|\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i))\||(X_1,\ldots,X_n)|^2\leqslant\mathbb{E}|\psi_{j,k}(\hat{G}_n(X'_i))f(X'_i)|^2|(X_1,\ldots,X_n)$$

$$=\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2|f(G^{-1}(y))|^2\,\mathrm{d}y\leqslant\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y.$$

So, if $2^j\leqslant n/\log n$, there exists some constant C such that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i)-\tilde{\beta}_{jk})\right|\geqslant\frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\middle|(X_1,\ldots,X_n)\right).$$

$$\leqslant 2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y+\kappa\|f\|_\infty}\right).$$

So

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}_n(X'_i))(f(X'_i)-\tilde{\beta}_{jk})\right|\geqslant\frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\right)$$

$$\leqslant\mathbb{E}_{(X_1,\ldots,X_n)}\left[2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y+\kappa\|f\|_\infty}\right)\right].$$

But by the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky *et al.* 1956),

$$\mathbb{P}\left(\sup_{y\in\mathbb{R}}|\hat{U}_n(y)-y|\geqslant\lambda\right)\leqslant K\exp(-2n\lambda^2),$$

with K a universal constant. Let

$$B_n(\alpha)=\left\{\sup_{y\in\mathbb{R}}|\hat{U}_n(y)-y|\geqslant\alpha\sqrt{\frac{\log n}{n}}\right\}. \tag{72}$$

It is clear that there exists, for all j, k, an interval $L_{j,k}$ homothetical to the support of ψ_{jk} with a ratio bounded by a constant independent of j, k (but depending on α) and $D(\alpha)$ such that, for $2^j\leqslant\sqrt{n/\log n}$,

$$I_{B_n^c(\alpha)}\int_0^1 |\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y = I_{B_n^c(\alpha)}\int_{L_{j,k}} |\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y \leqslant D(\alpha), \tag{73}$$

$$\begin{aligned}
&\mathbb{E}_{(X_1,\dots,X_n)}\left[2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y+\kappa\|f\|_\infty}\right)\right]\\
&\quad\leqslant \mathbb{E}_{(X_1,\dots,X_n)}I_{B_n(\alpha)}\left[2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y+\kappa\|f\|_\infty}\right)\right]\\
&\quad\quad+\ \mathbb{E}_{(X_1,\dots,X_n)}I_{B_n^c(\alpha)}\left[2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y+\kappa\|f\|_\infty}\right)\right]\\
&\quad\leqslant 2K\exp\left(-2n\left(\alpha\sqrt{\frac{\log n}{n}}\right)^2\right)+2\exp\left(-\frac{C\kappa\log n}{\|f\|_\infty^2+\kappa D(\alpha)\|f\|_\infty}\right)\leqslant Cn^{-\gamma},
\end{aligned}$$

with a suitable choice of α and then κ.

Again let us observe that conditionally on $(X_1',\dots,X_n')$, we have

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

and

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\psi_{j,k}(\hat{G}_n(X_i'))\epsilon_i\right|\geqslant\frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\right)\\
&\quad\leqslant \mathbb{E}_{(X_1,\dots,X_n)}\mathbb{E}_{(X_1',\dots,X_n')}\left(\exp\left(-\frac{\kappa^2\log n}{32n^{-1}\sum_{i=1}^n\psi_{j,k}^2(\hat{G}_n(X_i'))}\right)\right)\\
&\quad\leqslant 2K\exp\left(-2n\left(\alpha\sqrt{\frac{\log n}{n}}\right)^2\right)\\
&\quad\quad+\ \mathbb{E}_{(X_1,\dots,X_n)}I_{B_n^c(\alpha)}\mathbb{E}_{(X_1',\dots,X_n')}\left(\exp\left(-\frac{\kappa^2\log n}{32n^{-1}\sum_{i=1}^n\psi_{j,k}^2(\hat{G}_n(X_i'))}\right)\right).
\end{aligned}$$

But

$$\mathbb{E}_{(X_1',\dots,X_n')}\psi_{j,k}^2(\hat{G}_n(X_i'))=\int_0^1|\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y,$$

and on $B_n^c(\alpha)$ for $2^j\leqslant\sqrt{n/\log n}$, this quantity is less than $D(\alpha)$. Using the same argument as for (66), by the Hoeffding inequality,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}^2(\hat{G}_n(X_i')) - \int_0^1 |\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y\right| > \lambda\right) \leqslant 2\exp\left(-\frac{2n^2\lambda^2}{n(\|\psi\|_\infty^4 2^{2j} + D(\alpha))}\right).$$

But if

$$\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}^2(\hat{G}_n(X_i')) - \int_0^1 |\psi_{j,k}(\hat{U}_n(y))|^2\,\mathrm{d}y\right| \leqslant \lambda,$$

we have

$$\exp\left(-\frac{\kappa^2 \log n}{3n^{-1}\sum_{i=1}^{n}\psi_{j,k}^2(\hat{G}_n(X_i'))}\right) \leqslant \exp\left(-\frac{\kappa^2\log n}{32(\lambda + D(\alpha))}\right).$$

So

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}_n(X_i'))\epsilon_i\right| \geqslant \frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\right)$$

$$\leqslant 2K\exp\left(-2n\left(\alpha\sqrt{\frac{\log n}{n}}\right)^2\right) + 2\exp\left(-\frac{2n^2\lambda^2}{n(\|\psi\|_\infty^2 2^{2j} + D(\alpha))}\right) + 2\exp\left(-\frac{\kappa^2\log n}{32(\lambda + D(\alpha))}\right).$$

With a suitable choice of α, then λ, and finally κ, we obtain

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}_n(X_i')\epsilon_i\right| \geqslant \frac{\kappa}{4}\sqrt{\frac{\log n}{n}}\right) \leqslant C\frac{1}{n^\gamma}.$$

9.2.2. Bounds for $\tilde{\beta}_{jk} - \beta_{jk}$

Let us observe that

$$\tilde{\beta}_{jk} - \beta_{jk} = \int_0^1 (\psi_{j,k}(\hat{U}_n(y)) - \psi_{j,k}(y))f(G^{-1}(y))\mathrm{d}y.$$

In this subsection we will use refinements very specific to wavelets. Let us recall that if ψ is a compactly supported wavelet, then there exists a compactly supported function θ such that

$$\psi = \Delta_{-h}(\theta), \quad h = 2^{-1}, \qquad \Delta_h(\theta)(y) = \theta(y+h) - \theta(y).$$

So

$$\psi_{j,k} = \Delta_{-h}(\theta_{j,k}), \quad h = 2^{-j-1}, \qquad \theta_{j,k} = 2^{j/2}\theta(2^j y - k).$$

Let us prove the following lemma:

Lemma 8. *Let $h \in \mathbb{R}$ and θ be a compactly supported function. Then*

$$(\Delta_h\theta)(\hat{U}_n(y)) - (\Delta_h\theta)(y) = \Delta_h[\theta(\hat{U}_n(y)) - \theta(y)] + \theta(\hat{U}_n(y) + h) - \theta(\hat{U}_n(y+h)). \quad (74)$$

Proof. The left-hand side is equivalent to

$$\theta(\hat{U}_n(y)+h)-\theta(\hat{U}_n(y))-\theta(y+h)+\theta(y).$$

The right-hand side equals

$$\begin{aligned}
&\Delta_h[\theta(\hat{U}_n(y))-\theta(y)]+\theta(\hat{U}_n(y)+h)-\theta(\hat{U}_n(y+h))\\
&\quad=(\theta(\hat{U}_n(y+h))-\theta(y+h)-\theta(\hat{U}_n(y))+\theta(y))+\theta(\hat{U}_n(y)+h)-\theta(\hat{U}_n(y+h))\\
&\quad=-\theta(y+h)-\theta(\hat{U}_n(y))+\theta(y)+\theta(\hat{U}_n(y)+h),
\end{aligned}$$

which is equal to the left-hand side. □

As a consequence of this lemma we have

$$\begin{aligned}
\tilde{\beta}_{jk}-\beta_{jk}&=\int_0^1(\psi_{j,k}(\hat{U}_n(y))-\psi_{j,k}(y))f(G^{-1}(y))\mathrm{d}y\\
&=\int_0^1(\Delta_{-2^{-j-1}}\theta_{j,k})(\hat{U}_n(y))-(\Delta_{-2^{-j-1}}\theta_{j,k})(y)f(G^{-1}(y))\mathrm{d}y\\
&=\int_0^1\Delta_{-2^{-j-1}}[\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)]f(G^{-1}(y))\mathrm{d}y\\
&\quad+\int_0^1[\theta_{j,k}(\hat{U}_n(y)-2^{-j-1})-\theta_{j,k}(\hat{U}_n(y-2^{-j-1}))]f(G^{-1}(y))\mathrm{d}y\\
&=\int_0^1[\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)]\Delta_{-2^{-j-1}}(f\circ G^{-1})(y)\mathrm{d}y\\
&\quad+\int_0^1[\theta_{j,k}(\hat{U}_n(y)-2^{-j-1})-\theta_{j,k}(\hat{U}_n(y-2^{-j-1}))]f(G^{-1}(y))\mathrm{d}y.
\end{aligned}$$

So

$$\begin{aligned}
|\tilde{\beta}_{jk}-\beta_{jk}|&\leqslant 2^{-j/2}\|f\circ G^{-1}\|_{lip_{1/2}}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y\\
&\quad+\|f\|_\infty\int_0^1|\theta_{j,k}(\hat{U}_n(y)-2^{-j-1})-\theta_{j,k}(\hat{U}_n(y-2^{-j-1}))|\mathrm{d}y.
\end{aligned}$$

If we now recall that $\|f\|_\infty\leqslant D$, $\|f\circ G^{-1}\|_{lip_{1/2}}\leqslant D$, then

$$\begin{aligned}
|\tilde{\beta}_{jk}-\beta_{jk}|&\leqslant 2^{-j/2}D\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y))|\mathrm{d}y\\
&\quad+D\int_0^1|\theta_{j,k}(\hat{U}_n(y)-2^{-j-1})-\theta_{j,k}(\hat{U}_n(y-2^{-j-1}))|\mathrm{d}y.
\end{aligned}$$

We tackle the two terms on the right-hand side separately. First we obtain bounds for $2^{-j/2}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y))|\mathrm{d}y$, using definition (72). It is clear that there exists, for all j, k, an interval $L_{j,k}$ homothetical to the support of θ_{jk} with a ratio bounded by a constant independent of j, k (but depending on α) such that, for $2^j \leqslant \sqrt{n/\log n}$,

$$
\begin{aligned}
&2^{-j/2}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y\\
&\quad= 2^{-j/2}I\{B_n(\alpha)\}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y + 2^{-j/2}I\{B_n^c(\alpha)\}\int_{L_{jk}}|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y\\
&\quad\leqslant 2\|\theta\|_\infty I\{B_n(\alpha)\} + 2^j\|\theta'\|_\infty I\{B_n^c(\alpha)\}\int_{L_{jk}}|\hat{U}_n(y)-y|\mathrm{d}y.
\end{aligned}
$$

We obtain

$$
\mathbb{E}\left(2^{-j/2}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y\right)^p \leqslant (2\|\theta\|_\infty)^p\frac{K}{n^{2\alpha^2}} + 2^{jp}\|\theta'\|_\infty^p\mathbb{E}\left(\int_{L_{jk}}|\hat{U}_n(y)-y|\mathrm{d}y\right)^p.
$$

Using Hölder's inequality, and then Rosenthal's inequality applied to the binomial distribution,

$$
\begin{aligned}
2^{jp}\|\theta'\|_\infty^p\mathbb{E}\left(\int_{L_{jk}}|\hat{U}_n(y)-y|\mathrm{d}y\right)^p &\leqslant 2^{jp}\|\theta'\|_\infty^p C(\alpha)2^{-j(p-1)}\int_{L_{jk}}\mathbb{E}|\hat{U}_n(y))-y|^p\,\mathrm{d}y\\
&\leqslant \|\theta'\|_\infty^p C(\alpha)C_p\left(\frac{(1-y)^p y + y^p(1-y)}{n^{p-1}} + \frac{(y(1-y))^{p/2}}{n^{p/2}}\right)\\
&\leqslant C\frac{1}{n^{p/2}}.
\end{aligned}
$$

We also obtain

$$
\begin{aligned}
&P\left[2^{-j/2}\int_0^1|\theta_{j,k}(\hat{U}_n(y))-\theta_{j,k}(y)|\mathrm{d}y \geqslant \kappa\sqrt{\frac{\log n}{n}}\right]\\
&\quad\leqslant \frac{K}{n^{2\alpha^2}} + P\left[2^j\|\theta'\|_\infty\int_{L_{jk}}|\hat{U}_n(y)-y|\mathrm{d}y \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right].
\end{aligned}
$$

But

$$
\begin{aligned}
&P\left[2^j\|\theta'\|_\infty\int_{L_{jk}}|\hat{U}_n(y)-y|\mathrm{d}y \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right]\\
&\quad\leqslant P\left[C(\alpha)\|\theta'\|_\infty\sup_{y\in[0,1]}|\hat{U}_n(y)-y| \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right] \leqslant \frac{K}{n^\gamma},
\end{aligned}
$$

using the Dvoretzky–Kiefer–Wolfowitz inequality, with $\gamma = \kappa^2/2C(\alpha)^2\|\theta'\|_\infty^2$.

Now we obtain bounds for $\int_0^1 |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))|\mathrm{d}y$, again using (72). It is clear that there exists, for all j, k, an interval $L_{j,k}$ homothetical to the support of θ_{jk} with a ratio bounded by a constant independent of j, k (but depending on α) such that, for $2^j \leqslant \sqrt{n/\log n}$,

$$\begin{aligned}
&\int_0^1 |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))|\mathrm{d}y \\
&\quad = I\{B_n(\alpha)\}\int_0^1 |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))|\mathrm{d}y \\
&\qquad + I\{B_n^c(\alpha)\}\int_{L_{jk}} |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))|\mathrm{d}y \\
&\quad \leqslant 2^{j/2} I\{B_n(\alpha)\}2\|\theta'\|_\infty + 2^{3j/2}\|\theta'\|_\infty I\{B_n^c(\alpha)\}\int_{L_{jk}} |\hat{U}_n(y) - 2^{-j-1} - \hat{U}_n(y - 2^{-j-1})|\mathrm{d}y \\
&\quad = 2^{j/2} I\{B_n(\alpha)\}2\|\theta'\|_\infty + 2^{3j/2}\|\theta'\|_\infty I\{B_n^c(\alpha)\}\int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right|\mathrm{d}y.
\end{aligned}$$

We have

$$\begin{aligned}
&\mathbb{E}\left(\int_0^1 |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))|\mathrm{d}y\right)^p \\
&\quad \leqslant \left(\frac{n}{\log n}\right)^{p/4}(2\|\theta'\|_\infty)^p \frac{K}{n^{2\alpha^2}} + 2^{3pj/2}\|\theta'\|_\infty^p \mathbb{E}\left(\int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right|\mathrm{d}y\right)^p.
\end{aligned}$$

Again, using Hölder's inequality, and then Rosenthal's inequality applied to the binomial law,

$$\begin{aligned}
&2^{3pj/2}\|\theta'\|_\infty^p \mathbb{E}\left(\int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right|\mathrm{d}y\right)^p \\
&\quad \leqslant 2^{3pj/2}\|\theta'\|_\infty^p C(\alpha)2^{-j(p-1)}\int_{L_{jk}} \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right|^p \mathrm{d}y \\
&\quad \leqslant 2^{pj/2}\|\theta'\|_\infty^p C(\alpha)C_p\left(\frac{(1-2^{-j-1})^p 2^{-j-1} + 2^{-(j+1)p}(1-2^{-j-1})}{n^{p-1}} + \frac{(2^{-j-1}(1-2^{-j-1}))^{p/2}}{n^{p/2}}\right) \\
&\quad \leqslant 2^{pj/2}\|\theta'\|_\infty^p C(\alpha)C_p\left(\frac{2^{-j}}{n^{p-1}} + \frac{2^{-jp/2}}{n^{p/2}}\right) = \|\theta'\|_\infty^p C(\alpha)C_p\frac{1}{n^{p/2}}\left(\left(\frac{2^j}{n}\right)^{p/2-1} + 1\right) \leqslant C\frac{1}{n^{p/2}}.
\end{aligned}$$

We also have

$$P\left[\int_0^1 |\theta_{j,k}(\hat{U}_n(y) - 2^{-j-1}) - \theta_{j,k}(\hat{U}_n(y - 2^{-j-1}))| \mathrm{d}y \geqslant \kappa \sqrt{\frac{\log n}{n}}\right]$$

$$\leqslant \left(\frac{n}{\log n}\right)^{1/4} 2\|\theta'\|_\infty \frac{K}{n^{2\alpha^2}} + P\left[2^{3j/2}\|\theta'\|_\infty \int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right| \mathrm{d}y \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right].$$

But

$$P\left[2^{3j/2}\|\theta'\|_\infty \int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right| \mathrm{d}y \geqslant \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right]$$

$$= P\left[\frac{1}{|L_{jk}|}\int_{L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right| \mathrm{d}y \geqslant \frac{\kappa}{2C(\alpha)\|\theta'\|_\infty} 2^{-j/2}\sqrt{\frac{\log n}{n}}\right]$$

$$\leqslant P\left[\sup_{y \in L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right| \geqslant \frac{\kappa}{2C(\alpha)\|\theta'\|_\infty} 2^{-j/2}\sqrt{\frac{\log n}{n}}\right].$$

Let us recall Talagrand's inequality (see Ledoux 2001, p. 149): for

$$Z = \sup_{y \in L_{jk}} \left|\frac{1}{n}\sum_{i=1}^n 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right|, \qquad \sigma^2 = n2^{-j}(1 - 2^{-j}),$$

there exists a universal constant K such that

$$P\left[Z \geqslant 2\mathbb{E}(Z) + \frac{1}{n}\sigma\sqrt{Kr} + \frac{1}{n}2Kr\right] \leqslant \exp(-r).$$

So

$$P\left[Z \geqslant 2\mathbb{E}(Z) + \frac{1}{n}\sigma\sqrt{K\gamma}\log n + \frac{1}{n}2K\gamma\log n\right] \leqslant \frac{1}{n^\gamma}$$

and this implies

$$P\left[Z \geqslant 2\mathbb{E}(Z) + 2^{-j/2}\sqrt{K\gamma\frac{\log n}{n}} + \frac{1}{n}2K\gamma\log n\right] \leqslant \frac{1}{n^\gamma},$$

and as $2^j \leqslant \sqrt{n/\log n}$, this in turn clearly implies that there exists a new constant K such that

$$P\left[Z \geqslant 2\mathbb{E}(Z) + 2^{-j/2}\sqrt{K\gamma\frac{\log n}{n}}\right] \leqslant \frac{1}{n^\gamma}. \tag{75}$$

But now, by the Vapnik–Chervonenkis inequality (Devroye and Lugosi 1996),

$$\mathbb{E}(Z) \leqslant 2\sqrt{\frac{\log 2\mathcal{S}_{\mathcal{A}}(n)}{n}} \leqslant 2\sqrt{\frac{\log 2n^2}{n}},$$

where $\mathcal{S}_{\mathcal{A}}(n)$ is the shattering number of the class

$$\mathcal{A} = \{1_{]y-2^{-j-1},y]},\ y \in L_{jk}\},$$

which can be easily computed:

$$\mathcal{S}_{\mathcal{A}}(n) = 1 + \frac{n(n+1)}{2} \leqslant n^2.$$

Actually we need a slight improvement on this inequality, taking into account the fact that the variance is $2^{-j}(1-2^{-j})$. Using such an improvement (Lugosi 2003, p. 17), we obtain

$$\mathbb{E}(Z) \leqslant \frac{16\log 2\mathcal{S}_{\mathcal{A}}(2n)}{n} + 2^{-j/2}\sqrt{\frac{32\log 2\mathcal{S}_{\mathcal{A}}(2n)}{n}}.$$

So, as $2^j \leqslant \sqrt{n/\log n}$,

$$\mathbb{E}(Z) \leqslant C2^{-j/2}\sqrt{\frac{\log n}{n}}. \tag{76}$$

Now using (75) and (76), we obtain

$$P\left[\sup_{y\in L_{jk}}\left|\frac{1}{n}\sum_{i=1}^{n} 1_{]y-2^{-j-1},y]}(U_i) - 2^{-j-1}\right| \geqslant \frac{\kappa}{2C(\alpha)\|\theta'\|_\infty}2^{-j/2}\sqrt{\frac{\log n}{n}}\right] \leqslant \frac{1}{n^\gamma},$$

for κ suitably choosen.

9.3. Bounds for $\mathbb{E}(\|\hat{f}^{@} - \hat{f}'\|^p_{\mathbb{L}_p(\mathbb{R})})$

Our aim now is to prove Proposition 7. We have

$$\hat{f}^{@} - \hat{f}' = \sum_{J\geqslant j\geqslant -1}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk}I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\{\psi_{jk}(\hat{G}_n(x)) - \psi_{jk}(G(x))\}$$

and

$$\begin{aligned}\int|\hat{f}^{@}(x) - \hat{f}'(x)|^p\,\mathrm{d}x &= \int\left|\sum_{J\geqslant j\geqslant -1}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk}I\{|\hat{\beta}^{@}_{jk}| \geqslant \kappa t_n\}\{\psi_{jk}(\hat{G}_n(x)) - \psi_{jk}(G(x))\}\right|^p\mathrm{d}x \\ &= \int_0^1\left|\sum_{J\geqslant j\geqslant -1}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk}I\{|\hat{\beta}^{@}_{jk}| \geqslant \kappa t_n\}\{\psi_{jk}(\hat{G}_n(G^{-1}(y))) - \psi_{jk}(y)\}\right|^p\omega(y)\mathrm{d}y \\ &= \int_0^1\left|\sum_{J\geqslant j\geqslant -1}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk}I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\{\psi_{jk}(\hat{U}_n(y)) - \psi_{jk}(y)\}\right|^p\omega(y)\mathrm{d}y.\end{aligned}$$

Let

$$A(n, y) = \{|\hat{U}_n(y) - y| \geqslant \alpha t_n\}, \qquad t_n = \sqrt{\frac{\log n}{n}}.$$

So, as before, using Hoeffding's inequality, $P\{|\hat{U}_n(y) - y| \geqslant \lambda\} \leqslant 2\exp(-2n\lambda^2)$, and

$$P(A(n, y)) = P\{|\hat{U}'_n(y) - y| \geqslant \alpha t_n\} \leqslant \frac{2}{n^{2\alpha^2}},$$
$$\mathbb{E}|\hat{U}_n(y) - y|^p \leqslant \frac{1}{(2n)^{p/2}} p\Gamma\left(\frac{p}{2}\right).$$

Let us define

$$\Delta_{jk}(y) = \{\psi_{jk}(\hat{U}_n(y)) - \psi_{jk}(y)\}$$

and clearly

$$\mathbb{E}|\Delta_{jk}|^p \leqslant C_p \frac{2^{j3p/2}}{n^{p/2}}. \tag{77}$$

From now C denotes a constant depending only on the wavelet system, p, but not on the data, and may vary from line to line in the proof. We have

$$\sum_{j\leqslant J}\sum\nolimits_{k\in\Lambda_j}\hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\{\psi_{jk}(\hat{U}'_n(y)) - \psi_{jk}(y)\} = F_1 + F_2$$
$$= \sum_{j\leqslant J}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\Delta_{jk}(y) I\{A(n, y)\}$$
$$+ \sum_{j\leqslant J}\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\Delta_{jk}(y) I\{A^C(n, y)\}.$$

So

$$\mathbb{E}\int|\hat{f}^{@}(x) - \hat{f}'(x)|^p\,\mathrm{d}x \leqslant C\left(\int_0^1 \mathbb{E}(|F_1(y)|^p)\omega(y)\mathrm{d}y + \int_0^1 \mathbb{E}(|F_2(y)|^p)\omega(y)\mathrm{d}y\right).$$

We first bound $\mathbb{E}\int_0^1 |F_1(y)|^p\omega(y)\mathrm{d}y$. Using Hölder's inequality,

$$\int_0^1 |F_1(y)|^p\omega(y)\mathrm{d}y \leqslant (J2^J)^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}|\hat{\beta}^{@}_{jk}|^p I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}$$
$$\times \int_0^1 |\psi_{jk}(\hat{U}_n(y)) - \psi_{jk}(y)|^p I\{A(n, y)\}\omega(y)\mathrm{d}y,$$

and using Schwarz's inequality,

$$\mathbb{E}\int_0^1 |F_1(y)|^p \omega(y)\mathrm{d}y \leqslant (J2^J)^{p-1} \sum_{j\leqslant J}\sum_{k\in\Lambda_j} [\mathbb{E}(|\hat{\beta}^{@}_{jk}|^{2p} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\})]^{1/2}$$

$$\times \left[\mathbb{E}\left(\int_0^1 |\psi_{jk}(\hat{U}_n(y)) - \psi_{jk}(y)|^p I\{A(n, y)\}\omega(y)\mathrm{d}y\right)^2\right]^{1/2}.$$

But, using (67) and the a priori information, $f \circ G^{-1} \in Lip_{1/2}$,

$$[\mathbb{E}(|\hat{\beta}^{@}_{jk}|^{2p} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\})]^{1/2} \leqslant [\mathbb{E}(|\hat{\beta}^{@}_{jk} - \beta_{jk}|^{2p})]^{1/2} + |\beta_{jk}|^p$$

$$\leqslant \left\{C_p \frac{(1+D^{2p})}{n^p}\right\}^{1/2} + \{2^{-j}\|f \circ G^{-1}\|_{lip_{1/2}}\}^p \leqslant C(1+D^p)\left(\frac{1}{n^{p/2}} + 2^{-jp}\right)$$

and

$$\left[\mathbb{E}\left(\int_0^1 |\psi_{jk}(\hat{U}_n(y)) - \psi_{jk}(y)|^p I\{A(n, y)\}\omega(y)\mathrm{d}y\right)^2\right]^{1/2} \leqslant C2^{jp/2}\frac{1}{n^{\alpha^2}}.$$

So

$$\mathbb{E}\int_0^1 |F_1(y)|^p \omega(y)\mathrm{d}y \leqslant C(J2^J)^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}(1+D^p)\left(\frac{1}{n^{p/2}} + 2^{-jp}\right)2^{jp/2}\frac{1}{n^{\alpha^2}}$$

$$\leqslant CJ^{p-1}(1+D^p)\frac{1}{n^{\alpha^2}}\left(\frac{2^{Jp}}{n^{p/2}} + 1 \vee 2^{-J(p/2-1)J}\right) \leqslant C(1+D^p)\frac{1}{n^{p/2}}$$

if α is large enough.

We now estimate $\mathbb{E}\int_0^1 |F_2(y)|^p\omega(y)\mathrm{d}y$. Using Hölder's inequality,

$$\mathbb{E}\int_0^1 |F_2(y)|^p \omega(y)\mathrm{d}y \leqslant J^{p-1}\sum_{j\leqslant J}\mathbb{E}\int_0^1\left|\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk} I\{|\hat{\beta}^{@}_{jk}| \geqslant Kt_n\}\Delta_{jk}(y) I\{A^C(n, y)\}\right|^p \omega(y)\mathrm{d}y;$$

but obviously

$$\Delta_{jk}(y)I\{A^C(n, y)\} = \Delta_{jk}(y)I\{A^C(n, y)\}I\{L_{jk}\}(y),$$

where L_{jk} is some fixed homothetic interval of $[k/2^j, (k+1)/2^j]$. Let us observe that

$$\sum_{k\in\Lambda_j} I\{L_{jk}\} \leqslant CI\{[0, 1]\} \tag{78}$$

so, applying Hölder's inequality again,

$$\left|\sum_{k\in\Lambda_j}\hat{\beta}^{@}_{jk}I\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}\Delta_{jk}(y)I\{A^C(n,\,y)\}\right|^p$$

$$\leqslant\left(\sum_{k\in\Lambda_j}|\hat{\beta}^{@}_{jk}|^pI\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}|\Delta_{jk}(y)|^pI\{A^C(n,\,y)\}I\{L_{jk}\}(y)\right)\left(\sum_{k\in\Lambda_j}I\{L_{jk}\}\right)^{p-1}$$

$$\leqslant C\sum_{k\in\Lambda_j}(|\hat{\beta}^{@}_{jk}-\beta_{jk}|^p+|\beta_{jk}|^p)I\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)$$

$$\leqslant C\left\{\sum_{k\in\Lambda_j}|\beta_{jk}|^p|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)+\sum_{k\in\Lambda_j}|\hat{\beta}^{@}_{jk}-\beta_{jk}|^pI\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)\right\}.$$

Thus

$$\mathbb{E}\int_0^1|F_2(y)|^p\omega(y)\mathrm{d}y\leqslant C(A_1+A_2+A_3).$$

We conclude by bounding each of the A_i in turn. First,

$$A_1=J^{p-1}\sum_{j\leqslant J}\mathbb{E}\int_0^1\sum_{k\in\Lambda_j}|\beta_{jk}|^p|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}|\beta_{jk}|^p\int_0^1\mathbb{E}|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant J^{p-1}\frac{1}{n^{p/2}}\sum_{j\leqslant J}2^{jp}\sum_{k\in\Lambda_j}|\beta_{jk}|^p2^{jp/2}\omega(I_{jk})$$

$$\leqslant CJ^{p-1}\frac{1}{n^{p/2}}\sum_{j\leqslant J}2^{jp}2^{-jsp}\|f(G^{-1})\|_{B_{s,p,\infty}}\leqslant C\|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)}\left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)},$$

because if $s\geqslant 1$, then

$$J^{p-1}\frac{1}{n^{p/2}}\sum_{j\leqslant J}2^{jp}2^{-jsp}\leqslant J^p\frac{1}{n^{p/2}}\leqslant C(\log(n))^p\frac{1}{n^{p/2}}\leqslant C\left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)};$$

and if $1/2\leqslant s<1$, then

$$J^{p-1}\frac{1}{n^{p/2}}\sum_{j\leqslant J}2^{jp}2^{-jsp}\leqslant C(\log(n))^{p-1}\frac{1}{n^{p/2}}2^{Jp(1-s)}\leqslant C(\log(n))^{p-1}\frac{1}{n^{p/2}}\left(\sqrt{\frac{n}{\log n}}\right)^{p(1-s)}$$

$$\leqslant C\frac{1}{n^{sp/2}}(\log n)^{(p/2)(1-s)-1}\leqslant C\left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)}.$$

Next,

$$A_2=J^{p-1}\sum_{j\leqslant J}\mathbb{E}\int_0^1\sum_{k\in\Lambda_j}|\hat{\beta}^{@}_{jk}-\beta_{jk}|^pI\{|\beta_{jk}|<Kt_n/2\}I\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}\int_0^1\mathbb{E}|\hat{\beta}^{@}_{jk}-\beta_{jk}|^pI\{|\hat{\beta}^{@}_{jk}-\beta_{jk}|>Kt_n/2\}|\Delta_{jk}(y)|^pI\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$=J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}\int_0^1(\mathbb{E}|\hat{\beta}^{@}_{jk}-\beta_{jk}|^{3p})^{1/3}(P\{|\hat{\beta}^{@}_{jk}-\beta_{jk}|>Kt_n/2\})^{1/3}$$

$$\times(\mathbb{E}|\Delta_{jk}(y)|^{3p})^{1/3}I\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

using Hölder's inequality. But by (77) and (67) applied to $3p$ instead of p,

$$A_2\leqslant J^{p-1}\sum_{j\leqslant J}C\frac{1+D^p}{n^{p/2}}\frac{1}{n^{-\gamma p}}\frac{2^{3jp/2}}{n^{p/2}}\int_0^1\sum_{j\leqslant J}I\{L_{jk}\}(y)\omega(y)\mathrm{d}y.$$

So certainly

$$A_2\leqslant C\left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)}.$$

Finally,

$$A_3=J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}\int_0^1\mathbb{E}\Big[|\hat{\beta}^{@}_{jk}-\beta_{jk}|^pI\{|\beta_{jk}|\geqslant Kt_n/2\}I\{|\hat{\beta}^{@}_{jk}|\geqslant Kt_n\}|\Delta_{jk}(y)|^p\Big]$$

$$\times\ I\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}I\{|\beta_{jk}|\geqslant Kt_n/2\}\int_0^1\mathbb{E}\Big[|\hat{\beta}^{@}_{jk}-\beta_{jk}|^p|\Delta_{jk}(y)|^p\Big]I\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant J^{p-1}\sum_{j\leqslant J}\sum_{k\in\Lambda_j}I\{|\beta_{jk}|\geqslant Kt_n/2\}\int_0^1(\mathbb{E}|\hat{\beta}^{@}_{jk}-\beta_{jk}|^{2p})^{1/2}(\mathbb{E}|\Delta_{jk}(y)|^{2p})^{1/2}I\{L_{jk}\}(y)\omega(y)\mathrm{d}y,$$

using Schwarz's inequality. Again by (77) and (67) applied to $3p$ instead of p,

$$A_3 \leqslant CJ^{p-1} \sum_{j \leqslant J} \sum_{k \in \Lambda_j} I\{|\beta_{jk}| \geqslant Kt_n/2\} \int_0^1 C \frac{1+D^p}{n^{p/2}} \frac{2^{3jp/2}}{n^{p/2}} I\{L_{jk}\}(y)\omega(y)\mathrm{d}y$$

$$\leqslant C \frac{1+D^p}{n^p} J^{p-1} \sum_{j \leqslant J} 2^{jp} \sum_{k \in \Lambda_j} I\{|\beta_{jk}| \geqslant Kt_n/2\}\omega(I_{jk})2^{jp/2}.$$

But

$$\sum_{k \in \Lambda_j} I\{|\beta_{jk}| \geqslant Kt_n/2\}2^{jp/2}\omega(I_{jk}) \leqslant \left(\frac{1}{Kt_n/2}\right)^p \sum_{k \in \Lambda_j} |\beta_{jk}|^p 2^{jp/2}\omega(I_{jk})$$

$$\leqslant \left(\frac{n}{(K/2)\log n}\right)^{p/2} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} 2^{-jsp}.$$

So

$$A_3 \leqslant C \frac{1+D^p}{n^p} (\log n)^{p-1} \left(\frac{n}{(K/2)\log n}\right)^{p/2} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \sum_{j \leqslant J} 2^{jp(1-s)}$$

$$\leqslant C \frac{1+D^p}{n^{p/2}} (\log n)^{p/2-1} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \sum_{j \leqslant J} 2^{jp(1-s)}.$$

If $s \geqslant 1$, then clearly

$$A_3 \leqslant C \frac{1+D^p}{n^{p/2}} (\log n)^{p/2} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \leqslant C(1+D^p)\|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)}.$$

If $1/2 \leqslant s < 1$, then

$$A_3 \leqslant C \frac{1+D^p}{n^{p/2}} (\log n)^{p/2-1} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \left(\sqrt{\frac{n}{\log n}}\right)^{p(1-s)}$$

$$\leqslant C(1+D^p)(\log n)^{-1} \|f(G^{-1})\|^p_{B_{s,p,\infty}(\omega)} \left(\sqrt{\frac{\log n}{n}}\right)^{sp/2}$$

$$\leqslant C(1+D^p)\|f(G^{-1})\|^p_{B_{sp\infty}(\omega)} \left(\sqrt{\frac{\log n}{n}}\right)^{sp/(1+2s)}.$$

as obviously $s/2 \geqslant s/(1+2s)$.

References

Antoniadis, A. and Fan, J. (2001) Regularization of wavelet approximations (with discussion). *J. Amer. Statist. Assoc.*, **96**, 939–967.

Antoniadis, A., Grégoire, G. and Vial, P. (1997) Random design wavelet curve smoothing. *Statist. Probab. Lett.*, **35**, 225–232.

Cai, T.T. and Brown, L.D. (1998) Wavelet shrinkage for nonequispaced samples. *Ann. Statist.*, **26**, 1783–1799.

Cai, T.T. and Brown, L.D. (1999) Wavelet regression for random uniform design. *Statist. Probab. Lett.*, **42**, 313–321.

Clerc, M. and Mallat, S. (2003) Estimating deformations of stationary processes. *Ann. Statist.*, **31**, 1772–1821.

Cohen, A., DeVore, R., Kerkyacharian, G. and Picard, D. (2001) Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, **11**, 167–191.

Coifman, R. and Fefferman, C. (1974) Weighted norm inequality for maximal functions and singular integrals. *Studia Math*., **109**(3), 241–250.

Delouille, V., Franke, J. and Von Sachs, R. (2001) Nonparametric stochastic regression with design-adapted wavelets. *Sankhyā Ser. A*, **63**, 328–366.

Delouille, V., Simoens, J. and Von Sachs, R. (2004) Smooth design-adapted wavelets for nonparametric stochastic regression. *J. Amer. Statist. Assoc*., **99**, 643–658.

DeVore, R. (1998) Nonlinear approximation. In A. Iserles (ed.), *Acta Numerica, Vol. 7*, pp. 51–150. Cambridge: Cambridge University Press.

Devroye, L. and Lugosi, G. (1996) *A Probabilist Theory of Pattern Recognition.* New York: Springer-Verlag.

Donoho, D.L. (1993) Unconditional bases are optimal bases for data compression and statistical estimation. *Appl. Comput. Harmon. Anal.*, **1**, 100–115.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

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

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *J Roy. Statist. Soc. Ser. B*, **57**, 301–369.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642–669.

Foster, G. (1996) Wavelet for period analysis of unequally sampled time series. *Astronomy J.*, **112**, 1709–1729.

García-Cuerva, J. and Rubio de Francia, J.L. (1985) *Weighted Norm Inequalities and Related Topics*, North-Holland Math. Stud. 116. Amsterdam: North-Holland.

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

Hall, P. and Turlach, B.A. (1997) Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.*, **25**, 1912–1925.

Härdle W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation and Statistical Applications*, Lecture Notes in Statist. 129. New York: Springer-Verlag.

Kerkyacharian, G. and Picard, D. (2000) Thresholding algorithms, maxisets and well-concentrated bases (with discussion). *Test*, **9**(2), 283–345.

Kerkyacharian, G. and Picard, D. (2002) Non-linear approximation and Muckenhoupt weights. Technical report.

Kovac, A. and Silverman, B.W. (2000) Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Amer. Statist. Assoc.*, **95**, 172–183.

Ledoux, M. (2001) *The Concentration of Measure Phenomenon*, Math. Surveys Monogr. 89. Providence, RI: American Mathematical Society.

Le Pennec, E. and Mallat, S. (2003) Sparse geometric image representation with bandelets. http://www.cmap.polytechnique.fr/∼mallat/biblio.html.

Lugosi, G. (2003) Pattern classification and learning theory. In L. Györfi (ed.), *Principles of Nonparametric Learning*, pp. 1–56. Vienna: Springer-Verlag.

Maxim, V. (2002) Denoising signals observed on a random design. Paper presented to the Fifth AFA-SMAI Conference on Curves and Surfaces.

Meyer, Y. (1990) *Ondelettes et Opérateurs*. Paris: Hermann.

Muckenhoupt, B. (1972) Weighted norm inequalities for the Hardy maximal function. *Trans. Amer. Math. Soc*. **165**, 207–226.

Pensky, M. and Vidakovic, B. (2001) On non-equally spaced wavelet regression. *Ann. Inst. Statist. Math.*, **53**, 681–690.

Rivoirard, V. (2002) Estimation bayesienne non parametrique. Doctoral thesis, Université Paris VII.

Sardy, S., Percival, D.B., Bruce A.G., Gao, H.-Y. and Stuetzle, W. (1999) Wavelet shrinkage for unequally spaced data. *Statist. Comput.*, **9**, 65–75.

Stein, E. (1993) *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals.* Princeton, NJ: Princeton University Press.

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

Temlyakov, N.V. (1998) The best *m*-term approximation and greedy algorithms. *Adv. Comput. Math.*, **8**, 249–265.

Received January 2003 and revised September 2004