A maxiset approach of a Gaussian white noise model

Christophe Chesneau

Universités de Paris 6 & Paris 7 - CNRS (UMR 7599)

PRÉPUBLICATIONS DU LABORATOIRE

DE PROBABILITÉS & MODÈLES ALÉATOIRES

4, place Jussieu - Case 188 - 75 252 Paris cedex 05

http://www.proba.jussieu.fr

A maxiset approach of a
Gaussian white noise model

C. CHESNEAU

JUIN 2004

Prépublication n^o 920

Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599,
Université Paris VI & Université Paris VII,
4, place Jussieu, Case 188, F-75252 Paris Cedex 05.

A maxiset approach of a Gaussian white noise model

Chesneau Christophe

June 9, 2004

Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, 175 rue de Chevaleret, F-75013 Paris, France.

Abstract

This paper is devoted to the estimation of an unknown function f in the framework of a Gaussian white noise model. The noise process is represented by $t \to \frac{1}{\sqrt{n}} \int_0^t g(x)dB_x$, where the variance function g is assumed to be known. Adopting the maxiset point of view, we study the performance of two different hard thresholding estimators in $\mathbb{L}^p$ norm. In a first part, we expand f on a compactly supported wavelet basis $\{\psi_\lambda(.);\ \lambda \in \Lambda\}$. From this decomposition, we use some results about the heteroscedastic white noise model to construct a well adapted hard thresholding estimator and to exhibit the associated maxiset. In a second part, we introduce the classes of Muckenhoupt weights and we use this analytical tools to investigate the geometrical properties of warped wavelet basis $\{\psi_\lambda(T(.));\ \lambda \in \Lambda\}$ in $\mathbb{L}^p$ norm. Expanding f on such a basis and considering the associated hard thresholding estimator, we investigate the maxiset properties under some assumptions on g. We finally apply this result to find an upper bound over weighted Besov spaces.

Keywords: Non parametric estimation, maximal spaces, thresholding rules, Besov spaces, weak Besov spaces, weighted Besov spaces, Gaussian white noise model, Muckenhoupt weights, warped basis.

AMS: 42B25, 47B99, 62G05, 62G07, 62G20.

1 Introduction

For several years, the problem of signal recovery in Gaussian white noise is a paradigm for non parametric curve estimation. The general framework of this model is the following: For $n \geq 1$, one observes the stochastic process $Y_t^n = Y_t$ defined by:

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}} g(t)dB_t, \quad t \in [0,1] \tag{1}$$

where f is an unknown function which belongs to $\mathbb{L}^2([0,1])$, B_t is a Brownian motion with $B_0 = 0$ and g is a known function which belongs to $\mathbb{L}^2([0,1])$. The goal is to recover f from the noisy observations (1).

Among non parametric situations, the Gaussian white noise model has been studied in several papers starting from Ibragimov and Has'minskii (1977, 1981). This is a model very useful to understand some statistical questions. For instance, the precise mathematical relationship between white noise and curve estimation was established by Brown and Low (1996) who proved the asymptotic equivalence of model (1) and non parametric regression. Under smoothness restrictions on f, the homogeneous white noise model i.e $g(t) \equiv g$ in (1) is asymptotically equivalent to the following experiment: For $i = 1, ..., n$, one observes Y_i^n, where

$$Y_i^n = f(\frac{i}{n}) + \epsilon_i^n, \quad \epsilon_i^n \text{ i.i.d } \sim \mathcal{N}(0, g^2).$$

An analogous result for asymptotic equivalence of density estimation and homogeneous white noise was obtained by Nussbaum (1996). For the study of minimax properties, we refer the reader to the book of Tsybakov (2004).

More generally, introducing some inhomogeneity in the variance function (when $g(t) \not\equiv g$) may enrich significantly the curve estimation modelling. Under some assumptions of regularity on f and conditions of boundedness from above and below on g, Brown and Low (1996) proved that the model (1) is asymptotically equivalent to the following experiment: For $i = 1, ..., n$, one observes (X_i^n, Y_i^n), where

$$Y_i^n = f(X_i^n) + \sigma(X_i^n)\epsilon_i^n. \tag{2}$$

The random variables ϵ_i^n are i.i.d $\mathcal{N}(0, 1)$, the X_i^n are independent of the ϵ_i^n and are i.i.d with density μ. The equivalence between (1) and (2) is obtained under the calibration $g(t) = \frac{\sigma^2(t)}{\mu(t)}$.

The aim of this paper is to study the performance of hard thresholding procedures associated to the estimated function f in a $\mathbb{L}^p$ norm given a realization (1) with some inhomogeneity on g, more general that conditions of boundedness from above and below.

To measure the performance of a statistical procedure, we will adopt the maxiset approach which has been introduced by Cohen, De Vore, Kerkyacharian and Picard (2000). It consists in investigating the maximal space (or maxiset) where a procedure has a given rate of convergence. This point of view is less pessimist than the minimax one since it provides a functional set which is authentically connected to the procedure and the model. Precisely, this authors proved that the maxiset associated with hard thresholding procedures is the intersection of two well known spaces (see Theorem 3.1 below).

Focused on the model (1), this paper is organized in two parts which describe the maxisets obtained for two different hard thresholding estimators in $\mathbb{L}^p$ norm. In a first part, we expand the unknown function of interest f on a compactly supported wavelet basis $\xi = \{\psi_\lambda(.);\ \lambda \in \Lambda\}$. To construct a hard thresholding estimator well adapted to (1) and to identify the associated maxiset, we exploit two important geometrical properties of the basis ξ and we use some results of Picard and Kerkyacharian (2000) about the heteroscedastic white noise model:

$$y_\lambda = \beta_\lambda + \frac{1}{\sqrt{n}}\sigma_\lambda\epsilon_\lambda, \quad \epsilon_\lambda \text{ i.i.d} \sim \mathcal{N}(0, 1), \quad \lambda \in \Lambda,$$

where $(\beta_\lambda)_{\lambda\in\Lambda}$ is a sequence to be estimated (see Theorem 4.1 below).

In a second part, we investigate an another strategy. We expand f on a compactly supported warped wavelet basis $\{\psi_\lambda(T(.));\ \lambda\in\Lambda\}$ where the warping factor T is a known function only depending on g. This allows to perform a very stable and computable thresholding algorithm. To study the performance of the associated hard thresholding estimator, we introduce the *classes of Muckenhoupt weights* $\mathcal{A}_p$ which have been introduced by Muckenhoupt in (1972). They characterize the boundedness of some integral operators on weighted $\mathbb{L}^p$ spaces like the Hardy-Littlewood maximal operator or the Hilbert transform. Using the fact that there exists a link between geometrical properties of warped wavelet basis and Muckenhoupt weights theory as it shown by Garcia and Martell in (1999), we describe the performance of our procedure in terms of maxiset properties (see Theorem 6.1 below).

The above method has the advantage of giving interesting statistical results about some weighted spaces. For instance, observing the realisations (2) with $\sigma\equiv 1$, expanding the function f on the warped wavelet basis $\{\psi_\lambda(\mathbb{F}_{X_1}(.));\ \lambda\in\Lambda\}$ where $\mathbb{F}_{X_1}(x)=\mathbb{P}(X_1\le x)$ and taking the associated hard thresholding estimator, Picard and Kerkyacharian (2003) obtained an upper bound over weighted Besov spaces (see Definition 7.1 below) corresponding to the rate which proved to be minimax in a uniform design up to a logarithmic factors under some Muckenhoupt conditions on g. In the last section of this paper, we find a similar result from our observations (1) (see Proposition 7.1 below).

2 Maxiset point of view and hard thresholding procedures

Here we start by introducing the maxiset point of view and after briefly explaining the reason of this choice, we define the hard thresholding procedures. Throughout this paper, for non negative locally integrable function w on $\Omega\subseteq\mathbb{R}$, we define:

$$\mathbb{L}^p_w(\Omega)=\big\{f \text{ measurable on } \Omega \mid \|f\|^p_{w,p}=\int_\Omega |f(t)|^p w(t)dt<+\infty\big\}.$$

$\mathbb{L}^p(\Omega)=\mathbb{L}^p_1(\Omega)$ denotes the Lebesgue space.

2.1 Maxiset point of view

From an unknown function f of $\mathbb{L}^p(\Omega)$ (where Ω is a bounded interval of $\mathbb{R}$) randomly observed via n-statistical observations $\mathcal{O}^n$, the maxiset approach consists in investigating the maximal space (called maxiset) where an estimation procedure $\hat{f}$ constructed from $\mathcal{O}^n$ attains a given rate of convergence $c(n)$ for the $\mathbb{L}^p$ risk.

In other words, we want exhibit the space $\mathcal{A}_{\alpha,p}\subset\mathbb{L}^p(\Omega)$ such that for $p>1$ and $1>\alpha>0$:

$$\mathbb{E}(\|\hat{f}-f\|^p_p)\le Cc(n)^{\alpha p}\Longleftrightarrow f\in\mathcal{A}_{\alpha,p} \tag{3}$$

This way of measuring the performances of statistical procedures has been particularly successful in the nonparametric framework. It has often the advantage of giving less arbitrary and pessimistic comparaisons of procedures than the minimax approach. It is clear that we compare the performance of two procedures by comparing their respective maxisets.

The equivalence (3) is possible if and only if $\hat{f}$ is small as possible for the $\mathbb{L}^p$ risk without a priori knowledge of the regularity of f. Among others, this is one of the reasons why we work with *hard thresholding procedures* in our maxiset studied.

2.2 Hard thresholding procedures

First of all we consider $\mathcal{E} = \{e_{j,k}(.);\ j \geq -1\ ,\ k \in \mathbb{Z}\}$ a basis of $\mathbb{L}^p(\Omega)$ and we suppose that all function f of $\mathbb{L}^p(\Omega)$ have an expansion on $\mathcal{E}$ under the form:

$$f(x) = \sum_{\lambda \in \Lambda} \beta_\lambda e_\lambda(x), \quad \forall x \in \Omega$$

where $\lambda = (j,k)$, $\Lambda = \{j \geq -1\ ,\ k \in \mathbb{Z}\}$ and β_λ denote the associated coefficients.

The construction of hard thresholding procedures is accomplished in three steps:

1. A linear step corresponding to the estimation of the β_λ's by some estimators $\hat{\beta}_\lambda$.
2. A non-linear step consisting to consider the hard thresholding operator $\mathcal{T}(\hat{\beta}_\lambda) = w_\lambda \hat{\beta}_\lambda$ where w_λ is characterized by the equivalence:

   $$w_\lambda = 1 \Longleftrightarrow \lambda \in \Lambda_n \text{ and } |\hat{\beta}_\lambda| \geq \kappa \mathrm{c(n)}$$

   - κ denotes a fixed positive constant,
   - $c(n)$ a decreasing positive sequence such that $\lim_{n\to\infty} c(n) = 0$ and there exists a positive constant K verifying $c(n) \leq Kc(n+1)$,
   - Λ_n is a sequence of set such that $\Lambda_1 \subseteq \Lambda_2 \subseteq ... \subseteq \Lambda_n$ and $\cup_{n=1}^{\infty} \Lambda_n = \mathbb{N} \cup \{-1\} \times \mathbb{Z}$.
3. A reconstruction step for deriving the global estimator:

   $$\hat{f}(x) = \sum_{\lambda \in \Lambda_n} \mathcal{T}(\hat{\beta}_\lambda) e_\lambda(x) = \sum_{\lambda \in \Lambda_n} \hat{\beta}_\lambda w_\lambda e_\lambda(x), \quad \forall x \in \Omega \tag{4}$$

Remark 2.1. *The construction of hard thresholding procedures only depends on the data, doesn't require any a priori knowledge on the function f in its construction.*

In the third section, we focus on the maxiset theory applied to hard thresholding procedures for arbitrary bases.

Notations: Throughout this paper, C denotes positive constants only depending on p and that may change from one line to the other.

3 Maxiset associated with the hard thresholding procedures

Now we introduce determinant geometrical notions of *unconditional bases* and *p-Temlyakov's property* which are very useful in maxiset theory.

3.1 Geometrical bases properties

Following Mallat's heuristic, Donoho (1996) pointed out the importance of unconditional bases in statistics. Let us recall the definition:

Definition 3.1 (Unconditional bases of $\mathbb{L}^p(\Omega)$). *We say that $\mathcal{E} = \{e_\lambda(.);\ \lambda \in \Lambda\}$ is an unconditional basis of $\mathbb{L}^p(\Omega)$ if and only if there exists an absolute constant C such that if $|u_\lambda| \leq |v_\lambda|$ for all $\lambda \in \Lambda$, then:*

$$\|\sum_{\lambda\in\Lambda} u_\lambda e_\lambda\|_p^p \leq C\|\sum_{\lambda\in\Lambda} v_\lambda e_\lambda\|_p^p \tag{5}$$

The elementary consequence in Statistics is that shrinkage methods are not expected to produce an explosion of the norm. In the maxiset theory, we also need the notion of *p-Temlyakov's property*:

Definition 3.2 (p-Temlyakov's property). *We say that a basis $\mathcal{E} = \{e_\lambda(.);\ \lambda \in \Lambda\}$ of $\mathbb{L}^p(\Omega)$ satisfies the p-Temlyakov's property if and only if there exists two positive constants c and C such that for any finite set of integer $F \in \Lambda$ we have:*

$$c\sum_{\lambda\in F} \|e_\lambda\|_p^p \leq \int_\Omega (\sum_{\lambda\in F} |e_\lambda(x)|^2)^{\frac{p}{2}} dx \leq C\sum_{\lambda\in F} \|e_\lambda\|_p^p \tag{6}$$

Verified by a basis, this two properties allow to very simply transfer the arguments from $\mathbb{L}^2(\Omega)$ to $\mathbb{L}^p(\Omega)$ as it shown by Picard and Kerkyacharian (2000).

3.2 The key theorem

Now we recall the *key theorem* of this studied which determines the maxiset of hard thresholding procedures for a completely general basis. We refer to Cohen, De Vore, Kerkyacharian and Picard (2000) for its complete proof.

Theorem 3.1. *Here we suppose that the basis $\mathcal{E}$ satisfies the geometrical properties (5) and (6). Let us consider the unknown function $f = \sum_{\lambda\in\Lambda} \beta_\lambda e_\lambda \in \mathbb{L}^p(\Omega)$ which is randomly observed from $\mathcal{O}^n$ and the associated hard thresholding estimator (4) under the calibration:*

- $c(n) = \sqrt{\frac{\ln(n)}{n}}$,
- $\Lambda_n = \{(j,k); j \leq j_1(n), |k| \leq C2^j\}$ *where $j_1(n)$ is the integer verifying $2^{j_1(n)} \leq c(n)^{-r} < 2^{j_1(n)+1}$, r is a positive real number such that:*

$$\sup_{n\geq 1}[\omega(\Lambda_n)c(n)^p] \leq C \tag{7}$$

and ω is the weight defined by $\omega(\mathcal{S}) = \sum_{\lambda\in\mathcal{S}} \|e_\lambda\|_p^p \quad \forall \mathcal{S} \in \Lambda$.

For any $\lambda \in \Lambda_n$, if we suppose also that $\hat{\beta}_\lambda$ satisfies the two following statistical properties:

- *Concentration property:*

$$\mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| \geq \frac{kc(n)}{2}\right) \leq Cc(n)^{2p} \wedge c(n)^4, \tag{8}$$

- *Moment property:*

$$\mathbb{E}\left(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}\right) \leq Cc(n)^{2p}, \tag{9}$$

then we can identify the maxiset links to the hard thresholding estimator (4); for any $1 < p < \infty$, $0 < \alpha < 1$ and κ a large enough constant we have the following equivalence:

$$\mathbb{E}(\|\hat{f} - f\|_p^p) \leq Cc(n)^{\alpha p} \Longleftrightarrow f \in B_{p,\infty}^{\frac{\alpha}{r}}(\mathcal{E}) \cap l_{((1-\alpha)p,p,\infty)}(\mathcal{E})$$

We have denoted $B_{p,\infty}^{\gamma}(\mathcal{E})$ and $l_{(r,p,\infty)}(\mathcal{E})$ the function spaces defined by:

- *For $0 < \gamma < \infty$, we say that a function f of $\mathbb{L}^p(\Omega)$ belongs to the Besov space $B_{p,\infty}^{\gamma}(\mathcal{E})$ if and only if:*

$$\sup_{n \geq 1} c(n)^{-\gamma p} \|f - \sum_{\lambda \in \Lambda_n} \beta_\lambda e_\lambda\|_p^p < \infty \tag{10}$$

- *For $0 < r < p$, we say that a function f of $\mathbb{L}^p(\Omega)$ belongs to the weak Besov space $l_{(r,p,\infty)}(\mathcal{E})$ if and only if:*

$$\sup_{u > 0} u^r \omega(\{\lambda \in \Lambda;\ |\beta_\lambda| > u\}) < \infty. \tag{11}$$

Besov spaces (10) and weak Besov spaces (11) shall be useful throughout this paper with two kinds of wavelet basis. The first one shall be the *compactly supported wavelet basis* and the second one shall be the *compactly supported warped wavelet basis*. Theorem 3.1 is in the center of this studied and it will be applied to obtain the main results of this paper: Theorem 4.1 and Theorem 6.1.

Throughout the next of this studied, we consider the set $\Omega = [0, 1]$.

3.3 Compactly supported wavelet bases

Wavelet analysis requires a description of two basic functions, the scaling function φ and the wavelet ψ. A wavelet system is the finite collection of translated and scaled version of φ and ψ defined by:

$$\varphi_\lambda(.) = \varphi_{j,k}(.) = 2^{\frac{j}{2}} \varphi(2^j. - k), j \in \mathbb{N}, k \in \mathbb{Z}$$

$$\psi_\lambda(.) = \psi_{j,k}(.) = 2^{\frac{j}{2}} \psi(2^j. - k), j \in \mathbb{N}, k \in \mathbb{Z}$$

Here, we suppose ψ and ϕ are compactly supported on $[0, L]$ with $L > 0$, that's why there exists a positive constant C such that $|k| \leq C2^j\ \forall j \in \mathbb{N}$. For the theory of Multiresolution Analysis, see Meyer (1990), Daubechies (1992), Cohen, Daubechies and Vial (1993) or Donoho (1995).

Notation: We denote $\xi = \{\psi_\lambda(.);\ \lambda \in \Lambda\}$ the compactly supported wavelet basis adapted on $[0,1]$. Note that ξ is orthonormal by construction.

Remark 3.1. *It is a classical result that ξ is an unconditional basis of $\mathbb{L}^p([0,1])$ (5) (see Meyer (1990)) and satisfies the p-Temlyakov's property (6) (see Kerkyacharian and Picard (2002b).*

Using such a basis, all function f of $\mathbb{L}^p([0,1])$ have the following expansion:

$$f(x) = \sum_{\lambda \in \Lambda} \beta_\lambda \psi_\lambda(x), \quad \forall x \in [0,1]$$

where

$$\beta_\lambda = \int_0^1 f(t)\psi_\lambda(t)dt \tag{12}$$

and $\psi_{-1,k}(.) = \varphi_{0,k}(.)$ for all $\lambda \in \Lambda = \{\lambda = (j,k), -1 \leq j \leq \infty, |k| \leq C2^j\}$. The literature who talks about wavelet procedures is very impressive. See for instance Donoho and Jonhstone (1996), Donoho, Johnstone, Kerkyacharian and Picard (1994), Jonhstone (2000), Kerkyacharian and Picard (2000), Hardle, Kerkyacharian, Picard and Tsybakov (1998) or Johnstone (1998).

By definition of ξ, it is easy to see that (7) is satisfied for $r = 2$. Therefore, if the statistical properties (8) and (9) are satisfied, we can apply Theorem 3.1 with $\mathcal{E} = \xi$. Several papers investigate the performances of hard thresholding procedures (4) by using this result from some statistical models like density estimation, heteroscedastic Gaussian white noise.... See for instance Kerkyacharian and Picard (2000, 2002a, 2002b), Rivoirard (2002) or Autin (2003).

In the section 4, we expand the unknown function f on such a basis and we investigate the performance of the associated hard thresholding estimator.

4 First idea: Heteroscedastic framework

The classical idea consists of exhibiting the wavelet coefficients β_λ defined in (12) from our statistical model (1). Using the compactly supported wavelet basis ξ and the estimator:

$$\hat{\beta}_\lambda = \int_0^1 \psi_\lambda(t)dY_t \tag{13}$$

we obtain:

$$\hat{\beta}_\lambda = \int_0^1 f(t)\psi_\lambda(t)dt + \frac{1}{\sqrt{n}}\int_0^1 g(t)\psi_\lambda(t)dB_t.$$

Since:

$$\frac{1}{\sqrt{n}}\int_0^1 g(t)\psi_\lambda(t)dB_t \sim \mathcal{N}(0, \frac{\rho_\lambda^2}{n})$$

where $\rho_\lambda = \sqrt{\int_0^1 g^2(t)\psi_\lambda^2(t)dt}$, it can also be written as follows:

$$\hat{\beta}_\lambda = \beta_\lambda + \frac{1}{\sqrt{n}}\rho_\lambda\epsilon_\lambda$$

where ϵ_λ i.i.d $\sim \mathcal{N}(0,1)$. Therefore, we are in a heteroscedastic framework and it is obvious that observe the collection of the $\hat{\beta}_\lambda$ is equivalent to observe the whole trajectory. Using this remark, we construct in subsection 4.1 a *hard local thresholding estimator* well adapted to our statistical model (1).

Notation: From a sequence $v = (v_j)_{j\in\mathbb{N}}$ and the basis ξ, we denote $\xi_v = \{v_j\psi_\lambda(.);\ \lambda \in \Lambda\}$ the compactly supported weighted wavelet basis adapted on $[0,1]$.

4.1 Hard local thresholding estimator

In order to do the link between the basis ξ_m and the p-Temlyakov's property, let us consider the reverse Hölder inequality:

Definition 4.1 (Reverse Hölder inequality). *We say that a sequence $(t_j)_{j\in\mathbb{N}}$ verifies the reverse Hölder inequality if and only if for all $A \subseteq \mathbb{N}$ we have:*

$$(\sum_{j\in A}(2^{\frac{j}{2}}t_j)^{p\wedge 2})^{\frac{1}{p\wedge 2}} \leq C(\sum_{j\in A}(2^{\frac{j}{2}}t_j)^{p\vee 2})^{\frac{1}{p\vee 2}} \tag{14}$$

Let us also define the coefficients d_j and c_j by

$$d_j = 2^{\frac{j}{2}} \sup_{0\leq k\leq 2^j - L}\sqrt{\int_{\frac{k}{2^j}}^{\frac{k+L}{2^j}} g^2(t)dt}, \quad c_j = 2^{\nu_0 j}$$

where ν_0 is the smaller real number ν belongs to $]-\frac{1}{2},\frac{1}{2}]$ such that there exists a positive constant C verifying:

$$d_j \leq C2^{\nu j}$$

From d_j and c_j, we set:

$$m_j = \begin{cases} d_j & \text{if we can prove that } d_j \text{ verifies (14),} \\ c_j & \text{if we can not prove that } d_j \text{ verifies (14).} \end{cases} \tag{15}$$

Clearly, all function f of $\mathbb{L}^p([0,1])$ have the following atomic decomposition:

$$f(x) = \sum_{\lambda\in\Lambda}\beta_\lambda\psi_\lambda(x) = \sum_{\lambda\in\Lambda}(\frac{\beta_\lambda}{m_j})(m_j\psi_\lambda(x)), \quad \forall x\in[0,1] \tag{16}$$

Therefore, the hard thresholding estimator associated with the unknown function f of (1) can be written as:

$$\hat{f}^m(x) = \sum_{\lambda \in \Lambda_n} \mathcal{T}(\frac{\hat{\beta}_\lambda}{m_j}) m_j \psi_\lambda(x) = \sum_{\lambda \in \Lambda_n} \hat{\beta}_\lambda w^m_\lambda \psi_\lambda(x), \quad \forall x \in [0,1] \tag{17}$$

where

- $c(n)$ and Λ_n are defined like in Theorem 3.1 with $r = \frac{1}{\frac{1}{2}+\nu_0}$ (we will explain the reason of this choice in subsection 4.2.2),
- $\hat{\beta}_\lambda$ is defined in (13),
- w^m_λ is characterized by the equivalence: $w^m_\lambda = 1 \Longleftrightarrow \lambda \in \Lambda_n$ and $|\hat{\beta}_\lambda| \geq \kappa |\mathrm{m_j}| \mathrm{c(n)}$.

We are now in position to state the main theorem of this section:

Theorem 4.1. *Using the hard thresholding estimator (17) for κ a large enough constant, we have the following equivalence:*

$$\mathbb{E}(\|\hat{f}^m - f\|_p^p) \leq C \left(\frac{\ln(n)}{n} \right)^{\frac{\alpha p}{2}} \Longleftrightarrow f \in B^{(\frac{1}{2}+\nu_0)\alpha}_{p,\infty}(\xi_m) \cap l_{((1-\alpha)p,p,\infty)}(\xi_m) \tag{18}$$

where $0 < \alpha < 1$, $B^\gamma_{p,\infty}(\xi_m)$ and $l_{(r,p,\infty)}(\xi_m)$ are respectively (10) and (11) with the basis ξ_m where $m = (m_j)_{j \in \mathbb{N}}$ is defined in (15). In other words, $B^\gamma_{p,\infty}(\xi_m) = B^\gamma_{p,\infty}(\xi)$ and $l_{(r,p,\infty)}(\xi_m)$ is the set of functions f belonging to $\mathbb{L}^p([0,1])$ such that:

$$\sup_{u>0} u^r \omega(\{\lambda \in \Lambda;\ |\beta_\lambda| > u|m_j|\}) < \infty \tag{19}$$

where $\omega(\mathcal{S}) = \sum_{\lambda \in \mathcal{S}} |m_j|^p \|\psi_\lambda\|_p^p$.

Proof. In the next subsection, we focus on the proof of this theorem. The aim is to apply Theorem 3.1 so we need to prove the geometrical properties (5), (6) of the basis ξ_m, the weight condition (7) and the statistical properties (8), (9).

4.2 Proof of Theorem 4.1

4.2.1 Geometrical properties (5) and (6) of the basis ξ_m

From Remark 3.1 and Definition 3.1, it is easy to see that the basis ξ_m is unconditional on $\mathbb{L}^p([0,1])$. Now let us consider the following results:

Lemma 4.1. *Suppose that $t = (t_j)_{j \in \mathbb{N}}$ is a positive sequence who verifies the reverse Hölder inequality (14) then the p-Temlyakov's property (6) holds for the basis ξ_t.*

For a complete proof see Johnstone, Kerkyacharian, Picard and Raimondo (2004).

Remark 4.1. *In (2000), Kerkyacharian and Picard remark that geometrical dyadic sequence $(2^{vj})_{j \in \mathbb{N}}$ with $v > -\frac{1}{2}$ verifies (14) therefore, if σ_j is proportional to 2^{vj} with $v > -\frac{1}{2}$, we can apply Lemma 4.1 to conclude that the p-Temlyakov's property (6) holds for the basis ξ_σ.*

An immediate consequence is that ξ_m verifies p-Temlyakov's property (6) by direct construction of the m_j's.

Nowadays, we only have a result about general weighted wavelet basis $\{\sigma_\lambda \psi_\lambda(.);\ \lambda \in \Lambda\}$ and p-Temlyakov's property if and only if the σ_λ's depend only on the resolution level j, that's why we have introduced the m_j's in Theorem 4.1.

4.2.2 Weight condition (7)

First of all, by definition of the c_j's we have the inequality:

$$0 \leq m_j \leq C 2^{\nu_0 j}$$

and the orthonormality of ξ we get:

$$\|\psi_\lambda\|_p^p = \int_0^1 |\psi_\lambda(x)|^p dx \leq 2^{\frac{j}{2}(p-2)} \|\psi\|_\infty^{p-2} \int_0^1 |\psi_\lambda(x)|^2 dx = 2^{\frac{j}{2}(p-2)} \|\psi\|_\infty^{p-2}.$$

So using the definition of $j_1(n)$ with $r = \frac{1}{\frac{1}{2}+\nu_0}$, it is obvious that for all $n \geq 1$:

$$c(n)^p \omega(\Lambda_n) = c(n)^p \sum_{\lambda \in \Lambda_n} |m_j|^p \|\psi_\lambda\|_p^p \leq \|\psi\|_\infty^{p-2} c(n)^p \sum_{j \leq j_1(n)} 2^{jp(\frac{1}{2}+\nu_0)} \leq C$$

and the weight condition (7) holds.

4.2.3 Statistical properties (8) and (9)

Using the decomposition (16), we see that in the context of weighted basis ξ_m we focus on the statistical properties of $\frac{\hat{\beta}_\lambda}{m_j}$. Using the heteroscedastic framework and the inequality $\rho_\lambda \leq \|\psi\|_\infty d_j$, we obtain:

$$|\frac{\hat{\beta}_\lambda}{m_j} - \frac{\beta_\lambda}{m_j}| = \frac{\rho_\lambda}{m_j} \frac{1}{\sqrt{n}} |\epsilon_\lambda| \leq \|\psi\|_\infty \frac{1}{\sqrt{n}} |\epsilon_\lambda|. \tag{20}$$

To conclude, we will use the following lemma:

Lemma 4.2. *If $V_n \sim \mathcal{N}(0, \frac{1}{n})$ then for $\kappa \geq 2\sqrt{2p}$, $n \geq 3$ and $c(n) = \sqrt{\frac{\ln(n)}{n}}$ there exists a positive constant C depending only on p such that we have:*

- $\mathbb{P}(|V_n| \geq \frac{\kappa c(n)}{2}) \leq C c(n)^{2p}$,
- $\mathbb{E}(|V_n|^{2p}) \leq C c(n)^{2p}$.

Proof. It is well known that if $N \sim \mathcal{N}(0, \sigma^2)$ then we have the concentration inequality $\mathbb{P}(|N| \geq x) \leq 2\exp(-\frac{x^2}{2\sigma^2})$ so, for $\kappa \geq 2\sqrt{2p}$ and $n \geq 3$, we have:

$$\mathbb{P}\left(|V_n| \geq \frac{\kappa c(n)}{2}\right) \leq 2\exp(-\frac{\kappa^2 n c(n)^2}{8}) = 2n^{-\frac{\kappa^2}{8}} \leq 2n^{-p} \leq C c(n)^{2p}.$$

Moreover, it is well known that if $N \sim \mathcal{N}(0, \sigma^2)$ then $\mathbb{E}(|N|^{2p}) = K\sigma^{2p}$ where $K = \frac{2^p}{\sqrt{\pi}} \int_0^{+\infty} x^{p-\frac{1}{2}} e^{-x} dx$. Thus we get:

$$\mathbb{E}(|V_n|^{2p}) = Kn^{-p} \leq Cc(n)^{2p}$$

This ends the proof of the lemma. $\square$

A direct application of Lemma 4.2 with $V_n = \frac{1}{\sqrt{n}} \epsilon_\lambda$ in the inequality (20) give us the properties (8) and (9).

Therefore, all conditions are satisfied to apply Theorem 3.1 with $\mathcal{E} = \xi_m$ and the hard thresholding estimator (17) so the proof of Theorem 6.1 is completed. $\square$

Remark 4.2. *The property (19) is equivalent to*

$$\sup_{u>0} u^{r-q} \omega^m(\{\lambda \in \Lambda;\ |\beta_\lambda| \leq u|v_j|\}) < \infty$$

where $\omega^m(\mathcal{S}) = \sum_{\lambda \in \mathcal{S}} |\beta_\lambda|^p \|\psi_\lambda\|_p^p$, $\forall \mathcal{S} \in \Lambda$. *See Cohen, De Vore, Kerkyacharian, and Picard (2000).*

Using this remark, it is easy to see that for all sequences $v = (v_j)_{j \in \mathbb{N}}$ and $z = (z_j)_{j \in \mathbb{N}}$ such that $0 < v_j \leq z_j$ we have the embedding $l_{(r,p,\infty)}(\xi_z) \subseteq l_{(r,p,\infty)}(\xi_v)$, that's why if we want the biggest maxiset in (19), we must choose the smallest coefficients m_j. This is exactly what we have done in the construction (15).

4.3 Example

To illustrate this method, let us consider the function $g :]0,1] \to [1, +\infty[$ defined by $g(x) = \frac{1}{x^{\frac{\sigma}{2}}}$ with $1 > \sigma > 0$. Clearly, g belongs to $\mathbb{L}^2([0,1])$ and we have:

$$d_j = 2^{\frac{j}{2}} \sqrt{\int_0^{\frac{L}{2^j}} \frac{1}{x^\sigma} dx} = \frac{L^{\frac{1-\sigma}{2}}}{\sqrt{1-\sigma}} 2^{j\frac{\sigma}{2}}.$$

Since $2^{\frac{\sigma j}{2}}$ verifies (14) by Remark 4.1, if we using Theorem 4.1 with $m_j = \frac{L^{\frac{\sigma+1}{2}}}{\sqrt{1-\sigma}} 2^{j\frac{\sigma}{2}}$ then we obtain the equivalence (18) under the calibration $\nu_0 = \frac{\sigma}{2}$.

In the second part of this paper, we always consider the statistical model (1) and the maxiset point of view but we investigate a different strategy; we expand f on a compactly supported wavelet basis warped by a certain known function T and we study the performance of the associated hard thresholding estimator under some Muckenhoupt conditions on g.

5 Muckenhoupt weights and warped bases

5.1 Muckenhoupt weights

First of all, we start by recalling the definition of Muckenhoupt weights:

Definition 5.1 (Muckenhoupt weights). *For $1 < p < \infty$ and Θ a bounded interval of $\mathbb{R}$, a nonnegative measurable function w belongs to the Muckenhoupt $\mathcal{A}_p(\Theta)$ class if and only if $|\{w(x) = 0\} \cap \{w(x) = +\infty\}| = 0$ for all $x \in \Theta$ (where $|.|$ denotes the Lebesgue measure) and there exists a positive constant C_w such that for all subinterval I of Θ:*

$$\left(\frac{1}{|I|}\int_I w(x)dx\right)\left(\frac{1}{|I|}\int_I \frac{1}{(w(x))^{\frac{1}{(p-1)}}}dx\right)^{p-1} \leq C_w < \infty \qquad (21)$$

All functions belonging to these classes are called *Muckenhoupt Weights*. The infimum of the constant C_w is called the $\mathcal{A}_p$-constant of w. For example, $w(x) = |x|^\sigma$ or $w(x) = |x|^\sigma |\ln|x||$ belongs to $\mathcal{A}_p([-1,1])$ for $-1 < \sigma \leq p-1$. Remark that the inclusion $\mathcal{A}_p(\Theta) \subset \mathcal{A}_q(\Theta)$ holds for $p < q$. The concept of Muckenhoupt weights (21) has been introduced by Muckenhoupt in (1972) and widely used afterwards in the context of Calderón-Zygmund theory. The classes of Muckenhoupt weights characterize the boundedness of some integral operators on $\mathbb{L}^p_w$ spaces like the Hardy-Littlewood maximal operator or the Hilbert transform. For complete theory, see the book of Stein (1993).

Notation: From a known function $T : [0,1] \mapsto [0,1]$ and a compactly supported wavelet basis ξ, we denote $\xi^T = \{\psi_\lambda(T(.));\ \lambda \in \Lambda\}$ the compactly supported warped wavelet basis adapted to $[0,1]$.

5.2 Warped bases

Here we investigate the geometrical properties of the basis ξ^T which truly depend on the warping factor T.

Definition 5.2 ($\mathcal{H}_p$'s property). *Let $1 < p < \infty$ and $\mathcal{B}, \mathcal{D}$ two intervals of $\mathbb{R}$. We say that a measurable function $T : \mathcal{B} \mapsto \mathcal{D}$ verifies the property $\mathcal{H}_p$ if and only if there exists a measurable function $S : \mathcal{D} \mapsto \mathcal{B}$ and a measurable positive function $w \in \mathcal{A}_p(\mathcal{D})$ such that:*

- *$S(T(x)) = x$ a.e and $T(S(x)) = x$ a.e.*
- *For all measurable positive function z, $\int_{\mathcal{B}} z(T(x))dx = \int_{\mathcal{D}} z(x)w(x)dx$*

The following theorem shows the link between Muckenhoupt weights theory and geometrical properties of compactly supported warped wavelet basis.

Theorem 5.1 (Wavelet warped basis and geometrical properties). *If T verifies the $\mathcal{H}_p$'s property then the family ξ^T is an unconditional basis (5) (see Garcia-Cuerva and Martell (1999)) and satisfies the p-Temlyakov's property (6) (see Picard Kerkyacharian (2002a)).*

This result will be apply to obtain the determinant Theorem 6.1.

6 Applications to the Gaussian white noise model

6.1 Statistical context and notations

If we suppose that $\frac{1}{g}$ belongs to $\mathbb{L}^2([0,1])$ then model (1) can be rewritten as:

$$dY_t = f(t)dt + \frac{1}{\sqrt{C_1 n}} g_1(t)dW_t$$

where we denote $g_1(t) = \sqrt{C_1}g(t)$ with $C_1 = \int_0^1 \frac{1}{g(x)}dx$. For compatibility reasons between the definition domains, we will work in the sequel with this new expression until the end of this studied.
Let $G_1 : \ [0,1] \to [0,1]$ be the function defined by:

$$G_1(x) = \int_0^x \frac{1}{g_1^2(t)}dt. \tag{22}$$

In the following subsection, we study the atomic decomposition of f in the new basis ξ^{G_1} (which is no longer orthonormal) and we defined the associated hard thresholding estimator.

6.2 Warped basis and hard thresholding estimator

Since f belongs to $\mathbb{L}^p([0,1])$, we have $f(G_1^{-1}(.))$ which belongs to $\mathbb{L}^p_{w_1}([0,1])$ where $w_1(.) = g_1^2(G_1^{-1}(.))$ with $G^{-1}(y) = \inf\{x \in [0,1], \quad G(x) > y\} \ \forall y \in [0,1]$ so:

$$\int_0^1 |f(G_1^{-1}(x)) - \sum_{\lambda\in\Lambda} \beta_\lambda \psi_\lambda(x)|^p w(x)dx = \int_0^1 |f(x) - \sum_{\lambda\in\Lambda} \beta_\lambda \psi_\lambda(G_1(x))|^p dx$$

Therefore, all function f of $\mathbb{L}^p([0,1])$ have the following atomic decomposition:

$$f(x) = \sum_{\lambda\in\Lambda} \beta_\lambda \psi_\lambda(G_1(x)), \quad \forall x \in [0,1]$$

where β_λ denotes the coefficients:

$$\beta_\lambda = \int_0^1 f(G_1^{-1}(x))\psi_\lambda(x)dx = \int_0^1 f(x)\psi_\lambda(G_1(x))g(x)dx \tag{23}$$

we deduce that the thresholding estimator (4) associated with the unknown function f of noisy observations (1) is under the form:

$$\hat{f}^*(x) = \sum_{\lambda\in\Lambda_n} \hat{\beta}^*_\lambda w^*_\lambda \psi_\lambda(G_1(x)), \quad \forall x \in [0,1] \tag{24}$$

where:

- $c(n)$ and Λ_n are defined like in Theorem 4 with $r = 2$ (we will explain the reason of this choice in subsection 6.3.2),

- $\hat{\beta}^*_\lambda$ is the following estimator:

$$\hat{\beta}^*_\lambda = \int_0^1 \psi_\lambda(G_1(t)) \frac{1}{g_1^2(t)} dY_t, \tag{25}$$

- w^*_λ is characterized by the equivalence: $w^*_\lambda = 1 \Longleftrightarrow \lambda \in \Lambda_n$ and $|\hat{\beta}^*_\lambda| \geq \kappa\sqrt{\frac{\ln(n)}{n}}$.

We are now in position to state the main theorem of this part:

Theorem 6.1. *If g and $\frac{1}{g}$ belong to $\mathbb{L}^2([0,1])$ and $g(G_1^{-1}(.))$ belongs to $\mathcal{A}_p([0,1])$ where G_1 is defined in (22) then the maxiset links to the hard thresholding estimator (24) with $\kappa \geq 2\sqrt{\frac{2p}{C_1}}$ is characterized by the equivalence:*

$$\mathbb{E}(\|\hat{f}^* - f\|_p^p) \leq C\left(\frac{\ln(n)}{n}\right)^{\frac{\alpha p}{2}} \Longleftrightarrow f \in \; B^{\frac{\alpha}{2}}_{p,\infty}(\xi^{G_1}) \cap l_{((1-\alpha)p,p,\infty}(\xi^{G_1}) \tag{26}$$

where $0 < \alpha < 1$.

The spaces $B^\gamma_{p,\infty}(\xi^{G_1})$ and $l_{((1-\alpha)p,p,\infty)}(\xi^{G_1})$ are respectively (10) and (11) with the basis ξ^{G_1}. (Here let us recall that we work with the coefficients β_λ defined in (23)).

Proof. In the next subsection, we focus on the proof of this theorem. The aim is to apply Theorem 3.1 so we need to show the geometrical properties (5), (6) of the basis ξ^{G_1}, the weight condition (7) and the statistical properties (8), (9).

6.3 Proof of theorem 6.1

6.3.1 Geometrical properties (5) and (6) of the basis ξ^{G_1}

For all measurable positive function $m : [0,1] \mapsto \Theta$ with $\Theta \in \mathbb{R}$ we have:

$$\int_0^1 m(G_1(x))dx = \int_0^1 m(x) g_1^2(G_1^{-1}(x))dx$$

Therefore, if we choose $T = G_1$, $S = G_1^{-1}$ and we suppose that $w(.) = g_1^2(G_1^{-1}(.))$ belongs to the Muckenhoupt class $\mathcal{A}_p([0,1])$ then G_1 satisfies the $\mathcal{H}_p$'s property (see Definition 5.2) and Theorem 5.1 says that the geometrical properties (5) and (6) hold for the basis ξ^{G_1}.

6.3.2 Weight condition (7)

Under the notations and the conditions of Theorem 6.1, we can apply a result obtained by Kerkyacharian and Picard (2003) about wavelet compactly supported warped wavelet basis to deduce the following inequality:

$$\omega(\mathcal{S}) = \sum_{\lambda \in \mathcal{S}} \|\psi_\lambda(G_1)\|_p^p \leq C \sum_{\lambda \in \mathcal{S}} 2^{j\frac{p}{2}} \int_{\frac{k}{2^j}}^{\frac{k+1}{2^j}} g_1^2(G_1^{-1}(x))dx \quad \forall \mathcal{S} \in \Lambda.$$

Since $\sum_{k=0}^{2^j-1}\int_{\frac{k}{2^j}}^{\frac{k+1}{2^j}} g_1^2(G_1^{-1}(x))dx = \int_0^1 g_1^2(G_1^{-1}(x))dx = 1$ and $2^{\frac{j_1(n)}{2}}c(n)$ is bounded by definition of $j_1(n)$ for $r=2$, there exists a positive constant C such that

$$\sup_{n\geq 1}[\omega(\Lambda_n)c(n)^p] \leq C.$$

So the weight condition (7) holds.

6.3.3 Statistical properties (8) and (9)

First of all, we remark that:

$$\begin{aligned}
\hat{\beta}_\lambda^* &= \int_0^1 \psi_\lambda(G_1(t))\frac{1}{g_1^2(t)}dY_t \\
&= \int_0^1 f(t)\psi_\lambda(G_1(t))\frac{1}{g_1^2(t)}dt + \frac{1}{\sqrt{C_1 n}}\int_0^1 \psi_\lambda(G_1(t))\frac{1}{g_1(t)}dB_t
\end{aligned}$$

from (23), it comes:

$$\hat{\beta}_\lambda^* - \beta_\lambda = \frac{1}{\sqrt{C_1 n}}\int_0^1 \psi_\lambda(G_1(t))\frac{1}{g_1(t)}dB_t$$

and clearly:

$$\frac{1}{\sqrt{C_1 n}}\int_0^1 \psi_\lambda(G_1(t))\frac{1}{g_1(t)}dB_t \sim \mathcal{N}(0,\sigma_n^2)$$

where, by the change of variables $t = G^{-1}(x)$:

$$\begin{aligned}
\sigma_n^2 &= \mathbb{E}\left(\frac{1}{\sqrt{C_1 n}}\int_0^1 \psi_\lambda(G_1(t))\frac{1}{g_1(t)}dB_t\right)^2 = \frac{1}{C_1 n}\int_0^1 \psi_\lambda^2(G_1(t))\frac{1}{g_1^2(t)}dt \\
&= \frac{1}{C_1 n}\int_0^1 \psi_\lambda^2(x)dx = \frac{1}{C_1 n}.
\end{aligned}$$

Using Lemma 4.2, we immediately deduce that $\hat{\beta}_\lambda^*$ verifies the properties (8) and (9).

All conditions are satisfied to apply Theorem 3.1 with the hard thresholding estimator (24) so the proof of Theorem 6.1 is completed. □

7 Statistical applications and example

Along this section, we study the rate of convergence of the hard thresholding estimator (24) over weighted Besov spaces which truly link to the model (1) and we illustrate our result by an example.

7.1 Weighted Besov spaces

We start to define the weighted Besov spaces which are simply the natural expansion of well known Besov spaces in a warped basis ξ^G. The main particularity of this spaces is there are directly linked to the warping factor G.

Definition 7.1 (Weighted Besov spaces). *We define for any measurable function f:*

$$\Delta_{G,h}(f)(x) = f(G^{-1}(G(x) + h)) - f(x).$$

Recursively $\Delta^2_{G,h}(f)(x) = \Delta_{G,h}(\Delta_{G,h}(f))(x)$ and identically, for $N \in \mathbb{N}^*$, $\Delta^N_{G,h}(f)(x)$ again and again, and we consider:

$$\rho^N(t, f, G, p) = \sup_{|h|\le t} \left(\int |\Delta^N_{G,h}(f)(u)|^p du \right)^{\frac{1}{p}}.$$

Notice that ρ^N is defined with the standard uniform weight, the spacial inhomogeneity now lies in the definition of $\Delta_{G,h}$. Let us consider the following spaces calling weighted Besov spaces:

$$\mathcal{V}(G, s, \vartheta, r) = \{f \in \mathbb{L}^p([0, 1])| \quad \left(\int_0^1 \left(\frac{\rho^N(t, f, G, \vartheta)}{t^s} \right)^q \frac{1}{t} dt \right)^{\frac{1}{q}} < \infty\}$$

7.2 Upper bound result

In order to investigate an upper bound result over some weighted Besov spaces, we now state a result proved by Picard and Kerkyacharian (2003):

Proposition 7.1. *For $p > 1$, if $\vartheta \ge p$, $s \ge \frac{1}{2}$, $r \ge \frac{p}{1+2s}$, $\alpha = \frac{2s}{1+2s}$ and G satisfies the $\mathcal{H}_p$'s property then we have the following embedding:*

$$\mathcal{V}(G, s, \vartheta, r) \subset B^{\frac{\alpha}{2}}_{p,\infty}(\xi^G) \cap l_{((1-\alpha)p,p,\infty)}(\xi^G)$$

If we choose $G = G_1$ defined in (22) then the functional space $B^{\frac{\alpha}{2}}_{p,\infty}(\xi^G) \cap l_{((1-\alpha)p,p,\infty)}(\xi^G)$ is exactly the maxiset found in (26) so an immediate consequence of Proposition 7.1 is the following one:

Proposition 7.2. *Under the same hypothesis than Theorem 6.1, for $p > 1$, $\vartheta \ge p$, $s \ge \frac{1}{2}$, $r \ge \frac{p}{1+2s}$, $\alpha = \frac{2s}{1+2s}$ we have the following upper bound:*

$$f \in \mathcal{V}(G_1, s, \vartheta, r) \Rightarrow \mathbb{E}(\|\hat{f}^* - f\|_p^p) \le C \left(\frac{\ln(n)}{n} \right)^{\frac{sp}{1+2s}}.$$

Remark 7.1. *For $g(t) \equiv g$ and for the same hypothesis on s, p and q than the Proposition 7.2, we have the minimax rate:*

$$\inf_{\hat{f}} \sup_{\mathcal{V}(g.1,s,\vartheta,r)} \mathbb{E}(\|\hat{f}^* - f\|_p^p) \sim n^{-\frac{sp}{1+2s}}.$$

So if we choose the hard thresholding estimator (17) with $G_1(x) = g.x$, we rediscover the minimax upper bound up to a logarithmic factor.

See for instance the book of Härdle, Kerkyacharian, Picard and Tsybakov (1998).

7.3 Example using warped basis

To illustrate our result, we consider our Gaussian noise model (1) with $g(x) = \frac{1}{x^{\frac{\sigma}{2}}}$ for $x \in]0,1]$ and $1 > \sigma > 0$. It is obvious that g and $\frac{1}{g}$ belongs to $\mathbb{L}^2([0,1])$. Moreover:

$$G(y) = \int_0^y \frac{1}{g^2(x)} dx = \frac{1}{\sigma+1} y^{\sigma+1}$$

so $G_1(y) = y^{\sigma+1}$ and

$$G_1^{-1}(y) = y^{\frac{1}{\sigma+1}}.$$

This implies that $\forall y \in]0,1]$ we have:

$$w_1(y) = g_1^2(G_1^{-1}(y)) = C_1 y^{\frac{-\sigma}{\sigma+1}} = C_1 y^{\beta}$$

where $\beta = \frac{-\sigma}{\sigma+1}$.

Now, we remark that $\beta \in]-1, 0[$ for $1 > \sigma > 0$ so $w_1 \in \mathcal{A}_p([0,1])$.

Applying Theorem 6.1 with $\hat{f}^*$ the hard thresholding estimator defined in (24) with $G_1(x) = x^{\sigma+1}$, $1 > \sigma > 0$, $x \in [0,1]$,

$$\hat{\beta}_\lambda^* = (\sigma+1) \int_0^1 \psi_\lambda(t^{\sigma+1}) t^{\sigma} dY_t,$$

and $\kappa \geq 2\sqrt{(\sigma+1)2p}$ we obtain the equivalence:

$$\mathbb{E}\|\hat{f}^* - f\|_p^p \leq C \left(\frac{ln(n)}{n} \right)^{\frac{\alpha p}{2}} \Longleftrightarrow f \in B_{p,\infty}^{\frac{\alpha}{2}}(\xi^{G_1}) \cap l_{((1-\alpha)p,p,\infty)}(\xi^{G_1})$$

Moreover, under the same hypothesis that Proposition 7.1, we have the following upper bound:

$$f \in \mathcal{V}(G_1, s, \vartheta, r) \Rightarrow \mathbb{E}(\|\hat{f}^* - f\|_p^p) \leq C \left(\frac{\ln(n)}{n} \right)^{\frac{sp}{1+2s}}.$$

References

Autin, F. (2003). *Maxiset for density estimation.* Preprint du Laboratoire de probabilités et modèles aléatoires. Paris VI et Paris VII.

Brown, L.D. and Low, M.G. (1996). *Asymptotic equivalence of nonparametric regression and white noise.* Ann. Statist. 24, 2384-2398.

Cohen, A., De Vore, R., Kerkyacharian, G., and Picard, D. (2000). *Maximal q-paces with given rate of convergence for thresholding algorithms. Appl. Comput. Harmon. Anal.,* **11**, 167-191.

Cohen, A., Daubechies, I. and Vial, P. (1993). *Wavelets on the interval and fast wavelet transforms; Applied and Comp. Harmonic Analysis.* **1**(1), 54-81.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

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

Donoho, D.L., Johnstone, I.M. (1995). *Adaptating to unknown smoothness via wavelet shrinkage.* J . Am Stat. Assoc, **90**(432), 1200-1224.

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

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994). *Wavelet Shrinkage by optimal recovery.* Lucien LeCam Festschrift D. Pollard and G.Yang (eds), 183-215, Springer Verlag.

Garcia-Cuerva, J. and Martell, B. (1999). *Wavelet characterisation of weighted spaces.* Journal of Geometric Analysis.

Gobulev, G.K. (1991). *Local asymptotic normality in problem of nonparametric estimation of functions and lower bound for quadratic risk.* Theory Probab. Appliq. **36**, 1, 152-157.

Grama, I. and Nussbaum, M. (1996). *Asymptotics equivalence for non parametric generalized models.* Preprint No.289, Weierstrass Institute.

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

Ibragimov, I.A. and Khaminskii, R.Z. (1977). *On the estimation of an infinite dimensional parameter in Gaussian white noise.* Soviet Math. Dokl. 236, No.5, 1053-1055.

Ibragimov, I.A. and Khaminskii, R.Z. (1981). *Statistical estimation: Asymptotic Theory.* Springer, New York.

Johnstone, I.M. (1998). *Function estimation : White noise, sparsity and wavelets.* Lectures notes 130p.

Johnstone, I.M. (2000). *Wavelet and the theory of non parametric function estimation.* Phil.Trans.R. Soc. Lond. **A**.

Johnstone, I.M., Kerkyacharian, G., Picard, D., Raimondo, M. (2004). *Wavelet deconvolution in a periodic setting* to appear.

Kerkyacharian, G., Picard, D. (2000). *Thresholding algorithms, maxisets and well concentred bases*, with discussion, Test, vol9, No2,P 283-345.

Kerkyacharian, G., Picard, D. (2002a). *Non linear Approximation and Muckenhoupt weights.* Technical report.

Kerkyacharian, G., Picard, D. (2002b). *Minimax or Maxisets ?* Bernouilli, 8 N2, 219-253.

Kerkyacharian, G., Picard, D. (2003). *Regression in random design and warped wavelets.* Bernoulli, 8 N2, 219-253.

Le Cam, L. and Yang G.L. (1990). *Asymptotics in Statistics: Some basic concepts.* Springer. New York.

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

Muckenhoupt, B. (1972). *Weighted norm ineqalities for the hardy maximal function.*Trans. Amer. Math. Hencec. 165, 207-226.

Nussbaum, M. (1996). *Asymptotic equivalence of density estimation and white noise.* Annals od Statistics, **24**, 2399-2430.

Rivoirard, V. (2002). *Estimation baysienne non paramétrique.* Thèse de l'université Paris VII.

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

Tsybakov, A.B. (2004) *Introduction à l'estimation non-parametrique.* Springer.