Honest and adaptive confidence sets in L_p

Alexandra Carpentier

Statistical Laboratory
Center for Mathematical Sciences
Wilberforce Road
CB3 0WB Cambridge
United Kingdom
e-mail: a.carpentier@statslab.cam.ac.uk

Abstract: We consider the problem of constructing honest and adaptive confidence sets in L_p-loss (with $p \geq 1$ and $p < \infty$) over sets of Sobolev-type classes, in the setting of non-parametric Gaussian regression. The objective is to adapt the diameter of the confidence sets with respect to the smoothness degree of the underlying function, while ensuring that the true function lies in the confidence interval with high probability. When $p \geq 2$, we identify two main regimes, (i) one where adaptation is possible without any restrictions on the model, and (ii) one where critical regions have to be removed. We also prove by a matching lower bound that the size of the regions that we remove can not be chosen significantly smaller. These regimes are shown to depend in a qualitative way on the index p, and a continuous transition from $p = 2$ to $p = \infty$ is exhibited.

1. Introduction

We consider in this paper the problem of building *honest and adaptive* confidence sets around functions that belong to a L_p-Sobolev-type space in the non-parametric Gaussian regression setting.

This question was already investigated in L_∞ and L_2, see for instance the papers (Hoffmann and Lepski, 2002; Juditsky and Lambert-Lacroix, 2003; Baraud, 2004; Robins and Van Der Vaart, 2006; Cai and Low, 2006; Giné and Nickl, 2010; Hoffmann and Nickl, 2011; Bull and Nickl, 2013). In particular, the recent papers (Hoffmann and Nickl, 2011; Bull and Nickl, 2013) develop for respectively L_∞ and L_2 a minimax-optimal setting in which the construction of *honest and adaptive* confidence sets is possible.

In the present paper, we extend these results to general values of $p \in [1, \infty[$. We develop minimax-optimal settings in which the construction of honest and adaptive confidence sets is possible. Since the case $1 \leq p \leq 2$ is essentially equivalent to the case $p = 2$, we focus on this case $p \geq 2$ (and $p < \infty$). We prove that there is a continuous transition between the case $p = 2$ described in (Bull

and Nickl, 2013) and the case $p = \infty$ described in (Hoffmann and Nickl, 2011). While the main idea of this paper, i.e. to investigate the relationship between the problem of constructing adaptive and honest confidence sets and a certain infinite-dimensional composite testing problem, is similar to the one in (Hoffmann and Nickl, 2011; Bull and Nickl, 2013), the techniques required for the solution of this testing problem are significantly more involved. It appears that the approaches of (Hoffmann and Nickl, 2011; Bull and Nickl, 2013) (in particular the analysis of the so-called infimum test) could not have been generalised in a straightforward way to the settings $p \in]2, \infty[$. Also, our results imply that the curious dependence on whether p is an even integer or not, that appears in a related minimax estimation problem studied in (Lepski et al., 1999), is not relevant in the setting of confidence sets.

This paper is organised as follows. In Section 2, we present the general setting. In Section 3, we provide our results, which are (i) the existence of adaptive estimators in $\|.\|_p$ norm, and (ii) the existence of honest and adaptive confidence sets in $\|.\|_p$ norm on some maximal models. The other sections of the paper present detailed proofs of these results. The Supplementary Material contains the proof for the existence of adaptive estimators, and also some classical preliminary results.

2. Setting

Let $p \geq 1$ (and $p < \infty$). Let $\lfloor p \rfloor$ be the largest *even* integer smaller than p (this notation is not usual but we will need it in the course of the proofs). Let $s > t \geq 1/2$ be two degrees of smoothness.

Denote by $L_p([0,1]) = L_p$ the space of functions defined on $[0,1]$ such that $\|f\|_p^p = \int_0^1 |f(x)|^p dx < +\infty$, where $\|.\|_p$ is the usual L_p-norm.

For any functions $(f, g) \in L_p \times L_q$ where $1/p + 1/q = 1$, we consider the bilinear form $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.

2.1. Wavelet basis

We consider an orthonormal wavelet basis

$$\{\phi_k, k \in Z_0, \psi_{l,k}, l > 0, k \in Z_l\},$$

such that for any integer $l \geq 0$, $Z_l \subset \mathbb{Z}$ and $|Z_l| \leq c2^l$ (where c is a numerical constant). Also, we impose the usual conditions that for any $l > 0, k \in Z_l$, $\int_0^1 \psi_{l,k}(x)dx = 0$, and that $\psi_{l,k}(x) = 2^{l/2}\psi_k(2^l x)$. Such a basis exists (for instance the Cohen-Daubechies-Vial basis satisfies all these conditions, see Cohen et al. (1993)).

We assume that the wavelet basis we consider satisfies the following assumption, which is quite standard.

Assumption 1. *We assume that there is a universal constant C_p such that we have for all $x \in [0, 1]$ and any integer $J \geq 0$*

$$\sqrt{\sum_{k \in Z_0} \phi_k^2(x) + \sum_{0 < l \leq J, k \in Z_l} \psi_{l,k}^2(x)} \leq C_p 2^{J/2}.$$

It holds for any wavelet basis such that the mother wavelets ψ_k are uniformly bounded and have sufficiently "disjoint support", i.e. are well spread on the domain, see (Härdle et al., 1998). In other words, for uniformly bounded mother wavelets ψ_k defined on a compact, this property is necessary to ensure the conservation of the norm of signals in L_2. This assumption is in particular satisfied for Cohen-Daubechies-Vial wavelets with $S > 0$ first null moments (where the constant C_p in the definition depends on S), see Cohen et al. (1993).

For any function $f \in L_p$, we consider the sequence of coefficients $a(f) = a$ and the complementary sequence of coefficients $a'(f) = a'$ as

$$a_{l,k} = \int_0^1 \psi_{l,k}(x) f(x) dx = \langle \psi_{l,k}, f \rangle \quad and \quad a_k' = \langle \phi_k, f \rangle.$$

Consider the functions $f \in L_p$ that have the representation

$$f = \sum_{k \in Z_0} \phi_k \langle \phi_k, f \rangle + \sum_{l > 0} \sum_{k \in Z_l} \psi_{l,k} \langle \psi_{l,k}, f \rangle = \sum_{k \in Z_0} a_k' \phi_k + \sum_{l > 0} \sum_{k \in Z_0} a_{l,k} \psi_{l,k}. \quad (2.1)$$

We moreover write for any $J \geq 0$

$$\Pi_{V_J}(f) = \sum_{k \in Z_0} a_k' \phi_k + \sum_{0 < l \leq J} \sum_{k \in Z_l} a_{l,k} \psi_{l,k},$$

the projection of f onto $\mathrm{span}(\phi_k, k \in Z_0, \psi_{l,k}, 0 < l \leq J, k \in Z_l)$. We also write

$$\Pi_{W_J}(f) = \sum_{k \in Z_J} a_{J,k} \psi_{J,k}, \quad and \quad \Pi_{W_0}(f) = \sum_{k \in Z_0} a_k' \phi_k.$$

the projection of f onto $\mathrm{span}(\psi_{J,k}, k \in Z_J)$ or $\mathrm{span}(\phi_k, k \in Z_0)$.

2.2. Besov spaces

We consider for any $h \geq 1$, $p \geq 1$ (and $p < \infty$) and $r \geq 0$ the Besov norms

$$\|f\|_{r,p,h} = \left(|(\langle f, \phi_k \rangle)_k|_{l_p}^h + \sum_{l > 0} 2^{lh(r + 1/2 - 1/p)} |(\langle f, \psi_{l,k} \rangle)_k|_{l_p}^h \right)^{1/h},$$

where $|u|_{l_p} = (\sum_i u_i^p)^{1/p}$ (we extend this definition as $|u|_{l_\infty} = \sup_i |u_i|$ for $p = \infty$) is the sequential l_p norm. We extend this definition for $h = \infty$ as

$$\|f\|_{r,p,\infty} = \max \left(|(\langle f, \phi_k \rangle)_k|_{l_p}, \sup_{l > 0} 2^{l(r + 1/2 - 1/p)} |(\langle f, \psi_{l,k} \rangle)_k|_{l_p} \right).$$

The Besov-type spaces are defined for any $h \in [1, \infty]$, $p \geq 1$ (and $p < \infty$) and $r \geq 0$ as

$$B_{r,p,h} = \{f \in L^p : \|f\|_{r,p,h} < +\infty\}.$$

We write for a given $B > 0$ the $B_{r,p,\infty}$ Besov ball of smoothness r and radius B as

$$\Sigma(r, B) \equiv \Sigma(r, p, B) = \{f \in B_{r,p,\infty} : \|f\|_{r,p,\infty} \leq B\}.$$

For regular enough wavelets (e.g. Cohen-Daubechies-Vial wavelets with S first null moments), the defined Besov spaces correspond to the functional Besov spaces (Sobolev-type spaces) up to some smoothness $S \geq s$, see Meyer (1992); Härdle et al. (1998). We assume that our basis satisfies this property with $s \leq S$ where s is the largest smoothness that we wish to consider in our testing problem.

The spaces $B_{r,p,\infty}$ are slightly larger than the usual L_p-Sobolev spaces, see Bergh and Löfström (1976); Besov et al. (1978). They are however the natural objects to consider for the construction of honest and adaptive confidence sets, since they are the largest Besov spaces where the rate $n^{-r/(2r+1)}$ is minimax-optimal for functional estimation (see Section 3 for references and a precise statement of this assertion).

We will consider in this paper functions f that have a smoothness larger than $1/2$, which is a common assumption for the problem of the construction of adaptive and honest confidence sets, see (Bull and Nickl, 2013). This assumption is technical and wether or not the results in this paper could be generalised to rougher functions is an open question.

2.3. Observation scheme

The data is a realisation of a Gaussian process defined for any $x \in [0, 1]$ and for a given n as

$$dY^{(n)}(x) = f(x)dx + \frac{dB_x}{\sqrt{n}},$$

where $(B_x)_{x \in [0,1]}$ is a standard Brownian motion, and $f \in L_2$ is the function of interest.

Let us write for any $l > 0$ and $k \in Z_l$ the associated wavelet coefficients as

$$\hat{a}_{l,k} = \langle \psi_{l,k}, dY^{(n)} \rangle = \int_0^1 f(x)\psi_{l,k}(x)dx + \frac{1}{\sqrt{n}} \int_0^1 \psi_{l,k}(x)dB_x,$$

$$\text{and } a_{l,k} = \langle \psi_{l,k}, f \rangle,$$

and for any $k \in Z_0$ the complementary wavelet coefficients as

$$\hat{a}_k = \langle \phi_k, dY^{(n)} \rangle = \int_0^1 f(x)\phi_k(x)dx + \frac{1}{\sqrt{n}} \int_0^1 \phi_k(x)dB_x,$$

$$\text{and } a'_k = \langle \phi_k, f \rangle.$$

We consider the wavelet estimate of f:

$$\hat{f}_n = \sum_{k \in Z_0} \hat{a}'_k \phi_k + \sum_{l>0} \sum_{k \in Z_l} \hat{a}_{l,k} \psi_{l,k}.$$

Projected estimates up to frequency $J \geq 0$

$$\hat{f}_n(J) := \Pi_{V_J} \hat{f}_n,$$

and also the estimate of f at level J

$$\Pi_{W_J}(f) = \Pi_{W_J} \hat{f}_n.$$

are usually considered.

In the sequel, we write $\Pr_f$ (respectively $\mathbb{E}_f$) the probability (respectively expectation) under the law of $Y^{(n)}$ when the function underlying the data is f. When no confusion is likely to arise, we write simply $\Pr$ (respectively $\mathbb{E}$).

3. Main results

In this Section, we will consider the Cohen-Daubechies-Vial wavelet basis with $S > s$ first null moments (where s is the largest smoothness according to which we wish to adapt). As a matter of fact, any wavelet basis satisfying the conditions defined in Section 2 (in particular Assumption 1 and the S null moments condition) will work.

3.1. Adaptive estimation

We provide, when $p \geq 2$ (and $p < \infty$), a result for adaptive estimation, i.e. that adaptive estimators for functions in $B_{r,p,\infty}$ (adaptive to the smoothness r) exist. The technique that we use is closely related to what is proposed in the papers (Lepski, 1992; Giné and Nickl, 2009; Bull and Nickl, 2013). We do not need any assumptions on f except that it is in $B_{r,p,\infty}$ for a given smoothness r.

Theorem 3.1. *Assume that $p \geq 2$ (and $p < \infty$). There exists an adaptive estimator $\tilde{f}_n(dY^{(n)}, p)$ such that there are two constants u_p and N_p that depend only on p such that for every $B \geq 0$ and every $r > 0$, we have for any $n \geq N_p$ that*

$$\sup_{f \in \Sigma(r,B)} \mathbf{E}_f \|\tilde{f}_n - f\|_p \leq u_p \big(B^{1/(2r+1)} + 1\big) n^{-r/(2r+1)}$$

$$\equiv U_p(B) n^{-r/(2r+1)}.$$

The proof and also construction of this estimate are in Section 10 (it is a Corollary of Theorem 10.1 with $c_a = 0$).

This result is minimax-optimal in r and in B whenever $B \geq 1$ and $r > 1/p$ (see (Härdle et al., 1998)).

3.2. Previous results regarding honest and adaptive confidence sets

For $p \geq 1$ (and $p < \infty$), a confidence set is a random subset C_n of L_p that depends on the data and perhaps on some additional knowledge that is available. We define its diameter in $\|.\|_p$ norm as

$$|C_n| = \inf_{\tau \geq 0} \left\{ \tau : \exists g \in L_p : C_n \subset \{h : \|h - g\|_p \leq \tau\} \right\}. \tag{3.1}$$

We define honest and adaptive confidence sets as follows.

Definition 3.1 ((L_p, α)-honest and adaptive confidence set given $\mathcal{P}_n$, I and B)**.**
Let $0 < t < s$, $B > 0$, $\alpha > 0$. Let I be a subset of $[t, s]$. Let $\mathcal{P}_n$ be a non-empty subset included in $\Sigma(t, B)$. Let $C_n(Y^{(n)}, s, t, p, B, \alpha)$ be a random subset of L_p. C_n is called (L_p, α)-honest and adaptive given $\mathcal{P}_n$, I and B if there exists a constant $L := L(s, t, p, B, \alpha)$ such that for any $n \geq 0$

$$\sup_{f \in \Sigma(r,B) \bigcap \mathcal{P}_n} \mathrm{Pr}_f \left(|C_n| \geq L n^{-\frac{r}{2r+1}} \right) \leq \alpha \ \ \forall r \in I,$$

$$\text{and} \quad \inf_{f \in \mathcal{P}_n} \mathrm{Pr}_f \left(f \in C_n \right) \geq 1 - \alpha.$$

In this definition, the set I is the set of Besov indexes to which we wish to adapt. We will in this paper consider the case $I = \{s, t\}$, since it is not too involved to pass from this case to the case $[t, s]$ (see e.g. (Hoffmann and Nickl, 2011; Bull and Nickl, 2013)). The model $\mathcal{P}_n$ is the set of functions on which we want to build honest and adaptive confidence sets. Ideally, we would like this set to be $\Sigma(t, B)$, but it will be seen not to always be possible to consider the whole set: some functions of $\Sigma(t, B)$ that are very close to $\Sigma(s, B)$ but not in this set can be a source of problems for the existence of honest and adaptive confidence sets. In some cases, as we will explain later in this section, a subset of $\Sigma(t, B)$ has to be removed.

To the best of our knowledge, the question of building honest and adaptive confidence sets in L_p for $p \geq 1$ and $p < \infty$ has only been addressed in the case $p = 2$. The most recent paper on this topic is (Bull and Nickl, 2013). Also, it is noticeable that the case $p = \infty$ has been treated in the paper (Hoffmann and Nickl, 2011), but we are not going to present the results in this case here, since they are different in essence from the case $p \in]1, \infty[$.

The results in the paper (Bull and Nickl, 2013), although proved in the density estimation setting, apply as well in the Gaussian process setting (it is actually more technical to derive them in the setting of density estimation). One only needs to change slightly the test statistics used. In the case $2t \geq s$, one should use instead of the statistic in Equation 35 in (Bull and Nickl, 2013)

$$U_n(\hat{f}_n) = \|\Pi_{V_j} \hat{f}_n(1) - \Pi_{V_j} \hat{f}_n(2)\|_2^2 - \sum_{l \leq j} \frac{|Z_l|}{n},$$

where $\hat{f}_n(1), \hat{f}_n(2)$ are estimates of f as in the present paper but computed respectively on the first and second half of the paper, and j is such that $2^j \approx$

$n^{\frac{1}{2t+1/2}}$. When $2t < s$, one should use instead of the statistic in Equation 17 in (Bull and Nickl, 2013)

$$T_n(g) = \|\Pi_{V_j}\hat{f}_n - g\|_2^2 - \sum_{l \leq j} \frac{|Z_l|}{n},$$

where $\hat{f}_n$ is an estimate of f as in my paper. These statistics have similar properties than the ones in the paper (Bull and Nickl, 2013), and similar results hold for confidence sets in this setting.

The authors of the paper (Bull and Nickl, 2013) first prove the following result when $s \leq 2t$, i.e. that adaptive and honest confidence sets exist in this case on $\Sigma(t, B)$ itself.

Theorem 3.2 (Bull and Nickl (2013)). *Set $p = 2$. Let $1/2 \leq t < s$. Assume also that $s/2 \leq t$. Let $\mathcal{P}_n = \Sigma(t, B)$. Let $B > 0$ and $\alpha > 0$. There exists a (L_2, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*

In order to build such a confidence set, the authors measure the L_2 distance between the data and an estimate of the function f that we can think of as the orthogonal projection of $\hat{f}_n$ on $\Sigma(s, B)$. They then define (intuitively speaking) the confidence set as being the set of functions that are at a distance that is smaller than this estimated distance from an adaptive estimate of f (as defined in Theorem 3.1).

It becomes however more involved when $s > 2t$ (the authors actually need some additional formalism). For $G \subset L_p$, set $\|f - G\|_p = \inf_{g \in G} \|f - g\|_p$. We define for $\rho_n \geq 0$ and $B > 0$ the sets

$$\tilde{\Sigma}(t, B, \rho_n) = \tilde{\Sigma}(t, s, p, B, \rho_n) = \left\{ f \in \Sigma(t, B) : \|f - \Sigma(s, B)\|_p \geq \rho_n \right\}. \quad (3.2)$$

These sets are separated away from $\Sigma(s, B)$ whenever $\rho_n > 0$. They correspond to $\Sigma(t, B) \setminus \Sigma(s, B)$ where we have removed some critical functions very close to functions in $\Sigma(s, B)$ in $\|.\|_p$ norm. We now remind a simplified and slightly weaker version of the main Theorem in the paper (Bull and Nickl, 2013)'s when $s > 2t$ (in the paper (Bull and Nickl, 2013), the authors actually prove a stronger result, which includes adaptation also to the radius B of the Besov ball).

Theorem 3.3 (Bull and Nickl (2013)). *Set $p = 2$. Let $B > 0$ and $\alpha > 0$, and assume that $s/2 > t \geq 1/2$. Let $\mathcal{P}_n = \tilde{\Sigma}(t, B, \rho_n) \bigcup \Sigma(s, B)$ for some ρ_n.*

- *Let $\rho_n = Cn^{-t/(2t+1/2)}$, where $C := C(p, B, \alpha)$ is large enough. Then there exists a (L_2, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*
- *Let $\rho_n = \upsilon n^{-t/(2t+1/2)}$, where $\upsilon := \upsilon(p, B, \alpha)$ is small enough. Then there exists no (L_2, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*

In order to prove this theorem, the authors consider the following testing problem:

$$H_0 : f \in \Sigma(s, B) \qquad v.s. \qquad H_1 : f \in \tilde{\Sigma}(t, B, \rho_n). \quad (3.3)$$

As explained in Hoffmann and Lepski (2002); Juditsky and Lambert-Lacroix (2003); Hoffmann and Nickl (2011); Bull and Nickl (2013), this problem is very related to the problem of building honest and adaptive confidence sets. In the process of constructing honest and adaptive confidence sets, the authors of the paper (Bull and Nickl, 2013) construct a test for the testing problem (3.3) that is uniformly consistent over $\tilde{\Sigma}(t, B, \rho_n) \bigcup \Sigma(s, B)$ for

$$\rho_n = C \max(n^{-t/(2t+1/2)}, n^{-s/(2s+1)}),$$

where C large enough (ρ_n is of order $n^{-t/(2t+1/2)}$ whenever $s \geq 2t$). More precisely, they prove that for any $n > 0$ and $\alpha > 0$, there exists a test Ψ_n such that

$$\sup_{f \in \Sigma(s,B)} \mathbb{E}_f \Psi_n + \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{E}_f[1 - \Psi_n] \leq \alpha. \tag{3.4}$$

They then use this result to prove Theorem 3.3.

An indirect consequence of Theorem 3.3 is that the case $1 \leq p \leq 2$ is not relevant, since the following negative result applies.

Proposition 1. *Let* $1 \leq p \leq 2$. *Let* $B > 0$ *and* $\alpha > 0$, *and assume that* $s > 2t \geq 1/p$. *Let* $\mathcal{P}_n = \tilde{\Sigma}(t, B, \rho_n) \bigcup \Sigma(s, B) \subset L_2$ *for some* ρ_n. *Let* $\rho_n = \upsilon n^{-t/(2t+1/2)}$, *where* $\upsilon := \upsilon(p, B, \alpha)$ *is small enough. There exists no* (L_p, α)-*honest and adaptive confidence set given* $\mathcal{P}_n$, $\{s, t\}$ *and* B.

The proof is the same in (Bull and Nickl, 2013) for the L_2 case, since the L_2 norms of the functions constructed to prove the impossibility result in L_2 are equal to the L_p norm of these same functions. Since the L_2 norm dominates all L_p norms for $p \leq 2$, Proposition 1 implies in particular that the the confidence sets, and the regions one has to remove in order to be able to build honest and adaptive confidence sets, are *both* larger in L_p than in L_2, noting that $f \in L^2$ is a natural assumption in Gaussian white noise. It is thus preferable to use for L_p the same confidence sets as for L_2 when $p \leq 2$. The case $p \geq 2$ (and $p < \infty$) is more interesting and is the setting we shall consider more in depth.

Concerning lower bounds for general $p \geq 1$, the papers (Ingster, 1987, 1993; Ingster and Suslina, 2002) state that for a simpler but related testing problem

$$H_0 : f = 0 \quad vs. \quad H_1 : f \in \{f \in \Sigma(t, B) : \|f - 0\|_p \geq \rho_n\},$$

the minimax rate of separation is

$$\rho_n \geq D n^{-\frac{t}{2t+1-1/p}}, \tag{3.5}$$

for some $D > 0$, which coincides with the size of the region one has to remove in Theorem 3.3 for $p = 2$. One can wonder if it is still the minimax rate in the composite problem. The minimax order of the separation ρ_n in Theorem 3.3 is related to the results of paper (Lepski et al., 1999), where the authors prove that whenever p is an *even* integer, it is possible to construct an estimate of $\|f\|_p$ whose error is of order $n^{-\frac{t}{2t+1-1/p}}$. In L_2, as proved in the paper (Bull and Nickl, 2013), empirical process theory combined with this idea imply that it is

possible to estimate $\|f - \Sigma(s, B)\|_2$ with precision $\max(n^{-\frac{t}{2t+1/2}}, n^{-\frac{s}{2s+1}})$. As in the paper (Bull and Nickl, 2013), this estimate can be used to test whether f is in $\Sigma(s, B)$ or not, and then to construct honest and adaptive confidence sets. However, Lepski et al. (1999) also provide negative results whenever p is not an even integer and prove that in this case, it is not possible to estimate $\|f\|_p$ at a better rate than $(n \log(n))^{-\frac{t}{2t+1}}$. One might worry that this poor rate has repercussions on the construction of confidence sets. By developing a new technique with respect to what was achieved in the paper (Bull and Nickl, 2013), we will prove that it is not the case.

3.3. Honest and adaptive confidence sets

We first state our results in terms of the testing problem (3.3), and then apply these results to the construction of honest and adaptive confidence sets. As a matter of fact, the results we provide for the existence of honest and adaptive confidence sets are a direct consequence of the solution to the testing problem (3.3).

3.3.1. Testing bounds on the related testing problem

We state the following theorem for the existence of a uniformly consistent test (in the sense of Equation (3.4)) for the testing problem (3.3).

Theorem 3.4. *Let $B > 0$, $\alpha > 0$ and $n > 0$.*

- *Let $\rho_n = C(B+1) \max(n^{-t/(2t+1-1/p)}, n^{-s/(2s+1)})$, where $C := C(p, \alpha)$ is large enough. Then there exists a (L_p, α)-uniformly consistent test (in the sense of Equation (3.4)) for the testing problem (3.3).*
- *Let $\rho_n = \upsilon n^{-t/(2t+1-1/p)}$, where $\upsilon := \upsilon(p, B, \alpha)$ is small enough. Then there exists no (L_p, α)-uniformly consistent test (in the sense of Equation (3.4)) for the testing problem (3.3).*

Theorem 3.4 is proven in Section 4 (upper bound), and in Section 7 (lower bound).

There is a gap between the upper and lower bound for ρ_n, which is not relevant, as we will see in next paragraph, for the existence of honest and adaptive confidence sets, but which matters for the testing problem. As a matter of fact, this gap does not exist whenever p is an even integer and it is possible to prove that a consistent test with

$$\rho_n \geq D n^{-\frac{t}{2t+1-1/p}},$$

exists (see (Carpentier, 2013) for $p = 2$, and the results can be extended for any even integer p). The case when p is not an even integer is more involved and we conjecture that it is *not* possible to build a consistent test whenever

$$\rho_n \leq D' \max(n^{-\frac{s}{2s+1}}, n^{-\frac{t}{2t+1-1/p}}),$$

for D' small enough. This remains an open question.

3.3.2. *Consequences for confidence sets*

A first consequence of Theorem 3.4 is the following theorem when $s(1-1/p) \leq t$ (analog to $s \leq 2t$ when $p = 2$ in (Bull and Nickl, 2013), Theorem 3.2).

Theorem 3.5. *Let $1/2 \leq t < s$. Assume also that $s(1 - 1/p) \leq t$. Let $\mathcal{P}_n = \Sigma(t, B)$. Let $B > 0$ and $\alpha > 0$. There exists a (L_p, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*

The proof that we provide in this paper is different from the proof in paper (Bull and Nickl, 2013), and is in Section 6. We recover the results in the paper (Bull and Nickl, 2013) for $p = 2$ (Theorem 3.2).

Another direct consequence of Theorem 3.4 is, in the case $s(1 - 1/p) > t$, the minimax-optimal order of ρ_n for which the construction of confidence sets is made possible on the set $I = \{s, t\}$.

Theorem 3.6. *Let $B > 0$ and $\alpha > 0$, and assume that $s(1 - 1/p) > t \geq 1/2$. Let $\mathcal{P}_n = \tilde{\Sigma}(t, B, \rho_n) \bigcup \Sigma(s, B)$ for some ρ_n.*

- *Let $\rho_n = C(B + 1)n^{-t/(2t+1-1/p)}$, where $C := C(p, \alpha)$ is large enough. Then there exists a (L_p, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*
- *Let $\rho_n = \upsilon n^{-t/(2t+1-1/p)}$, where $\upsilon := \upsilon(p, B, \alpha)$ is small enough. Then there exists no (L_p, α)-honest and adaptive confidence set given $\mathcal{P}_n$, $\{s, t\}$ and B.*

The proof of this Theorem is in Section 5 (upper bound) and in Section 8 (lower bound), and it is a direct consequence of Theorem 3.4. The upper and lower bound match the results in the paper (Bull and Nickl, 2013) for $p = 2$ (Theorem 3.3). It is also remarkable that for general p, the lower and upper bound also match the lower bound in Equation (3.5) of the simpler testing problem.

3.4. *Discussion*

Adaptation to the smoothness in $\mathcal{P}_n$. On some specific model $\mathcal{P}_n \subset \Sigma(t, B)$ (with $\mathcal{P}_n = \Sigma(t, B)$ if $s(1 - 1/p) \leq t$), we have created honest and adaptive confidence sets given B. In this paper, we considered adaptation to the exponent r of the Besov spaces $B_{r,p,\infty}$, which are the natural classes to adapt to since adaptive estimation is on these spaces. We did not consider adaptation to the radius B of the Besov ball. If $t < (1 - 1/p)s$, the model $\mathcal{P}_n$ on which we adapt is strictly smaller than $\Sigma(t, B)$. We however state that $\mathcal{P}_n$ could not have been considered significantly larger.

Remark on the case $s(1 - 1/p) \leq t$. There is a strong relation between the testing problem (3.3), and the problem of creating adaptive and honest confidence sets. It is remarkable however that in the case $s(1 - 1/p) \leq t$, although an uniformly consistent test exists only on a constrained model $\mathcal{P}_n =$

$\Sigma(s, B) \bigcup \tilde{\Sigma}(t, B, \rho_n)$ (with $\rho_n = 2C(B+1)n^{-\frac{s}{2s+1}}$), honest and adaptive confidence sets exist on $\Sigma(t, B)$ itself (Theorem 3.5). The proof of this theorem is actually very enlightening for understanding what is happening. Its nice feature is that it emphasises the connection between the testing problem (3.3), and the problem of building adaptive and honest confidence sets, also in the case $s(1 - 1/p) \leq t$ (unlike the proof in the paper (Bull and Nickl, 2013) for $p = 2$ and $s \leq 2t$). First, on $\mathcal{P}_n$, the existence of adaptive and honest confidence sets is a consequence of Theorem 3.4 (the proof of this fact is similar to the proof of Theorem 3.6). Indeed, set for $\alpha > 0$

$$C_n = \left\{ f \in \Sigma(t, B) : \|f - \tilde{f}_n\|_p \leq \frac{D}{\alpha} n^{-\frac{s}{2s+1}} (1 - \Psi_n) + \frac{D}{\alpha} n^{-\frac{t}{2t+1}} \Psi_n \right\}$$

where $\tilde{f}_n$ is the adaptive estimate constructed for Theorem 10.1 (which is the same as the adaptive estimate for Theorem 3.1), Ψ_n is the test from Theorem 3.4 with level α, and D is a large enough constant depending only on C, B. This confidence set C_n will be α-honest and adaptive on $\mathcal{P}_n$ for $\{s, t\}$ given B, since the test Ψ_n is accurate with probability at least $1 - \alpha$ for any functions in $\mathcal{P}_n$ (since $\rho_n = 2C(B+1)n^{-\frac{s}{2s+1}}$). Second, the functions of $\Sigma(t, B) \setminus \mathcal{P}_n$ are at a distance smaller than $\rho_n = 2C(B+1)n^{-\frac{s}{2s+1}}$ from functions in $\Sigma(s, B)$. Theorem 10.1 applies to these functions, and the adaptive estimate $\tilde{f}_n$ is such that

$$\sup_{f \in \Sigma(t, B) \setminus \mathcal{P}_n} \mathbb{E}_f \|\tilde{f}_n - f\|_2 \leq E n^{-\frac{s}{2s+1}},$$

for $E > 0$ large enough and depending only on C, B. For this reason, misclassifying a function $f \in \Sigma(t, B) \setminus \mathcal{P}_n$ into $\Sigma(s, B)$ is not problematic for confidence sets: indeed the previous equation implies that C_n contains such an f with probability larger than $1 - \alpha$ provided that $D > E$, even though $f \notin \Sigma(s, B)$. We illustrate the idea of the proof in Figure 1.

Confidence sets for a general segment $[t, s]$. As in the paper (Bull and Nickl, 2013), it is possible to extend Theorem 3.5 to the case $I = [t, s]$ (see Theorem 11.1 in the supplementary material, Section 11) on $\mathcal{P}_n = \Sigma(t, B)$, provided that $s(1 - 1/p) \leq t$. One can then combine the results in Theorems 3.6 and 11.1 (eaxactly in the same way as in the paper (Bull and Nickl, 2013), Theorem 5) to construct honest and adaptive confidence sets over any segment $[t, s]$, with $1/2 \leq t < s$, and on a maximal model $\mathcal{P}_n$. We refer the reader to Theorem 5 and it's proof in the paper (Bull and Nickl, 2013), as the construction and proof for this fact in L_p with $p \geq 2$ (and $p < \infty$) is exactly the same as what is done in this paper for L_2, by just combining Theorems 3.6 and 3.5.

Extension to other settings. In the construction of confidence bands that we propose in Subsection 4.2, we first construct a test for the testing problem in Equation (3.3). In order to do that, we estimate the quantities $|a_{l,k}|^p$. The estimates we propose have good properties because the data is generated by an homocedastic Gaussian process. The main obstacle in more general settings is
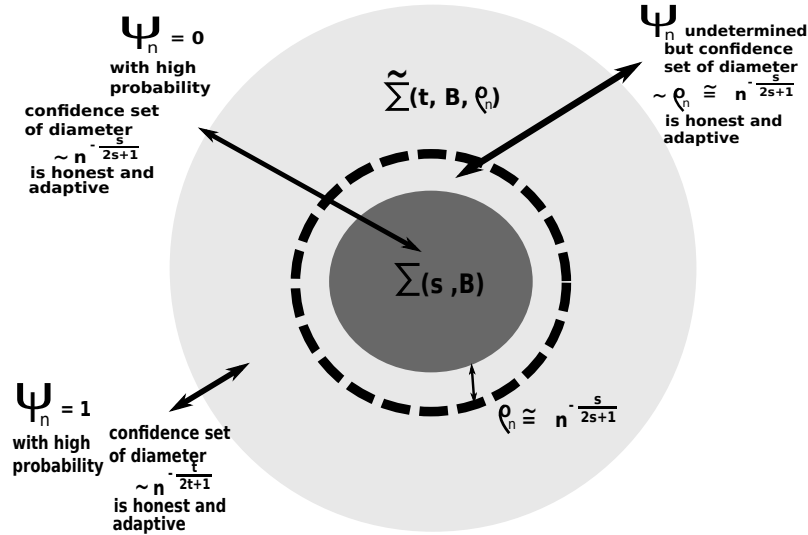
FIG 1. *Illustration of the proof of Theorem 3.5.*

that one does not know the distribution of the noise (and in particular its p first moments). Indeed, in the computation of the quantities $\hat{F}_p^p(l, k)$, we plug the p first moments of a Gaussian distribution in order to correct the bias of $|\hat{a}_{l,k}|^p$ toward $|a_{l,k}|^p$. If the distribution of the noise is not Gaussian, the bias is not going to be corrected by these (Gaussian) moments, and we would want to replace them with the moments of the noise, or rather by estimates of the moments computed on the empirical residuals. A more detailed discussion can be found in the supplementary material, Appendix 12.

4. Proof of Theorem 3.4 (upper bound)

The method that we propose for the construction of the test in Theorem 3.4 for the testing problem (3.3) is quite different from what was developed in (Hoffmann and Nickl, 2011; Bull and Nickl, 2013). The main idea is to prove that for any $f \in \Sigma(t, B)$, the quantity $\|f - \Sigma(s, B)\|_p$ is close to the quantity

$$\|\Pi_{V_{j_s}}(f) - \Sigma(s, B)\|_p + \sum_{l=j_s+1}^{j} \|\Pi_{W_l} f\|_p.$$

This finding is actually very useful in practice since it allows to eliminate the empirical minimisation over $\Sigma(s, B)$ for wavelet coefficients of high resolution (which are the difficult ones to estimate), and it is a way around the technical difficulties encountered when performing the infimum test (see e.g. papers (Hoffmann and Nickl, 2011; Bull and Nickl, 2013)). Then, one needs to estimate care-

fully $\|\Pi_{V_{j_s}}(f) - \Sigma(s, B)\|_p$ and the terms $\|\Pi_{W_l} f\|_p$. The first term is easy to control using Borell's inequality. The second terms are, for each l, approximated by a rescaled sum of proper Taylor expansions of the terms $|\hat{a}_{l,k}|^p$ (i.e. the quantities $\hat{F}_p^p(l, k)$, defined in next subsection). The variance of these estimates $\hat{F}_p^p(l, k)$ is not too difficult to bound in a proper way, since their variance is of same order than the variance of $|\hat{a}_{l,k}|^p$, which is bounded as $\tilde{C}^{(p)}(n^{-p} + |a_{l,k}|^{2(p-1)} n^{-1})$. The critical quantity is the mean of these terms. When p is even, the idea behind the construction of $\hat{F}_p^p(l, k)$ follows from the fact that

$$\mathbb{E}|\hat{a}_{l,k}|^p = \mathbb{E}(a_{l,k} + \hat{a}_{l,k} - a_{l,k})^p$$
$$= \sum_{u=0, u \ even}^{p} \mathbf{C}_p^u a_{l,k}^u \frac{\mathbf{E}_{G \sim \mathcal{N}(0,1)} |G|^{p-u}}{n^{(p-u)/2}}.$$

where $\mathbf{C}_p^u = \frac{p(p-1)\cdots(p-u+1)}{1 \cdots u}$ is the usual binomial coefficient. $\hat{F}_p^p(l, k)$ is then $|\hat{a}_{l,k}|^p$ minus an unbiased estimate (constructed by induction) of the sum in last equation up to $u = p - 2$. One can prove that

$$\mathbb{E}\hat{F}_p^p(l, k) = |a_{l,k}|^p.$$

Otherwise if p is not an even integer, but any positive real number larger than or equal to 2, the expectation of these Taylor expansions $\hat{F}_p^p(l, k)$ is such that

$$D^{(m)}|a_{l,k}|^p \le \mathbb{E}\hat{F}_p^p(l, k) \le D^{(M)}|a_{l,k}|^{\lfloor p \rfloor} \left(|a_{l,k}|^{p - \lfloor p \rfloor} + \frac{1}{n^{(p - \lfloor p \rfloor)/2}} \right),$$

where $D^{(m)}$ and $D^{(M)}$ are two strictly positive constants. Since under H_1, only the lower bound on $\mathbb{E}\hat{F}_p^p(l, k)$ matters, and under H_0, only the upper bound on $\mathbb{E}\hat{F}_p^p(l, k)$ matters (and under H_0, the sum of the $|a_{l,k}|^{\lfloor p \rfloor}$ is small enough to neutralise the effect of the disturbing sum of the terms $|a_{l,k}|^{\lfloor p \rfloor} \frac{1}{n^{(p - \lfloor p \rfloor)/2}}$), we will have satisfying concentration results for the sums of $\hat{F}_p^p(l, k)$ (i.e. $T_n(l)$). Controlling the mean and variance of these terms leads to large deviation results on the sums. This all enables us to construct an uniformly consistent test by considering if or if not these quantities (estimates of $\|\Pi_{V_{j_s}}(f) - \Sigma(s, B)\|_p$ and of the terms $\|\Pi_{W_l} f\|_p$) exceed given thresholds.

4.1. Definition of a related testing problem

Assume that $p \ge 2$ (and $p < \infty$) and $s > t \ge 1/2$.

4.2. Definition of the test statistic

Let $0 < j_s \le j$ be two integers such that

$$j_s = \lfloor \log(n^{\frac{1}{2s+1}}) \rfloor \quad and \quad j = \lfloor \log(n^{\frac{1}{2t+1-1/p}}) \rfloor.$$

For any $u > 0$ we define the following quantities

$$m_u^u = \mathbf{E}_{G \sim \mathcal{N}(0,1)} |G|^u.$$

We also define by convention, for any $l \geq j_s, k \in Z_l$, the following estimate of $(a_{l,k})^0$

$$\hat{F}_0^0(l,k) = 1.$$

We now define by induction for any $u \geq 2$ even, the following estimates of $(a_{l,k})^u$.

$$\hat{F}_u^u(l,k) = \hat{a}_{l,k}^u - \sum_{i=0, i \text{ even}}^{u-2} \mathbf{C}_u^i \left(\frac{m_{u-i}}{n^{1/2}}\right)^{(u-i)} \hat{F}_i^i(l,k),$$

where $\mathbf{C}_u^i = \frac{u(u-1)\cdots(u-i+1)}{1\cdots i}$ is the usual binomial coefficient. We extend this definition for $|a_{l,k}|^p$ if p non-even (and also non necessarily integer) by setting

$$\hat{F}_p^p(l,k) = |\hat{a}_{l,k}|^p - \sum_{u=0, u \text{ even}}^{\lfloor p \rfloor - 2} \mathbf{C}_p^u \left(\frac{m_{p-u}}{n^{1/2}}\right)^{(p-u)} \hat{F}_u^u(l,k),$$

where we set also for non-integer p that $\mathbf{C}_p^u = \frac{p(p-1)\cdots(p-u+1)}{1\cdots u}$ (and $\mathbf{C}_p^0 = 1$ by convention).

Consider the test statistics, for any $j_s \leq l \leq j$

$$T_n(l) = \sum_{k \in Z_l} 2^{lp(1/2 - 1/p)} \hat{F}_p^p(l,k),$$

and also

$$\tilde{T}_n = \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}} \hat{f}_n - g\|_p.$$

Consider positive constants $(t_n(l))_{j_s \leq l \leq j}$ and $\tilde{t}_n$. We consider the test:

$$\Psi_n = 1 - \mathbb{I}\left\{\tilde{T}_n \leq \tilde{t}_n\right\} \prod_{j_s \leq l \leq j} \mathbb{I}\left\{T_n(l) \leq (t_n(l))^p\right\}.$$

We set

$$t_n(l) = E_1 \left(2^{-ls(p-1)/p} \left(\frac{2^{l(1-1/p)}}{n}\right)^{1/(2p)} + 2^{-\lfloor p \rfloor ls/p} \left(\frac{2^l}{n}\right)^{(p-\lfloor p \rfloor)/(2p)}\right.$$
$$\left. + 2^{-ls} + \sqrt{\frac{2^{(p-1)(j+l)/(2p)}}{n}}\right),$$

and

$$\tilde{t}_n = E_2 \sqrt{\frac{2^{j_s}}{n}},$$

where E_1 and E_2 are some large enough constants that depend only on p, B and the desired level of the test. Then the test is uniformly consistent with

$$\rho_n = 4\left((B+1)C'2^{-jt} + 2\sum_{l=j_s}^{j} t_n(l) + 2\tilde{t}_n\right),$$

where C' is a large enough constant that depends only on p and the desired level of the test.

4.3. Decomposition of the problem

We justify here the test that we proposed.

Lemma 4.1. *Let $0 < j_s < j$ be two integers. Let $(\tau_l)_{j_s \le l \le j}$ be a sequence of positive numbers such that $\tau_l \ge 2 \times 2^{-ls}$. Assume that $\rho_n \ge 4C_p(B+1)(2^{-jt} + \sum_{l=j_s}^{j} \tau_l)$ where $C_p \ge 1$ is some positive constant that depends on p only. Then we have*

- $f \in \Sigma(s, B) \Rightarrow (\max_{j_s \le l \le j} \|\Pi_{W_l}(f)\|_{0,p,\infty} \le (B+1)\tau_l/2$ *AND* $\|\Pi_{V_{j_s}} f - \Sigma(s, B)\|_p = 0)$.
- $f \in \tilde{\Sigma}(t, B, \rho_n) \Rightarrow (\max_{j_s \le l \le j} \|\Pi_{W_l}(f)\|_{0,p,\infty} \ge (B+1)\tau_l$ *OR* $\|\Pi_{V_{j_s}} f - \Sigma(s, B)\|_p \ge 3\rho_n/8)$.

Proof. *Under the null Hypothesis H_0*
Assume that $f \in \Sigma(s, B)$. Then $\Pi_{V_{j_s}} f \in \Sigma(s, B)$, and

$$\|\Pi_{V_{j_s}} f - \Sigma(s, B)\|_p = 0.$$

If f is in $\Sigma(s, B)$ then by definition of the Besov spaces

$$\|\Pi_{V_j} f\|_{s,p,\infty} \le B,$$

which implies by definition of the $\|.\|_{0,p,\infty}$ norm that

$$\sup_{j_s \le l \le j} \|\Pi_{W_l} f\|_{0,p,\infty} - \frac{B}{2^{ls}} \le 0,$$

which implies that

$$\sup_{j_s \le l \le j} \|\Pi_{W_l} f\|_{0,p,\infty} \le B2^{-ls} \le \tau_l/2,$$

Under the alternative Hypothesis H_1
Assume that f is in $\tilde{\Sigma}(t, B, \rho_n)$. We have by triangular inequality

$$\inf_{g \in \Sigma(s,B)} \|f - g\|_p \le \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p + \|\Pi_{V_j \setminus V_{j_s}}(f)\|_p + \|\Pi_{V^-}(f)\|_p,$$

where we set $V^- = \mathrm{span}\big(\bigcup_{l=j+1}^{\infty} W_l\big)$. By definition of $\tilde{\Sigma}(t, B, \rho_n)$ and definition of the Besov spaces, we know that $\|\Pi_{V^-} f\|_p \leq C_p B 2^{-jt}$ (see Proposition 3, Supplementary Material). We thus have

$$\inf_{g \in \Sigma(s,B)} \|f - g\|_p \leq \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p + \|\Pi_{V_j \setminus V_{j_s}}(f)\|_p + C_p B 2^{-jt}.$$

We thus have (since $\rho_n \leq \inf_{g \in \Sigma(s,B)} \|f - g\|_p$)

$$3\rho_n/4 \leq \rho_n - C_p B 2^{-jt} \leq \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p + \|\Pi_{V_j \setminus V_{j_s}}(f)\|_p.$$

This implies that $\inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p \geq 3\rho_n/8$, or $\|\Pi_{V_j \setminus V_{j_s}}(f)\|_p \geq 3\rho_n/8$.

Note now that by imbrication of the Besov spaces (see Proposition 2) there exists a constant $C_p \geq 1$ that depends on p only such that

$$\|\Pi_{V_j - V_{j_s}}(f)\|_p \leq C_p \|\Pi_{V_j - V_{j_s}}(f)\|_{0,p,1}$$

$$= C_p \sum_{l=j_s}^{j} \|\Pi_{W_l}(f)\|_{0,p,\infty}.$$

Since $3/8\rho_n \geq C_p B \sum_{l=j_s}^{j} \tau_l$, the previous equation implies that if $\|\Pi_{V_j \setminus V_{j_s}}(f)\|_p \geq 3\rho_n/8$, then there exists $j_s \leq l \leq j$ such that $\|\Pi_{W_l}(f)\|_{0,p,\infty} \geq B\tau_l$. This concludes the proof. □

4.4. Large deviations for $\|\Pi_{V_{j_s}}(\hat{f}_n - f)\|_p$

Similarly to Lemma 10.2 (Supplementary Material), we have the following Lemma.

Lemma 4.2. *We have*

$$\sup_{f \in L_p} \Pr\left\{ \|\Pi_{V_{j_s}}(\hat{f}_n - f)\|_p \geq (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\delta)})\sqrt{\frac{2^{j_s}}{n}} \right\} \leq \delta,$$

where $C_{p/(p-1)}$ and D_p are positive constants that depend only on p.

Proof. The Lemma follows directly from Propositions 5 and 6 (Supplementary Material). □

4.5. Convergence tools for $T_n(l)$

Lemma 4.3. *There are constants $C^{(p)}, D^{(m)}, D^{(M)}, D^{(D)}$ that depend on p only such that for any $\Delta \in]0, 1[$ we have*

$$\Pr\left\{ \exists l : j_s \leq l \leq j, T_n(l) \leq \|\Pi_{W_l} f\|_{0,p,\infty}^{p-1} \Big(D^{(m)} \|\Pi_{W_l} f\|_{0,p,\infty} \right.$$

$$\left. - \sqrt{\frac{C^{(p)} D^{(D)}}{\Delta}} \sqrt{\frac{2^{l(1-1/p)}}{n}} \Big) - \sqrt{\frac{D^{(D)} C^{(p)}}{\Delta}} \frac{2^{(p-1)(j+l)/2}}{n^p} \forall j_s \leq l \leq j \right\} \leq \Delta.$$

$$(4.1)$$

Also if $f \in \Sigma(s)$ then

$$\Pr\left\{\exists l : j_s \leq l \leq j, T_n(l) \geq (B^p + 1)\left(D^{(M)} + \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\right)(\tau_{p,l}(s))^p\right.$$

$$\left. + \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\sqrt{\frac{2^{(p-1)(j+l)/2}}{n^p}}\forall j_s \leq l \leq j\right\} \leq \Delta, \tag{4.2}$$

where $(\tau_{p,l}(s))^p = 2^{-ls(p-1)}\sqrt{\frac{2^{l(1-1/p)}}{n}} + 2^{-\lfloor p\rfloor ls}\left(\frac{2^l}{n}\right)^{(p-\lfloor p\rfloor)/2} + 2^{-pls}$.

Proof. We remind that $\Pi_{W_l}\hat{f}_n = \sum_{k\in Z_l}\hat{a}_{l,k}\psi_{l,k}$ and that $\Pi_{W_l}f = \sum_{k\in Z_l}a_{l,k}\psi_{l,k}$.

Lemma 4.4. *There exists two strictly positive constants $D^{(m)}$ and $D^{(M)}$ that depend on p only such that*

$$D^{(M)}\left(\|\Pi_{W_l}f\|_{0,p,\infty}^{\lfloor p\rfloor}\left(\frac{2^l}{n}\right)^{(p-\lfloor p\rfloor)/2} + \|\Pi_{W_l}f\|_{0,p,\infty}^p\right) \geq \mathbf{E}[T_n(l)] \geq D^{(m)}\|\Pi_{W_l}f\|_{0,p,\infty}^p.$$

Proof. We provide bounds on $\mathbf{E}\hat{F}_p^p(l,k)$, and this implies bounds on $T_n(l)$ by definition of $T_n(l)$.

Step 1: Expression of $\mathbf{E}\hat{F}_u^u(l,k)$ for u even. Let $u > 0$ be an even integer. We first prove by induction that $\mathbf{E}\hat{F}_u^u(l,k) = a_{l,k}^u$.

For $u = 2$ we have

$$\begin{aligned}
\mathbf{E}(\hat{a}_{l,k})^2 &= \mathbf{E}(a_{l,k} + \hat{a}_{l,k} - a_{l,k})^2 \\
&= a_{l,k}^2 + 2a_{l,k}\mathbf{E}(\hat{a}_{l,k} - a_{l,k}) + \mathbf{E}(\hat{a}_{l,k} - a_{l,k})^2 \\
&= a_{l,k}^2 + \frac{1}{n} = a_{l,k}^2 + \frac{m_2^2}{n},
\end{aligned}$$

since $\hat{a}_{l,k} - a_{l,k} \sim \mathcal{N}(0, 1/n)$. This implies that $\mathbf{E}\hat{F}_2^2(l,k) = a_{l,k}^2$ by definition of $\hat{F}_2^2(l,k)$.

The induction assumption is as follows: we assume that for any i even such that $2 \leq i \leq u-2$, we have $\mathbf{E}\hat{F}_i^i(l,k) = a_{l,k}^i$. Since u even, we have by a binomial expansion

$$\begin{aligned}
\mathbf{E}|\hat{a}_{l,k}|^u &= \mathbf{E}(a_{l,k} + \hat{a}_{l,k} - a_{l,k})^u \\
&= \sum_{i=0}^u \mathbf{C}_u^i a_{l,k}^i \mathbf{E}(\hat{a}_{l,k} - a_{l,k})^{u-i} \\
&= \sum_{i=0, i\ even}^u \mathbf{C}_u^i a_{l,k}^i \frac{m_{u-i}^{u-i}}{n^{(u-i)/2}},
\end{aligned}$$

since $\hat{a}_{l,k} - a_{l,k} \sim \mathcal{N}(0, 1/n)$ (and thus for i odd, $\mathbf{E}(\hat{a}_{l,k} - a_{l,k})^{u-i} = 0$). This implies by the induction assumption

$$\mathbf{E}\hat{F}_u^u(l,k) = \mathbf{E}\hat{a}_{l,k}^u - \sum_{i=0,i\ even}^{u-2} \mathbf{C}_u^i \mathbf{E}\hat{F}_i^i(l,k)\left(\frac{m_{u-i}}{n^{1/2}}\right)^{(u-i)}$$

$$= \sum_{i=0,i\ even}^{u} \mathbf{C}_u^i a_{l,k}^i \frac{m_{u-i}^{u-i}}{n^{(u-i)/2}} - \sum_{i=0,i\ even}^{u-2} \mathbf{C}_u^i a_{l,k}^i \left(\frac{m_{u-i}}{n^{1/2}}\right)^{(u-i)}$$

$$= \sum_{i=0,i\ even}^{u} \mathbf{C}_u^i a_{l,k}^i \left(\frac{m_{u-i}}{n^{1/2}}\right)^{(u-i)} - \sum_{i=0,i\ even}^{u-2} \mathbf{C}_u^i a_{l,k}^i \left(\frac{m_{u-i}}{n^{1/2}}\right)^{(u-i)}$$

$$= a_{l,k}^u.$$

This concludes the induction.

Since for p even it holds that $\mathbf{E}\hat{F}_p^p(l,k) = a_{l,k}^p$, it also holds in particular by definition of $T_n(l)$ and of $\|\Pi_{W_l}f\|_{0,p,\infty}^p$ that

$$\mathbf{E}T_n(l) = \|\Pi_{W_l}f\|_{0,p,\infty}^p.$$

Step 2: Lower bound on $\mathbf{E}\hat{F}_p^p(l,k)$ *for p non-even.* Consider now the case p non-even.

We have since $\hat{a}_{l,k} - a_{l,k} \sim \mathcal{N}(0, 1/n)$, and since, if a and b in $\mathbb{R}$ are such that $ab > 0$, then $|a + b| > \max(|a|, |b|)$,

$$\mathbf{E}|\hat{a}_{l,k}|^p = \mathbf{E}\left|a_{l,k} + \hat{a}_{l,k} - a_{l,k}\right|^p$$

$$\geq |a_{l,k}|^p \Pr\left(a_{l,k}(\hat{a}_{l,k} - a_{l,k}) \geq 0\right)$$

$$= |a_{l,k}|^p \Pr_{G\sim\mathcal{N}(0,1)}\left(G \geq 0\right) = |a_{l,k}|^p/2,$$

which implies by definition of $F_p^p(l,k)$ and also since $\mathbf{E}\hat{F}_u^u(l,k) = a_{l,k}^u$ for u even

$$\mathbf{E}\hat{F}_p^p(l,k) \geq |a_{l,k}|^p/2 - \sum_{u=0,u\ even}^{\lfloor p \rfloor-2} \mathbf{C}_p^u a_{l,k}^u \left(\frac{m_{p-u}}{n^{1/2}}\right)^{p-u}$$

$$\geq |a_{l,k}|^p/2 - C'\left(\left(\frac{m_p}{n^{1/2}}\right)^p + a_{l,k}^{\lfloor p \rfloor}\left(\frac{m_{p-\lfloor p \rfloor}}{n^{1/2}}\right)^{(p-\lfloor p \rfloor)}\right), \qquad (4.3)$$

where C' is some constant that depends on p only.

Also, we have by a Taylor expansion since the function $h : m \to \mathbf{E}_{G\sim N(0,1)}|m + G|^p$ is in $C^\infty([0,1])$

$$\mathbf{E}|\hat{a}_{l,k}|^p = \mathbf{E}|a_{l,k} + \hat{a}_{l,k} - a_{l,k}|^p = \mathbf{E}_{G\sim\mathcal{N}(0,1)}\left|a_{l,k} + \frac{G}{\sqrt{n}}\right|^p$$

$$= \sum_{u=0,u\ even}^{\lfloor p \rfloor-2} \mathbf{C}_p^u a_{l,k}^u \left(\frac{m_{p-u}}{n^{1/2}}\right)^{p-u} + \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \mathbf{E}_{G\sim\mathcal{N}(0,1)}\left|y + \frac{G}{\sqrt{n}}\right|^{p-\lfloor p \rfloor},$$

since $\hat{a}_{l,k} - a_{l,k} \sim \mathcal{N}(0, 1/n)$, and where $|y| \leq |a_{l,k}|$.

This implies, together with the fact that $\mathbf{E}\hat{F}_u^u(l,k) = a_{l,k}^u$ for u even, that

$$\mathbf{E}\hat{F}_p^p(l,k) = \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \mathbf{E}_{G \sim N(0,1)} \Big| y + \frac{G}{\sqrt{n}} \Big|^{p - \lfloor p \rfloor} \tag{4.4}$$

$$\geq \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \Big(\frac{m_{p-\lfloor p \rfloor}}{n^{1/2}} \Big)^{p - \lfloor p \rfloor}. \tag{4.5}$$

By considering the bound in Equation (4.5) for $|a_{l,k}| \leq C'' n^{-1/2}$ and the bound in Equation (4.3) for $|a_{l,k}| \geq C'' n^{-1/2}$, for some C'' that depends on p only (through m_p, $m_{p-\lfloor p \rfloor}$ and C'), we obtain that there exists a constant $D^{(m)} > 0$ that depends on p only and such that

$$\mathbf{E}\hat{F}_p^p(l,k) \geq D^{(m)} \max \Big(|a_{l,k}|^p, a_{l,k}^{\lfloor p \rfloor} n^{-(p-\lfloor p \rfloor)/2} \Big) \geq D^{(m)} |a_{l,k}|^p.$$

This leads to the lower bound in Lemma 4.4 by definition of $\|\Pi_{W_l} f\|_{0,p,\infty}^p$ and $T_n(l)$.

Step 3: Upper bound on $\mathbf{E}\hat{F}_p^p(l,k)$ for p non-even. By Equation (4.4) there exists y such that $|y| \leq |a_{l,k}|$ and such that

$$\mathbf{E}\hat{F}_p^p(l,k) = \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \mathbf{E}_{G \sim N(0,1)} \Big| y + \frac{G}{\sqrt{n}} \Big|^{p - \lfloor p \rfloor}$$

$$\leq \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \Big(4|y|^{p-\lfloor p \rfloor} + 4 \mathbf{E}_{G \sim \mathcal{N}(0,1)} \Big| \frac{G}{\sqrt{n}} \Big|^{p-\lfloor p \rfloor} \Big)$$

$$\leq \mathbf{C}_p^{\lfloor p \rfloor} a_{l,k}^{\lfloor p \rfloor} \Big(4|y|^{p-\lfloor p \rfloor} + 4 m_{p-\lfloor p \rfloor}^{p-\lfloor p \rfloor} \Big(\frac{1}{\sqrt{n}} \Big)^{p-\lfloor p \rfloor} \Big)$$

$$\leq \mathbf{C}_p^{\lfloor p \rfloor} \Big(4|a_{l,k}|^p + 4 a_{l,k}^{\lfloor p \rfloor} m_{p-\lfloor p \rfloor}^{p-\lfloor p \rfloor} \Big(\frac{1}{\sqrt{n}} \Big)^{p-\lfloor p \rfloor} \Big). \tag{4.6}$$

since $p - \lfloor p \rfloor \leq 2$. By Hölder's inequality,

$$\sum_{k \in Z_l} a_{l,k}^{\lfloor p \rfloor} \leq |Z_l|^{\frac{p-\lfloor p \rfloor}{p}} \Big(\sum_{k \in Z_l} |a_{l,k}|^p \Big)^{\frac{\lfloor p \rfloor}{p}} \leq (c2^l)^{\frac{p-\lfloor p \rfloor}{p}} \Big(\sum_{k \in Z_l} |a_{l,k}|^p \Big)^{\lfloor p \rfloor / p},$$

and this implies together with Equation (4.6) and by definition of $\|\Pi_{W_l} f\|_{0,p,\infty}^p$ and $T_n(l)$ that

$$\mathbf{E}T_n(l) \leq \mathbf{C}_p^{\lfloor p \rfloor} \Bigg(4\|\Pi_{W_l} f\|_{0,p,\infty}^p$$

$$+ 4 m_{p-\lfloor p \rfloor}^{p-\lfloor p \rfloor} \Big(\frac{1}{\sqrt{n}} \Big)^{p-\lfloor p \rfloor} 2^{lp(1/2-1/p)} (c2^l)^{\frac{p-\lfloor p \rfloor}{p}} \Big(\sum_{k \in Z_l} |a_{l,k}|^p \Big)^{\lfloor p \rfloor / p} \Bigg),$$

which leads to the upper bound in Lemma 4.4. □

Lemma 4.5. *There is a constant $C^{(p)}$ that depends on p only and that is such that*

$$\mathbb{V}[T_n(l)] \leq C^{(p)}\left(\sqrt{\frac{2^{l(p-1)}}{n^p}} + \|\Pi_{W_l}f\|_{0,p,\infty}^{p-1}\sqrt{\frac{2^{l(1-2/p)}}{n}}\right)^2.$$

Proof. Step 1: Bound on $\mathbb{V}\hat{F}_u^u(l,k)$ for u even. Let $u > 0$ be an even integer. We prove by induction that for any such u, there exists a constant $C^{(u)}$ that depends on u only and such that

$$\mathbb{V}\hat{F}_u^u(l,k) \leq C^{(u)}\left(\frac{1}{n^u} + \frac{|a_{l,k}|^{2(u-1)}}{n}\right).$$

For $u = 2$, this follows from the fact that

$$\mathbb{V}\hat{F}_2^2(l,k) = \mathbb{V}(\hat{a}_{l,k}^2 - 1/n) = \mathbf{E}_{G\sim N(0,1)}\mathbb{V}(G^2)/n \leq C^{(2)}/n,$$

where $C^{(2)} > 0$ is a universal constant.

Assume now that it is true for any i even such that $0 \leq i \leq u - 2$. We have

$$\mathbb{V}\hat{F}_u^u(l,k) = \mathbb{V}\left[\hat{a}_{l,k}^u - \sum_{i=0,i \text{ even}}^{u-2} \mathbf{C}_u^i\hat{F}_i^i(l,k)\frac{m_{u-i}^{u-i}}{n^{(u-i)/2}}\right]$$

$$\leq u^2\mathbb{V}\hat{a}_{l,k}^u + u^2\sum_{i=0,i \text{ even}}^{u-2}(\mathbf{C}_u^i)^2\frac{m_{u-i}^{2(u-i)}}{n^{u-i}}\mathbb{V}(\hat{F}_i^i(l,k))$$

$$\leq u^2\mathbb{V}\hat{a}_{l,k}^u + u^2\sum_{i=0,i \text{ even}}^{u-2}(\mathbf{C}_u^i)^2\frac{m_{u-i}^{2(u-i)}}{n^{u-i}}C^{(i)}\left(\frac{1}{n^i} + \frac{|a_{l,k}|^{2(i-1)}}{n}\right). \quad (4.7)$$

By (Ingster and Suslina, 2002) (page 86), if $G \sim \mathcal{N}(0,1)$ and m is a real number, and $u \geq 2$, then there is a constant $D^{(u)}$ that depends only on u such that $\mathbb{V}(|G + m|^u) \leq D^{(u)}(1 + |m|^{2u-2})$. We thus have

$$\mathbb{V}\left(|\hat{a}_{l,k}|^u\right) \leq D^{(u)}\left(\frac{1}{n^u} + \frac{|a_{l,k}|^{2(u-1)}}{n}\right). \quad (4.8)$$

This implies together with Equation (4.7) that

$$\mathbb{V}\hat{F}_u^u(l,k) \leq u^2 D^{(u)}\left(\frac{1}{n^u} + \frac{|a_{l,k}|^{2(u-1)}}{n}\right)$$

$$+ u^2\sum_{i=0,i \text{ even}}^{u-2}(\mathbf{C}_u^i)^2\frac{m_{u-i}^{2(u-i)}}{n^{u-i}}C^{(i)}\left(\frac{1}{n^i} + \frac{|a_{l,k}|^{2(i-1)}}{n}\right)$$

$$\leq C^{(u)}\left(\frac{1}{n^u} + \frac{|a_{l,k}|^{2(u-1)}}{n}\right), \quad (4.9)$$

where $C^{(u)}$ depends on u only. This concludes the induction.

Step 2: Bound on $\mathbb{V}\hat{F}_p^p(l,k)$. If p is even, we consider the bound in Step 1. If p is not even, we know similarly as in Equation (4.8) that

$$\mathbb{V}|\hat{a}_{l,k}|^p \leq D^{(p)}\Big(\frac{1}{n^p} + \frac{|a_{l,k}|^{2(p-1)}}{n}\Big),$$

which implies by definition of $\hat{F}_p^p(l,k)$ and also since for any u even, Equation (4.9) holds, the following result

$$\mathbb{V}\hat{F}_p^p(l,k) = \mathbb{V}\Big[|\hat{a}_{l,k}|^p - \sum_{\substack{u=0,i\ even}}^{\lfloor p \rfloor -2} \mathbf{C}_p^u \hat{F}_u^u(l,k)\frac{m_{p-u}^{p-u}}{n^{(p-u)/2}}\Big]$$

$$\leq (\lfloor p \rfloor + 1)^2 D^{(p)}\Big(\frac{1}{n^p} + \frac{|a_{l,k}|^{2(p-1)}}{n}\Big) \tag{4.10}$$

$$+ (\lfloor p \rfloor + 1)^2 \sum_{\substack{u=0,i\ even}}^{\lfloor p \rfloor -2} (\mathbf{C}_p^u)^2 \frac{m_{p-u}^{2(p-u)}}{n^{p-u}} C^{(p)}\Big(\frac{1}{n^u} + \frac{|a_{l,k}|^{2(u-1)}}{n}\Big)$$

$$\leq C^{(p)}\Big(\frac{1}{n^p} + \frac{|a_{l,k}|^{2(p-1)}}{n}\Big), \tag{4.11}$$

where $C^{(p)}$ depends on p only.

Step 3: Conclusion

Now by definition of $T_n(l)$ and since the $\hat{a}_{l,k}$ are independent, we have by Equation (4.11)

$$\mathbb{V}T_n(l) = \sum_{k \in Z_l} 2^{2lp(1/2-1/p)}\mathbb{V}\hat{F}_u^u(l,k)$$

$$\leq C^{(p)}\Big(\frac{2^{lp(1-1/p)}}{n^p} + 2^{lp(1-2/p)}\sum_{k \in Z_l}\frac{|a_{l,k}|^{2(p-1)}}{n}\Big)$$

$$\leq C^{(p)}\Big(\frac{2^{lp(1-1/p)}}{n^p} + 2^{lp(1-2/p)}\frac{\big(\sum_{k \in Z_l}|a_{l,k}|^p\big)^{2(p-1)/p}}{n}\Big)$$

$$= C^{(p)}\Big(\frac{2^{lp(1-1/p)}}{n^p} + 2^{l(1-2/p)}\frac{\big(\sum_{k \in Z_l}2^{lp(1/2-1/p)}|a_{l,k}|^p\big)^{2(p-1)/p}}{n}\Big)$$

$$\leq C^{(p)}\Big(\frac{2^{lp(1-1/p)}}{n^p} + \frac{2^{l(1-2/p)}}{n}\|\Pi_{W_l}f\|_{0,p,\infty}^{2(p-1)}\Big).$$

since $\sum_{k \in Z_l}|a_{l,k}|^{2(p-1)} \leq (\sum_{k \in Z_l}|a_{l,k}|^p)^{2(p-1)/p}$ (since $2(p-1)/p \geq 1$). This concludes the proof. □

Lemma 4.5 implies by Chebyshev's inequality that for $0 < \delta_l \leq 1$

$$\Pr\Big\{\big|T_n(l) - \mathbf{E}[T_n(l)]\big| \geq \sqrt{\frac{1}{\delta_l}}\sqrt{C^{(p)}\Big(\frac{2^{lp(1-1/p)}}{n^p} + \frac{2^{l(1-2/p)}}{n}\|\Pi_{W_l}f\|_{0,p,\infty}^{2(p-1)}\Big)}\Big\} \leq \delta_l.$$

which implies since for any $a, b \geq 0$, we have $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$,

$$\Pr\left\{\left|T_n(l) - \mathbf{E}[T_n(l)]\right| \geq \sqrt{\frac{C^{(p)}}{\delta_l}}\left(\sqrt{\frac{2^{l(p-1)}}{n^p}} + \|\Pi_{W_l}f\|_{0,p,\infty}^{p-1}\sqrt{\frac{2^{l(1-2/p)}}{n}}\right)\right\} \leq \delta_l. \tag{4.12}$$

Consider $\frac{1}{\delta_l} = \frac{D^{(D)}}{\Delta}\min(2^{(j-l)(p-1)/2}, 2^{l/p})$ where $0 < \Delta \leq 1$ and $D^{(D)}$ is a positive constant such that $\sum_{j_s \leq l \leq j}\frac{1}{\min(2^{(j-l)(p-1)/2}, 2^{l/p})} = D^{(D)}$. $D^{(D)}$ is a constant that is bounded depending on p only as a sum of the max of two geometric series. We thus have by an union bound on Equation (4.12)

$$\Pr\left\{\forall j_s \leq l \leq j, \left|T_n(l) - \mathbf{E}[T_n(l)]\right|\right.$$
$$\left.\geq \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\left(\sqrt{\frac{2^{(p-1)(j+l)/2}}{n^p}} + \|\Pi_{W_l}f\|_{0,p,\infty}^{p-1}\sqrt{\frac{2^{l(1-1/p)}}{n}}\right)\right\} \leq \sum_{j_s \leq l \leq j}\delta_l \leq \Delta. \tag{4.13}$$

Equation (4.13) implies with Lemma 4.4 (right hand side) that

$$\Pr\left\{\forall j_s \leq l \leq j, T_n(l) \leq \|\Pi_{W_l}f\|_{0,p,\infty}^{p-1}\left(D^{(m)}\|\Pi_{W_l}f\|_{0,p,\infty}\right.\right.$$
$$\left.\left. - \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\sqrt{\frac{2^{l(1-1/p)}}{n}}\right) - \sqrt{\frac{D^{(D)}C^{(p)}}{\Delta}}\frac{2^{(p-1)(j+l)/2}}{n^p}\right\} \leq \Delta.$$

Assume now that $f \in \Sigma(s)$. Then we have by Proposition 3 (Supplementary Material)

$$\|\Pi_{W_l}f\|_{0,p,\infty} \leq B2^{-ls}.$$

This implies with Equation (4.12) and Lemma 4.4 (left hand side)

$$\Pr\left\{\exists l : j_s \leq l \leq j, T_n(l) \geq D^{(M)}\left(B^p 2^{-pls} + B^{\lfloor p \rfloor}2^{-\lfloor p \rfloor ls}\left(\frac{2^l}{n}\right)^{(p-\lfloor p \rfloor)/2}\right)\right.$$
$$\left. + \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\left(B^{p-1}2^{-ls(p-1)}\sqrt{\frac{2^{l(1-1/p)}}{n}} + \sqrt{\frac{2^{(p-1)(j+l)/2}}{n^p}}\right)\right\} \leq \Delta,$$

which implies since $l \geq j_s$

$$\Pr\left\{\exists l : j_s \leq l \leq j, T_n(l) \geq (B^p + 1)\left(D^{(M)} + \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\right)(\tau_{p,l}(s))^p\right.$$
$$\left. + \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\sqrt{\frac{2^{(p-1)(j+l)/2}}{n^p}}\right\} \leq \Delta,$$

where $(\tau_{p,l}(s))^p = 2^{-ls(p-1)}\sqrt{\frac{2^{l(1-1/p)}}{n}} + 2^{-\lfloor p \rfloor ls}\left(\frac{2^l}{n}\right)^{(p-\lfloor p \rfloor)/2} + 2^{-pls}$. $\qquad\square$

4.6. Study of the test Ψ_n

Let $1 \geq \Delta > 0$ be a probability. We remind that j and j_s are such that

$$j_s = \lfloor \log(n^{\frac{1}{2s+1}}) \rfloor \quad and \quad j = \lfloor \log(n^{\frac{1}{2t+1-1/p}}) \rfloor.$$

We set for $j_s \leq l \leq j$,

$$\tau_l \equiv \tau_l(n, p, s) = C'\Big(\tau_{p,l}(s) + \sqrt{\frac{2^{(p-1)(j+l)/(2p)}}{n}} \Big),$$

where

$$C' \geq 2\Big(D^{(M)} + \sqrt{\frac{C^{(p)} D^{(D)}}{\Delta}} \Big(\frac{1}{D^{(m)}} + 1\Big) + C_p \Big),$$

and where $\tau_{p,l}(s)$ is defined in Lemma 4.3.

As sums of a geometric series, the following terms entering in the composition of $(\tau_{p,l}(s))^p$ satisfy

$$\sum_{l=j_s}^{j} 2^{-pls} \equiv n^{-\frac{sp}{2s+1}}, \quad and \quad \sum_{l=j_s}^{j} \sqrt{\frac{2^{(p-1)(j+l)/(2p)}}{n}} \equiv n^{-\frac{tp}{2t+1-1/p}},$$

$$and \quad \sum_{l=j_s}^{j} 2^{-ls(p-1)}\sqrt{\frac{2^{l(1-1/p)}}{n}} = O\Big(\max(n^{-\frac{sp}{2s+1}}, n^{-\frac{tp}{2t+1-1/p}}) \Big).$$

Also, since $s > 1/2 > \frac{p - \lfloor p \rfloor}{2\lfloor p \rfloor}$, we have in the same way

$$\sum_{l=j_s}^{j} 2^{-\lfloor p \rfloor ls} \Big(\frac{2^l}{n}\Big)^{(p-\lfloor p \rfloor)/2} \equiv n^{-\frac{sp}{2s+1}}.$$

The last two equation blocks, together with the definition of τ_l, imply that

$$\sum_{l=j_s}^{j} \tau_l = C' A_p (n^{-\frac{s}{2s+1}} + n^{-\frac{t}{2t+1-1/p}}),$$

where $10 \geq A_p \geq 1$.

We set

$$t_n(l) = (B+1)\tau_l/2 \quad and \quad \tilde{t}_n = (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}},$$

We also write

$$\rho_n = 4\Big[(B+1)C'2^{-jt} + 2\sum_{j_s \leq l \leq j} t_n(l) + 2\tilde{t}_n \Big] = C(B+1)(n^{-\frac{s}{2s+1}} + n^{-\frac{t}{2t+1-1/p}}),$$

with C fixed accordingly (we remind that $D^{(M)}, C^{(p)}, D^{(D)}, D^{(m)}$ are strictly positive constants that depend on p only).

Null Hypothesis H_0

Assume that $f \in \Sigma(s, B)$. We have by Lemma 4.3 (Equation (4.2)) and by definition of $t_n(l)$

$$\Pr\left\{\exists l : j_s \leq l \leq j, T_n(l) \geq (t_n(l))^p/2^p\right\} \leq \Delta.$$

Also since $f \in \Sigma(s, B)$

$$\tilde{T}_n = \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}} \hat{f}_n - g\|_p \leq \|\Pi_{V_{j_s}}(\hat{f}_n - f)\|_p.$$

Thus by Lemma 4.2, we have

$$\Pr\left\{\tilde{T}_n \geq (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}}\right\}$$
$$\leq \Delta,$$

and note that by definition $(D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}} \leq \tilde{t}_n$.

So with probability $1 - 2\Delta$, we have $\Psi_n = 0$ under H_0.

Alternative hypothesis H_1

We identify the seqence $(\tau_l)_{j_s \leq l \leq j}$, and the quantity ρ_n, with the quantities in Lemma 4.1: they have all required properties by definition and for C' (and thus C) large enough.

If H_1 is satisfied, then either $\inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p \geq 3\rho_n/8$ or $\max_{j_s \leq l \leq j} \|\Pi_{W_l}(f)\|_{0,p,\infty} \geq (B+1)\tau_l$ (see Lemma 4.1).

Case 1: $\max_{j_s \leq l \leq j} \|\Pi_{W_l}(f)\|_{0,p,\infty} \geq (B+1)\tau_l$

Using the results of Lemma 4.3 (Equation (4.1)), we have

$$\Pr\left\{\exists l : j_s \leq l \leq j, T_n^{(l)} \leq (B+1)^{p-1}\tau_l^{p-1}\left(D^{(m)}\tau_l(B+1)\right.\right.$$
$$\left.\left. - \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}}\sqrt{\frac{2^{l(1-1/p)}}{n}}\right) - \sqrt{\frac{C^{(p)}D^{(D)}\Delta}{C}}^{(p)}\frac{2^{(p-1)(j+l)/2}}{n^p}\right\} \leq \Delta.$$

By defintion of τ_l (since C' is large enough), we know that

$$D^{(m)}\tau_l - \sqrt{C^{(p)}D^{(D)}\frac{2^{l(1-1/p)}}{n\Delta}} \geq 3\tau_l/4.$$

So by definition of $(t_n(l))^p$, we have

$$\frac{3}{4}(\tau_l)^p - \sqrt{\frac{C^{(p)}D^{(D)}}{\Delta}\frac{2^{(p-1)(j+l)/2}}{n^p}} > \frac{1}{2}(\tau_l)^p > t_n(l).$$

Case 2: $\inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p > 3\rho_n/8$

By triangular inequality, for any $g \in \Sigma(s, B)$

$$\|\Pi_{V_{j_s}}(\hat{f}_n) - g\|_p \geq \|\Pi_{V_{j_s}}(f) - g\|_p - \|\Pi_{V_{j_s}}(\hat{f}_n) - f\|_p,$$

which implies when combined with Lemma 4.2

$$\Pr\left\{\|\Pi_{V_{j_s}}(\hat{f}_n) - g\|_p\right.$$

$$\left. \leq \|\Pi_{V_{j_s}}(f) - g\|_p - (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}}\right\} \leq \Delta,$$

which implies since $\inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(f) - g\|_p > 3\rho_n/8$

$$\Pr\left\{\inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(\hat{f}_n) - g\|_p\right.$$

$$\left. \leq 3\rho_n/8 - (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}}\right\} \leq \Delta,$$

so since $\tilde{T}_n = \inf_{g \in \Sigma(s,B)} \|\Pi_{V_{j_s}}(\hat{f}_n) - g\|_p$,

$$\Pr\left\{\tilde{T}_n \leq 3\rho_n/8 - (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}}\right\} \leq \Delta,$$

and by definition we have $3\rho_n/8 - (D_p + 2C_{p/(p-1)}\sqrt{\log(1/\Delta)})\sqrt{\frac{2^{j_s}}{n}} \geq \tilde{t}_n$.

So with probability $1 - 2\Delta$, we have $\Psi_n = 1$ under H_1.

Conclusion on the test Ψ_n

All the inequalities developed earlier are true for any f in H_0 or H_1 with *universal* constants (independent of f) and the supremum over f in H_0 and H_1 of the error of type one and two are bounded by $2\Delta = \alpha/2$. Finally, the test Ψ_n of error of type 1 and 2 bounded by $2\Delta = \alpha/2$ distinguishes between H_0 and H_1 with condition $\rho_n = C(B+1)(n^{-\frac{s}{2s+1}} + n^{-\frac{t}{2t+1-1/p}})$ for a value C large enough (but depending only on p, α). This implies that for any $n > 0$ we have

$$\sup_{f \in \Sigma(s,B)} \mathbf{E}_f \Psi_n + \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbf{E}_f(1 - \Psi_n) \leq \alpha. \tag{4.14}$$

The test is α-consistent (see (Ingster and Suslina, 2002) for a definition).

5. Proof of Theorem 3.6 (upper bound)

Let $1 \geq \Delta > 0$. We know that for n and C (in the definition of ρ_n) large enough (depending only on p, B, Δ), there exists a test Ψ_n for the testing problem (3.3) that is consistent and with level Δ (see Theorem 3.4).

Set $U_p := U_p(B)$ where $U_p(B)$ is the constant in Theorem 3.1. Consider the confidence set around the adaptive estimate $\hat{f}_n(\hat{j}_n)$ (where $\hat{f}_n(\hat{j}_n)$ is constructed as in Theorem 3.1) as being

$$C_n = \left\{g : \|\hat{f}_n(\hat{j}_n) - g\|_p \leq \frac{1}{\Delta}U_p n^{-\frac{s}{2s+1}}\mathbb{I}\{\Psi_n = 0\} + \frac{1}{\Delta}U_p n^{-\frac{t}{2t+1}}\mathbb{I}\{\Psi_n = 1\}\right\}.$$

Then since the test Ψ_n is Δ-consistent

$$\sup_{f \in \Sigma(s,B)} \Pr_f\left(|C_n| > \frac{1}{\Delta} U_p n^{-\frac{s}{2s+1}}\right) = \sup_{f \in \Sigma(s,B)} \mathbf{E}_f \Psi_n \leq \Delta,$$

and

$$\sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \Pr_f\left(|C_n| > \frac{1}{\Delta} U_p n^{-\frac{t}{2t+1}}\right) = 0.$$

Also we have by Markov's inequality

$$\sup_{f \in \Sigma(s,B)} \Pr_f\left(f \in C_n\right) \geq 1 - \sup_{f \in \Sigma(s,B)} \Pr_f\left(\|\hat{f}_n(\hat{j}_n) - f\|_p \geq \frac{1}{\Delta} U_p n^{-\frac{s}{2s+1}}\right)$$

$$\geq 1 - \frac{\Delta n^{\frac{s}{2s+1}}}{U_p} \sup_{f \in \Sigma(s,B)} \mathbf{E}\|\hat{f}_n(\hat{j}_n) - f\|_p$$

$$\geq 1 - \frac{\Delta n^{\frac{s}{2s+1}}}{U_p} U_p n^{-\frac{s}{2s+1}}$$

$$\geq 1 - \Delta,$$

where we use Theorem 3.1 for the bound on $\mathbf{E}\|\hat{f}_n(\hat{j}_n) - f\|_p$, and we also have still by Markov's inequality

$$\sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \Pr_f\left(f \in C_n\right) \geq 1 - \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \Pr_f\left(\|\hat{f}_n(\hat{j}_n) - f\|_p \geq \frac{1}{\Delta} U_p n^{-\frac{t}{2t+1}}\right)$$

$$- \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbf{E}_f(1 - \Psi_n)$$

$$\geq 1 - \frac{\Delta n^{\frac{t}{2t+1}}}{U_p} \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbf{E}\|\hat{f}_n(\hat{j}_n) - f\|_p$$

$$- \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbf{E}_f(1 - \Psi_n)$$

$$\geq 1 - \frac{\Delta n^{\frac{t}{2t+1}}}{U_p} U_p n^{-\frac{t}{2t+1}} - \Delta$$

$$\geq 1 - 2\Delta.$$

These four inequalities imply that C_n is an $(L_p, 2\Delta)$-honest and adaptive confidence bound on $\Sigma(s,B) \cup \tilde{\Sigma}(t,B,\rho_n)$ for $\{s,t\}$ and B.

6. Proof of Theorem 3.5

Let Ψ_n be a test as defined in Theorem 3.4, and

$$\rho_n = C(B+1)\max(n^{-t/(2t+1-1/p)}, n^{-s/(2s+1)}) = C(B+1)n^{-s/(2s+1)},$$

where C defined as in Theorem 3.4. The confidence set we consider is the following.

$$C_n = \left\{ g \in \Sigma(t,B) : \|\tilde{f}_n - g\|_p \leq \frac{D}{\alpha} n^{-\frac{s}{2s+1}}(1 - \Psi_n) + \frac{D}{\alpha} n^{-\frac{t}{2t+1}} \Psi_n \right\},$$

where $\tilde{f}_n$ is the adaptive estimate of Theorem 10.1 (which is actually the same estimate than the one for Theorem 3.1) and where $D \geq \tilde{U}_p(B, 2C)$ where $\tilde{U}_p(B, 2C)$ is the constant defined in Theorem 10.1.

By Markov's inequality

$$\sup_{f \in \Sigma(t,B)} \mathbb{P}(f \notin C_n) \leq \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{P}_f(f \notin C_n) + \sup_{f \in \Sigma(t,B) \setminus \tilde{\Sigma}(t,B,\rho_n)} \mathbb{P}_f(f \notin C_n)$$

$$\leq \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{P}_f\left(\|\tilde{f}_n - f\|_p \geq \frac{D}{\alpha} n^{-\frac{t}{2t+1}}\right) + \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{E}_f[1 - \Psi_n]$$

$$+ \sup_{f \in \Sigma(t,B): \|f - \Sigma(s,B)\|_p \leq C(B+1)n^{-s/(2s+1)}} \mathbb{P}_f\left(\|\tilde{f}_n - f\|_p \geq \frac{D}{\alpha} n^{-\frac{s}{2s+1}}\right)$$

$$\leq \sup_{f \in \tilde{\Sigma}(t,B,\rho_n)} \mathbb{E}_f(\|\tilde{f}_n - f\|_p) \frac{\alpha}{D} n^{\frac{t}{2t+1}} + \alpha$$

$$+ \sup_{f \in \Sigma(t,B): \|f - \Sigma(s,B)\|_p \leq C(B+1)n^{-s/(2s+1)}} \mathbb{E}_f(\|\tilde{f}_n - f\|_p) \frac{\alpha}{D} n^{\frac{s}{2s+1}})$$

$$\leq 3\alpha,$$

by Theorem 3.4 (since $\rho_n = C(B+1)n^{-s/(2s+1)}$) and by Theorem 10.1 (since $D \geq \tilde{U}_p(B, 2C)$). Also it holds that

$$\sup_{f \in \Sigma(s,B)} \mathbb{P}\left[|C_n| \geq \frac{D}{\alpha} n^{-\frac{s}{2s+1}}\right] \leq \alpha,$$

by definition of Ψ_n and by Theorem 3.4, and

$$|C_n| \leq \frac{D}{\alpha} n^{-\frac{t}{2t+1}}.$$

This concludes the proof.

7. Proof of Theorem 3.4 (lower bound)

Let $1 > \upsilon > 0$, and $j \in \mathbb{N}^*$ such that $2^j \approx n^{\frac{1}{2t+1-1/p}}$.

Step 1: Definition of a testing problem.

We define the following prior $\tilde{\Pi}$ on for a sequence $(\alpha_{l,k})_{l \geq J_0, k \in Z_l}$:

$$\alpha \sim \tilde{\Pi} \Leftrightarrow \forall l \neq j, \forall k \in Z_l, \alpha_{l,k} = 0, \forall k \in Z_j, \alpha_{l,k} = B_k,$$

where the B_k are i.i.d. Bernoulli of parameter $2^{-j/2}$.

Consider the sequence of coefficients indexed by a given $\alpha \in I$ as

$$a_{l,k}^{(\alpha)} = va\alpha_{l,k},$$

where $a = 2^{-j(t+1/2-1/(2p))} = n^{-1/2}$. Consider $f^{(\alpha)}$ the function associated to $a^{(\alpha)}$, i.e.

$$f^{(\alpha)} = \sum_{l \geq J_0} \sum_{k \in Z_l} a_{l,k}^{(\alpha)}.$$

We write by a slight abuse of notations that $f \sim \tilde{\Pi}$ if $f = f^{(\alpha)}$ where $\alpha \sim \tilde{\Pi}$.

Consider the testing problem

$$H_0 : f = 0 \quad vs. \quad H_1 : f \sim \tilde{\Pi}. \tag{7.1}$$

Step 2: Quantity of interest.

Let Ψ be a test, that is to say a measurable function that takes values in $\{0,1\}$. Equivalently to having access to the process $Y^{(n)}$, we have access to the coefficients $(\hat{a}_{l,k})_{l,k}$ and each of these coefficients are independent $\mathcal{N}(a_{l,k}, 1/n)$.

We have for any $\eta > 0$

$$\mathbf{E}_0[\Psi] + \mathbf{E}_{\alpha \sim \tilde{\Pi}} \mathbf{E}_{f^{(\alpha)}}[1 - \Psi] \geq \mathbf{E}_0[\Psi] + \mathbf{E}_{\alpha \sim \tilde{\Pi}} \mathbf{E}_{f^{(\alpha)}}[1 - \Psi]$$
$$\geq \mathbf{E}_0\Big[\mathbb{I}\{\Psi = 1\}] + \mathbb{I}\{\Psi = 0\} Z\Big]$$
$$\geq (1 - \eta)\mathbb{P}_0(Z \geq 1 - \eta), \tag{7.2}$$

where $Z = \mathbf{E}_{\alpha \sim \tilde{\Pi}} \prod_{l,k} \frac{dP_{l,k}^{(\alpha)}}{dP_{l,k}^{(0)}}$, where $dP_{l,k}^{(\alpha)}$ is the distribution of $\hat{a}_{l,k}$ when the function generating the data is $f^{(\alpha)}$, and $dP_{l,k}^{(0)}$ is the distribution of $\hat{a}_{l,k}$ when the function generating the data is 0 (this holds since the $(\hat{a}_{l,k})_{l,k}$ are independent).

More precisely, we have since the $(\hat{a}_{l,k})_{l,k}$ are independent $\mathcal{N}(a_{l,k}, 1/n)$

$$Z((x_k)_k) = \mathbf{E}_{\alpha \sim \tilde{\Pi}}\left[\prod_{l>0, k \in Z_l} \frac{\exp(-\frac{n}{2}(x_{l,k} - a_{l,k}^{(\alpha)})^2)}{\exp(-\frac{n}{2}x_{l,k}^2)}\right]$$
$$= \mathbf{E}_{\alpha \sim \tilde{\Pi}}\left[\prod_{k \in Z_j} \exp(nx_k a_k^{(\alpha)}) \exp\left(-\frac{n}{2}(a_k^{(\alpha)})^2\right)\right],$$

where we simplify notations by setting $x_k \equiv x_{j,k}$ and $a_k^{(\alpha)} \equiv a_{j,k}^{(\alpha)}$. We also write later $\alpha_k \equiv \alpha_{j,k}$.

By Markov and Cauchy Schwarz's inequality

$$\mathbb{P}_0(Z \geq 1 - \eta) \geq 1 - \frac{\mathbf{E}_0|Z - 1|}{\eta} \geq 1 - \frac{\sqrt{\mathbf{E}_0(Z - 1)^2}}{\eta}. \tag{7.3}$$

We have by definition of Z

$$\mathbf{E}_0\big[(Z-1)^2\big]$$

$$= \int_{x_1,\dots,x_{2^j}} \left(\mathbf{E}_{\alpha\sim\tilde\Pi}\Big[\prod_k \exp(x_k n a_k^{(\alpha)})\exp\Big(-\frac{n}{2}(a_k^{(\alpha)})^2\Big)\Big]-1\right)^2$$

$$\prod_k \frac{1}{\sqrt{2n\pi}}\exp\Big(-\frac{n}{2}(x_k)^2\Big)dx_1\cdots x_{2^j}$$

$$= \mathbf{E}_{\alpha,\alpha'\sim\tilde\Pi}\left[\prod_k\int_{x_k}\exp(x_k n(a_k^{(\alpha)}+a_k^{(\alpha')}))\exp\Big(-\frac{n}{2}((a_k^{(\alpha)})^2+(a_k^{(\alpha')})^2)\Big)\right.$$

$$\left.\frac{1}{\sqrt{2n\pi}}\exp\Big(-\frac{n}{2}(x_k)^2\Big)dx_k\right]-1$$

$$= \mathbf{E}_{\alpha,\alpha'\sim\tilde\Pi}\left[\prod_k\Big((1-\mathbb{I}\{\alpha_k=\alpha'_k=1\})+\exp(nv^2a^2)\mathbb{I}\{\alpha_k=\alpha'_k=1\}\Big)\right]-1$$

$$= \mathbf{E}_{\alpha,\alpha'\sim\tilde\Pi}\left[\prod_k\Big(1+(\exp(nv^2a^2)-1)\mathbb{I}\{\alpha_k=\alpha'_k=1\}\Big)\right]-1. \qquad (7.4)$$

Since all $(\alpha_k)_k, (\alpha'_k)_k$ are i.i.d. Bernoulli of parameter $2^{-j/2}$, it implies that the $(\mathbb{I}\{\alpha_k=\alpha'_k=1\})_k$ are i.i.d. Bernoulli random variables of parameter 2^{-j}. This implies together with Equation (7.4) that

$$\mathbf{E}_0\big[(Z-1)^2\big] = \left[\prod_k\mathbf{E}_{B'_k\sim\mathcal{B}(2^{-j})}\Big(1+(\exp(nv^2a^2)-1)B'_k\Big)\right]-1$$

$$= \Big(1+(\exp(nv^2a^2)-1)\frac{1}{2^j}\Big)^{2^j}-1$$

$$\leq \exp((\exp(nv^2a^2)-1))-1.$$

where $\mathcal{B}(2^{-j})$ is the law of a Bernoulli of parameter 2^{-j}, and since for any $u\geq 0$, $1+u\leq\exp(u)$. Since $a^2=n^{-1}$, we get

$$\mathbf{E}_0\big[(Z-1)^2\big] \leq \exp((\exp(v^2)-1))-1$$

$$\leq \exp(2v^2)-1\leq 4v^2, \qquad (7.5)$$

since for $u\leq 1$, we have $\exp(u)\leq 1+2u$.

Step 3: Conclusion on the test 7.1

By combining Equations (7.2), (7.3), and 7.5 we know that

$$\mathbf{E}_0[\Psi]+\mathbf{E}_{\alpha\sim\tilde\Pi}\mathbf{E}_{f^{(\alpha)}}[1-\Psi]\geq 1-4v^2,$$

and since this holds with any Ψ, we have

$$\inf_\Psi\Big[\mathbf{E}_0[\Psi]+\mathbf{E}_{\alpha\sim\tilde\Pi}\mathbf{E}_{f^{(\alpha)}}[1-\Psi]\Big]\geq 1-4v^2, \qquad (7.6)$$

and this implies that there is no $1-4v^2$-consistent test for the testing problem (7.1).

Step 4: Extension of this result to a deterministic testing problem.
Define the set

$$I = \left\{ (\alpha_{l,k})_{l,k} : \forall l \neq j, \alpha_{l,k} = 0, \alpha_{j,k} \in \{0,1\}, \sum_{k \in Z_j} \alpha_{j,k} = S, \frac{2^{j/2}}{2} \leq S \leq \frac{3}{2} 2^{j/2} \right\}.$$

Consider the associated sequence of coefficients indexed by $\alpha \in I$, and the corresponding function $f^{(\alpha)}$. Consider the testing problem

$$H_0 : f = 0 \quad vs. \quad H_1 : f = f^{(\alpha)}, \alpha \in I. \tag{7.7}$$

Consider now $\alpha \sim \tilde{\Pi}$. By Hoeffding's inequality, we know that for $\lambda \leq 2^{j/2}$, we have

$$\mathrm{Pr}_{\alpha \sim \tilde{\Pi}} \left(\left| \sum_k \alpha_{j,k} - 2^{j/2} \right| \geq \lambda \right) \leq 2 \exp(-2^{-j/2} \lambda^2 / 2).$$

Let $\lambda = \frac{2^{j/2}}{2}$. Then the last equation implies

$$\mathrm{Pr}_{\alpha \sim \tilde{\Pi}} \left(\left| \sum_k \alpha_{j,k} - 2^{j/2} \right| \geq \frac{2^{j/2}}{2} \right) \leq 2 \exp(-2^{j/2} / 8),$$

so this implies in particular that with $\tilde{\Pi}$-probability larger than $1 - 2\exp(-2^{j/2}/8)$, we have $\alpha \in I$.

For any test Ψ, since $\Psi \leq 1$, we have

$$\mathbf{E}_{\alpha \sim \tilde{\Pi}} \mathbf{E}_{f^{(\alpha)}} [1 - \Psi] \leq \mathrm{Pr}_{\alpha \sim \tilde{\Pi}} (\alpha \notin I) + \sup_{\alpha \in I} \mathbf{E}_{f^{(\alpha)}} [1 - \Psi].$$

This implies since with $\tilde{\Pi}$-probability larger than $1 - 2\exp(-2^{j/2}/8)$, $\alpha \in I$, that

$$\mathbf{E}_{\alpha \sim \tilde{\Pi}} \mathbf{E}_{f^{(\alpha)}} [1 - \Psi] \leq \sup_{\alpha \in I} \mathbf{E}_{f^{(\alpha)}} [1 - \Psi] + 2\exp(-2^{j/2}/8).$$

This implies when combined to Equation (7.6) that

$$\inf_{\Psi} \left[\mathbf{E}_0[\Psi] + \sup_{\alpha \in I} \mathbf{E}_{f^{(\alpha)}} [1 - \Psi] \right] \geq 1 - 4\upsilon^2 - 2\exp(-2^{j/2}/8)$$

$$\leq 1 - 5\upsilon^2,$$

for n and thus j large enough. This implies that there is no $1 - 5\upsilon^2$-consistent test for the testing problem (7.7), and this holds for any $\upsilon > 0$.

8. Proof of Theorem 3.6 (lower bound)

Consider all the quantities defined in Section 7.
   Assume that $s(1 - 1/p) > t$. Let $\upsilon \leq B/2$. Set

$$\rho_n = C_p \frac{\upsilon n^{-\frac{t}{2t+1-1/p}}}{4}.$$

Let $\alpha \in I$.

By triangular inequality,

$$\|f^{(\alpha)}\|_{t,p,\infty} = \Big(\sum_k (a_{j,k}^{(\alpha)})^p 2^{jp(1/2-1/p)} \Big)^{1/p} = \frac{3}{2}\upsilon \leq B,$$

which implies that $f^{(\alpha)} \in \Sigma(t,B)$. Also that since only the $j-th$ first coefficients of f are non-zero, and since $\|.\|_p \geq C_p\|.\|_{0,p,p}$ (see Proposition 2 in the supplementary Material), we have

$$\begin{aligned}
\|f^{(\alpha)} - \Sigma(s,B)\|_p &\geq C_p\|f^{(\alpha)} - \Sigma(s,B)\|_{0,p,p} \\
&= C_p \min_{g \in \Sigma(s,B)} \|\Pi_{W_j}(f^{(\alpha)}) - \Pi_{W_j}(g)\|_{0,p,p} \\
&\geq C_p\|\Pi_{W_j}f^{(\alpha)}\|_{0,p,p}/2 - C_p B 2^{-js} \\
&\geq C_p \frac{\upsilon n^{-\frac{t}{2t+1-1/p}}}{4}.
\end{aligned}$$

by triangular inequality since for any $g \in \Sigma(s,B), \|g\|_{0,p,p} \leq B 2^{-js} \leq \upsilon n^{-\frac{t}{2t+1-1/p}}/4$ for n large enough, since $s(1-1/p) > t$. This implies in particular that $f^{(\alpha)} \in \tilde{\Sigma}(t,B,\rho_n)$

To sum up,

$$0 \in \Sigma(s,B) \quad and \quad \forall \alpha, f^{(\alpha)} \in \tilde{\Sigma}(t,B,\rho_n).$$

Assume that there exists some honest and adaptive confidence set C_n for $\mathcal{P}_n = \Sigma(s,B) \cup \Sigma(t,B,\rho_n)$, and $\{s,t\}$.

This implies that the confidence set C_n is in particular honest and adaptive over $\mathcal{P}'_n = \{0\} \cap \{f^{(\alpha)}, \alpha \in I\}$. So, for any $0 < \delta < 1$, there exists a constant L (that might depend on B,s,t,δ) such that for $n > 0$

$$\mathrm{Pr}_0\Big(\{|C_n| \geq Ln^{-s/(2s+1)}\} \cup \{0 \notin C_n\}\Big) \leq 2\delta \quad and \quad \sup_{\alpha \in I} \mathrm{Pr}_{f^{(\alpha)}}\big(f^{(\alpha)} \in C_n\big) \geq 1-\delta.$$

We define a test Ψ as follows. If $0 \in C_n$ and $|C_n| \leq Ln^{-s/(2s+1)}$, then $\Psi = 0$, otherwise $\Psi = 1$. We have for $n > 0$

$$\mathrm{Pr}_0(\Psi = 1) = \mathrm{Pr}_0\Big(\{|C_n| \geq Ln^{-s/(2s+1)}\} \cup \{0 \notin C_n\}\Big) \leq 2\delta.$$

Also

$$\begin{aligned}
\sup_{\alpha \in I} \mathrm{Pr}_{f^{(\alpha)}}(\Psi = 0) &\leq \sup_{\alpha \in I} \mathrm{Pr}_{f^{(\alpha)}}\Big(\|f^{(\alpha)} - 0\|_p \leq Ln^{-s/(2s+1)}\Big) \\
&\geq \sup_{\alpha \in I} \mathrm{Pr}_{f^{(\alpha)}}\Big(f^{(\alpha)} \notin C_n\Big) \geq \delta.
\end{aligned}$$

Combining both results imply that for n large enough, there exists a consistent test Ψ constructed using C_n, that is to say such that

$$\inf_\Psi \mathbf{E}_0[\Psi] + \sup_{f^{(\alpha)},\alpha \in I} \mathbf{E}_{f^{(\alpha)}}[1 - \Psi] \leq 3\delta,$$

and that for any $\delta > 0$. This is in contradiction with the result of Step 4 (no consistent test for the testing problem 7.7), and we deduce by contradiction that no honest and adaptive confidence set exists on $\mathcal{P}_n$. This concludes the proof.

Supplementary material

9. Technical preliminary results

In this Section, we remind some well-known preliminary results, which we sometimes extend, or adapt.

We first provide the following Assumption.

Assumption 2. *We assume that there is a universal constant C_p such that for any (l, k) we have*

$$\|\psi_{l,k}\|_p \le C_p 2^{l/2},$$

and

$$\|\phi_k\|_p \le C_p.$$

Note that Assumption 1 implies Assumption 2.

9.1. Properties of Besov spaces

We remind the following Proposition (see (Bergh and Löfström, 1976) or (Besov et al., 1978), volume 2, Chapter 18, page 68)

Proposition 2. *Assume that $p \ge 2$ (and $p < \infty$). Then*

$$B_{0,p,2} \subset L_p \subset B_{0,p,p}.$$

If $p' \le 2$, then we have

$$B_{0,p',p'} \subset L_{p'} \subset B_{0,p',2}.$$

We also remind the following Proposition (see also (Härdle et al., 1998)).

Proposition 3. *Let $s > 0$, $p \ge 2$ (and $p < \infty$) and $h \ge 2$. Assume that $f \in B_{s,p,\infty}$. Then*

$$\|f - \Pi_{V_j} f\|_{0,p,h} \le \|f\|_{s,p,\infty} 2^{-js},$$

and also

$$\|f - \Pi_{V_j} f\|_p \le C_p \|f\|_{s,p,\infty} 2^{-js}.$$

Note that this is also satisfied for the weaker condition $p' \ge 1$ and $h \ge 1$ (by just remarking that $L_{p'} \subset B_{0,p',1}$ also for any $p' \ge 1$).

Proof. We have

$$\|f - \Pi_{V_j} f\|_{0,p,h} = \left(\sum_{l \ge j} 2^{lh(1/2 - 1/p)} |\langle f, \psi_{l,.}\rangle|_{l_p}^h \right)^{1/h}$$

$$\le 2^{-js} \left(\sum_{l \ge j} 2^{lh(s + 1/2 - 1/p)} |\langle f, \psi_{l,.}\rangle|_{l_p}^h \right)^{1/h}$$

$$\le \|f\|_{s,p,\infty} 2^{-js}.$$

For the $\|.\|_p$ norm, it comes from the fact that there exists a constant C_p such that $\|f\|_p \le C_p \|f\|_{0,p,2}$ for any f (Proposition 2). $\square$

9.2. Behaviour of thresholded wavelet estimates

We also remind Rosenthal's inequality (see (Härdle et al., 1998), page 132)

Proposition 4. *Let $(X_1, \ldots, X_n)$ be n i.i.d. random variables such that $\mathbf{E}X_i = 0$ and for a given $p \geq 2$ (and $p < \infty$), $\mathbf{E}|X_i|^p < \infty$. Then there exists a universal constant $\tilde{C}_p^p$ such that*

$$\mathbf{E}\Big|\sum_{i=1}^n X_i\Big|^p \leq \tilde{C}_p^p\Big(\sum_{i=1}^n \mathbf{E}|X_i|^p + \Big(\sum_{i=1}^n \mathbf{E}X_i^2\Big)^{p/2}\Big)$$

We remind the following Proposition (see (Giné and Nickl, 2011), and here we provide an alternative proof).

Proposition 5. *If Assumption 2 is satisfied, there exists a universal constant D_p that depends on p only such that for any fixed $j \in \mathbb{N}^*$ we have*

$$\mathbf{E}\|\Pi_{V_j}\hat{f}_n - \Pi_{V_j}f\|_p^p \leq D_p^p \frac{2^{jp/2}}{n^{p/2}}.$$

Proof. Let $x \in [0, 1]$. We have

$$\mathbf{E}\Big|\Pi_{V_j}\hat{f}_n(x) - \Pi_{V_j}f(x)\Big|^p = \mathbf{E}\Big|\sum_k \frac{1}{\sqrt{n}}G_k'\phi_k(x) + \sum_{l\leq j,k}\frac{1}{\sqrt{n}}G_{l,k}\psi_{l,k}(x)\Big|^p,$$

where $G_{l,k} = \sqrt{n}(\hat{a}_{l,k} - a_{l,k})$ and $G_k' = \sqrt{n}(\hat{a}_k' - a_k')$ and the $(G_{k'}', G_{l,k})_{k',l,k}$ are thus i.i.d. gaussian random variables of mean 0 and variance 1. In order to simplify the notations, we abuse notations and set $\psi_{-1,k} = \phi_k$ and $G_{-1,k} = G_k'$.

We use Rosenthal's inequality (Proposition 4), and obtain

$$\mathbf{E}\Big|\Pi_{V_j}\hat{f}_n(x) - \Pi_{V_j}f(x)\Big|^p = \mathbf{E}\Big|\sum_{-1\leq l\leq j,k}\frac{1}{\sqrt{n}}G_{l,k}\psi_{l,k}(x)\Big|^p$$

$$\leq \tilde{C}_p^p\Bigg(\sum_{-1\leq l\leq j,k}\frac{1}{n^{p/2}}\mathbf{E}|G_{l,k}|^p|\psi_{l,k}(x)|^p$$

$$+ \Big(\sum_{-1\leq l\leq j,k}\frac{1}{n}\mathbf{E}G_{l,k}^2\psi_{l,k}(x)^2\Big)^{p/2}\Bigg)$$

$$= \tilde{C}_p^p\Bigg(\sum_{-1\leq l\leq j,k}\frac{1}{n^{p/2}}c_p^p|\psi_{l,k}(x)|^p + \Big(\frac{1}{n}\sum_{-1\leq l\leq j,k}\psi_{l,k}(x)^2\Big)^{p/2}\Bigg)$$

$$\leq \tilde{C}_p^p\Bigg(\sum_{-1\leq l\leq j,k}\frac{1}{n^{p/2}}c_p^p|\psi_{l,k}(x)|^p + \frac{C_p^p 2^{jp/2}}{n^{p/2}}\Bigg).$$

where c_p^p is the p-th moment of a $\mathcal{N}(0,1)$, and since Assumption 2 is satisfied.

By Assumption 2, we know that for any (l, k) we have

$$\int |\psi_{l,k}(x)|^p dx \leq C_p^p 2^{lp(1/2-1/p)},$$

and thus we have that

$$\int \mathbf{E}\Big|\Pi_{V_j}\hat{f}_n(x) - \Pi_{V_j}f(x)\Big|^p dx$$

$$\leq \int \tilde{C}_p^p \left(\sum_{-1\leq l\leq j,k} \frac{1}{n^{p/2}}c_p^p|\psi_{l,k}(x)|^p + \frac{C_p^p 2^{jp/2}}{n^{p/2}} \right) dx$$

$$\leq \tilde{C}_p^p \left(\sum_{-1\leq l\leq j,k} \frac{1}{n^{p/2}}c_p^p C_p^p 2^{lp(1/2-1/p)} + \frac{C_p^p 2^{jp/2}}{n^{p/2}} \right)$$

$$\leq \tilde{C}_p^p (c_p^p C_p^p + C_p^p) \frac{2^{jp/2}}{n^{p/2}}.$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

We finally state the following Proposition (it is a generalisation of what is done in (Giné and Nickl, 2011)).

Proposition 6. *Let* $2 \leq p \leq \infty$ *and* $2 \leq h \leq \infty$. *Then*

$$\Pr \left\{ \left| \|\Pi_{V_j}(\hat{f}_n - f)\|_{0,p,h} - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_{0,p,h}] \right| \right.$$

$$\left. \geq 2\sqrt{\log(1/\delta)\frac{2^{j(1-2/p)}}{n}} \right\} \leq \delta,$$

and in particular, this implies

$$\Pr \left\{ \left| \|\Pi_{V_j}(\hat{f}_n - f)\|_p - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_p] \right| \right.$$

$$\left. \geq 2C_{p/(p-1)}\sqrt{\log(1/\delta)\frac{2^{j(1-2/p)}}{n}} \right\} \leq \delta,$$

Proof. We first remind Borell's inequality:

Theorem 9.1 (Borell's inequality). *Let* $(G(t))_{t\in T}$ *be a centred Gaussian process indexed by a countable set* T *such that* $\sup_{t\in T} G(t) < +\infty$ *almost surely. Then* $\mathbf{E}\sup_{t\in T} G(t) < +\infty$ *and for every* $r \geq 0$, *we have*

$$\Pr\left(\left| \sup_{t\in T} G(t) - \mathbf{E}\sup_{t\in T} G(t) \right| \geq r \right) \leq 2\exp(-r^2/2\sigma^2),$$

where $\sigma^2 = \sup_{t\in T} \mathbf{E}G^2(t) < +\infty$.

We use the separability of the ball of radius 1 of $B_{0,p/(p-1),h/(h-1)}$ (that we write B_0 to prove that by Borell's inequality (since $(\Pi_{V_j}(\hat{f}_n - f) = \sum_{l\leq j,k}(\hat{a}_{l,k} - a_{l,k})\psi_{l,k}(x))_{x\in[0,1]}$ is a centered Gaussian process, and thus $(\langle \Pi_{V_j}(\hat{f}_n - f), g\rangle)_{g\in B_0}$ is a centred Gaussian process):

$$\Pr \left\{ \left| \sup_{g\in B_0} \langle \Pi_{V_j}(\hat{f}_n - f), g\rangle - \mathbf{E}[\sup_{h\in B_0} \langle \Pi_{V_j}(\hat{f}_n - f), g\rangle] \right| \geq 2\sqrt{\log(1/\delta)}\sigma \right\} \leq \delta,$$

where $\sigma^2 \leq \sup_{g\in B_0} \mathbf{E}[\langle \Pi_{V_j}(\hat{f}_n - f), g\rangle^2]$.

Note first that by Hahn-Banach's duality Theorem (since $\|.\|_{0,p,h}$ is the norm corresponding to B_0), we have that

$$\sup_{g \in B_0} \langle \Pi_{V_j}(\hat{f}_n - f), g \rangle = \|\Pi_{V_j}(\hat{f}_n - f)\|_{0,p,h},$$

and we can rewrite the previous equation as

$$\Pr\left\{ \left| \|\Pi_{V_j}(\hat{f}_n - f)\|_{0,p,h} - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_{0,p,h}] \right| \geq 2\sqrt{\log(1/\delta)}\sigma \right\} \leq \delta.$$

Concerning σ^2, we have (since the $(\hat{a}_{l,k} - a_{l,k})$ are independent centered Gaussian of variance $1/n$)

$$
\begin{aligned}
\sup_{u \in B_0} \mathbf{E}[\langle \Pi_{V_j}(\hat{f}_n - f), u \rangle^2] &= \sup_{u \in B_0} \mathbf{E}\left[\left\langle \sum_{l \leq j,k} (\hat{a}_{l,k} - a_{l,k})\psi_{l,k}, u \right\rangle^2 \right] \\
&= \sup_{u \in B_0} \mathbf{E}\left[\left(\sum_{l \leq j,k} (\hat{a}_{l,k} - a_{l,k})\langle \psi_{l,k}, u \rangle \right)^2 \right] \\
&= \sup_{u \in B_0} \sum_{l \leq j,k} \frac{1}{n}\left(<\psi_{l,k}, u> \right)^2 \\
&= \sup_{u \in B_0} \frac{1}{n}\|\Pi_{V_j}u\|_2^2.
\end{aligned}
$$

We are thus interested in computing $\sup_{u \in B_0} \|\Pi_{V_j}u\|_2^2$, i.e. the maximum squared L_2 norm of a vector of $\|.\|_{0,p/(p-1),h/(h-1)}$ norm of 1. We have by Plancherel's theorem

$$\sup_{u \in B_0} \|\Pi_{V_j}u\|_2^2 = \sup_{u \in B_0} \sum_{l \leq j,k} u_{l,k}^2.$$

Let us consider $u \in B_0$. We have, since $p \geq 2$ (and $p < \infty$)

$$\sum_k u_{l,k}^2 \leq |u_{l,.}|_{p/(p-1)}^2.$$

Also, since $h \geq 2$

$$
\begin{aligned}
\sum_{l \leq j} |u_{l,.}|_{p/(p-1)}^2 &= \sum_{l \leq j} |u_{l,.}|_{p/(p-1)}^2 2^{2l(1/2-(p-1)/p)} 2^{-2l(1/2-(p-1)/p)} \\
&\leq \left(\sum_{l \leq j} |u_{l,.}|_{p/(p-1)}^{\frac{h}{h-1}} 2^{l(1/2-(p-1)/p)\frac{h}{h-1}} \right)^{(h-1)/h} \Big)^2 2^{2j(1/2-1/p)} \\
&\leq \|\Pi_{V_j}u\|_{0,p/(p-1),h/(h-1)}^2 2^{j(1-2/p)}.
\end{aligned}
$$

When putting all this together, we obtain finally

$$\sigma^2 \leq \frac{1}{n} \sup_{u \in B_0} \|\Pi_{V_j}u\|_2^2 \leq \frac{2^{j(1-2/p)}}{n},$$

which concludes the proof for the $\|.\|_{0,p,h}$ norm.

For the $\|.\|_p$ norm, we apply as before Borell's inequality on the ball of radius 1 of $L_{\frac{p}{p-1}}$ (using the fact that it is separable for $p < \infty$, or that $[0,1]$ is separable for $p = \infty$), and also use Hahn Banach's theorem to obtain

$$\Pr\left\{\left|\|\Pi_{V_j}(\hat{f}_n - f)\|_p - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_p]\right| \geq 2\sqrt{\log(1/\delta)}\sigma\right\} \leq \delta,$$

where $\sigma^2 \leq \sup_{u:\|u\|_{p/(p-1)}\leq 1} \mathbf{E}[\langle \Pi_{V_j}(\hat{f}_n - f), u\rangle^2]$. Then we remark that $L_{\frac{p}{p-1}} \subset B_{0,p/(p-1),2}$ (see Proposition 2), which implies that there exists a universal constant $C_{p/(p-1)}$ such that $\|.\|_{\frac{p}{p-1}} \geq C_{p/(p-1)}\|.\|_{0,\frac{p}{p-1},2}$. This implies in particular that $\sup_{u:\|u\|_{p/(p-1)}\leq 1} \|\Pi_{V_j}u\|_2^2 \leq \sup_{u:\|u\|_{0,p/(p-1),2}\leq C_{p/(p-1)}} \|\Pi_{V_j}u\|_2^2$. This in particular implies, using previous results, that $\sigma^2 \frac{C_{p/(p-1)}^2 2^{j(1-2/p)}}{n}$, which leads finally to

$$\Pr\left\{\left|\|\Pi_{V_j}(\hat{f}_n - f)\|_p - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_p]\right|\right.$$
$$\left. \geq 2C_{p/(p-1)}\sqrt{\log(1/\delta)\frac{2^{j(1-2/p)}}{n}}\right\} \leq \delta,$$

$\square$

10. Adaptive estimation

We prove that adaptive estimators exist on sets that are slightly larger than $\Sigma(r, B)$. As a corollary, adaptive estimators exist on $\Sigma(r, B)$ (Theorem 3.1).

Theorem 10.1. *There exists an adaptive estimator $\hat{f}_n(dY^{(n)})$ such that there are two constants $u_p > 0$ and $N_p > 0$ that depend only on p such that for every $B > 0$, every $r > 0$ and every $c_a \geq 0$, we have*

$$\sup_{f\in L_2:\|f-\Sigma(r,B)\|_p\leq c_a n^{-r/(2r+1)}} \mathbf{E}_f\|\hat{f}_n - f\|_p \leq u_p\left((c_a + B)^{1/(2r+1)} + 1\right)n^{-r/(2r+1)}.$$

We can rewrite this as

$$\sup_{r>0} \sup_{f\in L_2:\|f-\Sigma(r,B)\|_p\leq c_a n^{-r/(2r+1)}} \left[\frac{1}{\tilde{U}_p(B,c_a)}n^{r/(2r+1)}\mathbf{E}_f\|\hat{f}_n - f\|_p\right] \leq 1,$$

where $\tilde{U}_p(B, c_a) = u_p((c_a + B) + 2)$.

10.1. Approximation and estimation errors of a thresholded estimator

The wavelet basis we use is the Cohen-Daubechies-Vial wavelet basis (it that satisfies Assumption 2).

We first remind the following Corollary of Proposition 5.

Corollary. *Consider $f \in L_2$. There exists a universal constant D_p that depends on p only such that for any fixed $j \in \mathbb{N}^*$ we have*

$$\mathbf{E}\|\Pi_{V_j}\hat{f}_n - \Pi_{V_j}f\|_p \leq D_p \frac{2^{j/2}}{n^{1/2}} := \sigma(j, n).$$

Proof. Since $p \geq 2$ (and $p < \infty$), we know by convexity that $\mathbf{E}\|.\|_p^p \geq (\mathbf{E}\|.\|_p)^{1/p}$, which concludes the proof together with Proposition 5. $\square$

We state the following Lemma, which is an extension of results in (Härdle et al., 1998; Giné and Nickl, 2011)).

Lemma 10.1. *Let $\epsilon > 0$. Let $f \in L_2$ such that $\|f - \Sigma(r, B)\|_p \leq \epsilon$. There exists a universal constant C_p that depends on p only such that for any fixed $j \in \mathbb{N}^*$ and any $f \in B_{r,p,\infty}$ such that $\|f\|_{r,p,\infty} \leq B$, we have*

$$\|f - \Pi_{V_j}f\|_p \leq 2\epsilon + C_p B 2^{-jr} := B(j, f, \epsilon).$$

Proof. We have

$$\|f - \Pi_{V_j}f\|_p = \inf_{g \in \Sigma(r,B)} \|f - (g - \Pi_{V_j}g) + (g - \Pi_{V_j}g) - \Pi_{V_j}f\|_p$$

$$\leq \inf_{g \in \Sigma(r,B)} \left[\|f - g\|_p + \|g - \Pi_{V_j}g\|_p + \|\Pi_{V_j}(f - g)\|_p \right]$$

$$\leq 2\epsilon + C_p \sup_{g \in \Sigma(r,B)} \|g\|_{r,p,\infty} 2^{-jr} \leq 2\epsilon + C_p B 2^{-jr} := B(j, f, \epsilon).$$

$\square$

10.2. Definition of a Lepski type estimator

Let $c_a > 0$, $r > 0$ and $B > 0$. Let $c > 0$ and $f \in L_2$ such that $\|f - \Sigma(r, B)\|_p \leq \epsilon$, where

$$\epsilon_n \equiv \epsilon = c_a n^{\frac{-r}{2r+1}}.$$

Set for $D'_p = D_p + 2C_{p/(p-1)}$

$$\hat{j}_n = \min \left\{ j \in \mathbb{N} : \|\Pi_{V_j}\hat{f}_n - \Pi_{V_l}\hat{f}_n\|_p \leq 4(D'_p + 1)\frac{2^{l/2}}{n^{1/2}}, \forall l > j, l \in \mathbb{N} \right\}.$$

We consider in the sequel the adaptive Lepski type estimator $\hat{f}_n(\hat{j}_n) = \Pi_{V_{\hat{j}_n}}\hat{f}_n$.

Set now the oracle threshold

$$j^* = j^*(f) = \inf \left\{ j \in \mathbb{N} : B(j, f, \epsilon) \geq \sigma(j, n) \right\}.$$

Note that by Corollary 10.1 and Lemma 10.1 we have

$$B(j^*, f, \epsilon) \leq \sigma(j^*, n) \leq 2\left(\frac{1}{D_p}(C_p B + c_a) \right)^{1/(2r+1)} n^{-r/(2r+1)}. \qquad (10.1)$$

10.3. Bound for the error on the event $\{\hat{j}_n \leq j^*\}$

We have by triangular inequality, Equation (10.1) and the definitions of $\hat{j}_n$ and j^* that (since $D'_p \geq D_p$)

$$
\mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n \leq j^*\Big\}\Big]
$$
$$
\leq \mathbf{E}\Big[\big(\|\hat{f}_n(\hat{j}_n) - \hat{f}_n(j^*)\|_p + \|\hat{f}_n(j^*) - f\|_p\big)\mathbb{I}\Big\{\hat{j}_n \leq j^*\Big\}\Big]
$$
$$
\leq 4(D'_p + 1)\frac{2^{j^*/2}}{n^{1/2}} + \sigma(j^*, n)
$$
$$
\leq 2\Big(\frac{5D'_p + 4}{D_p}\Big)\Big(\frac{1}{D_p}(C_p B + c_a)\Big)^{1/(2r+1)} n^{-r/(2r+1)}.
$$

10.4. Bound for the error on the event $\{\hat{j}_n > j^*\}$

We remind the following Lemma (see (Giné and Nickl, 2011)).

Lemma 10.2. *There is a constant D'_p that depends on p only and such that*

$$
\sup_{f \in L_p} \Pr\Big\{\|\Pi_{V_j}(\hat{f}_n - f)\|_p \geq D'_p\sqrt{\frac{2^j}{n}}\Big\} \leq 2^{-2^{2j/p}},
$$

Proof. Proposition 6 gives us that for the $\|.\|_p$ norm, we have

$$
\Pr\Big\{\big|\|\Pi_{V_j}(\hat{f}_n - f)\|_p - \mathbf{E}[\|\Pi_{V_j}(\hat{f}_n - f)\|_p|
$$
$$
\geq 2C_{p/(p-1)}\sqrt{\log(1/\delta)\frac{2^{j(1-2/p)}}{n}}\Big\} \leq \delta.
$$

By combining this with Corollary 10.1, we get

$$
\Pr\Big\{\|\Pi_{V_j}(\hat{f}_n - f)\|_p \geq D_p\sqrt{\frac{2^j}{n}} + 2C_{p/(p-1)}\sqrt{\log(1/\delta)\frac{2^{j(1-2/p)}}{n}}\Big\} \leq \delta,
$$

which implies by considering δ such that $\log(1/\delta) = 2^{2j/p}$

$$
\Pr\Big\{\|\Pi_{V_j}(\hat{f}_n - f)\|_p \geq (D_p + 2C_{p/(p-1)})\sqrt{\frac{2^j}{n}}\Big\} \leq \exp(-2^{2j/p}).
$$

$\square$

We have by Hölder's inequality that

$$
\mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n > j^*\Big\}\Big] = \sum_{j > j^*} \mathbf{E}\Big[\|f_n(j) - f\|_p \mathbb{I}\Big\{\hat{j}_n = j\Big\}\Big]
$$
$$
\leq \sum_{j > j^*} \big(\mathbf{E}\|f_n(j) - f\|_p^p\big)^{1/p}\big(\mathbf{E}\mathbb{I}\Big\{\hat{j}_n = j\Big\}^{p/(p-1)}\big)^{(p-1)/p},
$$

which implies by Corollary 10.1 that we have

$$\mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n > j^*\Big\}\Big] \leq \sum_{j>j^*} \sigma(j,n)\mathbb{P}(\hat{j}_n = j)^{(p-1)/p}, \qquad (10.2)$$

By an union bound and by definition of $\hat{j}_n$, we remark that

$$\mathbb{P}(\hat{j}_n = j) \leq \sum_{l\geq j} \mathbb{P}\Big(\|\hat{f}_n(j-1) - \hat{f}_n(l)\|_p \geq 4(D_p'+1)\frac{2^{l/2}}{n^{1/2}} = \frac{4(D_p'+1)}{D_p}\sigma(l,n)\Big),$$

$$(10.3)$$

and by triangle inequality we have

$$\|\hat{f}_n(j-1) - \hat{f}_n(l)\|_p$$
$$= \|\hat{f}_n(j-1) - \hat{f}_n(l) - \Pi_{V_{j-1}}f + \Pi_{V_{j-1}}f + \Pi_{V_l}f - \Pi_{V_l}f - f + f\|_p$$
$$\leq \|\hat{f}_n(j-1) - \Pi_{V_{j-1}}f\|_p + \|\hat{f}_n(l) - \Pi_{V_l}f\|_p + \|\Pi_{V_{j-1}}f - f\|_p + \|\Pi_{V_l}f - f\|_p$$
$$\leq \|\hat{f}_n(j-1) - \Pi_{V_j}f\|_p + \|\hat{f}_n(l) - \Pi_{V_l}f\|_p + 2\sigma(l,n),$$

since as $l > j - 1 \geq j^*$, we have $\|\Pi_{V_{j-1}}f - f\|_p + \|\Pi_{V_l}f - f\|_p \leq B(j-1,f,\epsilon) + B(l,f,\epsilon) \leq 2B(j^*,f,\epsilon) \leq 2\sigma(j^*,n) \leq 2\sigma(l,n)$ by Lemma 10.1. This implies that (since $D_p' \geq D_p$ by definition)

$$\mathbb{P}\Big(\|\hat{f}_n(j-1) - \hat{f}_n(l)\|_p \geq \frac{4(D_p'+1)}{D_p}\sigma(l,n)\Big)$$

$$\leq \mathbb{P}\Big(\|\hat{f}_n(j-1) - \Pi_{V_{j-1}}f\|_p \geq \Big(\frac{4(D_p'+1)}{2D_p} - 1\Big)\sigma(l,n)\Big)$$

$$+ \mathbb{P}\Big(\|\hat{f}_n(l) - \Pi_{V_l}f\|_p \geq \Big(\frac{4(D_p'+1)}{2D_p} - 1\Big)\sigma(l,n)\Big)$$

$$\leq \mathbb{P}\Big(\|\hat{f}_n(j-1) - \Pi_{V_{j-1}}f\|_p \geq \big(D_p'+2)\big)\sigma(l,n)\Big)$$

$$+ \mathbb{P}\Big(\|\hat{f}_n(l) - \Pi_{V_l}f\|_p \geq \big(D_p'+2\big)\sigma(l,n)\Big)$$

so we obtain by Lemma 10.2 that

$$\mathbb{P}\Big(\|\hat{f}_n(j-1) - \hat{f}_n(l)\|_p \geq \frac{4(D_p'+1)}{D_p}\sigma(l,n)\Big) \leq 2 \times \exp(-2^{2l/p}),$$

which implies when combined with Equations 10.2 and 10.3

$$\mathbf{E}\Big[\|\hat{f}_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n > j^*\Big\}\Big] \leq \sum_{j>j^*} D_p'\sqrt{\frac{2^j}{n}} \sum_{l\geq j} \big(2 \times \exp(-2^{2l/p})\big)^{(p-1)/p}$$

$$\leq \sum_{j>j^*} 8D_p'\sqrt{\frac{2^j}{n}} \exp(-2^{2j/p}(p-1)/p)$$

$$\leq \sum_{j>j^*} 16 D'_p \frac{\exp(-2^{2j/p}(p-1)/2p)}{\sqrt{n}}$$

$$\leq 16 D'_p \frac{\exp(-2^{2j^*/p}(p-1)/2p)}{\sqrt{n}}$$

$$\leq n^{-r/(2r+1)}.$$

for n (and thus j^*) large enough (but depending only on p, i.e. $n \geq N_p$).

10.5. Conclusion

By combining the results of the two precedent Subsections, we have

$$\mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p\Big] \leq \mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n \leq j^*\Big\}\Big]$$

$$+ \mathbf{E}\Big[\|f_n(\hat{j}_n) - f\|_p \mathbb{I}\Big\{\hat{j}_n > j^*\Big\}\Big]$$

$$\leq \left(2\Big(\frac{5D'_p + 4}{D_p}\Big)\Big(\frac{1}{D_p}(C_p B + c_a)\Big)^{1/(2r+1)} + 1\right) n^{-r/(2r+1)}.$$

which concludes the proof since all the constants in the bound depend only on p.

11. Extension of Theorem 3.5 to the entire segment $[t, s]$

We now state the analogue of Theorem 3.2, i.e. for the whole segment $I = [t, s]$ (still when $s(1 - 1/p) \leq t$). The proof of this Theorem is more technical than the proof of Theorem 3.5, but it is based on similar ideas.

Consider now the case where $s(1 - 1/p) \leq t$. In this case, full adaptation is possible without constraining the model $\mathcal{P}_n$ to be a strict subset of $\Sigma(t, B)$. We provide the following result, related to the case $s \leq 2t$ in (Bull and Nickl, 2013).

Theorem 11.1. *Let $1/2 \leq t < s$. Assume also that $s(1 - 1/p) \leq t$. Let $\mathcal{P}_n = \Sigma(t, B)$ and $I = [t, s]$. Let $B > 0$ and $\alpha > 0$. There exists a (L_p, α)-honest and adaptive confidence set given $\mathcal{P}_n$, I and B.*

Proof. Assume that $s(1-1/p) \leq t \leq s$ and let $B > 0$. Let $n > \max(N_p, \exp(2/t))$, where N_p defined as in Theorem 10.1.

Let $r \in [t, s]$. We define the following sets (similar to the sets defined in Equation (3.2), but separated from $\Sigma(r, B)$)

$$\tilde{\Sigma}(t, r, B, \rho_n(r)) = \Sigma(t, B) \setminus \{g \in \Sigma(t, B) : \|g - \Sigma(r, B)\|_p \leq \rho_n(r)\}.$$

where $\rho_n(r) \geq 0$. These sets are empty as $r \to t$, or when $\rho_n(r)$ is large, but are nevertheless defined.

Let $\alpha > 0$. Let us write, for $r \in [t, s]$, $\Psi_n(r)$ for the test described in Subsection 4.2, where the associated constants E_1 and E_2 are chosen large enough

(depending only on p, B, α) so that Lemma 11.1 holds. Set also

$$\rho_n(r) = 2C(B+1)n^{-\frac{r}{2r+1}} \geq C(B+1)(n^{-\frac{r}{2r+1}} + n^{-\frac{t}{2t+1-1/p}}),$$

for C as in Theorem 3.6 (depending only on α, p).

11.1. Step 1: Study of the process $(\Psi_n(r))_{r \in [t,s]}$

A first remark is that for any $r \in [t, s]$, the test $\Psi_n(r)$ is a measurable random variable from $(\mathcal{C}[0, 1], \mathcal{B}(\mathcal{C}[0, 1]))$ to $(\{0, 1\}, \{\{0\}, \{1\}, \{0, 1\}, \emptyset\})$ where $\mathcal{C}[0, 1])$ is the set of continuous functions from $[0, 1]$ to $\mathbb{R}$, and $\mathcal{B}(.)$ is the associated Borel set.

Lemma 11.1. *Consider the test $\Psi_n(r)$ described in Subsection 4.2. Assume that the associated constants E_1 and E_2 are large enough (depending only on p, B, α). The trajectories $r \in [t, s] \to \Psi_n(r)$ of the process $(\Psi_n(r))_{r \in [t,s]}$ are monotonously increasing, and caglad (left continuous right limit).*

Proof. Consider the tests $\Psi_n(r)$ described in Subsection 4.2. Since $\Psi_n(r)$ is either 1 or 0, increasing monotonicity is equivalent to $\forall(r_1, r_2), t \leq r_1 \leq r_2 \leq s$, $\Psi_n(r_2) = 0 \Rightarrow \Psi_n(r_1) = 0$.

The tests $\Psi_n(r)$ involve the statistics $T_n(l)$ (similar for any r), the statistics

$$\tilde{T}_n(r) = \inf_{g \in \Sigma(r,B)} \|\Pi_{V_{j_r}} \hat{f}_n - g\|_p,$$

where $j_r = \lfloor \log(n^{1/(2r+1)}) \rfloor$ is a decreasing function of r, the thresholds $(t_n(l, r))_l$ that are decreasing functions of r, and the threshold $\tilde{t}_n(r)$ that is a decreasing function of r. See Subsection 4.2 for a more complete definition of all these quantities. The test is defined as

$$\Psi_n(r) = 1 - \mathbb{I}\left\{\tilde{T}_n(r) \leq \tilde{t}_n(r)\right\} \prod_{j_r \leq l \leq j} \mathbb{I}\left\{T_n(l) \leq (t_n(l, r))^p\right\}.$$

Let $t \leq r_1 \leq r_2 \leq s$. Assume that $\Psi_n(r_2) = 0$, i.e. that

$$\tilde{T}_n(r_2) \leq \tilde{t}_n(r_2), \tag{11.1}$$

and

$$\forall l \in \mathbb{N} : j \geq l \geq j_{r_2}, T_n(l) \leq (t_n(l, r_2))^p. \tag{11.2}$$

Since $j_{r_1} \geq j_{r_2}$, and $\forall l, t_n(l, r_1) \geq t_n(l, r_2)$, we know by Equation (11.2) that

$$\forall l \in \mathbb{N} : j \geq l \geq j_{r_1}, T_n(l) \leq (t_n(l, r_1))^p. \tag{11.3}$$

If $j_{r_1} = j_{r_2}$, then $\mathbb{I}\{\tilde{T}_n(r_1) \leq \tilde{t}_n(r_1)\} = \mathbb{I}\{\tilde{T}_n(r_2) \leq \tilde{t}_n(r_2)\} = 1$. Otherwise, it implies that $j_{r_1} \geq 1 + j_{r_2}$, and by triangular inequality

$$\tilde{T}_n(r_1) = \inf_{g \in \Sigma(r_1, B)} \|\Pi_{V_{j_{r_1}}} \hat{f}_n - g\|_p \leq \inf_{g \in \Sigma(r_2, B)} \|\Pi_{V_{j_{r_1}}} \hat{f}_n - g\|_p$$

$$\leq \inf_{g \in \Sigma(r_2, B)} \| \Pi_{V_{j_{r_2}}} \hat{f}_n - g \|_p + \sum_{l=j_{r_2}}^{j_{r_1}} \| \Pi_{W_l} \hat{f}_n \|_p$$

$$\leq \tilde{T}_n(r_2) + C_p \sum_{l=j_{r_2}}^{j_{r_1}} \| \Pi_{W_l} \hat{f}_n \|_{0,p,\infty}$$

$$\leq \tilde{T}_n(r_2) + C_p E' \sum_{l=j_{r_2}}^{j_{r_1}} \Big(\frac{2^{l/2}}{\sqrt{n}} + \big(\max(T_n(l), 0) \big)^{1/p} \Big), \qquad (11.4)$$

for some $E' > 0$ large enough but depending only on p (see the proof of Lemma 4.4 for the argument on why $\| \Pi_{W_l} \hat{f}_n \|_{0,p,\infty}^p \leq E''(\frac{2^{lp/2}}{n^{p/2}} + \max(T_n(l), 0))$). Since the constants E_1 and E_2 defined in Subsection 4.2 can be chosen arbitrarily large, and since $s(1 - 1/p) \leq t$ (which implies that $\tilde{t}_n(r) \equiv \sum_{l=j_r}^{j} t_n(l, r) \equiv n^{-\frac{r}{2r+1}}$), we can choose E_1 and E_2 such that

$$\tilde{t}_n(r) = E_2 \sqrt{\frac{2^{j_r}}{n}} \geq E \sum_{l=j_r}^{j} t_n(l, r)$$

for some arbitrarily large $E_2 > 0$, and some arbitrarily large $E > E_2$ (by choosing E_2/E_1 large enough). Using this together with Equation (11.4), and the fact that $j_{r_1} \geq j_{r_2} + 1$, one obtains by Equations (11.1) and (11.2)

$$\tilde{T}_n(r_1) \leq \tilde{t}_n(r_2) + C_p E' \sum_{l=j_{r_2}}^{j_{r_1}} \Big(\frac{2^{l/2}}{\sqrt{n}} + t_n(l, r_2) \Big)$$

$$\leq 4 C_p E' \sqrt{\frac{2^{j_{r_1}}}{n}} + (E_2 + C_p E' E_2/E) \sqrt{\frac{2^{j_{r_2}}}{n}}$$

$$\leq (E_2/\sqrt{2} + C_p E' E_2/(\sqrt{2} E) + 4 C_p E') \sqrt{\frac{2^{j_{r_1}}}{n}}$$

$$\leq E_2 \sqrt{\frac{2^{j_{r_1}}}{n}} = \tilde{t}_n(r_1),$$

for E_2 and E/E_2 large enough. This implies together with Equation (11.3) that

$$\Psi_n(r_1) = 1 - \mathbb{I}\Big\{ \tilde{T}_n(r_1) \leq \tilde{t}_n(r_1) \Big\} \prod_{j_{r_1} \leq l \leq j} \mathbb{I}\{ T_n(l) \leq (t_n(l, r_1))^p \} = 0.$$

This concludes the proof of increasing monotonicity.

The trajectories $r \to \Psi_n(r)$ are increasing in $\{0, 1\}$. They are thus either caglad, or cadlag. By definition of the test, the sets $\{ r \in [t, s] : \Psi_n(r) = 1 \}$ are closed subsets of $[t, s]$. The trajectories are thus caglad. $\qquad\square$

Lemma 11.1, together with the fact that $\Psi_n(r)$ is measurable for any r, implies that the process $(\Psi_n(r))_{r \in [t,s]}$ is progressively measurable.

11.2. Step 2: Estimation of the Besov exponent

Consider $f \in \Sigma(t, B)$. As stated in Lemma 11.1, the trajectories $r \to \Psi_n(r)$ are increasing functions. More precisely, their value is 0 until some value $\hat{r}$ defined as

$$\hat{r} = \inf \left\{ r \in [t, s] : \Psi_n(r) = 1 \right\},$$

and then 1 for r large enough. $\hat{r} \in [t, s]$ is well defined since the trajectories are bounded by 1, and measurable since it is a stopping time on the progressively measurable process $(\Psi_n(r))_{r \in [t,s]}$ with caglad trajectories. Note also that, since the trajectories $\Psi_n(r)$ are of the form $x \in [t, s] \to \mathbf{1}\{x > c\}$

$$\hat{r} = \inf \left\{ r \in [t, s] : \Psi_n(r) = 1 \right\} = \sup \left\{ r \in [t, s] : \Psi_n(r) = 0 \right\}.$$

Consider the confidence set around $\hat{f}_n(\hat{j}_n)$, which is the adaptive estimate considered in Theorem 3.1, as being

$$C_n = \left\{ g : \|\hat{f}_n(\hat{j}_n) - g\|_p \leq \frac{1}{\alpha} U_p' n^{-\frac{\hat{r}}{2\hat{r}+1}} \right\},$$

where $U_p' = \tilde{U}_p(B, 2C(B + 1))$ is defined as in Theorem 10.1. Note that U_p' depends only on B and C, and thus only on B, α, p.

Write

$$r_f = \sup \left\{ r \in [t, s] : \|f - \Sigma(r, B)\|_p = 0 \right\} = \sup \left\{ r \in [t, s] : f \in \Sigma(r, B) \right\},$$

for the Besov exponent of f, and

$$r_f^+ \equiv r_f^+(n) = \sup \left\{ r \in [t, s] : \|f - \Sigma(r, B)\|_p \leq \rho_n(r) \right\}.$$

Note that r_f^+ exists since $\|f - \Sigma(t, B)\|_p = 0$.

Since $r \in [t, s] \to \|f - \Sigma(r, B)\|_p$ is a monotonously increasing function in r, we know that $\forall \epsilon > 0$,

$$f \in \Sigma(r_f - \epsilon, B). \tag{11.5}$$

Also, for the same reason and since $\rho_n(r)$ is a decreasing function of r,

$$\|f - \Sigma(r_f^+ + \epsilon, B)\|_2 \geq \rho_n(r_f^+) \geq \rho_n(r_f^+ + \epsilon), \tag{11.6}$$

and

$$\|f - \Sigma(r_f^+ - \epsilon, B)\|_2 \leq \rho_n(r_f^+) \leq \rho_n(r_f^+ - \epsilon), \tag{11.7}$$

where by convention, $\|f - \emptyset\|_2 = \|f\|_2$. We set

$$\epsilon \equiv \epsilon_n = 1/\log(n).$$

Since $f \in \Sigma(r_f - \epsilon, B)$ (Equation (11.5)), we have in particular by Equation (4.14) that for the n we fixed

$$\Pr_f(\Psi_n(r_f - \epsilon) = 0) \geq \inf_{g \in \Sigma(r_f - \epsilon, B)} \Pr_g(\Psi_n(r_f - \epsilon) = 0) \geq 1 - \alpha. \tag{11.8}$$

Since $\|f - \Sigma(r_f^+ + \epsilon, B)\|_2 \geq \rho_n(r_f^+ + \epsilon)$ (Equation (11.6)), and since thus $f \in \tilde{\Sigma}(t, r_f^+ + \epsilon, B, \rho_n(r_f^+ + \epsilon))$, we have in particular by Equation (4.14) that for the n we fixed

$$\Pr_f(\Psi_n(r_f^+ + \epsilon) = 1) \geq \inf_{g \in \tilde{\Sigma}(t, r_f^+ + \epsilon, B, \rho_n(r_f^+ + \epsilon))} \Pr_g(\Psi_n(r_f^+ + \epsilon) = 1) \geq 1 - \alpha.$$

$$(11.9)$$

By combining Equations (11.9) and (11.8), and since $r \to \Psi_n(r)$ is an increasing function (Lemma 11.1), we know that

$$\Pr_f(\hat{r} \in [r_f - \epsilon, r_f^+ + \epsilon]) \geq 1 - 2\alpha.$$

11.3. Step 3: Bound on the diameter of the confidence set

The bound of last Equation holds for any $f \in \Sigma(t, B)$ for the n we fixed, and thus by just considering the infimum over $\Sigma(t, B)$, we have

$$\inf_{f \in \Sigma(t,B)} \Pr_f(\hat{r} \in [r_f - \epsilon, r_f^+ + \epsilon]) \geq 1 - 2\alpha.$$

We thus have by definition of C_n that

$$\sup_{f \in \Sigma(t,B)} \Pr_f\left(|C_n| > \frac{1}{\alpha}U_p' n^{-\frac{r_f - \epsilon}{2(r_f - \epsilon)+1}}\right) \leq 1 - \inf_{f \in \Sigma(t,B)} \Pr_f(\hat{r} \in [r_f - \epsilon, r_f^+ + \epsilon]) \leq 2\alpha,$$

and since $\epsilon = 1/\log(n)$, this implies for the n we fixed

$$\sup_{f \in \Sigma(t,B)} \Pr_f\left(|C_n| > \frac{1}{\alpha}U_p' \exp(1) n^{-\frac{r_f}{2r_f + 1}}\right) \leq 1 - \inf_{f \in \Sigma(t,B)} \Pr_f(\hat{r} \in [r_f - \epsilon, r_f^+ + \epsilon]) \leq 2\alpha,$$

since $\exp(1/(2(r_f - \epsilon) + 1)) \leq \exp(1)$. Note now that

$$\sup_{f \in \Sigma(t,B)} \Pr_f\left(|C_n| > \frac{U_p'}{\alpha} \exp(1) n^{-\frac{r_f}{2r_f + 1}}\right)$$

$$= \sup_{r \in [t,s]} \sup_{f \in \Sigma(r,B)} \Pr_f\left(|C_n| > \frac{U_p'}{\alpha} \exp(1) n^{-\frac{r_f}{2r_f + 1}}\right)$$

$$\geq \sup_{r \in [t,s]} \sup_{f \in \Sigma(r,B)} \Pr_f\left(|C_n| > \frac{U_p'}{\alpha} \exp(1) n^{-\frac{r}{2r + 1}}\right),$$

since $f \in \Sigma(r, B)$ implies $r \leq r_f$. Finally

$$\sup_{r \in [t,s]} \sup_{f \in \Sigma(r,B)} \Pr_f\left(|C_n| > \frac{1}{\alpha}U_p' \exp(1) n^{-\frac{r}{2r + 1}}\right) \leq 2\alpha.$$

11.4. Step 4: Bound on the probability that the parameter is in the confidence set

Also we have by Markov's inequality

$$\inf_{f \in \Sigma(t,B)} \Pr_f\left(f \in C_n\right) \tag{11.10}$$

$$\geq 1 - \sup_{f \in \Sigma(t,B)} \Pr_f\left(\|\hat{f}_n(\hat{j}_n) - f\|_p \geq \frac{1}{\alpha}U_p'n^{-\frac{\hat{r}}{2\hat{r}+1}}\right)$$

$$\geq 1 - \sup_{f \in \Sigma(t,B)} \Pr_f(\hat{r} \notin [r_f - \epsilon, r_f^+ + \epsilon])$$

$$\quad - \sup_{f \in \Sigma(t,B)} \Pr_f\left(\|\hat{f}_n(\hat{j}_n) - f\|_p \geq \frac{1}{\alpha}U_p'n^{-\frac{r_f^+ + \epsilon}{2(r_f^+ + \epsilon)+1}}\right)$$

$$\geq 1 - 2\alpha - \sup_{f \in \Sigma(t,B)} \Pr_f\left(\|\hat{f}_n(\hat{j}_n) - f\|_p \geq \frac{1}{\alpha}U_p'\exp(-2)n^{-\frac{r_f^+ - \epsilon}{2(r_f^+ - \epsilon)+1}}\right)$$

$$\geq 1 - 2\alpha - \sup_{f \in \Sigma(t,B)} \left[\frac{\alpha}{U_p'}\exp(2)n^{\frac{r_f^+ - \epsilon}{2(r_f^+ - \epsilon)+1}}\mathbb{E}_f\|\hat{f}_n(\hat{j}_n) - f\|_p\right], \tag{11.11}$$

by definition of ϵ and since $\exp(-2/(2(r_f + \epsilon) + 1)) \geq \exp(-2)$.

We have for any $f \in \Sigma(t,B)$ that

$$\|f - \Sigma(r_f^+ - \epsilon, B)\|_p \leq \rho_n(r_f^+ - \epsilon) = 2C(B+1)n^{-(r_f^+ - \epsilon)/(2(r_f^+ - \epsilon)+1)}$$

by Equation (11.7). Since $r_f^+ - \epsilon > t - 1/\log(n) > t/2 > 0$ by definition of n, we have

$$\sup_{f \in \Sigma(t,B)} \left[\frac{1}{\tilde{U}_p(B, 2C(B+1))}n^{\frac{r_f^+ - \epsilon}{2(r_f^+ - \epsilon)+1}}\mathbb{E}_f\|\hat{f}_n(\hat{j}_n) - f\|_p\right]$$

$$\leq \sup_{r>0} \sup_{f \in L_2 : \|f - \Sigma(r,B)\|_2 \leq \rho_n(r)} \left[\frac{1}{\tilde{U}_p(B, 2C(B+1))}n^{r/(2r+1)}\mathbb{E}_f\|\hat{f}_n(\hat{j}_n) - f\|_p\right].$$

Combining this with Theorem 10.1 implies, since $n > N_p$, that

$$\sup_{f \in \Sigma(t,B)} \left[\frac{1}{\tilde{U}_p(B, 2C(B+1))}n^{\frac{r_f^+ - \epsilon}{2(r_f^+ - \epsilon)+1}}\mathbb{E}_f\|\hat{f}_n(\hat{j}_n) - f\|_p\right] \leq 1,$$

since $U_p' \geq \tilde{U}_p(B, 2C(B+1))$. We conclude by plugging this result into Equation (11.11) that

$$\inf_{f \in \Sigma(t,B)} \Pr_f\left(f \in C_n\right) \geq 1 - 3\exp(2)\alpha.$$

Conclusion   All these results hold for *any* $n > \max(N_p, \exp(2/t))$. We have thus proven that C_n is an honest and adaptive confidence set on $\Sigma(t, B)$ for the whole interval $[t, s]$ and B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

12. Discussion on the extension of the results to more general settings

The test statistic Ψ_n that is considered relies mostly on estimates of $|a_{l,k}|^p$, which are the $\hat{F}_p^p(l, k)$. The obstacle for generalising the method presented in the paper to the regression setting, is the adaptation of these estimates to the regression setting. But the construction of the $\hat{F}_p^p(l, k)$ depends crucially on the distribution of the noise (error) to the signal f. Indeed, in the computation of the quantities $\hat{F}_p^p(l, k)$, we plug the p first moments of a Gaussian distribution in order to correct the bias of $|\hat{a}_{l,k}|^p$ toward $|a_{l,k}|^p$. If the distribution of the noise is not Gaussian, the bias is not going to be corrected by these (Gaussian) moments, and we would want to replace them with the moments of the noise.

However, if we do not wish to assume that we know the distribution of ξ (or, moreover, if it is heterocedastic) the construction of L_p-adaptive and honest confidence sets in this setting is possible but slightly different from the construction proposed. We did not present it in the paper since it is rather technical but not fundamentally different from what happens in the Gaussian process setting.

We first remind how to estimate $\hat{a}_{l,k}$ in two classic settings, i.e. regression and density estimation. In the setting of density estimation, i.e. the data in this setting is n i.i.d. samples from a random variable of density f, we can estimate $a_{l,k}$ by

$$\hat{a}_{l,k} = \frac{1}{n} \sum_{j \leq n} \psi_{l,k}(X_j).$$

If the density f is bounded (and still defined on the compact $[0, 1]$) then the estimates $\hat{a}_{l,k}$ computed in this way will be unbiased and have a variance-covariance structure that is of same order than in the case of the Gaussian process model. In the setting of non-parametric regression, i.e. the data in this setting is n i.i.d. samples $(X_j, Y_j)_n$ such that $Y_j = f(X_j) + \xi_j(X_j)$ where $\xi_j(X_j)$ is the noise, we can estimate $a_{l,k}$ by

$$\hat{a}_{l,k} = \frac{1}{n} \sum_{j \leq n} \psi_{l,k}(X_j) Y_j.$$

If the function f is bounded (and still defined on the compact $[0, 1]$), the design X_j is uniformly random on $[0, 1]$ and the noise $\xi_j(X_j)$ is independent in j (although it might depend on X_j), of mean 0, and sub-Gaussian, then the estimates $\hat{a}_{l,k}$ computed in this way will be unbiased and have a variance-covariance structure that is of same order than in the case of the Gaussian process model.

Now, a first idea to adapt $\hat{F}_p^p(l, k)$ to these settings (if we have $2n$ data) is to divide the sample in two sub-samples of size n, estimate the p first moments

of the distribution of $\sqrt{n}\hat{a}_{l,k}$ on the first half (that we write $\hat{m}_u^u$), compute $\hat{a}_{l,k}$ on the second sub-sample, and then redefine the $\hat{F}_p^p(l,k)$ as

$$\hat{F}_p^p(l,k) = |\hat{a}_{l,k}|^p - \sum_{u=0,u\ even}^{\lfloor p \rfloor - 2} \mathbf{C}_p^u \Big(\frac{\hat{m}_{p-u}}{n^{1/2}}\Big)^{(p-u)} \hat{F}_u^u(l,k),$$

where

$$\hat{F}_u^u(l,k) = \hat{a}_{l,k}^u - \sum_{i=0,i\ even}^{u-2} \mathbf{C}_u^i \Big(\frac{\hat{m}_{u-i}}{n^{1/2}}\Big)^{(u-i)} \hat{F}_i^i(l,k).$$

These quantities will verify the same properties as the $hatF_p^p(l,k)$ analysed in the paper (but the proof is more technical).

Another idea is to redefine the estimates $\hat{F}_p^p(l,k)$ of $|a_{l,k}|^p$. The idea is to divide the data in $\lfloor p \rfloor + \mathbf{1}\{p - \lfloor p \rfloor \neq 0\}$ sub-samples of equal size, and to compute in each of these samples estimates of $a_{l,k}$ as described above. Let us denote by $\hat{a}_{l,k}^{(i)}$ the estimate of $a_{l,k}$ computed with the ith sub-sample. We propose to redefine the estimate $\hat{F}_p^p(l,k)$ of $|a_{l,k}|^p$ as

$$\hat{F}_p^p(l,k) = \prod_{i=1}^{\lfloor p \rfloor} \hat{a}_{l,k}^{(i)} \times \Big|\hat{a}_{l,k}^{(\lfloor p \rfloor + \mathbf{1}\{p - \lfloor p \rfloor \neq 0\})}\}\Big|^{p - \lfloor p \rfloor}. \tag{12.1}$$

The mean and variance of this estimate will verify the same inequalities as the estimate defined in the proof of Theorem 3.5 (Section 5), and similar results will hold.

Acknowledgments

References

BARAUD, Y. Confidence balls in gaussian regression. *Annals of Statistics*, pages 528–551, 2004. MR2060168

BARRON, A., BIRGÉ, L., and MASSART, P. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999. MR1679028

BERGH, J. and LÖFSTRÖM, J. *Interpolation Spaces: An Introduction*, volume 223. Springer-verlag Berlin, 1976. MR0482275

BESOV, O.V., IL'IN, V.P., and NIKOL'SKIĬ. *Integral Representations of Functions and Imbedding Theorems.*

L. Birgé and Massart, P. Gaussian model selection. *Journal of the European Mathematical Society*, 3(3):203–268, 2001. MR1848946

Bull, A.D. and Nickl, R. Adaptive confidence sets in L_2. *Probability Theory and Related Fields*, 156(3):889–919, 2013. MR3078289

Cai, T.T. and Low, M.G. Adaptive confidence balls. *The Annals of Statistics*, 34(1):202–228, 2006. MR2275240

Carpentier, A. Testing the regularity of a smooth signal. *To appear in Bernoulli*, 2013.

Cohen, A., Daubechies, I., and Vial, P. Wavelets on the interval and fast wavelet transforms. *Applied Computational Harmonic Analysis*, 1(1):54–81, 1993. MR1256527

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 301–369, 1995. MR1323344

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996. MR1394974

Efromovich, S. Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *The Annals of Statistics*, 36(3):1127–1155, 2008. MR2418652

Giné, E. and Nickl, R. Uniform limit theorems for wavelet density estimators. *The Annals of Probability*, 37(4):1605–1646, 2009b. MR2546757

Giné, E. and Nickl, R. Confidence bands in density estimation. *The Annals of Statistics*, 38(2):1122–1170, 2010b. MR2604707

Giné, E. and Nickl, R. Rates of contraction for posterior distributions in l_r-metrics, $\{1 \leq r \leq +\infty\}$. *The Annals of Statistics*, 39(6):2883–2911, 2011. MR3012395

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. *Wavelets, Approximation, and Statistical Applications*. Springer, New York, 1998. MR1618204

Hoffman, M. and Lepski, O. Random rates in anisotropic regression (with a discussion and a rejoinder by the authors). *The Annals of Statistics*, 30(2):325–396, 2002. MR1902892

Hoffmann, M. and Nickl, R. On adaptive inference and confidence bands. *The Annals of Statistics*, 39(5):2383–2409, 2011. MR2906872

Ingster, Y. and Suslina, I.A. *Nonparametric Goodness-of-fit Testing Under Gaussian Models*, volume 169. Springer, 2002. MR1991446

Ingster, Y.I. Minimax testing of nonparametric hypotheses on a distribution density in the l_p metrics. *Theory of Probability & Its Applications*, 31(2):333–337, 1987.

Ingster, Y.I. Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist.*, 2(2):85–114, 1993. MR1257978

Juditsky, A. and S. Lambert-Lacroix. Nonparametric confidence set estimation. *Mathematical Methods of Statistics*, 12(4):410–428, 2003. MR2054156

Lepski, O.V. On problems of adaptive estimation in white gaussian noise. *Topics in Nonparametric Estimation*, 12:87–106, 1992. MR1191692

Lepski, O., Nemirovski, A., and Spokoiny, V. On estimation of the l_r norm of a regression function. *Probability Theory and Related Fields*, 113(2):221–253, 1999. MR1670867

Lepski, O.V., Mammen, E., and Spokoiny, V.G. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, 25(3):929–947, 1997. MR1447734

Low, M.G. On nonparametric confidence intervals. *The Annals of Statistics*, 25(6):2547–2554, 1997. MR1604412

Meyer, Y. Wavelets and Applications. *Masson Paris*, 1992. MR1276521

Picard, D. and Tribouley, K. Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics*, 28(1):298–335, 2000. MR1762913

Robins, J. and Van Der Vaart, A. Adaptive nonparametric confidence sets. *The Annals of Statistics*, 34(1):229–253, 2006. MR2275241

Tsybakov, A.B. Introduction à l'estimation non paramétrique. Springer, volume 41, 2004. MR2013911