On nonparametric confidence set estimation

Anatoli Juditsky

Sophie Lambert-Lacroix

Université Joseph Fourier, St. Martin d'Heres, Sophie.Lambert@inrialpes.fr

October 2, 2001

Abstract: The problem of adaptive estimation of regression function from noisy observations is considered in the paper. We provide an adaptive confidence set $\hat{\mathcal{B}}_N$ of level $1-\alpha$, $0<\alpha<1$, for the unknown function f. Here $\hat{\mathcal{B}}_N$ is a L_2-ball of (random) diameter $\widehat{\tau}_N$, centered at the wavelet adaptive estimate $\widehat{f}_N$. We show that if it is known *a priori* that f belongs to a Besov functional class $\mathcal{F}^*$, then the proposed confidence ball cannot be improved in the minimax sense.

Key words: Adaptive estimation, nonparametric regression, confidence sets, wavelet estimators

1 Introduction

We consider the problem of recovering unknown function $f(x):\,[0,1]\to\mathbf{R}$ from noisy observations

$$y_i = f(\frac{i}{N}) + w_i, \quad i=1,...,N, \tag{1}$$

where (w_i), $i=1,...,N$ is the vector of independent and identically distributed Gaussian random variables with $Ew_1=0$ and $Ew_1^2=\sigma_w^2$. In what follows we suppose that f belongs to the Besov class $\mathcal{F}^*=\mathcal{F}(s^*,p^*,q^*,L^*)$ (refer to Section 2 for definitions), defined by the parameters (s^*,p^*,q^*,L^*) which are supposed to be known *a priori.*

It is well known how to construct the minimax (up to a constant) on the class $\mathcal{F}(s^*,p^*,q^*,L^*)$ estimator $\widehat{f}_N^*$, i.e the function $\widehat{f}_N^*$ which is the minimizer on the set of all measurable functions f_N of observations $y_1,...,y_N$ of the maximal on $\mathcal{F}$ risk

$$R(f_N,\mathcal{F}^*) = \sup_{f\in\mathcal{F}^*} E_f\|f_N-f\|^2. \tag{2}$$

The minimax (up to a constant) *rates of estimation* $\varphi_N^*=\varphi_N(\mathcal{F}^*)$,

$$\varphi_N(\mathcal{F}^*) = \left(R(\widehat{f}_N^*,\mathcal{F})\right)^{1/2}$$

are available for a variety of norms (or semi-norms) $\|\cdot\|$ when the parameters (s^*, p^*, q^*, L^*) of the class are known *a priori* (cf. [8], [1] or [4]). Furthermore, it is possible to provide *adaptive in order* estimation algorithms to estimate f (cf., for instance, [1], [7] and [13]).[1] For example, if $\|\cdot\|$ is an L_p-norm, $1 \leq p \leq \infty$, those estimators only use the observations but not the values of the parameters (s^*, p^*, q^*, L^*) to deliver the estimates $\widehat{f}_N$ of not worse quality than the parameter-dependent ones. That is the ratio of the maximal estimate risk $R(\widehat{f}_N, \mathcal{F}^*)$ and the minimax risk remains finite as $N \to \infty$.

However, those adaptive algorithms do not typically provide any information about the accuracy of the estimate $\widehat{f}_N$. Being a bit sloppy, one would say that the adaptive estimator can be very close to the unknown function (or quite far from it, if the underlying function is hard to estimate), but we would never know it. On the other hand, the latter information can be of paramount importance. For instance, one can be interested to point out a (random) set $\mathcal{B}$ in a functional space which covers f with high probability. A "natural" confidence set can be a (random) ball $\hat{\mathcal{B}}_N(g, \tau)$ in a functional space, of diameter τ, centered at g. We are interested to provide a confidence ball for the unknown function f.

In fact, little is known about this problem. For instance, there is an uncertainty as to which accuracy measures are relevant to confidence estimation, how to characterize the optimality of the confidence sets, how to construct the optimal confidence sets, etc. Therefore, our first objective is to provide meaningful definitions of the notions involved.

1.1 Problem statement

Let us fix some terminology. Let α be a real and $0 < \alpha < 1$, $\|\cdot\|$ be a functional norm or a semi-norm. We denote $\mathcal{B}(g, \tau)$,

$$\mathcal{B}(g, \tau) = \left\{ h : [0,1] \to \mathbf{R} \;\middle|\; \|g - h\| \leq \frac{\tau}{2} \right\},$$

a ball of diameter $\tau \in \mathbf{R}^+$, centered at the function g.

Definition 1 (Confidence ball) *Let $\mathcal{F}^*$ be a functional class. The (random) ball $\hat{\mathcal{B}}_N = \mathcal{B}(f_N, \tau_N)$ of diameter τ_N, centered at f_N is referred to as confidence ball of level $1 - \alpha$ on $\mathcal{F}^*$ if*

$$\inf_{f \in \mathcal{F}^*} P_f(f \in \hat{\mathcal{B}}_N) \geq 1 - \alpha. \tag{3}$$

Suppose that it is known *a priori* that the unknown function belongs to a (large) functional class $\mathcal{F}^*$ and let $\widehat{f}_N^*$ be a minimax on $\mathcal{F}^*$ estimate of f and φ_N^* be the minimax rate of estimation on $\mathcal{F}^*$. Then by the Chebychev inequality if $\tau(\alpha) = \frac{\varphi_N^*}{\sqrt{\alpha}}$, $\hat{\mathcal{B}}_N^* = \mathcal{B}(\widehat{f}_N^*, \tau(\alpha))$ is a confidence ball of level $1 - \alpha$ on $\mathcal{F}^*$. Such a confidence ball can be rather pessimistic because the class $\mathcal{F}^*$ contains also functions which can be estimated more accurately . In particular, $\mathcal{F}^*$ embeds "smaller" functional

[1] We use here the terminology introduced by Lepski in [16].

classes $\mathcal{F}' = \mathcal{F}(s', p^*, q^*, L^*)$ of higher regularity $s' > s^*$. If the information that the unknown function belongs to such an embedded class were available, one could point out a confidence ball of smaller diameter, which rather corresponds to the minimax rate of convergence on $\mathcal{F}'$. The question is if confidence balls for f which are better than $\hat{\mathcal{B}}_N^*$ can be provided using the available observations.

The answer to this question generally depends on the functional norm used. In the paper [19], Low has studied the problem of constructing confidence intervals for the value of $f(x)$, $x \in [0,1]$ in the context of the Hölder functional class $\mathcal{F}^* = \mathcal{F}(s^*, L^*)$ (the correspondent semi-norm is $\|g\| = |g(x)|$). He has shown that if the only *a priori* information available is that $f \in \mathcal{F}^*$, then the confidence interval of diameter $\tau_N \asymp N^{-\frac{s^*}{2s^*+1}}$, which corresponds to the minimax rate of convergence on $\mathcal{F}(s^*, L^*)$ is, in a certain sense, the best one. The same conclusion remains valid for the "natural" problem of constructing the confidence balls in the L_∞-space.

That negative result motivates our choice of the L_2-norm in the definition of the confidence ball. Though not the most natural, such confidence set provides an important information about the estimated function. Consider, e.g., the problem of predicting the value of $f(X)$ where X is randomly chosen from the uniform distribution U on $[0,1]$. Now let $\hat{\mathcal{B}}_N = \mathcal{B}(f_N, \tau_N)$ be a L_2-confidence ball of level $1-\alpha$, i.e.

$$P_f(\|f_N - f\|_2 > \frac{\tau_N}{2}) < \alpha.$$

Then by the Chebychev inequality

$$P(|f(X) - f_N(X)| \le \frac{\kappa\tau_N}{2}) \ge 1 - \frac{1}{\kappa^2} - \alpha.$$

So that the interval $[f_N(X) - \frac{\kappa\tau_N}{2}, f_N(X) + \frac{\kappa\tau_N}{2}]$ is a confidence interval for $f(X)$ of level $1-\kappa^{-2}-\alpha$.

From now on we consider only the L_2-confidence balls, i.e. the norm $\|\cdot\|$ in Definition 1 is the usual L_2-norm on $[0,1]$:

$$\|g\|_2 = \left(\int_0^1 g^2(x)dx\right)^{1/2}.$$

The next question to answer is how to compare confidence balls. One could suggest quite naturally that the confidence set $\hat{\mathcal{B}}_N = \mathcal{B}(f_N, \tau_N)$ on $\mathcal{F}^*$ is "better" than $\hat{\mathcal{B}}'_N = \mathcal{B}(f'_N, \tau'_N)$ if its "average size" $E_f\tau_N$ is "smaller" than $E_f\tau'_N$ for any $f \in \mathcal{F}^*$. And the optimal confidence ball $\hat{\mathcal{B}}_N^* = \mathcal{B}(f_N^*, \tau_N^*)$ of the smallest average size, i.e.

$$\sup_{f\in\mathcal{F}} \frac{E_f\tau_N^*}{E_f\tau'_N} \le 1$$

for any confidence ball $\hat{\mathcal{B}}'_N = \mathcal{B}(f'_N, \tau'_N)$ on $\mathcal{F}^*$.

However, here we are looking for something what does not exist. Let us consider, for instance the following procedure: let f_0 be a fixed function in $\mathcal{F}' \subset \mathcal{F}^*$. Then one can construct a nonparametric test of the hypothesis $H_0 :\ f = f_0$ *versus* the alternative $H_1 :\ \|f - f_0\|_2 \ge \rho_N^*$, $f \in \mathcal{F}$ such that the sum of the error probabilities of the first and the second type is bounded with α. Here $\rho_N^* = \rho_N(\mathcal{F}^*)$ is *the minimax rate of test* on $\mathcal{F}^*$ as in [11]. Let us now define a random function $\bar{f}_N$ and a random variable $\bar{\tau}_N$ as follows: $\bar{f}_N = f_0$, $\bar{\tau}_N = \rho_N$ if the hypothesis is accepted and, say, $\bar{f}_N = 0$ and $\bar{\tau}_N = C$, where C is such that $\|f\|_2 \le C$ for any $f \in \mathcal{F}$, if the hypothesis is rejected. One can easily see that $\bar{\mathcal{B}}_N = \mathcal{B}(\bar{f}_N, \bar{\tau}_N)$ is indeed a confidence ball on $\mathcal{F}^*$. Since the minimax rate of test ρ_N^* is typically much better than the minimax rate of estimation φ_N^* (cf. [17]), such a confidence ball would be hard to improve at f_0. Furthermore, it has been shown in [18], that by using the nonparametric testing technique the accuracy of a the minimax estimator can be significantly improved at any given function

$f_0 \in \mathcal{F}^*$ (or at a not too complex family of functions). So that a confidence ball can be constructed such that its diameter is particularly small (in fact, it corresponds to the minimax rate of the test) at any fixed function $f_0 \in \mathcal{F}^*$ and corresponds to the minimax rate of estimation on the class $\mathcal{F}^*$ otherwise. This is why we opt for the following definition:

Definition 2 (Adaptive in order confidence ball) *Let $\mathcal{F}(s,p,q,L)$ be a family of embedded Besov classes, $\mathcal{F}(s,p,q,L) \subseteq \mathcal{F}^* = \mathcal{F}(s^*,p^*,q^*,L^*)$, where the parameters s^*, p^*, q^* and L^* of the large class are* a priori *known.*

The confidence ball $\hat{\mathcal{B}}_N = \mathcal{B}(f_N^*, \tau_N^*)$ on $\mathcal{F}(s^*,p^*,q^*,L^*)$ is called adaptive in order (or minimax adaptive) on $\mathcal{F}(s^*,p^*,q^*,L^*)$ if there is a constant C (which can depend on the parameters s^*, p^*, q^* and L^* of $\mathcal{F}^*$) such that for any class $\mathcal{F}$ of the family and for any other confidence ball $\hat{\mathcal{B}}'_N = \mathcal{B}(f'_N, \tau'_N)$

$$\frac{\sup_{f\in\mathcal{F}} E_f \tau_N^*}{\sup_{f\in\mathcal{F}} E_f \tau'_N} \le C. \tag{4}$$

Comment: in the definition above we do not require the optimal confidence ball to be good at each function of $\mathcal{F}^*$, but only at "bad" representatives of each class $\mathcal{F}(s,p,q,L)$ in the parametric family of embedded classes. Note that the approach we adopt differs significantly from that, described in [10], [22] or [23], where several methods of constructing point-wise confidence intervals are discussed. In fact, in the present paper we consider the *minimax* on the class $\mathcal{F}^*$ confidence sets as in Definition 1. Therefore, our results can be hardly compared with those in the papers cited above.

We would like to mention that the problem of constructing confidence sets is closely related to those of sequential estimation (cf. [9], where that problem is studied for the family of Sobolev classes) and of providing data-driven normalizations (cf. [18]).

1.2 Result summary

The result in the current paper is twofold: first we establish a lower bound for the "average size" $E_f\tau_N$ of any confidence ball on $\mathcal{F}^*$: in particular, we show (Propositions 1 and 2 of Section 3) that

1. for any Besov class $\mathcal{F} = \mathcal{F}(s,p,q,L) \subseteq \mathcal{F}^*$ the "average size" $E_f\tau_N$ of a confidence set $\hat{\mathcal{B}}_N$ satisfies:
$$\lim_{N\to\infty} (\varphi_N(\mathcal{F}))^{-1} \sup_{f\in\mathcal{F}} E_f \tau_N > 0,$$
where $\varphi_N(\mathcal{F})$ is the minimax rate of estimation on $\mathcal{F}$;

2. if $\mathcal{F}' \subset \mathcal{F}^*$ is a "smaller" class than $\mathcal{F}^*$, then

$$\lim_{N\to\infty} \inf_{f\in\mathcal{F}'} (\rho_N^*)^{-1} E_f \tau_N > 0,$$

where $\rho_N^* = \rho_N(\mathcal{F}^*)$ is the minimax rate of test on $\mathcal{F}^*$.

This implies, in particular, that if a confidence ball $\hat{\mathcal{B}}_N^* = \mathcal{B}(f_N, \tau_N)$ on $\mathcal{F}^*$ which satisfies for N large enough

$$\begin{array}{lll} \sup_{f\in\mathcal{F}} E_f \tau_N^* \le C\varphi_N(\mathcal{F}), & \text{if the class } \mathcal{F} \text{ is such that} & \varphi_N(\mathcal{F}) \ge \rho_N^* \\ \sup_{f\in\mathcal{F}} E_f \tau_N^* \le C'\rho_N^*, & \text{if the class } \mathcal{F} \text{ is such that} & \varphi_N(\mathcal{F}) < \rho_N^*, \end{array} \tag{5}$$

can be constructed, then it would be minimax adaptive on $\mathcal{F}^*$. Note that it is rather evident that the maximal on the class $\mathcal{F}$ average size of the confidence ball cannot be less than the minimax rate of estimation on $\mathcal{F}$. However, the second inequality above is quite surprising: it states that the minimax rate of test $\rho_N^* = \rho_N(\mathcal{F}^*)$ on the large class $\mathcal{F}^*$ provides the lower bound for the size of the confidence ball.

At this point we would like to stress a spectacular difference between adaptive function estimation and adaptive confidence set estimation. Recall that a minimax adaptive estimator $\widehat{f}_N$ of the unknown function f can be constructed using very limited information on f. It is roughly sufficient to know, for instance, that f is continuous. Such an estimate can attain quasi-parametric rates of convergence if the function f is very smooth. On the other hand, in order to provide an adaptive confidence set $\hat{\mathcal{B}}_N$ one is to be sure that f belongs to some "not too large" functional class $\mathcal{F}^*$ with known *a priori* parameters. Moreover, the best attainable performance, even at very smooth functions, is limited by the minimax rate of test on $\mathcal{F}^*$.

Next we provide an upper bound, i.e. an adaptive in order confidence ball $\hat{\mathcal{B}}_N = \mathcal{B}(\widehat{f}_N, \hat{\tau}_N)$ which satisfies the inequalities in (5). One can guess that an adaptive in order estimate $\widehat{f}_N$ would be a nice candidate for the center of $\hat{\mathcal{B}}_N$.

Then we construct in Section 4 a confidence ball $\hat{\mathcal{B}}_N = \mathcal{B}(\widehat{f}_N, \widehat{\tau}_N)$ of diameter $\widehat{\tau}_N$, computed using a minimax on $\mathcal{F}^*$ estimate $\widehat{m}_N$ of the loss $m_N = \|\widehat{f}_N - f\|_2$ of the adaptive estimate $\widehat{f}_N$. We show that $\hat{\mathcal{B}}_N$ is indeed adaptive in order on $\mathcal{F}^*$. We start with the definition of functional classes used.

2 Besov functional family

We fix an integer $m > 0$. Let ϕ_k, ψ_{jk} be a system of compactly supported orthogonal wavelets ($\text{supp}\phi \subseteq [-A, A]$ and $\text{supp}\psi \subseteq [-A, A]$), i.e. $\phi_k(x)$ and $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $j = 1, \ldots,$ constitute (inhomogeneous) orthonormal wavelet basis of $L_2(0,1)$ [20], [3]. We suppose that ϕ and $\psi \in C^m$. This implies (see Ch. 7, [3]) that $\psi(x)$ has $m-1$ vanishing moments (here $[\cdot]$ is an integer part). We just note that wavelet basis on $[0,1]$ with such properties can be constructed (see, for instance, [2]). Since the regression function and the wavelets are compactly supported, there are at most $(2^j + 2A - 1)$ nonzero coefficients at each resolution level j of the wavelet expansion of f. We suppose with some stretch that this number is exactly 2^j, thus

$$f(x) = \alpha\phi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x),$$

where

$$\alpha = \int f(x)\phi(x)dx, \ \ \beta_{jk} = \int f(x)\psi_{jk}(x)dx.$$

From now on we suppose that the unknown function f belongs to some set $\mathcal{F} \in L_2(0,1)$ which is defined through the coefficients α and β_{jk} of the wavelet decomposition of f:

$$f(x) = \alpha\phi(x) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x). \tag{6}$$

Let $\mathcal{F}$ be a "Besov body"[2] of wavelet coefficients:

$$\mathcal{F} \quad = \quad \mathcal{F}(s,p,q,L) = \{f \text{ such that } \|f\|_{spq} \leq L\} \tag{7}$$

where

$$\|f\|_{spq} = |\alpha| + \left(\sum_{j=0}^{\infty} (2^{j(s+1/2-1/p)}\|\beta_{j\cdot}\|_p)^q\right)^{1/q}. \tag{8}$$

Following [6], [8] we choose the Besov classes because of their exceptional expressive power: the Hölder and Sobolev classes often referred to in the statistical literature can be obtained for a particular choice of parameters s, p, q [20].

Note that if B^s_{pq}, $m \geq s \geq (p^{-1}-1)_+$, $0 < p, q \leq \infty$, is the Besov space (see [24]), then there is $C > 0$ such that

$$\|f\|_{B^s_{pq}} \geq C\|f\|_{spq}, \tag{9}$$

where $\|f\|_{B^s_{pq}}$ is the norm of the Besov space. On the other hand, for any $(p^{-1}-1)_+ < s < m$, there exists $C < \infty$ such that

$$C\|f\|_{spq} \geq \|f\|_{B^s_{pq}},$$

(cf. Theorem 2 in [5]. See also [8] for a discussion and useful references). In what follows with a slight abuse of notations we refer to $\mathcal{F}(s,p,q,L)$ as the Besov class.

Let now $\mathcal{F}* = \mathcal{F}(s^*, p^*, q^*, L^*)$, $0 < s^* \leq m$, $1/s^* < p^*$, $q^* \geq 1$, $L^* > 0$, be a Besov class as in (7). Using (8) one can easily verify that if $s > s^*$, $p \geq p^*$, $q \geq 1$ and $L \leq L^*$, then $\mathcal{F}(s,p,q,L) \subset \mathcal{F}^*$. Moreover, as in the classical embedding theorems (see, for instance, [24]), one has $\mathcal{F}(s,p,q,cL^*) \subset \mathcal{F}^*$ (here c depends on the wavelet used and on the parameters s, p, ..., q^* but not on L^*) with $s^* - \frac{1}{p^*} + \frac{1}{p} < s \leq m$, $p \geq 1$ and $q \geq 1$. For instance, the class $\mathcal{F}(s^*+1, 1, \infty, L^*)$ is embedded in the Hölder class $H(s^*, L^*) = \mathcal{F}(s^*, \infty, \infty, L^*)$, and so on.

3 Lower bound for confidence set estimation

Suppose that the observations $y_1, ..., y_N$ in the model (1) are available and $f \in \mathcal{F}^* = \mathcal{F}(s^*, p^*, q^*, L^*)$ with the parameters s^*, p^*, q^* and L^* which are known *a priori*. Let us fix some $\alpha \leq 1/8$. We start with the following simple proposition:

[2] we borrow the terminology of D. Donoho and I. Johnstone [6].

Proposition 1 *Let $\mathcal{F} = \mathcal{F}(s, p, q, L) \subseteq \mathcal{F}^*$, be a Besov class and*

$$\varphi_N(\mathcal{F}) = L^{\frac{1}{2s+1}} \left(\frac{\sigma_w^2}{N} \right)^{\frac{s}{2s+1}}.$$

Then the diameter τ_N of a confidence ball $\hat{\mathcal{B}}_N$ of level $1-\alpha$ on $\mathcal{F}^*$ satisfy

$$\lim_{N\to\infty} (\varphi_N(\mathcal{F}))^{-1} \sup_{f\in\mathcal{F}} E_f \tau_N > 0.$$

Now let for some $0 < \delta < 1$, $\mathcal{F}' = \mathcal{F}(s^*, p^*, q^*, (1-\delta)L^*)$ be a sub-class of $\mathcal{F}^*$ and let $\rho_N^{(\delta)}$ be defined as follows:

$$\rho_N^{(\delta)} = \begin{cases} (L^*\delta)^{\frac{1}{4s^*+2-2/p^*}} \left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*+1/2-1/p^*}{4s^*+2-2/p^*}}, & \text{for} \quad p^* < 2; \\ (L^*\delta)^{\frac{1}{4s^*+1}} \left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*}{4s^*+1}}, & \text{for} \quad p^* \geq 2,\ s^* \geq 1/4, \\ L^*\delta N^{-s^*}, & \text{for} \quad p^* \geq 2,\ s^* < 1/4. \end{cases} \tag{10}$$

Proposition 2 *Suppose that a random variable $\tau_N' = \tau_N'(y_1, ..., y_N)$ satisfy for some function $f_0 \in \mathcal{F}'$*

$$P_{f_0}(\tau_N' \leq a_N) \geq \frac{1}{2},$$

where a_N is such that

$$\lim_{N\to\infty} \frac{a_N}{\rho_N^{(\delta)}} = 0.$$

Then there is no confidence ball on $\mathcal{F}^*$ of diameter τ_N'.

The proofs of the statements are collected in Section 5

Note that the sequence $\rho_N^{(\delta)}$, defined in (10) coincides (when $\delta = 1$) with the minimax rate of test on $\mathcal{F}^*$ (see, for instance, Theorem 1 in [17]).

As a consequence of the above statements we have the following characterization of adaptive in order confidence ball: Let $\nu_N = \max(\varphi_N(\mathcal{F}), \rho_N^*)$. The confidence ball $\mathcal{B}^* = \mathcal{B}(f_N^*, \tau_N^*)$ on $\mathcal{F}^*$, if constructed, would be adaptive in order on $\mathcal{F}^*$ if for some $C = C(s^*, p^*, q^*, L^*)$ and all N large enough, τ_N^* satisfies for any Besov class $\mathcal{F}(s, p, q, L) \subseteq \mathcal{F}^*$

$$\sup_{f\in\mathcal{F}} E_f \tau_N^* \leq C\nu_N. \tag{11}$$

We are up to prove that such a confidence ball can indeed be provided.

4 Adaptive in order confidence ball

We present in this section a construction of minimax adaptive confidence ball $\hat{\mathcal{B}}_N$. It is based on joint estimation of the function f using a wavelet adaptive estimate $\widehat{f}_N$ and of the loss $m_N = \|f - \widehat{f}_N\|_2$.

Let $\phi_k,\ \psi_{jk}$ be a system of compactly supported wavelets with the regularity parameter $m \geq \rfloor 2s^* \lfloor$. Here $l = \rfloor x \lfloor$ stands for the smallest integer $l \geq x$.

Now, let $\lambda^a = (\lambda_1^a, ..., \lambda_{j_0}^a)$, $N/4 \leq 2^{j_0} \leq N/2$, be the thresholds for adaptive wavelet estimation, obtained using the methods, proposed in [7] (or [13]) (note that the vector λ^a depends on the data sample). Let (ξ_{jk}) be a $\mathbf{R}^{2^{j_0+1}-1}$-vector of independent and identically distributed Gaussian random variables with $E\xi_{jk} = 0$ and $E\xi_{jk}^2 = \frac{\sigma_w^2}{N}$. We suppose that (ξ_{jk}) is independent of (w_i) and that $f \in \mathcal{F}(s^*, p^*, q^*, L^*)$. Consider the following estimation algorithm for the diameter $\widehat{\tau}_N$ of the confidence ball $\hat{\mathcal{B}}_N$:

Algorithm 1 `Put` $\sigma^2 = \frac{\sigma_w^2}{N}$, `take` j_0 `such that` $N/2 < 2^{j_0} \leq N$ `and`

$$\begin{aligned} \left(\frac{L^*}{\sigma^2}\right)^{\frac{2}{2s^*+1-1/p^*}} &\leq 2^{j^*} < 2\left(\frac{L^*}{\sigma^2}\right)^{\frac{2}{2s^*+1-1/p^*}}, \quad \texttt{for} \quad p^* < 2 \\ \left(\frac{L^*}{\sigma^2}\right)^{\frac{4}{4s^*+1}} &\leq 2^{j^*} < 2\left(\frac{L^*}{\sigma^2}\right)^{\frac{4}{4s^*+1}}, \quad \texttt{for} \quad p^* \geq 2, \end{aligned} \tag{12}$$

`If` $j^* > j_0$ `set` $j^* = j_0$. `Define`

$$\lambda_j = \begin{cases} \kappa\sqrt{(j-j^*)_+} \ \texttt{with} \ \kappa = 4\sqrt{\log 2} & \texttt{for} \quad p^* < 2, \\ \infty & \texttt{for} \quad p^* \geq 2. \end{cases} \tag{13}$$

1. `Compute the empirical wavelet coefficients`

$$\widehat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i \phi(\frac{i}{N}) \ \texttt{and} \ y_{jk} = \frac{1}{N}\sum_{i=1}^{N} y_i \psi_{jk}(\frac{i}{N}) \ \ 0 \leq k \leq 2^j - 1, \ \ j = 0, ..., j_0.$$

2. `Set`

$$y'_{jk} = y_{jk} + \xi_{jk}, \quad y''_{jk} = y_{jk} - \xi_{jk}$$

`and`

$$\lambda_j^* = \max(\lambda_j^a, \lambda_j \sigma); \tag{14}$$

`and compute the estimates of wavelet coefficients`

$$\widehat{\beta}_{jk} = y'_{jk} 1_{|y'_{jk}| \geq \lambda_j^*} \tag{15}$$

3. Set

$$\widehat{f}_N(x) = \widehat{\alpha}\phi(x) + \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1}\widehat{\beta}_{jk}\psi_{jk}(x) \tag{16}$$

and

$$\widehat{m}_N^2 \;=\; \left[\sum_{j=0}^{j^*}\left(\|y''_{j\cdot} - \widehat{\beta}_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2\right) + \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1}\left((y''_{jk} - \widehat{\beta}_{jk})^2 - 2\sigma^2\right) 1_{|y'_{jk}|\geq\lambda_j\sigma}\right]_+ . \tag{17}$$

4. Finally, compute

$$\rho_N^* = \begin{cases} (L^*)^{\frac{1}{4s^*+2-2/p^*}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*+1/2-1/p^*}{4s^*+2-2/p^*}}, & \texttt{for} \quad p^* < 2, \\ (L^*)^{\frac{1}{4s^*+1}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*}{4s^*+1}} + L^*N^{-s^*}, & \texttt{for} \quad p^* \geq 2, \end{cases} \tag{18}$$

and

$$\widehat{\tau}_N(\alpha) = 2\left(\widehat{m}_N + \frac{C_1}{\sqrt{\alpha}}\rho_N^*\right). \tag{19}$$

Comments: Note that $\widehat{f}_N$ in Algorithm 1 is a wavelet threshold estimate, which is slightly modified, with respect to those, proposed in [7] and [13]. As we shall see, such a modification does not alter the properties of the estimate on $\mathcal{F}^*$.

The quantity $\widehat{m}_N$, defined in (17), is a minimax on $\mathcal{F}$ estimate of the loss $m_N = \|f - \widehat{f}_N\|_2$ of the estimate $\widehat{f}_N$ (see the correspondent lower bound in [14]).

When $p^* \geq 2$, we know *a priori* that the unknown function f is in the Sobolev class $\mathcal{H}^*(s^*, L^*)$. We still use the threshold estimator in order to compute the center $\widehat{f}_N$ of the confidence set since the function f can belong to a smaller Besov class $\mathcal{F} \subset \mathcal{H}^*$. However, in this case we do not need to re-estimate wavelet coefficients beyond the level j^*, and the estimator $\widehat{m}_N$ of m_N becomes (cf. (17))

$$\widehat{m}_N^2 \;=\; \left[\sum_{j=0}^{j^*}\left(\|y''_{j\cdot} - \widehat{\beta}_{j\cdot}\|_2^2 - 2^j\sigma^2\right)\right]_+ .$$

The constant C_1 in the algorithm does not depend on N, L^* and σ_w^2, and can be computed explicitly for a given value of s^*, p^* and wavelet ψ. The properties of the objects $\widehat{f}_N$ and $\widehat{\tau}_N$, computed in Algorithm 1 are described in the following theorem:

Theorem 1 *Let $\mathcal{F}(s,p,q,L)$ be a Besov class with $s > 1/p$ such that $\mathcal{F}(s,p,q,L) \subset \mathcal{F}(s^*,p^*,q^*,L^*)$ with $s^* > 1/p^*$, $1 \leq q^* \leq \infty$. Then*

1. *the estimate $\widehat{f}_N$ satisfies:*

$$\sup_{f\in\mathcal{F}(s,p,q,L)} [E_f\|f-\widehat{f}_N\|_2^2]^{1/2} \le C_0\varphi_N(\mathcal{F})+\epsilon(N), \tag{20}$$

where $\epsilon(N)=O\left(\frac{\sigma_w \log N}{\sqrt{N}}\right)$, $\varphi_N(\mathcal{F})=L^{1/(2\bar{s}+1)}\left(\frac{\sigma_w^2}{N}\right)^{\bar{s}/(2\bar{s}+1)}$ *and* $\bar{s}=\min(s,m)$.

2. *it holds for $\widehat{m}_N$:*

$$\sup_{f\in\mathcal{F}(s^*,p^*,\infty,L^*)} \left[E_f(\widehat{m}_N-m_N)^2\right]^{1/2} \le C_1\rho_N^*+\epsilon(N). \tag{21}$$

3. *the ball $\hat{\mathcal{B}}_N=\mathcal{B}(\widehat{f}_N,\widehat{\tau}_N)$ is a minimax adaptive confidence ball on $\mathcal{F}^*$ of level $1-\alpha$. In particular,*

$$\inf_{f\in\mathcal{F}^*} P_f(m_N\le\tau_N)\ge 1-\alpha,$$

and the diameter $\widehat{\tau}_N$ in (19) satisfies:

$$\sup_{f\in\mathcal{F}(s,p,q,L)} E_f\widehat{\tau}_N \le C_0\varphi_N(\mathcal{F})+\left(1+\frac{1}{\sqrt{\alpha}}\right)C_1\rho_N^*+\epsilon(N). \tag{22}$$

Comment: we observe that the estimate $\widehat{f}_N$ is adaptive in order over the family of Besov classes with the regularity s bounded with the regularity of the wavelet used. Moreover, the estimate $\widehat{m}_N$ of the loss $m_N=\|f-\widehat{f}_N\|_2$ is minimax on $\mathcal{F}^*$ (cf. Theorem 1 in [14]). One can see from the upper bound (22) that the average diameter $E_f\widehat{\tau}_N$ satisfies the requirements in (11) and is minimax adaptive on $\mathcal{F}^*$.

5 Proof of Theorems

In what follows C, C', C'', C''' stand for positive constants which values may depend only on the parameters s, p and q of Besov classes.

5.1 Proof of Proposition 1

Suppose that the statement of the proposition does not hold, i.e. that there is a confidence ball $\hat{\mathcal{B}}_N=\mathcal{B}(f_N,\tau_N)$ on $\mathcal{F}$ such that

$$\lim_{N\to\infty}\frac{\sup_{f\in\mathcal{F}}E_f\tau_N}{\varphi_N(\mathcal{F})}=0.$$

This implies that there exist an infinite sequence $\beta_N \to 0$ such that

$$\sup_{f\in\mathcal{F}} P_f(\tau_N \le \beta_N \varphi_N(\mathcal{F})) \ge 7/8.$$

Since $\hat{\mathcal{B}}_N$ is a confidence ball, we have from the definition (cf. (3)) that

$$\sup_{f\in\mathcal{F}} P_f(\|f_N - f\|_2 \le \beta_N \varphi_N(\mathcal{F})) \ge \frac{7}{8} - \alpha \ge 3/4,$$

what is clearly impossible, for instance, due to the lower bound in [21]. ∎

5.2 Proof of Proposition 2

Our objective is to reduce the problem of confidence ball estimation to that of testing a simple hypothesis $H_0:\ f = f_0$ versus a complex alternative $H_1 : f \in F_M$, where the family $F_M = \{f_1, ..., f_M\}$ is constructed by small perturbation of f_0, so that the functions $f_1, ..., f_M$ are at a small distance $\asymp c_N \rho_N^*$, $c_N \to 0$, from f_0 (we use the construction, described in Section 5.1 of [17]). We are to show that if a confidence ball $\hat{\mathcal{B}}_N$ with desired properties can be constructed, then one could provide a consistent test $\bar{T}_N$ of H_0 *versus* H_1, i.e., such that the sum of probabilities of the error of the first and second type is less than 1. Then we apply the bound on the error probabilities, obtained in [17], which states that for any test T_N that sum is ≥ 1 in that setting to obtain a contradiction.

Lemma 1 *One can point out a finite family $F_M \in \mathcal{F}^*$ with the following properties:*

- *for any $f \in F_M$*

$$\|f - f_0\|_2 \ge 3a_N;$$

- *in the problem of testing the hypothesis H_0 versus H_1 in the following setting:*

$$\begin{aligned} H_0 &: f = f_0, \\ H_1 &: f \in F_M; \end{aligned} \tag{23}$$

  for any test T_N of the hypothesis H_0 (a measurable function of observations with values in $\{0, 1\}$, $T_N = 0$ means that H_0 is accepted and $T_N = 1$ means that H_0 is rejected) and any N large enough it holds:

$$P_{f_0}(T_N = 1) + \sup_{f\in F_M} P_f(T_N = 0) \ge 7/8. \tag{24}$$

Proof: Let us construct the functional family F_M. Let G be a smooth function on $[0,1]$ and let h be a small parameter. We denote by $\mathcal{I}$ a partition of $[0,1]$ into m intervals of length h. We choose

$$h = h_N = \begin{cases} \left(\frac{c_N^2 \sigma_w^2}{NL}\right)^{\frac{1}{2s^*+1-1/p^*}} & \text{for} \quad p^* < 2, \\ \left(\frac{c_N^2 \sigma_w^2}{NL}\right)^{\frac{1}{2s^*+1/2}} & \text{for} \quad p^* \geq 2, \end{cases}$$

where $c_N = \frac{4a_N}{\rho_N^*}$, so that $c_N \to 0$ as $N \to \infty$. Without loss of generality we may assume that $mh = 1$. For each interval $I \in \mathcal{I}$ we denote by t_I its center.

Consider the family of functions $\psi_I(\cdot)$ on $[0,1]$, $I \in \mathcal{I}$, where

$$\psi_I(t) = \frac{1}{\sqrt{h}\|G\|_2} G\left(\frac{t - t_I}{h}\right).$$

Now consider the random function

$$f(t) = f_0(t) + \frac{c_N}{\sqrt{N}} \sum_{I \in \mathcal{I}} \xi_I \psi_I(t),$$

where ξ_I, $I \in \mathcal{I}$ are independent and identically distributed random variables with values in the set $\{-1, 0, 1\}$ with

$$P(\xi_I = 0) = 1 - \sqrt{h}, \quad P(\xi_I = \pm 1) = \sqrt{h}/2, \quad I \in \mathcal{I}.$$

Let π_N be the distribution of f. Let Z_{π_N},

$$Z_{\pi_N} = \frac{dP_{\pi_N}}{dP_0},$$

be the likelihood ration. Here the measure P_0 of observations corresponds to $f = f_0$ and P_{π_N} is the Bayes measure for model (1). In the same way as it is done in Section 5.1 of [17] we verify that

$$Z_{\pi_N} \stackrel{P_0}{\to} 1. \tag{25}$$

On the other hand, following the proof of Lemmas 5.1 and 5.2 of [17] one show that as $N \to \infty$

$$P(f \in \mathcal{F}^*) \to 1, \quad \text{and} \quad P(\|f - f_0\|_2 \geq \frac{3}{4} c_N \rho_N^*) \to 1. \tag{26}$$

Let now F_M be the set of realizations of random function f such that

$$F_M = \{f : \ f \in \mathcal{F}^* \text{ and } \|f - f_0\|_2 \geq 3a_N\}.$$

We observe that (26) implies that

$$P(f \in F_M) \to 1 \text{ as } N \to \infty. \tag{27}$$

Now consider the problem of testing the hypothesis $H_0 : \ f = f_0$ against the alternative $H_1 : \ f \in F_M$. As a consequence to (25) and (27), we obtain using the result by Ingster, [12], that for any test T_N,

$$\lim_{N \to \infty} \left[P_{f_0}(T_N = 1) + \sup_{f \in F_M} P_f(T_N = 0) \right] \geq 1,$$

what implies the lemma. ∎

Let us return to the proof of the proposition. Suppose that a confidence interval $\hat{\mathcal{B}}'_N = \mathcal{B}(f'_N, \tau'_N)$ of level $1-\alpha$ on $\mathcal{F}^*$ with some f'_N and τ'_N which satisfy the premises of Proposition 2 can be constructed. Consider the following test in the problem (23):

$$T_N = 1_{\tau_N \le a_N} 1_{f'_N \in \mathcal{B}(f_0, \tau'_N)}.$$

Let us study the error probabilities of this test. First, we have

$$\begin{aligned} P_{f_0}(T_N = 1) &= P_{f_0}\left(\{\tau'_N > a_N\} \cup \{\|f'_N - f_0\|_2 > \tau'_N\}\right) \\ &\le P_{f_0}(\tau'_N > a_N) + P_{f_0}(\|f'_N - f_0\|_2 > \tau'_N) < 1/2 + \alpha < 5/8 \end{aligned}$$

(to obtain the second inequality we have used the fact that $\hat{\mathcal{B}}_N$ is a confidence ball of level $1-\alpha$ with $\alpha \le 1/8$). By (24) this implies that

$$\sup_{f \in F_M} P_f(T_N = 0) > \frac{7}{8} - \frac{5}{8} > 1/4. \tag{28}$$

On the other hand, if $T_N = 0$ is realized, i.e. $\tau'_N \le a_N$ and $\|f'_N - f_0\| \le \tau'_N$, then for any $f \in F_M$,

$$\|f - f'_N\|_2 \ge \|f - f_0\|_2 - \|f'_N - f_0\|_2 \ge 3a_N - \tau'_N \ge 2a_N > \tau'_N.$$

Thus for any $f \in F_M$

$$P_f(\|f - f'_N\|_2 > \tau'_N) \ge P_f(T_N = 0),$$

and, due to (28),

$$\sup_{f \in \mathcal{F}^*} P_f(\|f - f'_N\|_2 > \tau'_N) \ge \sup_{f \in F_M} P_f(\|f - f'_N\|_2 > \tau'_N) \ge \sup_{f \in F_M} P_f(T_N = 0) > 1/4.$$

The latter bound implies that $\hat{\mathcal{B}}_N$ cannot be a confidence ball of level $1-\alpha$ with $\alpha \le 1/8$. This contradiction proves the proposition. ∎

5.3 Proof of Theorem 1

Proof of the bound (20). This is a simple consequence of the properties of wavelet adaptive estimator, presented in [7] (or [13]). Indeed, let $\widehat{f}_N^{(a)}$ be a wavelet adaptive estimate, constructed in [7]. Note that $\widehat{f}_N^{(a)}$ is minimax on any class $\mathcal{F}(s,p,q,L)$ with $0 < s \le m$, $s > 1/p$. Suppose that λ_j^m, $j = 1, ..., j_0$ are the thresholds of the minimax on $\mathcal{F}^*$ estimator, defined in [6]. One can easily check that the thresholds λ_j in (13) verify $\lambda_j\sigma \le \lambda_j^m$ for $0 \le j \le j_0$. Recall that higher thresholds correspond to the estimation of functions of higher regularity. This implies that the estimate $\widehat{f}_N$ with the thresholds $\lambda_j^* = \max(\lambda_j^{(a)}, \lambda_j\sigma)$, still attains the minimax rate of convergence on any Besov class $\mathcal{F}(s,p,q,L)$ with $1/p < s \le m$, embedded in $\mathcal{F}^*$.

Proof of the bound (21). We start with the translation of our estimation problem into the space of the sequences of wavelet coefficients. For the sake of simplicity we suppose that $N = 2^{j_0}$. For the computation of wavelet coefficients in the case $N \neq 2^{j_0}$ the reader can refer to [5].

We set

$$f_{j_0}(x) = \alpha'\phi(x) + \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1}\beta'_{jk}\psi_{jk}(x), \tag{29}$$

where

$$\alpha' = \frac{1}{N}\sum_{i=1}^{N} f(\frac{i}{N})\phi(\frac{i}{N}), \quad \beta'_{jk} = \frac{1}{N}\sum_{i=1}^{N} f(\frac{i}{N})\psi_{jk}(\frac{i}{N}).$$

Then the empirical wavelet coefficients satisfy:

$$\widehat{\alpha} = \alpha' + \zeta, \quad y_{jk} = \beta'_{jk} + \zeta_{jk},$$

with

$$\zeta = \frac{1}{N}\sum_{i=1}^{N} w_i\phi(\frac{i}{N}), \quad \zeta_{jk} = \frac{1}{N}\sum_{i=1}^{N} w_i\psi_{jk}(\frac{i}{N}).$$

We present here a summary of properties of the sequence of empirical wavelet coefficients. The following statement is an immediate corollary of Proposition 1 in [5].

Proposition 3 *Suppose that $f \in \mathcal{F}(s,p,q,L)$ with $s > 1/p$. Then there is a constant C_0 (which depends on the wavelet used) such that the sequence $\beta' = (\alpha', \beta_{jk})$ satisfies*

$$\beta' \in \mathcal{F}(s,p,q,C_0L) \textit{ and } \|f - f_{j_0}\|_2 = O(L2^{-j_0s'}) = O\left(LN^{-s'}\right), \tag{30}$$

where $s' = s - 1/p + 1/2$ for $p < 2$ and $s' = s$ for $p \geq 2$.

Denote

$$m'_N = \|\widehat{f}_N - f_{j_0}\|_2 = \left[\sum_{j=0}^{j_0}\|\widehat{\beta}_{j\cdot} - \beta'_{j\cdot}\|_2^2\right]^{1/2}. \tag{31}$$

Then due to (30) we can bound $m_N = \|\widehat{f}_N - f_{j_0}\|_2$ as follows:

$$|m_N - m'_N| \leq \|f - f_{j_0}\|_2 \leq CLN^{-s'}.$$

This implies immediately that

$$\left|[E_f(\widehat{m}_N - m_N)^2]^{1/2} - [E_f(\widehat{m}_N - m'_N)^2]^{1/2}\right| \leq CLN^{-s'}. \tag{32}$$

So to show the second statement of Theorem 1 it suffices to control the value $[E_f(\widehat{m}_N - m'_N)^2]^{1/2}$. Furthermore, it follows from (30) that the coefficient β'_{jk} satisfies (up to a known "absolute constant") the same norm relation (7) as the true coefficients β_{jk}. Since this is the only property of wavelet coefficients used in the study of the estimate $\widehat{m}_N$, with some abuse of notations we substitute in the sequel β'_{jk} for β_{jk}, what gives the model

$$y_{jk} = \beta_{jk} + \zeta_{jk} \tag{33}$$

for empirical wavelet coefficients.

Now note random variable ζ and ζ_{jk} have Gaussian distribution with $E\zeta = E\zeta_{jk} = 0$. Since the sequences $\psi_{jk}\left(\frac{i}{N}\right)$, $i = 1, \dots, N$ are orthonormal for different j and k, the variables ζ_{jk} are mutually independent and $E\zeta_{jk}^2 = \frac{\sigma_w^2}{N}$.

In what follows we consider here only the case $p^* < 2$. The proof in the case $p^* \geq 2$ follows the lines of that of Theorem 2 in [15].

We denote

$$\gamma_{jk} = \beta_{jk} - \widehat{\beta}_{jk}, \quad z_{jk} = y''_{jk} - \widehat{\beta}_{jk}$$

and

$$\zeta'_{jk} = y'_{jk} - \beta_{jk} = \zeta_{jk} + \xi_{jk} \quad \zeta''_{jk} = y''_{jk} - \beta_{jk} = \zeta_{jk} - \xi_{jk}.$$

We conclude from (33) that (ζ'_{jk}) and (ζ''_{jk}) are two non-correlated (and thus mutually independent) sequences of independent and identically distributed Gaussian random variables with $E\zeta'_{jk} = E\zeta''_{jk} = 0$ and $E(\zeta'_{jk})^2 = E(\zeta''_{jk})^2 = \frac{2\sigma_w^2}{N}$.

Note that the quantity m'_N, defined in (31) can be expressed as $m'_N = \left[\sum_{j=0}^{j_0} \|\gamma_{j\cdot}\|_2^2\right]^{1/2}$. Then the difference $\widehat{m}_N^2 - (m'_N)^2$ can be rewritten as:

$$\begin{aligned}
|\widehat{m}_N^2 - (m'_N)^2| &\leq \left| \sum_{j=0}^{j^*} (\|z_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2 - \|\gamma_{j\cdot}\|_2^2) + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} (z_{jk}^2 - 2\sigma^2 - \gamma_{jk}^2) 1_{|y'_{jk}| \geq \lambda_j \sigma} \right. \\
&\quad \left. - \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \gamma_{jk}^2 1_{|y'_{jk}| < \lambda_j \sigma} \right| \\
&= \left| \sum_{j=0}^{j^*} \|\zeta''_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2 \right| + \left| \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} ({\zeta''_{jk}}^2 - 2\sigma^2) 1_{|y'_{jk}| \geq \lambda_j \sigma} \right| \\
&\quad + 2 \left| \sum_{j=0}^{j^*} \gamma_{j\cdot}^T \zeta''_{j\cdot} + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \gamma_{jk}^T \zeta''_{jk} 1_{|y'_{jk}| \geq \lambda_j \sigma} \right| + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|y'_{jk}| < \lambda_j \sigma} \\
&= \sum_{i=1}^{4} \delta_N^{(i)}.
\end{aligned} \tag{34}$$

We have the following immediate inequality:

$$\left[E(\delta_N^{(1)})^2\right]^{1/2} \leq C 2^{j^*/2} \sigma^2. \tag{35}$$

In order to continuer we need some technical results.

Lemma 2

$$\begin{aligned} 1_{|y_{jk}|\geq\lambda\sigma} &\leq 1_{|\beta_{jk}|\geq\lambda\sigma/2} + 1_{|\zeta_{jk}|>\lambda\sigma/2}; && (36)\\ 1_{|y_{jk}|<\lambda\sigma} &\leq \sum_{l=0}^{\infty} 1_{|\beta_{jk}|<(2+l)\lambda\sigma} 1_{|\zeta_{jk}|\geq l\lambda\sigma}; && (37) \end{aligned}$$

Proof: The proof of (36) is immediate. To show (37) we decompose $1_{|y_{jk}|<\lambda\sigma}$:

$$\begin{aligned} 1_{|y_{jk}|<\lambda\sigma} &= 1_{|y_{jk}|<\lambda\sigma} 1_{|\beta_{jk}|<2\lambda\sigma} + 1_{|y_{jk}|<\lambda\sigma} \sum_{l=2}^{\infty} 1_{l\lambda\sigma\leq|\beta_{jk}|<(l+1)\lambda\sigma} \\ &\leq 1_{|\beta_{jk}|<2\lambda\sigma} + \sum_{l=1}^{\infty} 1_{|\beta_{jk}|<(l+2)\lambda\sigma} 1_{|\zeta_{jk}|>l\lambda\sigma} \leq \sum_{l=0}^{\infty} 1_{|\beta_{jk}|<(l+2)\lambda\sigma} 1_{|\zeta_{jk}|\geq l\lambda\sigma} \end{aligned}$$

∎

Lemma 3 $\left[E(\delta_N^{(2)})^2\right]^{1/2} \leq C\left(2^{j^*/2}\sigma^2 + (L^*)^{p^*/2}\sigma^{2-p^*/2}2^{-j^*(s^*p^*+p^*/2-1)/2}\right)$.

Proof: In the decomposition below we use (36) and first take the expectation over the distribution of (ζ''_{jk}) and then over that of (ζ'_{jk}):

$$\begin{aligned} E(\delta_N^{(2)})^2 &= 8\sigma^4 E\left(\sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1} 1_{|y'_{jk}|\geq\lambda_j\sigma}\right) \leq 8\sigma^4 \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1}\left(1_{|\beta_{jk}|\geq\frac{\lambda_j\sigma}{2}} + P(|\zeta'_{jk}| > \frac{\lambda_j\sigma}{2})\right) \\ &\leq 2^{p^*+3}\sigma^4 \sum_{j=j^*+1}^{j_0} \frac{\|\beta_{j\cdot}\|_{p^*}^{p^*}}{(\lambda_j\sigma)^{p^*}} + 16\sigma^4 \sum_{j=j^*+1}^{j_0} 2^j \exp(-\frac{\lambda_j^2}{16}) = I_N^{(1)} + I_N^{(2)}. && (38) \end{aligned}$$

Let us estimate $I_N^{(1)}$. Recall that $\lambda_j = \kappa\sqrt{j-j^*}$. Due to the definition of the class $\mathcal{F}(s^*, p^*, \infty, L)$,

$$\begin{aligned} I_N^{(1)} &\leq \frac{2^{p^*+3}\sigma^{4-p^*}(L^*)^{p^*}}{\kappa^{p^*}} \sum_{j=j^*+1}^{j_0} \frac{2^{-j(s^*p^*+p^*/2-1)}}{(j-j^*)^{p^*/2}} \\ &\leq \frac{2^{p^*+3}\sigma^{4-p^*}(L^*)^{p^*}}{\kappa^{p^*}} 2^{-j^*(s^*p^*+p^*/2-1)} \sum_{l=1}^{\infty} \frac{2^{-l(s^*p+p^*/2-1)}}{l^{p^*/2}} \\ &\leq C(L^*)^{p^*}\sigma^{4-p^*}2^{-j^*(s^*p^*+p^*/2-1)}. && (39) \end{aligned}$$

Since κ in the definition (13) of λ_j satisfy $\kappa^2 > 16\log 2$, we can bound $I_N^{(2)}$ as follows:

$$I_N^{(2)} < 16\sigma^4 2^{j^*} \sum_{l=1}^{\infty} 2^l \exp(-\frac{\kappa^2 l}{16}) \leq 16\sigma^4 2^{j^*} \sum_{l=1}^{\infty} \exp\left(l(\log 2 - \frac{\kappa^2}{16})\right) \leq C\sigma^4 2^{j^*}. \quad (40)$$

When substituting (39) and (40) into (38) we obtain the bound announced in the lemma. ∎

Lemma 4 *Suppose that N is large enough, so that $L^* \geq \sigma$. Then*

$$\left[E(\delta_N^{(3)})^2\right]^{1/2} \leq C'((L^*)^{\frac{2}{4s^*+1}}\sigma^{\frac{8s^*}{4s^*+1}} + \sqrt{\log N}\sigma^2).$$

Recall that it holds for the adaptive estimate $\widehat{\beta}_N$

$$\sup_{f\in\mathcal{F}(s^*,p^*,\infty,L^*)} E_f\|\widehat{\beta}_N - \beta\|_2^2 \leq C\left[(L^*)^{\frac{2}{2s^*+1}}\sigma^{\frac{4s^*}{2s^*+1}} + \sigma^2\log N\right]$$

(cf, for instance Theorem 1 in [7]). When taking the expectation over the distribution of (ζ''_{jk}) and then over the distribution of (ζ'_{jk}) we obtain:

$$\begin{aligned}\left[E(\delta_N^{(3)})^2\right]^{1/2} &\leq 2\left[E\sum_{j=0}^{j_0}\|\gamma_{j\cdot}\|_2^2\sigma^2\right]^{1/2} \leq 2\left[E\|\gamma\|_2^2\sigma^2\right]^{1/2} \\ &\leq C\left[(L^*)^{\frac{1}{2s^*+1}}\sigma^{\frac{4s^*+1}{2s^*+1}} + \sqrt{\log N}\sigma^2\right].\end{aligned}$$

If $L^* \geq \sigma$ then

$$\left[E(\delta_N^{(3)})^2\right]^{1/2} \leq C'\left[(L^*)^{\frac{2}{4s^*+1}}\sigma^{\frac{8s^*}{4s^*+1}} + \sqrt{\log N}\sigma^2\right].$$

∎

Lemma 5 $\left[E(\delta_N^{(4)})^2\right]^{1/2} \leq C(L^*)^{p^*}\sigma^{2-p^*}2^{-j^*(s^*p^*+p^*/2-1)}$.

Proof: The decomposition (37) yields

$$\begin{aligned}\left[E(\delta_N^{(4)})^2\right]^{1/2} &\leq \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1}\sum_{l=0}^{\infty}\beta_{jk}^2 1_{|\beta_{jk}|<(2+l)\lambda_j\sigma}P^{1/2}\left(|\zeta_{jk}| \geq l\lambda_j\sigma\right) \\ &\leq \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1}|\beta_{jk}|^{p^*}\sum_{l=0}^{\infty}|\beta_{jk}|^{2-p^*}1_{|\beta_{jk}|<(2+l)\lambda_j\sigma}P^{1/2}\left(|\zeta'_{jk}| \geq l\lambda_j\sigma\right) \\ &\leq \sqrt{2}\sum_{j=j^*+1}^{j_0}L^{p^*}2^{-j(s^*p^*+p^*/2-1)}\lambda_j^{2-p^*}\sigma^{2-p^*}\sum_{l=0}^{\infty}(2+l)^{2-p^*}\exp\left(-\frac{l^2\lambda_j^2}{4}\right) \\ &\leq C(L^*)^{p^*}\sigma^{2-p^*}\sum_{j=j^*+1}^{\infty}2^{-j(s^*p^*+p^*/2-1)}(j-j^*)^{(2-p^*)/2} \\ &\leq C'(L^*)^{p^*}\sigma^{2-p^*}2^{-j^*(s^*p^*+p^*/2-1)}.\end{aligned}$$

∎

From (34), using the results of (35) and Lemmas 3 – 5, we obtain

$$\begin{aligned}\left[E(\widehat{m}_N^2-(m_N')^2)^2\right]^{1/2} \quad &\leq \quad C\Big(2^{j^*/2}\sigma^2+(L^*)^{p^*/2}\sigma^{2-p^*/2}2^{-j^*(s^*p^*+p^*/2-1)/2}\\ &+(L^*)^{p^*}\sigma^{2-p^*}2^{-j^*(s^*p^*+p^*/2-1)}+(L^*)^{2/(4s^*+1)}\sigma^{8s^*/(4s^*+1)}+\sqrt{\log N}\sigma^2\Big).\end{aligned}$$

The choice of j^* in (12) now results in the bound

$$\left[E(\widehat{m}_N^2-(m_N')^2)^2\right]^{1/2} \quad \leq \quad C\left((L^*)^{\frac{1}{2s^*+1-1/p^*}}\sigma^{\frac{4s^*+1-2/p^*}{2s^*+1-1/p^*}}+\sqrt{\log N}\sigma^2\right). \tag{41}$$

Note that (cf. proof of Theorem 2 in [15]) for any $\gamma>0$

$$(\widehat{m}_N-m_N)^2\leq\frac{(\widehat{m}_N^2-m_N^2)^2}{\gamma^2}+2(\widehat{m}_N^2-m_N^2)+\gamma^2.$$

We set

$$\gamma=\sqrt{(L^*)^{\frac{1}{2s^*+1-1/p^*}}\sigma^{\frac{4s^*+1-2/p^*}{2s^*+1-1/p^*}}+\sigma^2\log N}$$

Then (21) follows from (32) and (41). The last statement of the theorem is an immediate consequence of (20), (21) and the Chebychev inequality. ∎

References

[1] L. Birgé and P. Massart, *From model selection to adaptive estimation*, in Festschrift for Lucien Le Cam: research papers in probability and statistics, ed. D. Pollard, New York, Springer, 55-87, 1997.

[2] A. Cohen, I. Daubechies, B. Jawerth and P. Vial, Multiresolution analysis, wavelets and fast algorithms on an interval, *C.R.* **316**, 417-421, 1992.

[3] I. Daubechies, *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[4] B. Delyon, A. Juditsky, On Minimax wavelet estimators, *Applied Computational Harmonic Analysis*, 3, 215-228, 1996

[5] B. Delyon, A. Juditsky, On the Computation of wavelet coefficients, *Journal of Approximation Theory*, **88**, 1, 47-79, 1997

[6] D. Donoho, I. Johnstone, Minimax estimation via wavelet shrinkage, *Ann. of Stat.*, **26**, 3, 879-921, 1998.

[7] D. Donoho, I. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, Technical Report, Department of Statistics, Stanford University, 1993.

[8] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: asymptotia? *J. R. Statist. Soc. B*, **57**, 2, 301-369, 1995.

[9] S. Efromovich, Sequential nonparametric estimation with assigned risk, *Ann. of Stat.*, **23**, 4, 1376-1392, 1995.

[10] P. Hall, On bootstrap confidence intervals in nonparametric regression, *Ann. of Stat.*, **20**, 2, 695-711, 1992.

[11] Yu. Ingster, Minimax nonparametric detection of signals in white Gaussian noise, *Probl. Inform. Transm.*, **18**, 130-140, 1982.

[12] Yu. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Stat.*, **2**, No 2, 85-114, 1993.

[13] A. Juditsky, Wavelet estimators: adapting to unknown smoothness, *Math. Methods of Stat.*, **6**, 1, 1-20, 1997.

[14] A. Juditsky, S. Lambert-Lacroix, *Confidence Intervals for Adaptive Regression Estimation*, INRIA Technical Report, RR-3643, 1999.

[15] A. Juditsky, O. Lepski, *Confidence Intervals for Adaptive Regression Estimation*, INRIA Technical Report, RR-3580, INRIA, 1998.

[16] O. Lepski, Asymptotically minimax adaptive estimation I. *TVIP*, **36**, 4, 645-659, 1991 (in Russian). Translated as *SIAM J. Theory of Probab. and Appl.*.

[17] O. Lepski and V. Spokoiny, Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative, *Bernoulli* **5**, 2, 333-358 (1999).

[18] O. Lepski, How to improve the accuracy of estimation. to appear in *Math. Methods of Stat.*.

[19] M. Low, On nonparametric confidence intervals. *Ann. of Stat.*, **25**, 2547-2554, 1997.

[20] Y. Meyer, *Ondelettes* Hermann, Paris, 1990.

[21] A. Nemirovski, Nonparametric estimation of smooth regression functions, *Sov. J. Comput. Syst. Sci.* 23, No.6, 1-11 (1985); translation from *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, No.3, 30-50, (1985).

[22] M. Neumann, Automatic bandwith choice and confidence intervals in nonparametric regression, *Ann. of Stat.*, **23**, 6, 1937-1959, 1995.

[23] D. Picard, K. Tribouley, Adaptive confidence interval for pointwise curve estimation *Ann. of Stat.*, **28**, 1, 298-335, 2000.

[24] H. Triebel, *Theory of Function Spaces II.* Birkhäuser Verlag, Berlin, 1992.