Confidence Intervals for Adaptive Regression Estimation on the Besov Spaces

Anatoli B. Juditsky, Sophie Lambert-Lacroix

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Confidence intervals for adaptive regression estimation on the Besov spaces

Anatoli Juditsky, Sophie Lambert-Lacroix

No 3643

mars 99

THÈME 4

apport de recherche

ISSN 0249-6399

INRIA
RHÔNE-ALPES

Confidence intervals for adaptive regression estimation on the Besov spaces

Anatoli Juditsky, Sophie Lambert-Lacroix

Thème 4 — Simulation et optimisation

de systèmes complexes

Projet IS2

Rapport de recherche n°3643 — mars 99 — 25 pages

Abstract: The problem of adaptive estimation of the regression function f from noisy observations is concerned. A confidence interval for the L_2-error for wavelet adaptive estimator is provided. We show that if f belongs to a Besov class, the proposed confidence interval is minimax.

Key-words: Adaptive estimation, nonparametric regression, confidence intervals, wavelet estimators.

(Résumé : tsvp)

Unité de recherche INRIA Rhône-Alpes

655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)

Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00

Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Intervalles de confiance pour des estimateurs adaptatifs sur des classes de Besov

Résumé : Le thème de ce rapport se situe dans le cadre de l'estimation adaptative d'une fonction de régression f à partir d'observations bruitées. On s'intéresse à la norme L_2 de l'erreur d'estimation pour des estimateurs adaptatifs par ondelettes. Nous proposons un estimateur pour cette norme et nous montrons que ce dernier est minimax sur les classes de Besov. Finalement un intervalle de confiance est associé à cet estimateur.

Mots-clé : Estimation adaptative, régression non paramétrique, intervalles de confiance, estimateurs par ondelettes.

1 Introduction

We consider the problem of recovering of 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 body $\mathcal{F}(s,p,q,L)$ (refer to Section 2 for definitions), defined with the parameters (s,p,q,L). We consider a classical L_2-risk

$$\rho(\widehat{f}_N, f) = E_f\|\widehat{f}_N - f\|_2^2 = \int_0^1 (\widehat{f}_N(x) - f(x))^2 dx.$$

It is well known how to construct the minimax on the class $\mathcal{F}(s,p,q,L)$ estimator $\widehat{f}_N$, i.e the function $\widehat{f}_N^*$ which is the minimizer of

$$R(\widehat{f}_N, \mathcal{F}) = \sup_{f\in\mathcal{F}} E_f \rho(\widehat{f}_N, f),$$

when the parameters (s,p,q,L) of the class are known *a priori.*

Furthermore, it is possible to provide "adaptive" estimation algorithms (cf. for instance [1], [6] and [10]) which only use the observations but not the values of the parameters (s,p,q,L) and deliver an estimate $\widehat{f}_N$ of not worse quality than the parameter-dependent ones. That is the ratio of the estimate risk $R(\widehat{f}_N, \mathcal{F})$ and the minimax risk $R(\widehat{f}_N^*, \mathcal{F})$ remains finite as $N \to \infty$. However, those adaptive estimation algorithms do not typically provide any information about the error $m_N = \|\widehat{f}_N - f\|_2$.

If it is known *a priori* that $f \in \mathcal{F}(s,p,q,L)$ then one can take $\widehat{m}_N^*$ as the minimax rate of convergence on the class $\mathcal{F}(s,p,q,L)$. On the other hand, this class contains also functions which can be estimated with better rate. Indeed, this class contains, for instance, the balls $\mathcal{F}(s',p,q,L)$ with $s' > s$. Then if the information that the unknown function belongs to such an embedded class was available, one can expect to find an estimate $\widehat{f}_N$ which attains the minimax rate of convergence which corresponds to the parameters (s',p,q,L) of this smaller class and the bound $\widehat{m}_N^*$ would be rather pessimistic. On the other hand, it is also known that if the accuracy of estimation is characterized with the L_∞-norm of the error the bound $\widehat{m}_N^*$ cannot be improved in the minimax sense (cf. [13]). This motivate our choice of the L_2-norm of the error $m_N = \|\widehat{f}_N - f\|_2$ as the measure of the quality of the estimate to be accessed. In fact we want to point out the value $\tau_N(\alpha)$ (a confidence interval) such that for any $\alpha > 0$ the ball $\mathcal{B}_{L_2}(\hat{f}_N, \tau_N(\alpha))$ in the L_2-space, centered at $\hat{f}_N$ with radius $\tau_N(\alpha)$, satisfies

$$P_f(f \in \mathcal{B}_{L_2}(\hat{f}_N, \tau_N(\alpha)) \geq 1 - \alpha. \tag{2}$$

In order to characterize the quality of the bound τ_N we use the quadratic error:

$$[E_f(\tau_N - m_N)^2]^{1/2}. \tag{3}$$

The problem of constructing the confidence interval for adaptive regression estimation has been studied in [11] in the context of the Sobolev classes[1]. Our objective here is to provide a method to construct confidence intervals $\tau_N(\alpha)$ for adaptive threshold estimators proposed in [5], [6] or [10]. We show that if no *a priori* knowledge of the parameters (s, p, q, L) of the functional class is available, the problem of construction of confidence intervals cannot be solved in the minimax sense. Indeed, in Theorem 1 below, a minimax on the class $\mathcal{F}(s, p, q, L)$ lower bound for the rate of convergence of an estimate $\widehat{m}_N$ of the error m_N is established. We observe that the lower bound in (9) does not vanish if, for instance, the only information available is that the unknown function is bounded (note that the adaptive estimator $\widehat{f}_N$ is a good performer in this case). On the other hand, we show in Theorem 2 that if it is *a priori* known that $f \in \mathcal{F}(s^*, p^*, \infty, L)$, one can construct an estimate $\widehat{m}_N$ which is minimax optimal (up to a constant). Finally, a confidence interval is associated with this estimate.

The paper is organized as follows: in Section 2 we recall some basic properties of Besov classes and adaptive wavelet estimators. Then in Section 3 the lower bound for the rate of convergence for L_2-error estimators is established. Next in Section 4 we provide an confidence interval for adaptive estimators $\widehat{f}_N$ on the Besov class.

2 Adaptive wavelet estimators

We start with the definition of functional classes used.

2.1 Besov body

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 $L_2(0, 1)$ [14], [3]. Let $m = \max(1, s_{\max})$. We suppose that ϕ and $\psi \in C^m$. This implies (see Ch. 7, [3]) that $\psi(x)$ has $l = [s_{\max}]$ 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

[1] Note that this problems is closely related to that of sequential estimation (cf. [1])

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

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

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:

$$f(x) = \alpha\phi(x) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x). \tag{4}$$

We suppose that $\mathcal{F}$ is a "Besov body"[2] of wavelet coefficients:

$$\begin{aligned} \mathcal{F} &= \mathcal{F}(s,p,q,L) \\ &= \{f \text{ such that } (\sum_k |\alpha_k|^p)^{1/p} + \left(\sum_{j=0}^{\infty}(2^{j(s+d/2-d/p)}(\sum_{k\in\mathbf{Z}} |\beta_{jk}|^p)^{q/p}\right)^{1/q} \le L\}. \end{aligned} \tag{5}$$

Following [5], [7] 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 [14].

Note that if B^s_{pq}, $s \ge (p^{-1}-1)_+$, $0 < p, q \le \infty$, is the Besov space (see [15]), then there is $C > 0$ such that

$$\|f\|_{B^s_{pq}} \ge C\|f\|_{spq}, \tag{6}$$

where $\|f\|_{B^s_{pq}}$ is the norm of the Besov space and

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

On the other hand, for any $(p^{-1}-1)_+ < s < m$, there exists $C < \infty$ such that

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

(cf. Theorem 2 in [4]. See also [7] for a discussion and useful references). In what follows with some abuse of notations we refer to $\mathcal{F}(s,p,q,L)$ as the Besov class.

[2] we borrow the terminology of D. Donoho and I. Johnstone [5].

2.2 Wavelet estimator

Consider the following problem: given the observations (1) to design an estimate $\widehat{f}_N$ of f which uses only the observations $y_1, ..., y_N$ (but not the knowledge of parameters s, p, q and L of the class), such that for any class $\mathcal{F}(s, p, q, L)$ the ratio of the estimate risk

$$R(\widehat{f}_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} E_f \|\widehat{f}_N - f\|_2^2$$

to the minimax risk

$$R(\mathcal{F}) = \inf_{\widehat{f}_N} R(\widehat{f}_N, \mathcal{F})$$

remains finite as $N \to \infty$. Following [12] we call such estimates *adaptive in order.*

In the above problem the minimax rates of convergence were established in [5]. These rates are attained by adaptive wavelet estimates $\widehat{f}_N$, designed in [6], [10] and [1]. For our model these estimates are constructed as follows: first we compute the coefficients

$$\widehat{\alpha}_k = N^{-1} \sum_{i=1}^N y_i \phi_k(\frac{i}{N}), \quad y_{jk} = N^{-1} \sum_{i=1}^N y_i \psi_{jk}(\frac{i}{N}), \text{ for } j = 0, \ldots, j_0,$$

where j_0 is such that $\frac{N}{4} < 2^{j_0} \le \frac{N}{2}$. Then y_{jk} are shrinked to zero using the thresholding rule:

$$\widehat{\beta}_{jk} = \delta(y_{jk}, \lambda_j). \tag{7}$$

Here $\delta(\cdot)$ can be hard- or soft-thresholding rule, respectively,

$$\delta(x, \lambda) = x 1_{|x| \ge \lambda} \quad \text{or} \quad \delta(x, \lambda) = \text{sign}(x)(x - \lambda)_+.$$

The threshold λ_j is selected from the observations using a kind of cross-validation procedure which is different for the estimates provided in the papers cited above. Finally one put

$$\widehat{f}_N(x) = \widehat{\alpha}\phi_k(x) + \sum_{j=0}^{j_0} \sum_{k=0}^{2^j - 1} \widehat{\beta}_{jk} \psi_{jk}(x). \tag{8}$$

The risk R of these estimates satisfy

$$R(\widehat{f}_N, \mathcal{F}) \le C L^{2/(2s+1)} \left(\frac{\sigma_w^2}{N}\right)^{\frac{2s}{2s+1}} + O\left(\frac{\sigma_w^2 \log^2 N}{N}\right).$$

3 Lower bound for confidence interval estimation

Suppose that the observations $y_i = f(\frac{i}{N}) + w_i$, $i = 1, ..., N$ of the function f are available. It is known *a priori* that $f \in \mathcal{F}(s,p,q,L)$.

Let $\widehat{f}_N$ be an adaptive estimate of f. Our objective here is to establish the lower bound on the rate of convergence of the estimate of the error $\|\widehat{f}_N - f\|_2$. This bound cannot be given for all estimates $\widehat{f}_N$; indeed, the error of a trivial estimate $\widehat{f}_N(i/N) = y_i$ can be estimated with parametric rate. However, if we limit our consideration to a class of "not-trivial" estimates, which, of course, are the only estimates being of interest, such a bound can be established. Such a class of "reasonably good" estimates can be defined in many ways. We consider here the following

Assumption 1 *hyp The estimate $\hat{f}_N$ is "almost minimax" on $\mathcal{F}(s,p,q,L)$, i.e. if*

$$\nu_N = L^{\frac{1}{2s+1}} \left(\frac{\sigma_w^2 \log N}{N} \right)^{\frac{s}{2s+1}},$$

then for some $C < \infty$

$$\sup_{f \in \mathcal{F}(s,p,q,L)} \nu_N^{-1} \left[E_f \|\hat{f}_N - f\|_2^2 \right]^{1/2} \leq C.$$

ote that this assumption holds for known adaptive estimates (cf. the estimate proposed in [5], [6] or [10])

Let now for some $0 < \delta < 1$, $f_0 \in \mathcal{F}(s,p,q,(1-\delta)L)$. We say that f belongs to $\delta\mathcal{F}_{\beta^0}(s,p,q,L)$ if $f - f_0 \in \mathcal{F}(s,p,q,\delta L)$.

Theorem 1 *Suppose that Assumption 1 holds for the estimate $\hat{f}_N$ of f. Then there is an absolute constant c_0 such that for any $0 < \delta < 1$, $f_0 \in \mathcal{F}(s,p,q,(1-\delta)L)$, $s \geq 1/p$, and any estimate $\hat{m}_N$ of $m_N = \|f - \hat{f}_N\|_2$ it holds*

$$\sup_{f \in \mathcal{F}_{f_0}(s,p,q,L)} \left[E_f (m_N - \hat{m}_N)^2 \right]^{\frac{1}{2}} \geq \begin{cases} c_0 (L\delta)^{\frac{1}{4s+2-2/p}} \left(\frac{\sigma_w^2}{N} \right)^{\frac{2s+1/2-1/p}{4s+2-2/p}}, & \text{for} \quad p < 2; \\ c_0 (L\delta)^{\frac{1}{4s+1}} \left(\frac{\sigma_w^2}{N} \right)^{\frac{2s}{4s+1}}, & \text{for} \quad p \geq 2,\ s \geq 1/4, \\ c_0 L\delta N^{-s}, & \text{for} \quad p \geq 2,\ s < 1/4, \end{cases} \tag{9}$$

The proof of the theorem is put in Section 5.

4 Adaptive threshold estimate with confidence interval

We present in this section an algorithm to estimate m_N for estimates $\hat{f}_N$ proposed in [5], [6] or [10]. Let $\widehat{f}_N$ be an adaptive estimator in the papers cited. We denote by $\lambda^a = (\lambda_1^a, ..., \lambda_{j_0}^a)$, $N/4 \le 2^{j_0} \le N/2$, the adaptive thresholds used in its construction. 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 $f \in \mathcal{F}(s^*, p^*, \infty, L^*)$.

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

$$\begin{array}{llll} \left(\frac{L^*}{\sigma^2}\right)^{\frac{2}{2s^*+1-1/p^*}} \le 2^{j^*} < 2\left(\frac{L^*}{\sigma^2}\right)^{\frac{2}{2s^*+1-1/p^*}}, & \texttt{for} & p^* < 2 \\ \left(\frac{L^*}{\sigma^2}\right)^{\frac{4}{4s^*+1}} \le 2^{j^*} < 2\left(\frac{L^*}{\sigma^2}\right)^{\frac{4}{4s^*+1}}, & \texttt{for} & p^* \ge 2, \end{array} \tag{10}$$

with $j^* = j_0$ when $j^* > j_0$. Moreover define

$$\lambda_j = \begin{cases} \kappa\sqrt{(j-j^*)_+} \ \texttt{with} \ \kappa > 4\sqrt{\log 2} & \texttt{for} \quad p^* < 2, \\ \infty & \texttt{for} \quad p^* \ge 2. \end{cases} \tag{11}$$

1. Compute the empirical wavelet coefficients

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

2. Set

$$y'_{jk} = y_{jk} + \xi_{jk}, \quad y''_{jk} = y_{jk} - \xi_{jk}, \quad \lambda_j^* = \max(\lambda_j^a, \lambda_j);$$

and compute the estimates of wavelet coefficients

$$\widehat{\beta}_{jk} = y'_{jk} 1_{|y'_{jk}| \ge \lambda_j^*} \tag{12}$$

3. To terminate set

$$\widehat{f}_N(x) = \widehat{\alpha}\phi(x) + \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1} \widehat{\beta}_{jk}\psi_{jk}(x) \tag{13}$$

and

$$\widehat{m}_N^2 = \left[\sum_{j=0}^{j^*}\left(\|y''_{j\cdot} - \widehat{\beta}_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2\right) + \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j-1}\left((y''_{jk} - \widehat{\beta}_{jk})^2 - 2\sigma^2\right)1_{|y'_{jk}|\geq\lambda_j\sigma^2}\right]_+ \tag{14}$$

Theorem 2 *Let $\mathcal{F}(s,p,q,L)$ be a Besov class with $s \geq 1/p$ such that $\mathcal{F}(s,p,q,L) \subset \mathcal{F}(s^*,p^*,\infty,L^*)$ with $s^* \geq 1/p^*$. Then*

$$\sup_{f\in\mathcal{F}(s,p,q,L)} [E_f\|f - \widehat{f}_N\|_2^2]^{1/2} \leq C_0 L^{1/(2s+1)}\left(\frac{\sigma_w^2}{N}\right)^{s/(2s+1)} + \epsilon(N), \tag{15}$$

where $\epsilon(N) = O\left(\frac{\sigma_w \log N}{\sqrt{N}}\right)$. It also holds

$$\sup_{f\in\mathcal{F}(s^*,L^*)} \left[E_f(\widehat{m}_N - m_N)^2\right]^{1/2} \leq \begin{cases} C_1 (L^*)^{\frac{1}{4s^*+2-2/p^*}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*+1/2-1/p^*}{4s^*+2-2/p^*}} + \epsilon(N), & p^* < 2, \\ C_1\left((L^*)^{\frac{1}{4s^*+1}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*}{4s^*+1}} + L^* N^{-s^*}\right) + \epsilon(N) & p^* \geq 2. \end{cases} \tag{16}$$

Here C_0 and C_1 are constants which do not depend on N, L^* and σ_w^2 and can be computed explicitly for a given value of s^*, p^* and wavelet ψ.

Remark: Using the bounds in (16) for the error $\widehat{m}_N - m_N$ we can modify the estimation algorithm above to construct a confidence interval τ_N. Indeed, for example when $p^* < 2$, if we set for $\alpha > 0$,

$$\tau_N(\alpha) = \hat{m}_N + \frac{1}{\sqrt{\alpha}}\left[C(L^*)^{\frac{1}{4s^*+2-2/p^*}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*+1/2-1/p^*}{4s^*+2-2/p^*}} + \epsilon(N)\right], \tag{17}$$

we obtain the following evident

Corollary 1 *The quantity $\tau_N(\alpha)$, delivered by Algorithm 1 and (17), satisfies*

$$\left[E(\tau_N(\alpha) - m_N)^2\right]^{1/2} \geq \left(1 + \frac{1}{\sqrt{\alpha}}\right)\left[C(L^*)^{\frac{1}{4s^*+2-2/p^*}}\left(\frac{\sigma_w^2}{N}\right)^{\frac{2s^*+1/2-1/p^*}{4s^*+2-2/p^*}} + \epsilon(N)\right],$$

and for $\alpha < 1$

$$P\left(m_N > \tau_N(\alpha)\right) \leq \alpha.$$

Note that the analogous results hold in the case $p^* \geq 2$.

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 Theorem 1

We consider here only the case $p < 2$. The case $p > 2$ follows the same lines as that of Theorem 1 in [11]. We transfer the problem in the space of wavelet coefficients. We say that $\beta \in \mathbf{R}^N$ belongs to $\mathcal{F}(s,p,q,L)$ if

$$\left(\sum_{j\geq 0} 2^{jq(s+1/2-1/p)}\|\beta_{j\cdot}\|_p^q\right)^{\frac{1}{q}} \leq L.$$

Let $\beta^0 \in \mathbf{R}^N$ be a vector of wavelet coefficients. We say that $\beta \in \delta\mathcal{F}_{\beta^0}(s,p,q,\,L)$ if $\beta-\beta^0 \in \mathcal{F}(s,p,q,\delta L)$. Now suppose that for some $0<\delta\leq 1$, β^0 belongs to $\mathcal{F}(s,p,q,(1-\delta)L)$. Let $y=(y_{jk})_{j=0,\ldots,j_0,k=0,\ldots,2^j-1} \in \mathbf{R}^N$,

$$y_{jk} = \beta_{jk} + \sigma\zeta_{jk} \tag{18}$$

be the observation of the vector $\beta=(\beta_{jk}) \in \mathbf{R}^N$, $\beta \in \delta\mathcal{F}_{\beta^0}(s,p,q,L)$. Note that this also implies that $\beta \in \mathcal{F}(s,p,q,L)$. The vector $\zeta = (\zeta_{jk})$ in (18) consists of independent and identically distributed Gaussian random variables, $E\zeta_1 = 0$, $E\zeta_1^2 = 1$.

Let θ_N an estimate of the quantity $\|\hat{\beta}-\beta\|_2$. The proof of Theorem 1 results from the following

Proposition 1 *For any $0<\delta\leq 1$ and any $\beta^0 \in \mathcal{F}(s,p,q,(1-\delta)L)$, there exists $C>0$*

such that for all N sufficiently large, any estimate θ_N and any estimate $\hat{\beta}$

$$\sup_{\beta\in\mathcal{F}_{\beta^0}(s,p,q,L)} E_{\beta^0}(\theta_N - \|\hat{\beta}-\beta\|_2)^2 \geq C\rho_N^2 \frac{\rho_N^2 - \sigma\left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2}$$

where

$$\rho_N = (\delta L)^{1/(4s+2-2/p)} \sigma^{\frac{4s+1-2/p}{2s+1-1/p}}.$$

Indeed, if we set

$$f^0(x) = \sum_{j=0}^{j_0} \beta^0_{jk}\psi_{jk}(x), \quad \text{and} \quad f(x) = \sum_{j=0}^{j_0} \beta_{jk}\psi_{jk}(x).$$

Then Assumption 1 implies that $\sigma E_{\beta^0}(\|\hat{\beta} - \beta^0\|_2^2)^{1/2} = o(\rho_N^2)$, and

$$\sup_{\beta \in \mathcal{F}_{\beta^0}(s,p,q,L)} \left[E_{\beta^0}(\theta_N - \|\hat{\beta} - \beta\|_2)^2\right]^{1/2} \geq c_0(\delta L)^{1/(4s+2-2/p)} \left(\frac{\sigma_e^2}{N}\right)^{\frac{4s+1-2/p}{4s+2-2/p}}.$$

5.2 Proof of Proposition 1

Let $\beta^0 = (\beta^0_{jk}) \in \mathbf{R}^N$, $\beta^0 \in \mathcal{F}(s,p,q,(1-\delta)L)$ and j^* satisfy

$$\left(\frac{L\delta}{\sigma}\right)^{\frac{2}{2s+1-1/p}} \leq 2^{j^*} < 2\left(\frac{L\delta}{\sigma}\right)^{\frac{2}{2s+1-1/p}}.$$

Note that for N sufficiently large, $2^{j^*} \leq N$ since $s \geq 1/p$. We define

$$\tilde{\beta} = \lambda\sigma = \lambda(\delta L)2^{-j^*(s+1/2-1/2p)}, \quad \text{with} \quad \lambda \leq \min\left(\frac{1}{4}, \left(\frac{2}{3}\right)^{\frac{1}{p}}\right).$$

Next we set

$$\xi_{jk} = \begin{cases} 0, & \text{if} \quad j \neq j^* \\ \xi_k, & \text{if} \quad j = j^*, \end{cases}$$

where (ξ_k), $k = 0, \dots, 2^{j^*}-1$ is a sequence of independent and identically distributed random variables, such that $P(\xi_0 = 1) = P(\xi_0 = -1) = \frac{r}{2}$ and $P(\xi_0 = 0) = 1 - r$ with

$$r = 2^{-j^*/2}. \tag{19}$$

Finally, we define the vector $\beta^{(\xi)}$ in the following way: make another independent drawing such that

$$\beta^{(\xi)} = \begin{cases} \beta^0 + \tilde{\beta}\xi, & \text{with probability} \quad 1/2 \\ \beta^0, & \text{with probability} \quad 1/2. \end{cases}$$

Let

$$\Delta = \{\omega : \|\xi\|_2^2 \le 3r2^{j^*-1}\}. \tag{20}$$

Note that due to the definition of $\tilde{\beta}$, the vector $\beta^{(\xi)}$ belongs to $\delta\mathcal{F}_{\beta^0}(s,p,q,L)$ on Δ. This implies immediately $\beta^{(\xi)} \in \mathcal{F}(s,p,q,L)$ on Δ. Consider the observation $y = (y_{jk})$ of $\beta^{(\xi)}$,

$$y_{jk} = \beta_{jk}^{(\xi)} + \sigma\zeta_{jk},$$

where (ζ_{jk}) is a sequence of independent and identically distributed Gaussian random variables (independent of ξ) $\zeta_{jk} \sim N(0,1)$.

We can write down

$$\sup_{\beta \in \mathcal{F}_{\beta^0}(s,p,q,L)} E_{\beta^0}(\theta_N - \|\hat{\beta} - \beta\|_2)^2 \ge r_{\beta^0}(\delta, N),$$

where

$$r_{\beta^0}(\delta,N) = \frac{1}{2}E_\xi\left\{\left(E_{\beta^{(\xi)}}(\theta_N - \|\hat{\beta}-\beta^{(\xi)}\|_2)^2 + E_{\beta^0}(\theta_N - \|\hat{\beta}-\beta^0\|_2)^2\right)1_\Delta\right\},$$

and E_ξ stands for the expectation with respect to the distribution of ξ.

Let us denote $\delta\beta = \hat{\beta} - \beta^0$ and let Z_ξ stand for the likelihood ratio

$$Z_\xi = \frac{dP_{\beta^{(\xi)}}}{dP_{\beta^0}} = \prod_{k=0}^{2^{j^*}-1} \exp\left(\frac{\tilde{\beta}\zeta_{j^*k}\xi_k}{\sigma} - \frac{\tilde{\beta}^2\xi_k^2}{2\sigma^2}\right) = \prod_{k=0}^{2^{j^*}-1} \exp\left(\lambda\zeta_{j^*k}\xi_k - \frac{\lambda^2\xi_k^2}{2}\right).$$

We set

$$\Xi = \frac{1}{r^2}\sum_{(k,l)\in S} E_\xi\left\{\xi_k^2\xi_l^2 Z_\xi\right\}, \tag{21}$$

where S is a 2-dimensional array of indices

$$S = \left\{(k,l) \in [0,\ldots,2^{j^*}-1]^2 : k \ne l\right\}.$$

We define the events:

$$\begin{aligned}
A &= \{\omega : \quad E_\xi Z_\xi \le 2e^{\lambda^4/2}\},\\
B &= \{\omega : \quad \sum_{k=0}^{2^{j^*}-1} \zeta_{j^*k}^2 \le 22^{j^*}\},\\
D &= \{\omega : \quad \|\delta\beta\|_2 \le 4e^{\frac{\lambda^4}{2}}(E_{\beta^0}\|\delta\beta\|_2^2)^{1/2}\},\\
F &= \{\omega : \quad \Xi \ge \frac{1}{2}2^{j^*}(2^{j^*}-1)\},
\end{aligned}$$

and

$$\Gamma = A \cap B \cap D \cap F.$$

Lemma 1 *There is $C > 0$ such that*

$$r_{\beta^0}(\delta, N) \geq C \frac{E_{\beta^0}\left\{E_\xi[Z_\xi(\|\tilde\beta\xi\|_2^2 - 2\tilde\beta\delta\beta^T\xi)^2 1_\Delta] 1_\Gamma\right\}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2}. \tag{22}$$

Proof: Since

$$\theta_N - \|\hat\beta - \beta^{(\xi)}\|_2 = \theta_N - \|\hat\beta - \beta^0\|_2 - (\|\hat\beta - [\beta^0 + \tilde\beta\xi]\|_2 - \|\hat\beta - \beta^0\|_2),$$

when changing the integration measure, we have for $r_{\beta^0}(\delta, N)$:

$$2r_{\beta^0}(\delta, N) = E_{\beta^0}\left\{E_\xi\left(Z_\xi 1_\Delta(\Delta_N - \|\delta\beta - \tilde\beta\xi\|_2 + \|\delta\beta\|_2)^2\right) + \Delta_N^2 P(\Delta)\right\}, \tag{23}$$

where $\Delta_N = \theta_N - \|\hat\beta - \beta^0\|_2$. Note that

$$\Delta_N^* = \frac{E_\xi\left(Z_\xi[\|\delta\beta - \tilde\beta\xi\|_2 - \|\delta\beta\|_2]1_\Delta\right)}{P(\Delta) + E_\xi Z_\xi 1_\Delta}$$

is the minimizer of (23) with respect to Δ_N. When substituting Δ_N^* into (23) we get

$$\begin{aligned} 2r_{\beta^0}(\delta, N) &\geq E_{\beta^0}\left\{\frac{1}{1 + E_\xi Z_\xi} E_\xi\left[Z_\xi(\|\delta\beta - \tilde\beta\xi\|_2 - \|\delta\beta\|_2)^2 1_\Delta\right]\right\} \\ &= E_{\beta^0}\left\{\frac{1}{1 + E_\xi Z_\xi} E_\xi\left[Z_\xi\left(\frac{\|\delta\beta - \tilde\beta\xi\|_2^2 - \|\delta\beta\|_2^2}{\|\delta\beta - \tilde\beta\xi\|_2 + \|\delta\beta\|_2}\right)^2 1_\Delta\right]\right\}. \end{aligned} \tag{24}$$

Since the expression under the expectation E_{β^0} in (24) is positive, we can bound r_{β^0} from below as follows:

$$2r_{\beta^0}(\delta, N) \geq E_{\beta^0}\left\{\frac{1}{1 + E_\xi Z_\xi} E_\xi\left[Z_\xi\left(\frac{\|\delta\beta - \tilde\beta\xi\|_2^2 - \|\delta\beta\|_2^2}{\|\delta\beta - \tilde\beta\xi\|_2 + \|\delta\beta\|_2}\right)^2 1_\Delta\right] 1_\Gamma\right\}. \tag{25}$$

We use the bounds for $E_\xi Z_\xi$ and $\|\delta\beta\|_2$ on Γ and the one of $\|\tilde\beta\xi\|_2$ on Δ to obtain from (25):

$$\begin{aligned} 2r_{\beta^0}(\delta, N) &\geq E_{\beta^0}\left\{\frac{1}{1 + E_\xi Z_\xi} E_\xi\left[Z_\xi\left(\frac{\|\tilde\beta\xi\|_2^2 - 2\tilde\beta\delta\beta^T\xi}{\|\tilde\beta\xi\|_2 + 2\|\delta\beta\|_2}\right)^2 1_\Delta\right] 1_\Gamma\right\} \\ &\geq \frac{E_{\beta^0} E_\xi\left\{Z_\xi(\|\tilde\beta\xi\|_2^2 - 2\tilde\beta\delta\beta^T\xi)^2 1_\Delta 1_\Gamma\right\}}{(1 + 2e^{\lambda^4})\left(\lambda\sqrt{3/2}\rho_N + 8e^{\frac{\lambda^4}{2}}\left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2}. \end{aligned}$$

∎

In order to continuer we need some notations. We denote

$$\begin{aligned} I_k &= \tfrac{1}{2}\exp\left(\lambda\zeta_{j^*k} - \tfrac{\lambda^2}{2}\right) + \tfrac{1}{2}\exp\left(-\lambda\zeta_{j^*k} - \tfrac{\lambda^2}{2}\right), \\ J_k &= \tfrac{1}{2}\exp\left(\lambda\zeta_{j^*k} - \tfrac{\lambda^2}{2}\right) - \tfrac{1}{2}\exp\left(-\lambda\zeta_{j^*k} - \tfrac{\lambda^2}{2}\right), \\ K_k &= \tfrac{r}{2}\exp\left(\lambda\zeta_{j^*k} - \tfrac{\lambda^2}{2}\right) - \tfrac{r}{2}\exp\left(-\lambda\zeta_{j^*k} + \tfrac{\lambda^2}{2}\right) + 1 - r. \end{aligned} \tag{26}$$

We have the following straightforward expressions

$$E_{\beta^0}(I_k^2) = \frac{1}{2}\left(e^{\lambda^2} + e^{-\lambda^2}\right), \tag{27}$$

$$E_{\beta^0}(K_k^2) = \frac{r^2}{2}\left(e^{\lambda^2} + e^{-\lambda^2}\right) + 1 - r^2, \tag{28}$$

$$E_{\beta^0}(J_k^2) = \frac{1}{2}\left(e^{\lambda^2} - e^{-\lambda^2}\right), \tag{29}$$

$$E_{\beta^0}(I_k K_k) = \frac{r}{2}\left(e^{\lambda^2} + e^{-\lambda^2}\right) + 1 - r. \tag{30}$$

Lemma 2 *For $\lambda \le 1/4$, we have*

$$1) \quad E_{\beta^0}\left[E_\xi Z_\xi\right]^2 \le e^{\lambda^4},$$

$$2) \quad P_{\beta^0}(E_\xi Z_\xi \le 2e^{\lambda^4/2}) \ge 3/4.$$

Proof: 1) By (28) we have

$$E_{\beta^0}\left[E_\xi Z_\xi\right]^2 = \prod_{k=0}^{2^{j^*}-1} E_{\beta^0}(K_k^2) = \left[\frac{r^2}{2}\left(e^{\lambda^2} + e^{-\lambda^2}\right) + 1 - r^2\right]^{2^{j^*}}.$$

On the other hand, since $\lambda < 1/4$ and $(\frac{e^x + e^{-x}}{2})'' < 2$ for $0 \le x \le 1$, we have

$$\frac{e^{\lambda^2} + e^{-\lambda^2}}{2} < 1 + \lambda^4,$$

and

$$E_{\beta^0}\left[E_\xi Z_\xi\right]^2 < [1+\lambda^4 r^2]^{2^{j^*}} < e^{\lambda^4}.$$

Now 2) follows by the Tchebychev inequality. ∎

Lemma 3 *For $\lambda \le 1/4$, we have*

$$P_{\beta^0}(F) \ge 1 - 6\lambda^4 - 640\, 2^{-j^*/2}.$$

Proof: When using the notations, introduced in (26), we have the following decomposition of Ξ in (21)

$$\Xi = \sum_{(k,l)\in S} \Xi_{kl}, \quad \text{where} \Xi_{kl} = I_k I_l \prod_{n=0, n\neq (k,l)}^{2^{j^*}-1} K_n.$$

This gives immediately $E_{\beta^0}\Xi_{kl} = 1$ and

$$E_{\beta^0}\Xi = 2^{j^*}(2^{j^*}-1). \tag{31}$$

Now we have

$$Var\{\Xi\} = \sum_{(k,l)\in S}\sum_{(n,m)\in S}\left[E_{\beta^0}\{\Xi_{kl}\Xi_{nm}\} - 1\right].$$

Consider the following decomposition of the set $S\times S$:

$$\begin{array}{rcl} (S\times S)_0 &=& \{((k,l),(m,n)) \in S\times S : k \neq n, m, l \neq n, m\}, \\ (S\times S)_1 &=& \{((k,l),(m,n)) \in S\times S : k = n, l = m \text{ or } k = m, l = n\}, \\ (S\times S)_2 &=& S\times S \setminus \{(S\times S)_0 \cup (S\times S)_1\}. \end{array}$$

When using the results of (27) - (30), we obtain on this decomposition (as in the proof of Lemma 2):

$$E_{\beta^0}\{\Xi_{kl}\Xi_{nm}\} \le \left\{ \begin{array}{ll} e^{\lambda^4}(1+\lambda^4 r)^4 & ((k,l),(m,n)) \in (S\times S)_0, \\ e^{\lambda^4}(1+\lambda^4)^2 & ((k,l),(m,n)) \in (S\times S)_1, \\ e^{\lambda^4}(1+\lambda^4)(1+\lambda^4 r)^2 & ((k,l),(m,n)) \in (S\times S)_2. \end{array}\right.$$

On the other hand, we have for the cardinality of the subsets

$$\begin{array}{rcl} \text{card}(S\times S)_0 &=& 2^{j^*}(2^{j^*}-1)(2^{j^*}-2)(2^{j^*}-3), \\ \text{card}(S\times S)_1 &=& 22^{j^*}(2^{j^*}-1), \\ \text{card}(S\times S)_2 &=& 42^{j^*}(2^{j^*}-1)(2^{j^*}-2). \end{array}$$

Finally, we conclude that for $\lambda < 1$,

$$\begin{aligned}\frac{Var\{\Xi\}}{[2^{j^*}(2^{j^*}-1)]^2} &\leq e^{\lambda^4}(1+\lambda^4 r)^4 - 1 + 2r^2[e^{\lambda^4}(1+\lambda^4)^2 - 1] + 4r^2[e^{\lambda^4}(1+\lambda^4)^3 - 1], \\ &\leq e^{\lambda^4} - 1 + 160r.\end{aligned}$$

However, due to (31), by the Tchebychev inequality

$$\begin{aligned}P_{\beta^0}\left(\Xi < 2^{j^*-1}(2^{j^*}-1)\right) &= P_{\beta^0}\left(\Xi - E_{\beta^0}\Xi > 2^{j^*-1}(2^{j^*}-1)\right) \\ &\leq 4(e^{\lambda^4}-1) + 640r \leq 6\lambda^4 + 640r\end{aligned}$$

for $\lambda \leq 1/4$. When substituting the value of r from (19) we obtain the lemma. ∎

On the other hand, by Assumption 1

$$P_{\beta^0}\left(\|\delta\beta\|_2 \geq 4e^{\frac{\lambda^4}{2}}(E_{\beta^0}\|\delta\beta\|_2^2)^{1/2}\right) \leq \frac{1}{16e^{\lambda^4}}.$$

Along with the results of Lemmae 2 and 3 this gives

$$\begin{aligned}P_{\beta^0}(\Gamma) &\geq P_{\beta^0}(A) - P_{\beta^0}(B^C) - P_{\beta^0}(D^C) - P_{\beta^0}(F^C), \\ &\geq 3/4 - \frac{1}{16e^{\lambda^4}} - 6\lambda^4 - 640r - P_{\beta^0}\left(2^{-j^*}\sum_{k=0}^{2^{j^*}-1}\zeta_{j^*k}^2 > 2\right), \\ &\geq c_0 > 0 \qquad (32)\end{aligned}$$

for N large enough. We are to study the cross term in the numerator of the right-hand side of (22).

Lemma 4 *Let*

$$M = E_{\beta^0}\left\{\left|E_\xi Z_\xi \tilde{\beta}\delta\beta^T\xi\|\tilde{\beta}\xi\|_2^2\right| 1_\Gamma\right\}.$$

Then there is $C > 0$ such that

$$M \leq C\rho_N^2\sigma\left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}.$$

Proof: We have

$$\begin{aligned} M &= \tilde{\beta}^3 E_{\beta^0}\left\{\left|\sum_{k=0}^{2^{j^*}-1} \delta\beta_{j^*k} E_\xi(\xi_k Z_\xi) + \sum_{k=0}^{2^{j^*}-1}\sum_{l=0,l\neq k}^{2^{j^*}-1} \delta\beta_{j^*k} E_\xi(\xi_k \xi_l^2 Z_\xi)\right| 1_\Gamma\right\}, \\ &\leq M_1 + M_2. \end{aligned}$$

With the previous notations,

$$\begin{aligned} M_1 &= \tilde{\beta}^3 E_{\beta^0}\left\{\left|\sum_{k=0}^{2^{j^*}-1} \delta\beta_{j^*k} \frac{r}{2} J_k \prod_{l=0,l\neq k}^{2^{j^*}-1} K_l\right| 1_\Gamma\right\}, \\ &\leq \tilde{\beta}^3 E_{\beta^0}\left\{\sum_{k=0}^{2^{j^*}-1} |\delta\beta_{j^*k}| \frac{r}{2} |J_k| \prod_{l=0,l\neq k}^{2^{j^*}-1} K_l 1_\Gamma\right\}. \end{aligned}$$

On the other hand, note that $|sh(x)| \leq |x| ch(x)$, thus $|J_k| \leq \lambda |\zeta_{j^*k}| I_k$. Since $\frac{r}{2} I_k < K_k$, we obtain

$$\begin{aligned} M_1 &\leq \lambda\tilde{\beta}^3 E_{\beta^0}\left\{\left(\sum_{k=0}^{2^{j^*}-1} |\delta\beta_{j^*k}||\zeta_{j^*k}|\right) E_\xi Z_\xi 1_\Gamma\right\}, \\ &\leq \lambda\tilde{\beta}^3 E_{\beta^0}\left\{\|\delta\beta\|_2 \left(\sum_{k=0}^{2^{j^*}-1} \zeta_{j^*k}^2\right)^{1/2} E_\xi Z_\xi 1_\Gamma\right\}. \end{aligned}$$

Due to the bounds on Γ and the values of $\tilde{\beta}$ and 2^{j^*}, the right hand side of the latter inequality can be bounded by

$$\lambda^4 8\sqrt{2} e^{\frac{3}{2}\lambda^4} \rho_N^2 \sigma \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}.$$

On the other hand, we get for M_2

$$\begin{aligned} M_2 &= \tilde{\beta}^3 r^2 \sum_{k=0}^{2^{j^*}-1}\sum_{l=0,l\neq k}^{2^{j^*}-1} E_{\beta^0}\left\{|\delta\beta_{j^*k}| 1_\Gamma J_k I_l \prod_{n=0,n\neq k,l}^{2^{j^*}-1} K_n\right\}, \\ &\leq \tilde{\beta}^3 r^2 \sum_{k=0}^{2^{j^*}-1}\sum_{l=0,l\neq k}^{2^{j^*}-1} \left(E_{\beta^0}\delta\beta_{j^*k}^2 1_\Gamma\right)^{\frac{1}{2}} \left(E_{\beta^0}\left\{(J_k I_l \prod_{n=0,n\neq k,l}^{2^{j^*}-1} K_n)^2\right\}\right)^{\frac{1}{2}}. \end{aligned}$$

Now using (27) - (30) we obtain as in the proof of Lemma 2

$$\begin{aligned} M_2 &\leq \frac{1}{2}(e^{2\lambda^2} - e^{-2\lambda^2})^{\frac{1}{2}} e^{\frac{\lambda^4}{2}} \tilde{\beta}^3 2^{j^*/2} \left(E_{\beta^0}\|\delta\beta\|_2^2 1_\Gamma\right)^{\frac{1}{2}}, \\ &\leq 2\lambda^3(e^{2\lambda^2} - e^{-2\lambda^2})^{\frac{1}{2}} e^{\lambda^4} \rho_N^2 \sigma \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}, \end{aligned}$$

what finishes the proof of the lemma. ∎

Lemma 5 *There is a constant $C > 0$ such that for any $\lambda \leq 1/4$,*

$$E_{\beta^0} E_\xi \left\{ Z_\xi 1_{\Delta^C} 1_\Gamma \right\} \leq C \exp\left(-\frac{9}{32} 2^{j^*/2} \right).$$

Proof: By the Cauchy inequality, we have

$$E_{\beta^0} E_\xi \left\{ Z_\xi 1_{\Delta^C} 1_\Gamma \right\} \leq P_\xi(\Delta^C)^{\frac{1}{2}} P_{\beta^0}(\Gamma)^{\frac{1}{2}} \left(E_{\beta^0} E_\xi \left\{ Z_\xi^2 \right\} \right)^{\frac{1}{2}}.$$

We decompose

$$E_{\beta^0} E_\xi \left\{ Z_\xi^2 \right\} = \prod_{k=0}^{2^{j^*}-1} M_k,$$

where

$$M_k = E_{\beta^0} \left[\frac{r}{2} \exp\left(2\lambda \zeta_{j^* k} - \lambda^2 \right) + \frac{r}{2} \exp\left(-2\lambda \zeta_{j^* k} - \lambda^2 \right) + 1 - r \right].$$

When taking the expectation, we obtain $M_k = r \exp(\lambda^2) + 1 - r$. Since $e^{\lambda^2} < 1 + 3\lambda^2$ for $\lambda < 1$, we have

$$E_{\beta^0} E_\xi \left\{ Z_\xi^2 \right\} \leq \exp\left(3\lambda^2 2^{j^*/2} \right).$$

On the other hand by the Bernstein inequality,

$$\begin{aligned} P_\xi(\Delta^C) &= P_\xi\left(\frac{1}{2^{j^*}} \sum_{k=0}^{2^{j^*}-1} \xi_k^2 > 3r/2 \right) \leq 2\exp\left(-\frac{9r^2 2^{j^*-2}}{2r(1-r)+r} \right), \\ &\leq 2\exp\left(-\frac{3}{4} 2^{j^*/2} \right). \end{aligned}$$

Due to the bound on λ, this implies the lemma. ∎

Note now that

$$E_{\beta^0}\left[Z_\xi \|\tilde{\beta}\xi\|_2^4 1_\Delta 1_\Gamma \right] \geq E_{\beta^0}\left[r^2 \tilde{\beta}^4 \Xi 1_\Delta 1_\Gamma \right] \geq \rho^4 P(\Gamma \cap \Delta).$$

Hence, when using (32) and the bound of lemma 5, from (22)

$$
\begin{aligned}
r_{\beta^0}(\delta, N) &\geq C \frac{E_{\beta^0}\left\{E_\xi[Z_\xi(\|\tilde\beta\xi\|_2^4 - 4\tilde\beta\delta\beta^T\xi\|\tilde\beta\xi\|_2^2)1_\Delta]1_\Gamma\right\}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2}, \\
&\geq C' \frac{E_{\beta^0}\left\{(P_{\beta^0}(\Gamma\cap\Delta)\rho_N^4 - \rho_N^3(\rho_N + \|\delta\beta\|_2)2^{\frac{3}{4}j^*}E_\xi[Z_\xi 1_{\Delta^C}] - |E_\xi[Z_\xi\tilde\beta\delta\beta^T\xi\|\tilde\beta\xi\|_2^2]|)1_\Gamma\right\}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2} \\
&\geq C'' \frac{\rho_N^4 - \rho_N^3(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}})2^{\frac{3}{4}j^*}e^{-\frac{9}{32}2^{j^*/2}} - \rho_N^2\sigma\left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2} \\
&\geq C'''\rho_N^2 \frac{\rho_N^2 - \sigma\left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}}{\left(\rho_N + \left(E_{\beta^0}(\|\delta\beta\|_2^2)\right)^{\frac{1}{2}}\right)^2}
\end{aligned}
$$

for N large enough. ∎

5.3 Proof of Theorem 2

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 [4].

We set

$$f_{j_0}(x) = \alpha'\phi(x) + \sum_{j=0}^{j_0}\sum_{k=0}^{2^j-1}\beta'_{jk}\psi_{jk}(x), \tag{33}$$

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 next lemma is an immediate corollary of Proposition 1 in [4].

Proposition 2 *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*

$$\beta' \in \mathcal{F}(s,p,q,C_0L) \text{ and } \|f - f_{j_0}\|_2 = O(L2^{-j_0 s'}) = O\left(LN^{-s'}\right), \tag{34}$$

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

if we denote

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

Then due to (34) 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 if $m'_N = \|\widehat{f}_N - f_{j_0}\|_2$, then

$$\left|[E_f(\widehat{m}_N - m_N)^2]^{1/2} - [E_f(\widehat{m}_N - m'_N)^2]^{1/2}\right| \leq CLN^{-s'}. \tag{35}$$

So to show the upper bounds in Theorem 2 it suffices to control the value $[E_f(\widehat{m}_N - m'_N)^2]^{1/2}$. Furthermore it follows from (34) that the coefficient β'_{jk} satisfies (up to an "absolute constant") the same norm relation (5) 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}. this give the model

$$y_{jk} = \beta_{jk} + \zeta_{jk} \tag{36}$$

for empirical wavelet coefficients.

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

Let us now turn to the proof of Theorem 2 itself. We consider here only the case $p^* < 2$. The case $p^* \geq 2$ follows the same lines as that of Theorem 2 in [11].

Let λ_j^m be the "minimax" thresholds for the estimation of the function f on the class $\mathcal{F}(s^*, p^*, q^*, L^*)$, defined in [5]. It can be easily verified that $\lambda_j \leq \lambda_j^m$ for $0 \leq j \leq j_0$. This implies that the adaptive estimate $\widehat{f}_N$ with "corrected" thresholds λ_j^* still possess minimax rate of convergence on any Besov class embedded in $\mathcal{F}(s^*, p^*, q^*, L^*)$. Due to the results of [5], [6] and [10] gives the bound (15) of Theorem 2.

We denote

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

and

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

Then we conclude from (36) that (ζ'_{jk}) and (ζ''_{jk}) are two uncorrelated (and thus mutually independent) sequences of independent and identically distributed Gaussian random variables with $E\zeta'_{jk} = E\zeta''_{jk} = 0$ and $E(\zeta'_{jk})^2 = E(\zeta''_{jk})^2 = \frac{2\sigma_w^2}{N}$.

Let $m'_N = \sqrt{\sum_{j=0}^{j_0} \|\gamma_{j\cdot}\|_2^2}$ then the difference $\widehat{m}_N^2 - (m'_N)^2$ can be rewritten as:

$$\begin{aligned}
|\widehat{m}_N^2 - (m'_N)^2| \leq & \left| \sum_{j=0}^{j^*} (\|z_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2 - \|\gamma_{j\cdot}\|_2^2) + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} (z_{jk}^2 - 2\sigma^2 - \gamma_{jk}^2) 1_{|y'_{jk}| \geq \lambda_j \sigma} \right. \\
& \left. - \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \gamma_{jk}^2 1_{|y'_{jk}| < \lambda_j \sigma} \right| \\
= & \left| \sum_{j=0}^{j^*} \|\zeta''_{j\cdot}\|_2^2 - 2^{j+1}\sigma^2 \right| + \left| \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} ({\zeta''_{jk}}^2 - 2\sigma^2) 1_{|y'_{jk}| \geq \lambda_j \sigma} \right| \\
& +2 \left| \sum_{j=0}^{j^*} \gamma_{j\cdot}^T \zeta''_{j\cdot} + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \gamma_{jk}^T \zeta''_{jk} 1_{|y'_{jk}| \geq \lambda_j \sigma} \right| + \sum_{j=j^*+1}^{j_0} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|y'_{jk}| < \lambda_j \sigma} \\
= & \sum_{i=1}^{4} \delta_N^{(i)}. \qquad (37)
\end{aligned}$$

We have the following immediate estimate:

$$\left[E(\delta_N^{(1)})^2 \right]^{1/2} \leq C 2^{j^*/2} \sigma^2. \qquad (38)$$

In order to continuer we need some technical results.

Lemma 6

$$1_{|y_{jk}| \geq \lambda\sigma} \leq 1_{|\beta_{jk}| \geq \lambda\sigma/2} + 1_{|\zeta_{jk}| > \lambda\sigma/2}; \qquad (39)$$

$$1_{|y_{jk}| < \lambda\sigma} \leq \sum_{l=0}^{\infty} 1_{|\beta_{jk}| < (2+l)\lambda\sigma} 1_{|\zeta_{jk}| \geq l\lambda\sigma}; \qquad (40)$$

Proof: The proof of (39) is immediate. To show (40) we decompose $1_{|y_{jk}|<\lambda\sigma}$:

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

∎

Lemma 7 $\left[E(\delta_N^{(2)})^2\right]^{1/2} \le C\left(2^{j^*/2}\sigma^2 + (L^*)^{p^*/2}\sigma^{2-p^*/2}2^{-j^*(s^*p^*+p^*/2-1)/2}\right).$

Proof: In the decomposition below we use (39) and first take the expectation over the distribution of (ζ''_{jk}) and then over that of (ζ'_{jk}):

$$\begin{aligned}E(\delta_N^{(2)})^2 &= 2\sigma^4 E\left(\sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j+1} 1_{|y'_{jk}|\ge\lambda_j\sigma}\right) && (41)\\ &\le 2\sigma^4 \sum_{j=j^*+1}^{j_0}\sum_{k=0}^{2^j+1}\left(1_{|\beta_{jk}|\ge\frac{\lambda_j\sigma}{2}} + P(|\zeta'_{jk}| > \frac{\lambda_j\sigma}{2})\right) \\ &\le 2^{p^*+1}\sigma^4\sum_{j=j^*+1}^{j_0}\frac{\|\beta_{j\cdot}\|_{p^*}^{p^*}}{(\lambda_j\sigma)^{p^*}} + 4\sigma^4\sum_{j=j^*+1}^{j_0} 2^j\exp(-\frac{\lambda_j^2}{16}) = I_N^{(1)} + I_N^{(2)}. && (42)\end{aligned}$$

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

$$\begin{aligned}I_N^{(1)} &\le \frac{2^{p^*+1}\sigma^{4-p^*}(L^*)^{p^*}}{\kappa^{p^*}}\sum_{j=j^*+1}^{j_0}\frac{2^{-j(s^*p^*+p^*/2-1)}}{(j-j^*)^{p^*/2}} \\ &\le \frac{2^{p^*+1}\sigma^{4-p^*}(L^*)^{p^*}}{\kappa^{p^*}}2^{-j^*(s^*p^*+p^*/2-1)}\sum_{l=1}^{\infty}\frac{2^{-l(s^*p+p^*/2-1)}}{l^{p^*/2}} \\ &\le C(L^*)^{p^*}\sigma^{4-p^*}2^{-j^*(s^*p^*+p^*/2-1)}. && (43)\end{aligned}$$

Since κ in the definition (11) of λ_j satisfies $\kappa^2 > 16\log 2$, we can bound $I_N^{(2)}$ as follows:

$$I_N^{(2)} < 4\sigma^4 2^{j^*}\sum_{l=1}^{\infty}2^l\exp(-\frac{\kappa^2 l}{16}) \le 4\sigma^4 2^{j^*}\sum_{l=1}^{\infty}\exp\left(l(\log 2 - \frac{\kappa^2}{16})\right) \le C\sigma^4 2^{j^*}. \quad (44)$$

When substituting (43) and (44) into (42) we obtain the bound announced in the lemma.

∎

Lemma 8 *Suppose that there is $C<\infty$ such that $L^* > C\sigma(\log N)^{(4s^*+1)^2/2}$. Then*

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

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

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

(cf, for instance Theorem 1 in [5]). When taking the expectation over the distribution of (ζ''_{jk}) and then over the distribution of (ζ'_{jk}) we obtain:

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

One can easily verify that if $L^* > C\sigma(\log N)^{(4s^*+1)^2/2}$ then

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

∎

Lemma 9 $\left[E(\delta_N^{(4)})^2\right]^{1/2} \le C(L^*)^{p^*}\sigma^{2-p^*}2^{-j^*(s^*p^*+p^*/2-1)}$.

Proof: The decomposition (40) yields

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

$$
\begin{aligned}
&\le \sqrt{2} \sum_{j=j^*+1}^{j_0} L^{p^*} 2^{-j(s^*p^*+p^*/2-1)} \lambda_j^{2-p^*} \sigma^{2-p^*} \sum_{l=0}^{\infty} (2+l)^{2-p^*} \exp\left(-\frac{l^2\lambda_j^2}{4}\right) \\
&\le C(L^*)^{p^*} \sigma^{2-p^*} \sum_{j=j^*+1}^{\infty} 2^{-j(s^*p^*+p^*/2-1)} (j-j^*)^{(2-p^*)/2} \\
&\le C'(L^*)^{p^*} \sigma^{2-p^*} 2^{-j^*(s^*p^*+p^*/2-1)}.
\end{aligned}
$$

∎

From (37), using the results of (38) and Lemmas 7 – 9, we obtain

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

The condition $L^* \ge C\sigma(\log N)^{(4s^*+1)^2/2}$ and the choice of j^* in (10) now results in the bound

$$
\left[E(\widehat{m}_N^2 - (m'_N)^2)^2\right]^{1/2} \le C\left((L^*)^{\frac{1}{2s^*+1-1/p^*}} \sigma^{\frac{4s^*+1-2/p^*}{2s^*+1-1/p^*}}\right). \tag{45}
$$

Note that (cf. proof of Theorem 2 in [11]) for any $\gamma > 0$

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

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 (16) follows from (35) and (45). ∎

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 the Computation of Wavelet Coefficients, *Journal of Approximation Theory*, 88, N.1, 47-79, 1997

[5] D. Donoho, I. Johnstone, *Minimax Estimation via Wavelet Shrinkage*, *An. of Statist.* 26, No3, 879-921, 1998.

[6] D. Donoho, I. Johnstone, *Adapting to Unknown Smoothness via Wavelet Shrinkage*, Technical Report, Department of Statistics, Stanford University, 1993.

[7] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, Wavelet Shrinkage: Asymptotia? *J. R. Statist. Soc. B* 57, No2, 301-369, 1995.

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

[9] Ibragimov, R. Khasminskij, *Statistical Estimation, Asymptotic Theory.* Moscow, Nauka, 1979 (in Russian). Translation: Springer-Verlag, Berlin, 1981.

[10] A. Juditsky Wavelet Estimators: Adapting to Unknown Smoothness, *Math. Methods of Stat.*, 1, 1-20, 1997.

[11] A. Juditsky, O. Lepski, Confidence intervals for adaptive regression estimation, Technical report 3580, INRIA, 1998.

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

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

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

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

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,

615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY

Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex

Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN

Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex

Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

http://www.inria.fr

ISSN 0249-6399