



HAL
open science

Sharp estimation in sup norm with random design

Stéphane Gaïffas

► **To cite this version:**

Stéphane Gaïffas. Sharp estimation in sup norm with random design. *Statistics and Probability Letters*, 2009, 77 (8), pp.782. 10.1016/j.spl.2006.11.017 . hal-00538008

HAL Id: hal-00538008

<https://hal.science/hal-00538008>

Submitted on 20 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Sharp estimation in sup norm with random design

Stéphane Gaïffas

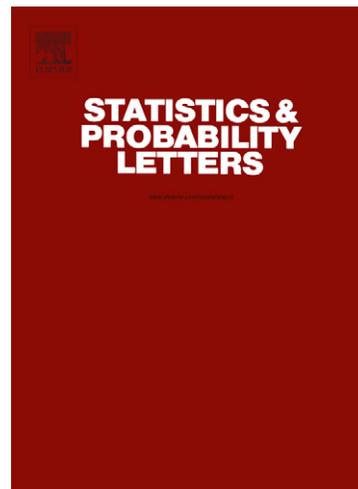
PII: S0167-7152(07)00003-X
DOI: doi:10.1016/j.spl.2006.11.017
Reference: STAPRO 4542

To appear in: *Statistics & Probability Letters*

Received date: 19 April 2006
Accepted date: 26 November 2006

Cite this article as: Stéphane Gaïffas, Sharp estimation in sup norm with random design, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2006.11.017

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

SHARP ESTIMATION IN SUP NORM WITH RANDOM DESIGN

STÉPHANE GAÏFFAS

*Laboratoire de Probabilités et Modèles Aléatoires
 Université Paris 7 Denis-Diderot
 email: gaiffas@math.jussieu.fr*

ABSTRACT. In this paper, we study the estimation of a function based on noisy inhomogeneous data (the amount of data can vary on the estimation domain). We consider the model of regression with random design, where the design density is unknown. We construct an asymptotically sharp estimator which converges, for sup norm error loss, with a spatially dependent normalisation which is sensitive to the variations in the local amount of data. This estimator combines both kernel and local polynomial methods, and it does not depend within its construction on the design density. Then, we prove that the normalisation is optimal in an appropriate sense.

1. INTRODUCTION

In most cases, the models considered in curve estimation do not allow situations where the data is inhomogeneous, in so far as the amount of data is implied to be constant over space. This is the case in regression with equispaced design and white noise models, for instance. In many situations, the data can happen to be concentrated at some points and to be little elsewhere. In such cases, an estimator shall behave better at a point where there is much data than where there is little data. In this paper, we propose a theoretical study of this phenomenon.

The available data $[(X_i, Y_i), 1 \leq i \leq n]$ is modeled by

$$Y_i = f(X_i) + \xi_i, \quad (1.1)$$

where ξ_i are i.i.d. centered Gaussian with variance σ^2 and independent of X_i . The design variables X_i are i.i.d. of unknown density μ on $[0, 1]$, which is bounded away from 0 and continuous. We want to recover f . When μ is not the uniform law, the information is spatially inhomogeneous. We are interested in recovering f globally, with sup norm loss $\|g\|_\infty := \sup_{x \in [0, 1]} |g(x)|$. An advantage of this norm is that it is exacting: it forces an estimator to behave well at every point simultaneously. A commonly used benchmark for the complexity of estimation over some fixed class Σ is the minimax risk, which is given by

$$\mathcal{R}_n(\Sigma) := \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbf{E}_f^n \{ \|\hat{f}_n - f\|_\infty \}, \quad (1.2)$$

where the infimum is taken over all estimators. We say that ψ_n is the minimax convergence rate over Σ if $\mathcal{R}_n(\Sigma) \asymp \psi_n$, where $a_n \asymp b_n$ means $0 < \liminf_n a_n/b_n \leq \limsup_n a_n/b_n < +\infty$. In the regression model (1.1) with Σ a Hölder ball with smoothness $s > 0$ and μ positive and bounded, we have $\psi_n = (\log n/n)^{s/(2s+1)}$, see Stone (1982). Thus, in this case, the minimax rate is not sensitive to the variations in the amount of data. Indeed, such global minimax

Date: December 4, 2006.

2000 Mathematics Subject Classification. 62G05, 62G08, 62G15.

Key words and phrases. random design, nonparametric regression, sharp estimation, inhomogeneous data.

benchmarks cannot assess the *design-adaptation* property of an estimator. Instead of (1.2), an improvement is to consider the spatially dependent risk

$$\sup_{f \in \Sigma} \mathbf{E}_f^n \left\{ \sup_{x \in [0,1]} r_n(x)^{-1} |\hat{f}_n(x) - f(x)| \right\}$$

of some estimator \hat{f}_n , where $r_n(\cdot) > 0$ is a family of spatially dependent normalisations. If this quantity is bounded as $n \rightarrow +\infty$, we say that $r_n(\cdot)$ is an upper bound over Σ . Necessarily, the "optimal" normalisation satisfies $r_n(x) \asymp (\log n/n)^{s/(2s+1)}$ for any x (note that the optimality requires an appropriate definition here). Therefore, in order to exhibit such an optimal normalisation, we need to consider the sharp asymptotics of the minimax risk.

2. RESULTS

If $s, L > 0$, we define the Hölder ball $\Sigma(s, L)$ as the set of all the functions $f : [0, 1] \rightarrow \mathbb{R}$ such that

$$|f^{(k)}(x) - f^{(k)}(y)| \leq L|x - y|^{s-k}, \quad \forall x, y \in [0, 1],$$

where $k = \lfloor s \rfloor$ is the largest integer $k < s$. If $Q > 0$, we define $\Sigma^Q(s, L) := \Sigma(s, L) \cap \{f \text{ s.t. } \|f\|_\infty \leq Q\}$, and we denote simply $\Sigma := \Sigma^Q(s, L)$ (the constant Q needs not to be known). All along this study, we suppose:

Assumption D. *There is $\nu \in (0, 1]$ and $\varrho, q > 0$ such that*

$$\mu \in \Sigma(\nu, \varrho) \text{ and } \mu(x) \geq q, \text{ for all } x \in [0, 1].$$

In the following, we consider a continuous, non-negative and nondecreasing loss function $w(\cdot)$ such that $w(x) \leq A(1 + |x|^b)$ for some $A, b > 0$ (typically a power function). Let us consider

$$r_{n,\mu}(x) := \left(\frac{\log n}{n\mu(x)} \right)^{s/(2s+1)}. \quad (2.1)$$

We prove in theorem 1 below that this normalisation is, up to the constants, an upper bound over Σ , and that it is indeed optimal in theorem 2. We denote by $\mathbb{E}_{f,\mu}^n$ the integration with respect to the joint law $\mathbb{P}_{f,\mu}^n$ of the observations (X_i, Y_i) , $1 \leq i \leq n$. The estimator used in theorem 1 does not depend, within its construction, on μ .

Theorem 1 (Upper bound). *Under assumption D, if \hat{f}_n is the estimator defined in section 4 below, we have for any $s, L > 0$,*

$$\limsup_n \sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \left\{ w \left(\sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\hat{f}_n(x) - f(x)| \right) \right\} \leq w(P), \quad (2.2)$$

where

$$P := \sigma^{2s/(2s+1)} L^{1/(2s+1)} \varphi_s(0) \left(\frac{2}{2s+1} \right)^{s/(2s+1)} \quad (2.3)$$

and φ_s is defined as the solution of the optimisation problem

$$\varphi_s := \operatorname{argmax}_{\substack{\varphi \in \Sigma(s, 1; \mathbb{R}), \\ \|\varphi\|_2 \leq 1}} \varphi(0), \quad (2.4)$$

where $\Sigma(s, L; \mathbb{R})$ is the extension of $\Sigma(s, L)$ to the whole real line.

In the same fashion as in Donoho (1994), the constant P is defined via the solution of an optimisation problem which is connected to optimal recovery. We discuss this result in section 3, where further details about optimal recovery can be found. The next theorem shows that $r_{n,\mu}(\cdot)$ is indeed optimal in an appropriate sense. In what follows, the notation $|I|$ stands for the length of an interval I .

Theorem 2 (Lower bound). *Under assumption D, if $I_n \subset [0, 1]$ is any interval such that for some $\varepsilon \in (0, 1)$,*

$$|I_n| n^{\varepsilon/(2s+1)} \rightarrow +\infty \quad \text{as } n \rightarrow +\infty, \quad (2.5)$$

we have

$$\liminf_n \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \left\{ w \left(\sup_{x \in I_n} r_{n, \mu}(x)^{-1} |\hat{f}_n(x) - f(x)| \right) \right\} \geq w((1 - \varepsilon)P),$$

where P is given by (2.3) and the infimum is taken among all estimators. A consequence is that if I_n is such that (2.5) holds for any $\varepsilon \in (0, 1)$, we have

$$\liminf_n \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \left\{ w \left(\sup_{x \in I_n} r_{n, \mu}(x)^{-1} |\hat{f}_n(x) - f(x)| \right) \right\} \geq w(P). \quad (2.6)$$

This result says that the normalisation $r_{n, \mu}(\cdot)$ cannot be strongly improved: no normalisation is uniformly better than $r_{n, \mu}(\cdot)$ within a "large" interval. This result is discussed in the following section.

3. DISCUSSION

Literature. When the design is equidistant, that is $X_i = i/n$, we know from Korostelev (1993) the exact asymptotic value of the minimax risk for sup norm error loss. If $\psi_n := (\log n/n)^{s/(2s+1)}$, we have for any $s \in (0, 1]$ and $\Sigma = \Sigma(s, L)$

$$\lim_{n \rightarrow +\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_f \left\{ w(\psi_n^{-1} \|\hat{f}_n - f\|_\infty) \right\} = w(C),$$

where

$$C := \sigma^{2s/(2s+1)} L^{1/(2s+1)} \left(\frac{s+1}{2s^2} \right)^{s/(2s+1)}. \quad (3.1)$$

This result was the first of its kind for sup norm error loss. In the white noise model

$$dY_t^n = f(t)dt + n^{-1/2}dW_t, \quad t \in [0, 1], \quad (3.2)$$

where W is a standard Brownian motion, Donoho (1994) extends the result by Korostelev (1993) to any $s > 1$. In this paper, the author makes a link between statistical sup norm estimation and the theory of optimal recovery (see below). It is shown for any $s > 0$ and $\Sigma = \Sigma(s, L)$ that the minimax risk satisfies

$$\lim_{n \rightarrow +\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbf{E}_f \left\{ w(\psi_n^{-1} \|\hat{f}_n - f\|_\infty) \right\} = w(P_1), \quad (3.3)$$

where P_1 is given by (2.3) with $\sigma = 1$. When $s \in (0, 1]$, we have $P = C$, see for instance in Leonov (1997). Since the results by Korostelev and Donoho, many other authors worked on the problem of sharp estimation (or testing) in sup norm. On testing, see Lepski and Tsybakov (2000), see Korostelev and Nussbaum (1999) for density estimation and Bertin (2004a) for white noise in an anisotropic setting. The paper by Bertin (2004b) works in the model of regression with random design (1.1). When μ satisfies assumption D and $\Sigma = \Sigma^Q(s, L)$ for $s \in (0, 1]$, it is shown that

$$\lim_{n \rightarrow +\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \left\{ w(v_{n, \mu}^{-1} \|\hat{f}_n - f\|_\infty) \right\} = w(C), \quad (3.4)$$

where C is given by (3.1) and $v_{n, \mu} := [\log n / (n \inf_x \mu(x))]^{s/(2s+1)}$. Note that the rate $v_{n, \mu}$ differs from (and is larger than) ψ_n when μ is not uniform. A disappointing fact is that $v_{n, \mu}$ depends on μ via its infimum only, which corresponds to the point in $[0, 1]$ where we have the least information. Therefore, this rate does not take into account all the other regions with more data.

As a consequence, the results presented here are extensions of both the papers by Donoho (1994) and Bertin (2004b): our results are stated in the regression model with random

design, where the design density is unknown. In particular, we provide the exact asymptotic value of the minimax risk in regression with random design for any $s > 0$, which was known only for $s \in (0, 1]$ beforehand. Nevertheless, the main novelty is, in our sense, the introduction of a spatially dependent normalisation factor for the assessment of an estimator, with an appropriate optimality criterion. The asymptotically sharp minimax framework is considered here only by necessity.

Optimal recovery. The problem of optimal recovery consists in recovering f from

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}, \quad (3.5)$$

where $\varepsilon > 0$, z is an unknown deterministic function such that $\|z\|_2 \leq 1$ and $f \in C(s, L; \mathbb{R}) := \Sigma(s, L; \mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. The link between this deterministic problem and estimation with sup norm loss in white noise model was made by Donoho (1994), see also Leonov (1999). The minimax risk for the optimal recovery of f at 0 from (3.5) is defined by

$$E_s(\varepsilon, L) := \inf_T \sup_{\substack{f \in C(s, L; \mathbb{R}) \\ \|f - y\|_2 \leq \varepsilon}} |T(y) - f(0)|,$$

where \inf_T is taken among all continuous and linear forms on $\mathbb{L}^2(\mathbb{R})$. We know from Micchelli and Rivlin (1977), Arestov (1990) that

$$E_s(\varepsilon, L) = \inf_{K \in \mathbb{L}^2(\mathbb{R})} \left(\sup_{f \in C(s, L; \mathbb{R})} \left| \int K(t)(f(t) - f(0)) \right| + \varepsilon \|K\|_2 \right) = \sup_{\substack{f \in \Sigma(s, L; \mathbb{R}) \\ \|f\|_2 \leq \varepsilon}} f(0).$$

Note that φ_s satisfies $\varphi_s(0) = E_s(1, 1)$. To our knowledge, the function φ_s is known only for $s \in (0, 1] \cup \{2\}$. The kernel K_s for $s \in (0, 1]$ was found by Korostelev (1993) and by Fuller (1961) for $s = 2$. For any $s > 0$, we know from Leonov (1997) that φ_s is well defined and unique, that it is even and compactly supported and that $\|\varphi_s\|_2 = 1$. A renormalisation argument from Donoho (1994) shows that $E_s(\varepsilon, L) = E_s(1, 1)L^{1/(2s+1)}\varepsilon^{2s/(2s+1)}$, thus it suffices to know $E_s(1, 1)$. If we define

$$B(s, L) := \sup_{f \in C(s, L; \mathbb{R})} \left| \int K_s(t)(f(t) - f(0)) \right|, \quad (3.6)$$

we have the decomposition $E_s(1, 1) = B(s, 1) + \|K\|_2$, and in particular, if P is given by (2.3) and

$$c_s := \left(\frac{\sigma}{L} \right)^{2/(2s+1)} \left(\frac{2}{2s+1} \right)^{1/(2s+1)}, \quad (3.7)$$

we have

$$P = Lc_s^s(B(s, 1) + \|K\|_2). \quad (3.8)$$

About theorem 1. We can understand the result of theorem 1 heuristically. Following Brown and Low (1996) and Brown et al. (2002), we can say that an "idealised" statistical experiment which is equivalent (in the sense that the LeCam deficiency goes to 0) to the model (1.1) is given by the heteroscedastic white noise model

$$dY_t^n = f(t)dt + \frac{\sigma}{\sqrt{n\mu(t)}}dB_t, \quad t \in [0, 1], \quad (3.9)$$

where B is a Brownian motion. In view of the result (3.3) by Donoho (1994), which is stated in the model (3.2), and comparing the noise levels in the models (3.2) and (3.9) (with $\sigma = 1$), we can explain informally that our rate $r_{n,\mu}(\cdot)$ comes from the former rate ψ_n where we "replace" n by $n\mu(x)$.

About theorem 2. From Bertin (2004b), we know when $s \in (0, 1]$ that

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \{w(v_{n, \mu}^{-1} \|\widehat{f}_n - f\|_\infty)\} \geq w(P),$$

where $v_{n, \mu} = [\log n / (n \inf_x \mu(x))]^{s/(2s+1)}$. An immediate consequence is

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \left\{ w \left(\sup_{x \in [0, 1]} r_{n, \mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \right) \right\} \geq w(P), \quad (3.10)$$

where it suffices to use the fact that $r_{n, \mu}(x) \leq v_{n, \mu}$ for any $x \in [0, 1]$. This entails that $r_{n, \mu}(\cdot)$ is optimal in the classical minimax sense. However, this lower bound is much weaker than the one considered in theorem 2: it does not exclude the existence of another normalisation $\varrho_n(\cdot)$ such that $\varrho_n(x) < r_{n, \mu}(x)$ for "many" x . Therefore, to prove the optimality of $r_{n, \mu}(\cdot)$, we need to localise the lower bound. Indeed, in theorem 2, if we choose $I_n = [0, 1]$ we find back (3.10) and if $I_n = [\bar{x} - (\log n)^\gamma, \bar{x} + (\log n)^\gamma] \cap [0, 1]$ for any $\gamma > 0$ and $\bar{x} \in [0, 1]$ such that $\mu(\bar{x}) \neq \inf_{x \in [0, 1]} \mu(x)$, then obviously $v_{n, \mu}$ does not satisfy (2.6).

About assumption D. In assumption D, μ is supposed to be bounded from below, and from above since it is continuous over $[0, 1]$. When μ is vanishing or exploding at a fixed point, we know from Gaïffas (2005a) that a wide range of pointwise minimax rates can be achieved, depending on the behaviour of μ at this point. In this case, we expect the optimal normalisation (whenever it exists) to differ from the classical minimax rate ψ_n not only up to the constants, but in order.

Adaptation to the smoothness. The estimator used in theorem 1 depends on the smoothness s of f (see below). In practice, such a parameter is unknown. Therefore, this estimator cannot be used directly: some smoothness-adaptive technique, like Lepski's method (see Lepski et al. (1997)) can be applied. However, this estimator is considered here for theoretical purposes only, and note that even in the white noise model, the problem of sharp adaptive estimation in sup norm over Hölder classes remains open when $s > 1$.

4. CONSTRUCTION OF AN ESTIMATOR

The estimator \widehat{f}_n described below uses both kernel and local polynomial methods. Its construction is divided into two parts: first, at some well-chosen discretization points, we use a Nadaraya-Watson estimator with optimal kernel and a design data driven bandwidth. This part of the estimator is used to attain the minimax constant. Then, between the discretization points, the estimator is defined by a Taylor expansion where the derivatives are estimated by local polynomial estimation. We define the empirical design sample distribution

$$\bar{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

where δ is the Dirac mass, and for $h > 0$, $x \in [0, 1]$, we consider the intervals

$$I(x, h) := \begin{cases} [x, x + h] & \text{when } 0 \leq x \leq 1/2, \\ [x - h, x] & \text{when } 1/2 < x \leq 1. \end{cases} \quad (4.1)$$

The choice of non-symmetrical intervals allows to skip boundaries effects. Then we define, when it makes sense, the "bandwidth" at x by

$$H_n(x) := \operatorname{argmin}_{h \in [0, 1]} \left\{ h \text{ s.t. } h^{2s} \bar{\mu}_n(I(x, h)) \geq \log n/n \right\}, \quad (4.2)$$

which makes the balance between the bias h^s and the variance $[\log n/(n\bar{\mu}_n(I))]^{1/2}$ of the kernel estimator. When the event in (4.2) is empty (which occurs with a very small probability for large n), we take simply $H_n(x) := \max(1-x, x)$. We consider the sequence of points

$$x_j := j\Delta_n, \quad \Delta_n := (\log n)^{-2s/(2s+1)} n^{-1/(2s+1)}, \quad (4.3)$$

for $j \in \mathcal{J}_n := \{0, \dots, [\Delta_n^{-1}]\}$ where $[a]$ is the integer part of a with $x_{M_n} = 1$, $M_n = |\mathcal{J}_n|$ (the notation $|A|$ stands also for the size of a finite set A). We define $H_n^M := \max_{j \in \mathcal{J}_n} H_n(x_j)$. From Leonov (1997, 1999) we know that the function φ_s defined by (2.4) is even and compactly supported. We denote by $[-T_s, T_s]$ its support and $\tau_n := \min(2c_s T_s H_n^M, \delta_n)$ where $\delta_n = (\log n)^{-1}$ and c_s is given by (3.7).

As usual with the estimation of a function over an interval, there is a boundary correction. We decompose the unit interval into three parts $[0, 1] = J_{n,1} \cup J_{n,2} \cup J_{n,3}$ where $J_{n,1} := [0, \tau_n]$, $J_{n,2} := [\tau_n, 1 - \tau_n]$ and $J_{n,3} := [1 - \tau_n, 1]$. We also define $\mathcal{J}_{a,n} := \{j | x_j \in J_{a,n}\}$ for $a \in \{1, 2, 3\}$. If φ_s is defined by (2.4), we consider the kernel

$$K_s := \varphi_s / \int \varphi_s. \quad (4.4)$$

The "sharp" part of the estimator is defined as follows: at the points x_j , we define \hat{f}_n by

$$\hat{f}_n(x_j) := \begin{cases} \frac{\frac{1}{nH_n(x_j)} \sum_{i=1}^n Y_i K_s\left(\frac{X_i - x_j}{c_s H_n(x_j)}\right)}{\max\left[\delta_n, \frac{1}{nH_n(x_j)} \sum_{i=1}^n K_s\left(\frac{X_i - x_j}{c_s H_n(x_j)}\right)\right]} & \text{if } j \in \mathcal{J}_{2,n}, \\ \bar{f}_n(x_j) & \text{if } j \in \mathcal{J}_{1,n} \cup \mathcal{J}_{3,n}. \end{cases} \quad (4.5)$$

This estimator is (up to the correction near the boundaries) a Nadaraya-Watson estimator with the optimal kernel K_s and a bandwidth fitted to the local amount of data. The boundary estimator \bar{f}_n is defined below.

We recall that $k = \lfloor s \rfloor$ where s is the smoothness of the unknown signal f . For any interval $I \subset [0, 1]$ such that $\bar{\mu}_n(I) > 0$, we define the inner product

$$\langle f, g \rangle_I := \frac{1}{\bar{\mu}_n(I)} \int_I f g d\bar{\mu}_n,$$

where $\int_I f d\bar{\mu}_n = \sum_{X_i \in I} f(X_i)/n$. If $I = I(x, h)$ (see (4.1)), we define $\phi_{I,m}(y) := (y - x)^m$ and we introduce the matrix \mathbf{X}_I and vector \mathbf{Y}_I with entries

$$(\mathbf{X}_I)_{p,q} := \langle \phi_{I,p}, \phi_{I,q} \rangle_I \quad \text{and} \quad (\mathbf{Y}_I)_p := \langle Y, \phi_{I,p} \rangle_I,$$

for $0 \leq p, q \leq k$. Then, we consider

$$\bar{\mathbf{X}}_I := \mathbf{X}_I + \frac{1}{\sqrt{n\bar{\mu}_n(I)}} \mathbf{I}_{k+1} \mathbf{1}_{\Omega_{n,I}^c},$$

where $\Omega_{n,I} := \{\lambda(\mathbf{X}_I) > (n\bar{\mu}_n(I))^{-1/2}\}$, where $\lambda(M)$ is the smallest eigenvalue of a matrix M and where \mathbf{I}_{k+1} is the identity matrix on \mathbb{R}^{k+1} . Note that the correction term in $\bar{\mathbf{X}}_I$ entails $\lambda(\bar{\mathbf{X}}_I) \geq (n\bar{\mu}_n(I))^{-1/2}$. When $\bar{\mu}_n(I) > 0$, the solution $\hat{\theta}_I$ of the system

$$\bar{\mathbf{X}}_I \theta = \mathbf{Y}_I,$$

is well defined. If $\bar{\mu}_n(I) = 0$, we take $\hat{\theta}_I = 0$. Then, for any $1 \leq m \leq k$, a natural estimate of $f^{(m)}(x_j)$ is

$$\tilde{f}_n^{(m)}(x_j) := m! (\hat{\theta}_{I(x_j, h_n)})_m,$$

where $h_n := (\sigma/L)^{2/(2s+1)}(\log n/n)^{1/(2s+1)}$. The boundary estimator is given by $\bar{f}_n(x_j) := (\hat{\theta}_{I(x_j, t_n)})_0$, where $t_n := (\sigma/L)^{2/(2s+1)}n^{-1/(2s+1)}$. If we define

$$\Gamma_{n,I} := \left\{ \min_{1 \leq m \leq k} \|\phi_{I,m}\|_I \geq n^{-1/2} \right\}, \quad (4.6)$$

where $\|\cdot\|_I^2 = \langle \cdot, \cdot \rangle_I$, then for $x \in [x_j, x_{j+1}]$, $j \in \mathcal{J}_n$, we take

$$\hat{f}_n(x) := \hat{f}_n(x_j) + \left(\sum_{m=1}^k \frac{\tilde{f}_n^{(m)}(x_j)}{m!} (x - x_j)^m \right) \mathbf{1}_{\Gamma_{n,I(x_j, h_n)}}. \quad (4.7)$$

5. PROOF OF THEOREM 1

The whole section is dedicated to the proof of theorem 1. For the sake of brevity, we skip some elements of the proof. A self-contained proof can be found in Gaïffas (2005b), any proof missing here (of the lemmas, mainly) can be found therein. We denote by \mathfrak{X}_n the sigma-algebra generated by X_1, \dots, X_n and by \mathbb{P}_μ^n the joint law of X_1, \dots, X_n . We recall that the discretization points x_j are given by (4.3). We introduce $h_{n,\mu}(x) := [\log n / (n\mu(x))]^{1/(2s+1)}$, and it is convenient to introduce for $j \in \mathcal{J}_n$: $H_j := H_n(x_j)$, $h_j := h_{n,\mu}(x_j)$, $\mu_j := \mu(x_j)$ and $r_j := r_{n,\mu}(x_j)$.

Step 1: approximation by the discretized risk. We introduce the uniform risk

$$\mathcal{E}_{n,f} := \sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\hat{f}_n(x) - f(x)|,$$

and its discretized version $\mathcal{E}_{n,f}^\Delta := \sup_{j \in \mathcal{J}_n} r_j^{-1} |\hat{f}_n(x_j) - f(x_j)|$. In view of assumption D, we obtain $\sup_{x \in [x_j, x_{j+1}]} |r_{n,\mu}(x)^{-1} - r_j^{-1}| = o(1)r_j^{-1}$. Thus, since $f \in \Sigma^Q(s, L)$, writing the Taylor expansion of f at $x \in [x_j, x_{j+1}]$, we obtain:

$$\mathcal{E}_{n,f} \leq (1 + o(1)) \left(\mathcal{E}_{n,f}^\Delta + \max_{j \in \mathcal{J}_n} r_j^{-1} \sum_{m=1}^k |\tilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \frac{\Delta_n^m}{m!} \right) + O(\delta_n^s),$$

where we recall that $\delta_n = (\log n)^{-1}$. In this step, we need the following lemma, which provides a control over the local polynomial estimator uniform risk. Its proof can be found in Gaïffas (2005b).

Lemma 1. *There is an event $\mathcal{C}_n \in \mathfrak{X}_n$ such that, under assumption D,*

$$\mathbb{P}_\mu^n \{ \mathcal{C}_n^c \} \leq \exp(-D_C n^{s/(2s+1)}), \quad (5.1)$$

where $D_C > 0$, and a centered Gaussian vector $W \in \mathbb{R}^{(k+1)M_n}$ with $\mathbb{E}_{f,\mu}^n \{ W_p^2 \} = 1$, $0 \leq p \leq (k+1)M_n$, such that on \mathcal{C}_n , one has for any $0 \leq m \leq k$ and $f \in \Sigma(s, L)$:

$$\max_{j \in \mathcal{J}_n} |\tilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| = O(h_n^{s-m})(1 + (\log n)^{-1/2} W^M), \quad (5.2)$$

where $W^M := \max_{0 \leq p \leq (k+1)M_n} |W_p|$. For the estimator near the boundaries, we have on \mathcal{C}_n , for $a = 1$ (the case $a = 3$ is similar):

$$\max_{j \in \mathcal{J}_{1,n}} |\bar{f}_n(x_j) - f(x_j)| = O(t_n^s)(1 + W^{(1)}), \quad (5.3)$$

where $W^{(1)} = \max_{0 \leq p \leq (k+1)|\mathcal{J}_{1,n}|} |W_p|$.

In view of (5.2), we have on \mathcal{C}_n , for any $1 \leq m \leq k$:

$$\max_{j \in \mathcal{J}_n} r_j^{-1} |\tilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \Delta_n^m / m! = O(\delta_n^m)(1 + (\log n)^{-1/2} W^M),$$

then $\mathcal{E}_{n,f} \mathbf{1}_{C_n} \leq (1 + o(1)) \mathcal{E}_{n,f}^\Delta \mathbf{1}_{C_n} + O(\delta_n)(1 + \delta_n^{1/2} W^M) + o(1)$. Let us define the event $\mathcal{W}_n := \{|W^M - \mathbb{E}_{f,\mu}^n \{W^M\}| \leq \delta_n^{-1}\}$. Since W is a centered Gaussian vector such that $\mathbb{E}_{f,\mu}^n \{W_p^2\} = 1$ for $0 \leq p \leq (k+1)M_n$, it is well known (see for instance in Ledoux and Talagrand (1991)) that $\mathbb{E}_{f,\mu}^n \{W^M\} \leq [2 \log((k+1)M_n)]^{1/2} = O(\delta_n^{-1/2})$ and

$$\mathbb{P}_{f,\mu}^n \{\mathcal{W}_n^c\} \leq 2 \exp(-\delta_n^{-2}/2). \quad (5.4)$$

Thus,

$$\mathcal{E}_{n,f} \mathbf{1}_{C_n \cap \mathcal{W}_n} \leq (1 + o(1)) \mathcal{E}_{n,f}^\Delta \mathbf{1}_{C_n} + o(1). \quad (5.5)$$

Step 2: some events study. In what follows, it is convenient to write K instead of K_s , and to introduce

$$K_{ij} := K[(X_i - x_j)/(c_s h_j)], \quad \bar{K}_{ij} := K[(X_i - x_j)/(c_s H_j)],$$

and $q_j := n c_s h_j \mu_j$, $\bar{q}_j := n c_s H_j \mu_j$, where c_s is given by (3.7). We introduce also

$$\bar{Q}_j := \sum_{i=1}^n \bar{K}_{ij}, \quad Q_j := \sum_{i=1}^n K_{ij}, \quad S_j := \sum_{i=1}^n \bar{K}_{ij}^2,$$

and the events

$$\begin{aligned} A_{n,j} &:= \{|\bar{Q}_j/\bar{q}_j - 1| \leq L_A \delta_n^{\min(s,1)}\}, & B_{n,j} &:= \{|Q_j/q_j - 1| \leq \delta_n\}, \\ C_{n,j} &:= \{|H_j/h_j - 1| \leq \delta_n\}, & E_{n,j} &:= \{|S_j/q_j - \|K\|_2^2| \leq L_E \delta_n^{\min(s,1)}\}, \\ \mathcal{B}_n &:= \bigcap_{j \in \mathcal{J}_{2,n}} (A_{n,j} \cap B_{n,j} \cap E_{n,j}) \cap \bigcap_{j \in \mathcal{J}_n} C_{n,j}, \end{aligned} \quad (5.6)$$

where L_A and L_E are some fixed positive constants, $\delta_n = (\log n)^{-1}$, and the sets of indices $\mathcal{J}_{a,n}$ are defined in section 4. In this step, we control the probabilities of these events.

For $j \in \mathcal{J}_{2,n}$, we consider the sequence of i.i.d variables $\zeta_{ij} := K_{ij} - \mathbb{E}_\mu^n \{K_{ij}\}$, $1 \leq i \leq n$. Since $\mu \in \Sigma_q(\nu, \varrho)$ and $\int K = 1$, we have for n large enough $|\mathbb{E}_\mu^n \{K_{1j}\}/q_j - 1| \leq \delta_n/2$, thus $B_{n,j}^c \subset \{|\sum_{i=1}^n \zeta_{ij}|/q_j \leq \delta_n/2\}$. Since $|\zeta_{ij}| \leq 2\|K\|_\infty$ and $\sum_{i=1}^n \mathbb{E}_\mu^n \{\zeta_{ij}^2\} \leq (1 + \delta_n)q_j \int K^2$ for n large enough, Bernstein inequality entails $\mathbb{P}_\mu^n \{B_{n,j}^c\} \leq 2 \exp(-D_1 \delta_n^2 n^{2s/(2s+1)})$, for any $j \in \mathcal{J}_{2,n}$, where D_1 is a positive constant. Since $\varphi_s \in \Sigma(s, 1; \mathbb{R})$, we have $K \in \Sigma(\min(s, 1), L_K; \mathbb{R})$ where $L_K := (\int \varphi_s)^{-1}$ if $0 < s \leq 1$ and $L_K := \|K'\|_\infty$ if $s > 1$. Since $\text{Supp } K = [-T_s, T_s]$, we have on $C_{n,j}$

$$|\bar{K}_{ij} - K_{ij}| \leq L_K T_s^{\min(s,1)} [\delta_n/(1 - \delta_n)]^{\min(s,1)} \mathbf{1}_{M_{ij}} = o(1) \mathbf{1}_{M_{ij}}, \quad (5.7)$$

where $M_{ij} := \{|X_i - x_j| \leq c_s T_s (1 + \delta_n) h_j\}$. We define $\eta_{ij} := \mathbf{1}_{M_{ij}} - \mathbb{P}_\mu^n \{M_{ij}\}$. Then, we obtain that on $C_{n,j}$ and for n large enough, taking $L_A := 4(L_K T_s^{\min(s,1)+1} + 1)$,

$$\mathbb{P}_{f,\mu}^n \{A_{n,j}^c \cap C_{n,j}\} \leq \mathbb{P}_\mu^n \{|\sum_{i=1}^n \eta_{ij}| > \delta_n^{\min(s,1)} q_j\} + \mathbb{P}_\mu^n \{|\sum_{i=1}^n \zeta_{ij}| > \delta_n^{\min(s,1)} q_j/2\}.$$

Then, applying Bernstein inequality to the sum of variables η_{ij} and ζ_{ij} , $1 \leq i \leq n$, we obtain that for any $j \in \mathcal{J}_{n,2}$, $\mathbb{P}_\mu^n \{A_{n,j}^c \cap C_{n,j}\} \leq 2 \exp(-D_2 \delta_{2,n}^2 n^{2s/(2s+1)})$, where D_2 is a positive constant and $\delta_{2,n} := \delta_n^{\min(s,1)}$. We can prove $\mathbb{P}_\mu^n \{E_{n,j}^c \cap C_{n,j}\} \leq 2 \exp(-D_3 \delta_{2,n}^2 n^{2s/(2s+1)})$ where D_3 is a positive constant in the same way as previously, with an appropriate choice for L_E . If $I = I(x, h)$ (see (4.1)) and $\delta_{1,n} := 1 - (1 + \delta_n)^{-(2s+1)}$, we define the event $N_{n,I} := \{|\bar{\mu}_n(I)/(\mu(x)h) - 1| \leq \delta_{1,n}\}$. From the definitions of H_j and h_j , we obtain

$$\begin{aligned} \{(1 - \delta_n)h_j < H_j\} &= \{(1 - \delta_n)^{2s} h_j^{2s} < \log n / (n \bar{\mu}_n(I(x_j, (1 - \delta_n)h_j)))\} \\ &= \left\{ \frac{\bar{\mu}_n(I(x_j, (1 - \delta_n)h_j))}{\mu_j(1 - \delta_n)h_j} \leq (1 - \delta_n)^{-(2s+1)} \right\}, \end{aligned}$$

and then $N_{n,I(x_j, (1 - \delta_n)h_j)} \subset \{(1 - \delta_n)h_j < H_j\}$. We can prove in the same way that on the other hand $N_{n,I(x_j, (1 + \delta_n)h_j)} \subset \{(1 + \delta_n)h_j \geq H_j\}$, hence

$$N_{n,I(x_j, (1 - \delta_n)h_j)} \cap N_{n,I(x_j, (1 + \delta_n)h_j)} \subset C_{n,j}. \quad (5.8)$$

If $I = I(x, h)$, we have in view of assumption D that $|\int_I \mu(t)dt - h\mu(x)| = O(h^{\nu+1})$, thus, if $Z_i := \mathbf{1}_{X_i \in I} - \int_I \mu(t)dt$, we have $\{|\sum_{i=1}^n Z_i| \leq n\mu(x)h\delta_{n,1}/2\} \subset N_{n,I}$ for n large enough. Then, using Bernstein inequality to the sum of Z_i , $1 \leq i \leq n$, we obtain $\mathbb{P}_\mu^n\{C_{n,j}^c\} \leq 2\exp(-D_C\delta_{1,n}^2n^{2s/(2s+1)})$, for n large enough, where $D_C > 0$ is fixed. Using together the previous inequalities, we obtain

$$\mathbb{P}_\mu^n\{\mathcal{B}_n^c\} \leq \exp(-D_B n^{s/(2s+1)}) \quad (5.9)$$

for n large enough, where $D_B > 0$ is fixed.

Step 3: controls on $\mathcal{E}_{n,f}^\Delta$. We need the following lemma, which is proven in Gaïffas (2005b).

Lemma 2. *There is an event $\mathcal{A}_n \in \mathfrak{X}_n$ such that, under assumption D,*

$$\mathbb{P}_\mu^n\{\mathcal{A}_n^c\} \leq \exp(-D_A n^{s/(2s+1)}) \quad (5.10)$$

for n large enough, where $D_A > 0$ and such that

$$\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n \cap \Gamma_n, \quad (5.11)$$

where \mathcal{B}_n is given by (5.6), \mathcal{C}_n by lemma 1 and $\Gamma_n := \bigcap_{j \in \mathcal{J}_n} \Gamma_{n,I}(x_j, h_n)$ where $\Gamma_{n,I}$ is defined by (4.6).

In this step, we prove that for any $\varepsilon > 0$, when n is large enough, the following inequality holds:

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{P}_{f,\mu}^n\{\mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n} > (1 + \varepsilon)P\} \leq \exp(-D_\mathcal{E} \varepsilon(1 \wedge \varepsilon)(\log n)^{2s/(2s+1)}), \quad (5.12)$$

where $D_\mathcal{E} > 0$, and we prove that

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n\{w^2(\mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n})\} = O(1). \quad (5.13)$$

We decompose the risk into $\mathcal{E}_{n,f}^\Delta = \mathcal{E}_{n,f}^{\Delta,1} + \mathcal{E}_{n,f}^{\Delta,2} + \mathcal{E}_{n,f}^{\Delta,3}$ where $\mathcal{E}_{n,f}^{\Delta,a} := \sup_{j \in \mathcal{J}_{a,n}} r_j^{-1} |\widehat{f}_n(x_j) - f(x_j)|$. For $a = 1$ and $a = 3$, $\mathcal{E}_{n,f}^{\Delta,a}$ is the risk at the boundaries of $[0, 1]$. Since on \mathcal{B}_n , $\bar{Q}_j/\bar{q}_j \geq 1 - L_A \delta_n^{\min(s,1)} > \delta_n$ for n large enough, the denominator in (4.5) is larger than δ_n . Hence, we can decompose on \mathcal{B}_n the middle risk into bias and variance terms as follows:

$$\mathcal{E}_{n,f}^{\Delta,2} \leq b_{n,f} + U_{n,f} + Z_n, \quad (5.14)$$

where the variance term is given by $Z_n := \max_{j \in \mathcal{J}_{2,n}} |Z_{n,j}|$, $Z_{n,j} := r_j^{-1} \sum_{i=1}^n \xi_i W_{ij}$ with $W_{ij} := \bar{K}_{ij}/\bar{Q}_j$, and the bias terms are $b_{n,f} := \max_{j \in \mathcal{J}_{2,n}} |b_{n,f,j}|$, $U_{n,f} := \max_{j \in \mathcal{J}_{2,n}} |U_{n,f,j}|$, where $b_{n,f,j} := \mathbb{E}_{f,\mu}^n\{B_{n,f,j} \mathbf{1}_{\mathcal{B}_n}\}$, $U_{n,f,j} := B_{n,f,j} - b_{n,f,j}$ with $B_{n,f,j} := r_j^{-1} \bar{P}_j/\bar{Q}_j$, $\bar{P}_j := \sum_{i=1}^n (f(X_i) - f(x_j)) \bar{K}_{ij}$. We use the three following inequalities: we have

$$\limsup_n \sup_{f \in \Sigma(s,L)} b_{n,f} \leq Lc_s^s B(s, 1), \quad (5.15)$$

where $B(s, L)$ is defined by (3.6), and there is a constant $D_U > 0$ such that for any $\varepsilon > 0$,

$$\sup_{f \in \Sigma(s,L)} \mathbb{P}_{f,\mu}^n\{U_{n,f} \mathbf{1}_{\mathcal{B}_n} > \varepsilon\} \leq \exp(-D_U \varepsilon(1 \wedge \varepsilon)n^{2s/(2s+1)}). \quad (5.16)$$

Moreover, we have for any $\varepsilon > 0$,

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{P}_{f,\mu}^n\{Z_n \mathbf{1}_{\mathcal{B}_n} > (1 + \varepsilon)Lc_s^s \|K\|_2\} \leq 2(\log n)^{2s/(2s+1)} n^{-\varepsilon/(2s+1)}. \quad (5.17)$$

Inequalities (5.15) and (5.17) are proven below in the section, the proof of (5.16) can be found in Gaïffas (2005b). In view of (5.15), we have $b_{n,f} \leq (1 + 2\varepsilon)Lc_s^s B(s, 1)$ for n large enough, and using (3.8) we obtain

$$\{\mathcal{E}_{n,f}^{\Delta,2} \mathbf{1}_{\mathcal{B}_n} > (1 + 2\varepsilon)P\} \subset \{Z_n \mathbf{1}_{\mathcal{B}_n} > (1 + \varepsilon)Lc_s^s \|K\|_2\} \cup \{U_{n,f} \mathbf{1}_{\mathcal{B}_n} > \varepsilon Lc_s^s \|K\|_2\}.$$

Then, in view of (5.16) and (5.17), it is easy to find $D_2 > 0$ such that uniformly for $f \in \Sigma^Q(s, L)$ and n large enough,

$$\mathbb{P}_{f,\mu}^n \{\mathcal{E}_{n,f}^{\Delta,2} \mathbf{1}_{\mathcal{B}_n} > (1 + 2\varepsilon)P\} \leq \exp(-D_2 \varepsilon (1 \wedge \varepsilon) \log n). \quad (5.18)$$

Now, we consider the boundary risk $\mathcal{E}_{n,f}^{\Delta,1}$ (the result is the same for $\mathcal{E}_{n,f}^{\Delta,3}$). In view of (5.3), we obtain $\mathcal{E}_{n,f}^{\Delta,1} \mathbf{1}_{\mathcal{C}_n} = O(\delta_n^{s/(2s+1)})(1 + W^{(1)})$, and we have as previously $\mathbb{E}_{f,\mu}^n \{W^{(1)}\} = O((\log \log n)^{1/2})$, since $|\mathcal{J}_{1,n}| = O(\log n)$, and for any $\lambda > 0$, $\mathbb{P}_{f,\mu}^n \{W^{(1)} - \mathbb{E}_{f,\mu}^n \{W^{(1)}\} > \lambda\} \leq 2 \exp(-\lambda^2/2)$. Then, for some $D_3 > 0$, we obtain when n is large enough

$$\mathbb{P}_{f,\mu}^n \{\mathcal{E}_{n,f}^{\Delta,1} \mathbf{1}_{\mathcal{C}_n} > 2\varepsilon P\} \leq 2 \exp(-D_3 \varepsilon^2 \delta_n^{-2s/(2s+1)}).$$

This inequality, together with (5.18) and the fact that $\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n$ (see lemma 2) entails (5.12). To prove (5.13), since $w(x) \leq A(1 + |x|^b)$, it suffices to use (5.12) and the fact that $\mathbb{E}_{f,\mu}^n \{(\mathcal{E}_{n,f}^{\Delta})^p \mathbf{1}_{\mathcal{A}_n}\} = p \int_0^{+\infty} t^{p-1} \mathbb{P}_{f,\mu}^n \{\mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{A}_n} > t\} dt$ for any $p > 0$.

Step 4: conclusion of the proof. We need the following inequality, which is proven in Gaïffas (2005b):

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n \{w^2(\mathcal{E}_{n,f})\} = O(n^{2b(1+s/(2s+1))}). \quad (5.19)$$

Since $w(\cdot)$ is nondecreasing, we have for any $\varepsilon > 0$

$$\begin{aligned} \mathbb{E}_{f,\mu}^n \{w(\mathcal{E}_{n,f})\} &\leq \mathbb{E}_{f,\mu}^n \{w(\mathcal{E}_{n,f}) \mathbf{1}_{\mathcal{A}_n \cap \mathcal{W}_n}\} + \mathbb{E}_{f,\mu}^n \{w(\mathcal{E}_{n,f}) \mathbf{1}_{\mathcal{A}_n^c \cup \mathcal{W}_n^c}\} \\ &\leq w((1 + 2\varepsilon)P) + (\mathbb{E}_{f,\mu}^n \{w^2(\mathcal{E}_{n,f})\} \mathbb{P}_{f,\mu}^n \{\mathcal{A}_n^c \cup \mathcal{W}_n^c\})^{1/2} \\ &\quad + (\mathbb{E}_{f,\mu}^n \{w^2((1 + 2\varepsilon)\mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{A}_n})\} \mathbb{P}_{f,\mu}^n \{\mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{A}_n} > (1 + \varepsilon)P\})^{1/2} \\ &\leq w((1 + 2\varepsilon)P) + O(n^{b(1+s/(2s+1))} \exp(-(\log n)^2/4)) \\ &\quad + O(\exp(-D\varepsilon \varepsilon (1 \wedge \varepsilon) (\log n)^{2s/(2s+1)})) = w((1 + 2\varepsilon)P) + o(1), \end{aligned}$$

where we used together lemma 2, equations (5.4), (5.12), (5.13), (5.19), and the fact that $w(\cdot)$ is continuous. Thus, $\limsup_n \sup_{f \in \Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n \{w(\mathcal{E}_{n,f})\} \leq w((1 + 2\varepsilon)P)$, which concludes the proof of theorem 1 since ε can be chosen arbitrarily small. \square

Proof of (5.15). On $\mathcal{A}_{n,j} \cap \mathcal{C}_{n,j}$ we have $(1 - o(1))q_j \leq \bar{Q}_j \leq (1 + o(1))q_j$ and since $\mathcal{B}_n \subset \mathcal{A}_{n,j} \cap \mathcal{C}_{n,j}$ for any $j \in \mathcal{J}_{2,n}$, we have

$$|b_{n,f,j}| = r_j^{-1} |\mathbb{E}_{f,\mu}^n \{(\bar{P}_j/\bar{Q}_j) \mathbf{1}_{\mathcal{B}_n}\}| \leq (1 + o(1))(r_j q_j)^{-1} |\mathbb{E}_{f,\mu}^n \{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}|.$$

Using (5.7), and introducing $\nu_{f,j}(x) := \mathbf{1}_{f(x) \geq f(x_j)} - \mathbf{1}_{f(x) < f(x_j)}$, $P_j := \sum_{i=1}^n (f(X_i) - f(x_j))K_{ij}$, $R_{ij} := \nu_{f,j}(X_i)(f(X_i) - f(x_j))\mathbf{1}_{M_{ij}}$ and $R_j := \sum_{i=1}^n R_{ij}$, we obtain

$$|\mathbb{E}_{f,\mu}^n \{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}| \leq |\mathbb{E}_{f,\mu}^n \{P_j\}| + o(1)|\mathbb{E}_{f,\mu}^n \{R_j\}|.$$

Since $b_{n,f}$ and $U_{n,f}$ only depend on f via its values in $[0, 1]$, we have $\sup_{f \in \Sigma(s,L)} b_{n,f} = \sup_{f \in \Sigma(s,L;\mathbb{R})} b_{n,f}$ and $\sup_{f \in \Sigma(s,L)} U_{n,f} = \sup_{f \in \Sigma(s,L;\mathbb{R})} U_{n,f}$. Thus, together with the fact

that $\Sigma(s, L; \mathbb{R})$ is invariant by translation and since $\mu \in \Sigma_q(\nu, \varrho)$, we obtain

$$\begin{aligned} \sup_{f \in \Sigma(s, L; \mathbb{R})} b_{n, f; j} &\leq (1 + o(1)) \sup_{f \in \Sigma(s, L; \mathbb{R})} \max_{j \in \mathcal{J}_{2, n}} \frac{1}{r_j} \left(\left| \int (f(c_s h_j y) - f(0)) K(y) dy \right| \right. \\ &\quad \left. + o(1) \int_{|y| \leq 2T} |f(c_s h_j y) - f(0)| dy \right). \end{aligned}$$

Now, we use an argument which is known as *renormalisation*, see Donoho and Low (1992). We introduce the functional operator $\mathcal{U}_{a, b} f(\cdot) := af(b\cdot)$. We have that $f \in \Sigma(s, L; \mathbb{R})$ is equivalent to $\mathcal{U}_{a, b} f \in \Sigma(s, Lab^s; \mathbb{R})$. Then, choosing $a = (Lc_s^s h_j^s)^{-1}$ and $b = c_s h_j$ entails

$$\sup_{f \in \Sigma(s, L; \mathbb{R})} b_{n, f} \leq (1 + o(1)) Lc_s^s B(s, 1) + o(1),$$

and the result follows. \square

Proof of (5.17). Conditionally on \mathfrak{X}_n , $Z_{n, j}$ is centered Gaussian with variance $v_j^2 := \sigma^2 r_j^{-2} \sum_{i=1}^n W_{ij}^2$. On \mathcal{B}_n , we have for any $j \in \mathcal{J}_{2, n}$ and n large enough

$$\sum_{i=1}^n W_{ij}^2 = \frac{S_j}{(\bar{Q}_j)^2} \leq (1 + o(1)) \frac{\|K\|_2^2}{q_j} \leq (1 + \varepsilon) \frac{\|K\|_2^2 r_j^2}{c_s \log n},$$

where we used the definition of $h_{n, \mu}(x)$, hence $v_j^2 \leq (1 + \varepsilon) \sigma^2 \|K\|_2^2 / (c_s \log n)$. Using the fact that $P(|N(0, v^2)| \geq \lambda) \leq 2 \exp(-\lambda^2 / (2v^2))$, we obtain

$$\mathbb{P}_{f, \mu}^n \{ |Z_{n, j}| \mathbf{1}_{\mathcal{B}_n} > (1 + \varepsilon) Lc_s^s \|K\|_2 \} \leq 2 \exp\left(-\frac{(1 + \varepsilon)}{2s + 1} \log n\right) = 2n^{-(1 + \varepsilon)/(2s + 1)},$$

and the result follows, since $|\mathcal{J}_{2, n}| \leq M_n \leq (\log n)^{2s/(2s + 1)} n^{1/(2s + 1)}$. \square

6. PROOF OF THEOREM 2

The proof of the lower bound consists in a classical reduction to the Bayesian risk over an hardest cubical subfamily of functions, see Korostelev (1993), Donoho (1994), Korostelev and Nussbaum (1999) and Bertin (2004b). A more detailed proof can be found in Gaïffas (2005b). The main difference with the former proofs is that the subfamily of functions depends on the design via the bandwidth $h_{n, \mu}(x)$, and that we work within a "small" interval I_n . We recall that φ_s is defined by (2.4) and that it has compact support $[-T_s, T_s]$. Let $h_n^I := \max_{x \in I_n} h_{n, \mu}(x)$ and $\Xi_n := 2T_s c_s (2^{1/(s-k)} + 1) h_n^I$. If $I_n = [a_n, b_n]$, $M_n := \lfloor |I_n| \Xi_n^{-1} \rfloor$, we define the points

$$x_j := a_n + j \Xi_n, \quad j \in \mathcal{J}_n := \{1, \dots, M_n\}. \quad (6.1)$$

We denote again $\mu_j = \mu(x_j)$, $h_j = h_{n, \mu}(x_j)$. Let us define the event

$$\mathbf{H}_{n, j} := \left\{ \left| \frac{1}{nc_s h_j \mu_j} \sum_{i=1}^n \varphi_s^2\left(\frac{X_i - x_j}{c_s h_j}\right) - 1 \right| \leq \varepsilon \right\},$$

and $\mathbf{H}_n := \bigcap_{j \in \mathcal{J}_n} \mathbf{H}_{n, j}$. Together with the fact that $\|\varphi_s\|_2 = 1$, we obtain using Bernstein inequality that

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\mu^n \{\mathbf{H}_n\} = 1. \quad (6.2)$$

The subfamily of functions is defined as follows: we consider an hypercube $\Theta \subset [-1, 1]^{M_n}$, and for $\theta \in \Theta$, we define $f(x; \theta) := \sum_{j \in \mathcal{J}_n} \theta_j f_j(x)$, where $f_j(x) := Lc_s^s h_j^s \varphi_s\left(\frac{x - x_j}{c_s h_j}\right)$, so that $f(\cdot; \theta) \in \Sigma(s, L)$. For any $j \in \mathcal{J}_n$, we define the statistics

$$y_j := \frac{\sum_{i=1}^n Y_i \varphi_s(X_i)}{\sum_{i=1}^n \varphi_s^2(X_i)}.$$

Since the f_j have disjoint supports, we have that conditionally on \mathfrak{X}_n , the y_j are Gaussian independent with $\mathbb{E}_{f,\mu}^n\{y_j|\mathfrak{X}_n\} = \theta_j$. Moreover, if $v_j^2 := \mathbb{E}_{f,\mu}^n\{y_j^2|\mathfrak{X}_n\}$, we have on \mathbb{H}_n

$$\frac{2s+1}{2(1+\varepsilon)\log n} \leq v_j^2 \leq \frac{2s+1}{2(1-\varepsilon)\log n}. \quad (6.3)$$

In the model (1.1) with $f(\cdot) = f(\cdot; \theta)$, conditionally on \mathfrak{X}_n , the likelihood function of (Y_1, \dots, Y_n) can be written on \mathbb{H}_n in the form

$$\frac{d\mathbb{P}_{f,\mu}^n}{d\lambda^n} |_{\mathfrak{X}_n}(Y_1, \dots, Y_n) = \prod_{i=1}^n g_\sigma(Y_i) \prod_{j \in \mathcal{J}_n} \frac{g_{v_j}(y_j - \theta_j)}{g_{v_j}(y_j)}, \quad (6.4)$$

where g_v is the density of $N(0, v^2)$, and λ^n is the Lebesgue measure over \mathbb{R}^n . In the following, we denote $\Sigma = \Sigma(s, L)$ and $\mathcal{E}_{n,f,T}^I := \sup_{x \in I} r_{n,\mu}(x)^{-1} |T(x) - f(x)|$. Since $w(\cdot)$ is nondecreasing and $f(\cdot; \theta) \in \Sigma$ for any $\theta \in \Theta$, we have for any probability distribution \mathcal{B} on Θ , by a minoration of the minimax risk by the Bayesian risk,

$$\inf_T \sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \{w(\mathcal{E}_{n,f,T}^I)\} \geq w((1-\varepsilon)P) \inf_T \int_{\Theta} \mathbb{P}_{\theta}^n \{ \mathcal{E}_{n,f,T}^I \geq (1-\varepsilon)P \} \mathcal{B}(d\theta),$$

where $\mathbb{P}_{\theta}^n := \mathbb{P}_{f(\cdot; \theta), \mu}^n$. Since by construction $f(x_j; \theta) = r_j \theta_j P$ and $x_j \in I_n$, we obtain

$$\inf_T \int_{\Theta} \mathbb{P}_{\theta}^n \{ \mathcal{E}_{n,f,T}^I \geq (1-\varepsilon)P \} \mathcal{B}(d\theta) \geq \int_{\mathbb{H}_n} \inf_{\hat{\theta}} \int_{\Theta} \mathbb{P}_{\theta}^n \{ \max_{j \in \mathcal{J}_n} |\hat{\theta}_j - \theta_j| \geq 1 - \varepsilon | \mathfrak{X}_n \} \mathcal{B}(d\theta) d\mathbb{P}_{\mu}^n,$$

where $\inf_{\hat{\theta}}$ is taken among any measurable vector (with respect to the observations (1.1)) in \mathbb{R}^{M_n} . Then, we note that theorem 2 follows from (6.2) if we prove that on \mathbb{H}_n ,

$$\sup_{\hat{\theta}} \int_{\Theta} \mathbb{P}_{\theta}^n \{ \max_{j \in \mathcal{J}_n} |\hat{\theta}_j - \theta_j| < 1 - \varepsilon | \mathfrak{X}_n \} \mathcal{B}(d\theta) = o(1). \quad (6.5)$$

We choose $\Theta := \Theta_{\varepsilon}^{M_n}$ where $\Theta_{\varepsilon} := \{-(1-\varepsilon), 1-\varepsilon\}$ and $\mathcal{B} := \bigotimes_{j \in \mathcal{J}_n} b_{\varepsilon}$ where $b_{\varepsilon} := (\delta_{-(1-\varepsilon)} + \delta_{1-\varepsilon})/2$. Note that using (6.4), the left hand side of (6.5) is smaller than

$$\int \frac{\prod_{i=1}^n g_{\sigma}(Y_i)}{\prod_{j \in \mathcal{J}_n} g_{v_j}(y_j)} \left(\prod_{j \in \mathcal{J}_n} \sup_{\hat{\theta}_j \in \mathbb{R}} \int_{\Theta_{\varepsilon}} \mathbf{1}_{|\hat{\theta}_j - \theta_j| < 1 - \varepsilon} g_{v_j}(y_j - \theta_j) db_{\varepsilon}(\theta_j) \right) dY_1 \dots dY_n,$$

and $\hat{\theta}_j = (1-\varepsilon)\mathbf{1}_{y_j \geq 0} - (1-\varepsilon)\mathbf{1}_{y_j < 0}$ are strategies attaining the maximum. Thus, it suffices to prove the lower bound among estimators $\hat{\theta}$ with coordinates $\hat{\theta}_j \in \Theta_{\varepsilon}$ and measurable with respect to y_j only. Since the y_j are independent with density $g_{v_j}(\cdot - \theta_j)$, the left hand side of (6.5) is smaller than

$$\prod_{j \in \mathcal{J}_n} \left(1 - \inf_{\hat{\theta}_j \in \Theta_{\varepsilon}} \int_{\Theta_{\varepsilon}} \int_{\mathbb{R}} \mathbf{1}_{|\hat{\theta}_j(u) - \theta_j| \geq 1 - \varepsilon} g_{v_j}(u - \theta_j) du db_{\varepsilon}(\theta_j) \right),$$

and if $\Phi(x) := \int_{-\infty}^x g_1(t) dt$ and D_1 is a positive constant,

$$\begin{aligned} & \inf_{\hat{\theta}_j \in \Theta_{\varepsilon}} \int_{\Theta_{\varepsilon}} \int_{\mathbb{R}} \mathbf{1}_{|\hat{\theta}_j(u) - \theta_j| \geq 1 - \varepsilon} g_{v_j}(u - \theta_j) du db_{\varepsilon}(\theta_j) \\ & \geq \inf_{\hat{\theta}_j \in \Theta_{\varepsilon}} \frac{1}{2} \int_{\mathbb{R}} (\mathbf{1}_{\hat{\theta}_j \geq 0} + \mathbf{1}_{\hat{\theta}_j < 0}) g_{v_j}(u - (1-\varepsilon)) \wedge g_{v_j}(u + (1-\varepsilon)) du \\ & = \frac{1}{v_j} \int_{-\infty}^0 g_1\left(\frac{y - (1-\varepsilon)}{v_j}\right) dy = \Phi\left(-\frac{1-\varepsilon}{v_j}\right) \geq \frac{D_1}{\sqrt{\log n}} n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}, \end{aligned}$$

where we used (6.3) and the fact that for $x > 0$, $\Phi(-x) = (1 + o(1)) \exp(-x^2/2)/(x\sqrt{2\pi})$. If $L_n := n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}(\log n)^{-1/2}$, it follows that the left hand side of (6.5) is smaller than $(1 - D_1 L_n)^{M_n} \leq \exp(|I_n| \Xi_n^{-1} \log(1 - D_1 L_n))$, and if D_2 is a positive constant,

$$|I_n| \Xi_n^{-1} L_n = D_2 |I_n| n^{\varepsilon/(2s+1)} \times n^{\varepsilon^2(1-\varepsilon)/(2s+1)} (\log n)^{-1/2-1/(2s+1)} \rightarrow +\infty$$

as $n \rightarrow +\infty$, since $|I_n| n^{\varepsilon/(2s+1)} \rightarrow +\infty$, thus the theorem. \square

REFERENCES

- ARESTOV, V. V. (1990). Optimal recovery of operators and related problems. *Proc. Steklov Inst. Math.*, **4** 1–20.
- BERTIN, K. (2004a). Asymptotically exact minimax estimation in sup-norm for anisotropic h older classes. *Bernoulli*, **10** 873–888.
- BERTIN, K. (2004b). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, **123** 225–242.
- BROWN, L. D., CAI, T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics*, **30** 688 – 707.
- BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, **24** 2384–2398.
- DONOHO, D. L. (1994). Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probability Theory and Related Fields*, **99** 145–170.
- DONOHO, D. L. and LOW, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *The Annals of Statistics*, **20** 944–970.
- FULLER, A. T. (1961). Relay control systems optimized for various performance criteria,. *Automatic and remote control*, **1**.
- GAIFFAS, S. (2005a). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*, **1** 1–27. Available at <http://hal.ccsd.cnrs.fr/ccsd-00003086/en/> .
- GAIFFAS, S. (2005b). Sharp estimation in sup norm with random design. Preprint LPMA no 1010 available at <http://hal.ccsd.cnrs.fr/ccsd-00009135/en/> .
- KOROSTELEV, A. and NUSSBAUM, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, **5** 1099–1118.
- KOROSTELEV, V. (1993). An asymptotically minimax regression estimator in the uniform norm up to exact constant. *Theory of Probability and its Applications*, **38** 737–743.
- LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.
- LEONOV, S. (1997). On the solution of an optimal recovery problem and its applications in nonparametric regression. *Mathematical Methods of Statistics*, **6** 476–490.
- LEONOV, S. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statistics and Probability Letters*, **43** 169–178.
- LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **25** 929–947.
- LEPSKI, O. V. and TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, **117** 17–48.
- MICCHELLI, C. A. and RIVLIN, T. J. (1977). A survey of optimal recovery. *Optimal estimation in approximation theory* 1 – 54.
- STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10** 1040–1053.

LABORATOIRE DE PROBABILIT ES ET MOD ELES AL EATOIRES - UMR 7599, UNIVERSIT E PARIS 7 DENIS-DIDEROT - BO TE COURRIER 7012, 2, PLACE JUSSIEU, 75251 PARIS CEDEX 05 (FRANCE)
E-mail address: gaiffas@math.jussieu.fr