



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. 2005. hal-00009135v1

HAL Id: hal-00009135

<https://hal.science/hal-00009135v1>

Preprint submitted on 27 Sep 2005 (v1), last revised 6 Apr 2006 (v2)

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.

SHARP ESTIMATION IN SUP NORM WITH RANDOM DESIGN

STÉPHANE GAÏFFAS

*Laboratoire de Probabilités et Modèles Aléatoires
 Université Paris 7, 175 rue du Chevaleret, 75013 Paris
 email: gaiffas@math.jussieu.fr*

ABSTRACT. The aim of this paper is to recover the regression function with sup norm loss. We construct an asymptotically sharp estimator which converges with the spatially dependent rate

$$r_{n,\mu}(x) = P(\log n / (n\mu(x)))^{s/(2s+1)},$$

where μ is the design density, s the regression smoothness, n the sample size and P is a constant expressed in terms of a solution to a problem of optimal recovery as in Donoho (1994). We prove this result under the assumption that μ is positive and continuous. This estimator combines kernel and local polynomial methods, where the kernel is given by optimal recovery, which allows to prove the result up to the constants for any $s > 0$. Moreover, the estimator does not depend on μ . We prove that $r_{n,\mu}(x)$ is optimal in a sense which is stronger than the classical minimax lower bound. Then, an inhomogeneous confidence band is proposed. This band has a non constant length which depends on the local amount of data.

1. INTRODUCTION & MAIN RESULTS

1.1. **The model.** Suppose we observe $(X_i, Y_i), 1 \leq i \leq n$, from

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where ξ_i are i.i.d. centered Gaussian with variance σ^2 and independent of X_i , with X_i i.i.d. with density μ on $[0, 1]$, which is bounded away from 0. We want to recover f . In this model, when μ is not the uniform law, we say that the information is spatially inhomogeneous.

1.2. **Methodology.** There are several ways to assess the quality of an estimation procedure. A first approach is local: we focus on recovering f at a fixed point $x_0 \in [0, 1]$. Over a function class Σ , the minimax risk is given by

$$\mathcal{R}_n(\Sigma, x_0) = \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbf{E}_f^n \{ |\hat{f}_n(x_0) - f(x_0)| \},$$

Date: 27th September 2005.

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

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

where the infimum is taken among all estimators. We say that $\rho_n(x_0) > 0$ is the minimax convergence rate at x_0 if

$$0 < \liminf_n \frac{\mathcal{R}_n(\Sigma, x_0)}{\rho_n(x_0)} \leq \limsup_n \frac{\mathcal{R}_n(\Sigma, x_0)}{\rho_n(x_0)} < +\infty.$$

In this paper, we are interested in recovering f globally. We consider the loss with sup norm defined by $\|g\|_\infty = \sup_{x \in [0,1]} |g(x)|$. In this case, the minimax risk is

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

and we say that ψ_n is the minimax convergence rate if

$$0 < \liminf_n \frac{\mathcal{R}_n(\Sigma)}{\psi_n} \leq \limsup_n \frac{\mathcal{R}_n(\Sigma)}{\psi_n} < +\infty.$$

An advantage of this norm is that it is exacting: it forces an estimator to behave well at every point simultaneously. In the regression model (1.1) with Σ a Hölder ball with smoothness $s > 0$, we have when μ is positive and bounded that $\psi_n \asymp (\log n/n)^{s/(2s+1)}$ (see Stone (1982)), where $a_n \asymp b_n$ means $0 < \liminf_n a_n/b_n \leq \limsup_n a_n/b_n < +\infty$.

However, when μ is positive and bounded, ψ_n is not sensitive to the variations in the amount of data. An improvement is to consider instead of (1.2) the spatially dependent risk

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

where \hat{f}_n is some estimator and $r_n(\cdot) > 0$ a family of spatially dependent normalisation factors. If this quantity is bounded as n goes to infinity, we say that $r_n(\cdot)$ is an upper bound over Σ . If we look for such upper bounds, we clearly find that $r_n(x) \asymp \psi_n$ for any x , thus we must sharp this upper bound up to constants. Here, we consider indeed the latter approach in the asymptotic minimax context. In this paper, we develop the consequences of inhomogeneous data within this framework.

1.3. Upper and lower bounds. If $s, L > 0$, we define the Hölder ball $\Sigma(s, L)$, which is the set of all the functions $f : [0, 1] \rightarrow \mathbb{R}$ such that for any $x, y \in [0, 1]$,

$$|f^{(k)}(x) - f^{(k)}(y)| \leq L|x - y|^{s-k},$$

where $k = \lfloor s \rfloor$ is the largest integer $k < s$. If $Q > 0$, we denote by $\Sigma^Q(s, L)$ the set of functions $f \in \Sigma(s, L)$ such that $\|f\|_\infty \leq Q$, and we denote simply $\Sigma = \Sigma^Q(s, L)$. All along this study, we suppose:

Assumption D. For some $0 < \nu \leq 1$ and $\varrho, q > 0$, we have

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

In the following, a loss function $w(\cdot)$ is any non negative and nondecreasing function such that $w(x) \leq A(1 + |x|^b)$ for some $A, b > 0$ (an example is $w(\cdot) = |\cdot|^p$ for $p > 0$). Let us consider

$$r_{n,\mu}(x) = \left(\frac{\log n}{n\mu(x)} \right)^{s/(2s+1)}. \quad (1.3)$$

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) , $1 \leq i \leq n$. Our first result shows that $r_{n,\mu}(\cdot)$ is, up to the constants, an upper bound over Σ .

Theorem 1 (Upper bound). *Under assumption D, if \widehat{f}_n is the estimator defined in section 3, we have for any $s, L > 0$,*

$$\limsup_n \sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \left\{ w \left(\sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \right) \right\} \leq w(P), \quad (1.4)$$

where

$$P = \sigma^{2s/(2s+1)} L^{1/(2s+1)} \varphi_s(0) \left(\frac{2}{2s+1} \right)^{s/(2s+1)} \quad (1.5)$$

and φ_s is defined as the solution of the optimisation problem

$$\varphi_s \triangleq \operatorname{argmax}_{\substack{\varphi \in \Sigma(s,1;\mathbb{R}), \\ \|\varphi\|_2 \leq 1}} \varphi(0), \quad (1.6)$$

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. For further details, see in sections 2 and A. 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 $\varepsilon \in (0, 1)$,*

$$|I_n| n^{\varepsilon/(2s+1)} \rightarrow +\infty \quad \text{as } n \rightarrow +\infty, \quad (1.7)$$

we have

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

where P is given by (1.5) and the infimum is taken among all estimators. A consequence is that if I_n is such that (1.7) holds for any $\varepsilon \in (0, 1)$, we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \left\{ w \left(\sup_{x \in I_n} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \right) \right\} \geq w(P). \quad (1.8)$$

This result is discussed in details in section 2.4. Now, we construct a confidence band which is adapted to inhomogeneous data. Indeed, its length varies depending on the local amount of data.

1.4. An inhomogeneous confidence band. 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

$$I(x, h) = \begin{cases} [x, x + h] & \text{when } 0 \leq x \leq 1/2, \\ [x - h, x] & \text{when } 1/2 < x \leq 1. \end{cases} \quad (1.9)$$

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

$$H_n(x) \triangleq \operatorname{argmin}_{h \in [0, 1]} \left\{ h^s \geq \left(\frac{\log n}{n \bar{\mu}_n(I(x, h))} \right)^{1/2} \right\}, \quad (1.10)$$

which makes the balance between the bias and the variance of a certain kernel estimator (more in section 3 below). We consider the sequence of points

$$x_j = j \Delta_n, \quad \Delta_n = (\log n)^{-2s/(2s+1)} n^{-1/(2s+1)}, \quad (1.11)$$

for $j \in \mathcal{J}_n \triangleq \{0, \dots, [\Delta_n^{-1}]\}$ 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). If $x \in [x_j, x_{j+1})$, we define

$$R_n(x) = H_n(x_j)^s,$$

and for any $x \in [0, 1]$, $\beta > 0$, we consider the band

$$C_{n,\beta}(x) = [\hat{f}_n(x) - (1 + \beta)P R_n(x), \hat{f}_n(x) + (1 + \beta)P R_n(x)], \quad (1.12)$$

where P is defined by (1.5). The next proposition provides a control over the coverage probability of this band, uniformly over $[0, 1]$.

Proposition 1. *Given a confidence level $\alpha \in (0, 1)$, $C_{n,\beta}$ with*

$$\beta = \beta(n, \alpha) = \left(\frac{\log(1/\alpha)}{D_c (\log n)^{2s/(2s+1)}} \right)^{1/2}$$

(where D_c is some positive constant), is under assumption D , a confidence band of asymptotic level $1 - \alpha$, namely:

$$\inf_{f \in \Sigma} \mathbb{P}_{f,\mu}^n \{ f(x) \in C_{n,\beta}(x), \text{ for all } x \in [0, 1] \} \geq 1 - \alpha, \quad (1.13)$$

for n large enough. Moreover, we have for any $x \in [0, 1]$,

$$\sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \{ |C_{n,\beta}(x)| \} / r_{n,\mu}(x) \rightarrow 2P \text{ as } n \rightarrow +\infty. \quad (1.14)$$

In figures 1 and 2, we give a numerical illustration of this confidence band. We consider the function $f(x) = 0.3(1 - |x - 0.5|/0.3)_+$, where $a_+ = \max(a, 0)$. The first dataset is simulated with an uniform design and the second dataset with design density $\mu(x) = 0.05 + 11.4|x - 0.5|^2$. In this example $s = L = 1$, the sample size is $n = 500$ and the root-signal-to-noise ratio is 7.

When the data is homogeneous (uniform design), the length of the confidence band is almost constant, see figure 1. In the non-uniform case, the band is confined at the boundaries of $[0, 1]$ and more spaced at the middle, see figure 2.

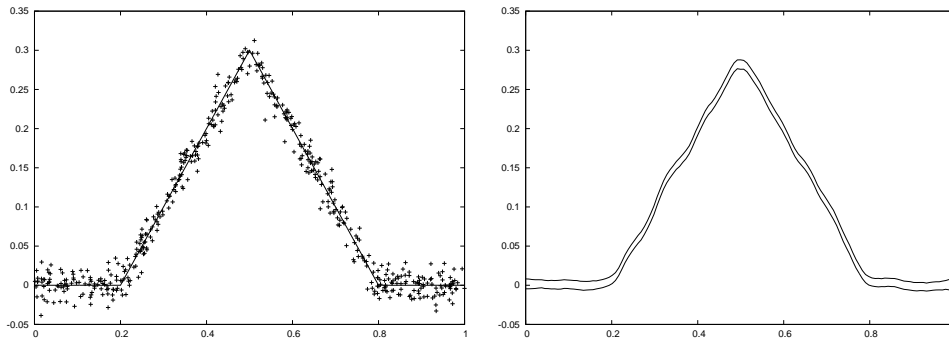


FIGURE 1. Confidence band with homogeneous data.

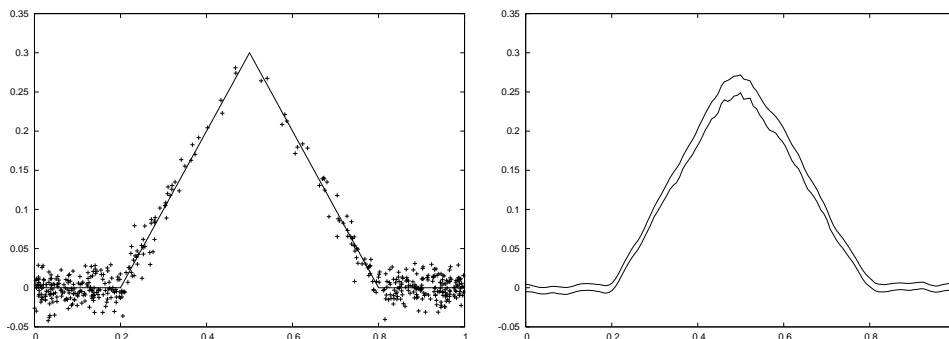


FIGURE 2. Confidence band with inhomogeneous data.

1.5. **Outline.** The remainder of the paper is organised as follows. In section 2 we discuss our results in details and compare them with former results. In section 3, we construct the estimator used in theorem 1. The proofs are delayed until sections 4 and 5. In section A, we recall some well known facts on optimal recovery, which are useful for the construction of our estimator and for the proofs.

2. DISCUSSION

2.1. **Motivation.** 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 implicitly assumed constant over space (or time). However, an increasing literature works in models where the data can be inhomogeneously distributed. Recent results deal with the estimation of the regression function when the observation points are not equispaced or random, see for instance Antoniadis et al. (1997), Brown and Cai (1998), Wong and Zheng (2002), Maxim (2003), among others. The estimators proposed in these papers present good minimax properties, but the results are always stated in a way in which the following basic principle does not appear: *an estimator shall behave better at a point where there is much data than where there is little data*. For instance, upper bounds are usually stated with the minimax rate, which

is not sensitive to the variations in the local amount of data nor to the information distribution in the considered model.

At this stage, it is also natural to look for confidence bands when the data is inhomogeneous, and especially distributed with an unknown density. Following the above principle, a striking question is that of the construction of a confidence band with a length which depends on the local amount of data: such a band should be more confined where there is much data than where there is little data. The aim of this paper is to develop this new approach.

2.2. 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

$$\psi_n = \left(\frac{\log n}{n}\right)^{s/(2s+1)},$$

we have for any $0 < s \leq 1$ and $\Sigma = \Sigma(s, L)$,

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

where

$$C = \sigma^{2s/(2s+1)} L^{1/(2s+1)} \left(\frac{s+1}{2s^2}\right)^{s/(2s+1)}. \quad (2.1)$$

This result was the first of its kind for sup norm error loss. The exact asymptotic value of the minimax risk was only known for square integrated norm error loss, see Pinsker (1980).

In the white noise model

$$dY_t^n = f(t)dt + n^{-1/2}dW_t, \quad t \in [0, 1], \quad (2.2)$$

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 section A). It is shown for any $s > 0$ and $\Sigma = \Sigma(s, L)$ that the minimax risk satisfies

$$\lim_{n \rightarrow +\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbf{E}_f \{\psi_n^{-1} \|\hat{f}_n - f\|_\infty\} = w(P_1), \quad (2.3)$$

where P_1 is given by (1.5) 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.

While most papers on sharp estimation work in models with homogeneous information, the paper by Bertin (2004c) works in the model of regression with random design (1.1). When μ satisfies assumption D and $\Sigma = \Sigma^Q(s, L)$ for $0 < s \leq 1$, it is shown that

$$\lim_{n \rightarrow +\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \{w(v_{n, \mu}^{-1} \|\hat{f}_n - f\|_\infty)\} = w(C), \quad (2.4)$$

where C is given by (2.1) and

$$v_{n,\mu} = \left(\frac{\log n}{n \inf_x \mu(x)} \right)^{s/(2s+1)}. \quad (2.5)$$

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 only via its infimum, which corresponds to the point in $[0, 1]$ where we have the least information. This rate does not take into account the regions with more data. It seems natural to wonder if we can improve this result, namely: *can we replace $\inf \mu$ by $\mu(x)$* ? Note that in section 1, we have answered positively to this question.

In this paper, we extend the result by Donoho (1994) to the model of regression with random design and we improve the result by Bertin (2004c) in several ways: our result holds for any $s > 0$, we construct an estimator which does not depend on μ , and when the design is not uniform, our convergence rate $r_{n,\mu}(\cdot)$ is better (smaller) than $v_{n,\mu}$ at the order of constants. More importantly, this rate is adapted to the local amount of information of the model.

2.3. About theorem 1. We can understand the result of theorem 1 heuristically. Following Brown and Low (1996) and Brown et al. (2002) we can find an "idealised" statistical experiment which is equivalent (in the sense that the LeCam deficiency goes to 0) to the model (1.1). The model (1.1) is clearly equivalent to

$$Y_i = f(G_\mu^{-1}(U_i)) + \xi_i, \quad 1 \leq i \leq n,$$

with independent and uniform U_i where $G_\mu(x) = \int_0^x \mu(t)dt$. Under appropriate conditions on f and μ , we know from Brown et al. (2002) that this model is equivalent to

$$dZ_t^n = f(G_\mu^{-1}(t))dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in [0, 1],$$

where W is a Brownian motion. Informally, if μ is known we obtain by the time change $t = G_\mu(u)$,

$$d\tilde{Z}_u^n = f(u)\mu(u)du + \sigma\sqrt{\frac{\mu(u)}{n}}d\tilde{W}_u, \quad u \in [0, 1],$$

where $\tilde{Z}_u = Z_{G_\mu(u)}$ and \tilde{W} is a Brownian motion. Finally, we obtain that (1.1) is equivalent to the heteroscedastic white noise model

$$dY_u^n = f(u)du + \frac{\sigma}{\sqrt{n\mu(u)}}dB_u, \quad u \in [0, 1], \quad (2.6)$$

where B is a Brownian motion. In view of the result by Donoho (1994) (see (2.3)) which is stated in the model (2.2) and comparing the noise levels in the models (2.2) and (2.6) (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 $n\mu(x)$.

2.4. About theorem 2. From Bertin (2004c), 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 $v_{n, \mu}$ is given by (2.5). An immediate consequence is

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}_{f, \mu}^n \{w(\sup_{x \in [0, 1]} r_{n, \mu}(x)^{-1} |\widehat{f}_n(x) - f(x)|)\} \geq w(P), \quad (2.7)$$

where it suffices to use $r_{n, \mu}(x) \leq v_{n, \mu}$ for any $x \in [0, 1]$. This entails that $r_{n, \mu}(\cdot)$ is optimal in the classical minimax sense, but this notion of optimality is weaker than ours. Indeed, to prove the optimality of $r_{n, \mu}(\cdot)$ we need a more "localised" version of the lower bound, hence theorem 2.

In theorem 2, if we choose $I_n = [0, 1]$ we find back (2.7) 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 (1.8).

2.5. About proposition 1. The confidence band $C_{n, \beta}(\cdot)$ is "design adaptive", in the sense that it does not depend on μ , but it depends on the smoothness of f via the parameters s and L . The construction of adaptive confidence bands is more involved. We know from Low (1997) that the construction of an adaptive confidence band without extra assumption is not feasible. However, if extra assumptions on the smoothness of f are supposed, it is possible to construct such confidence bands, see Picard and Tribouley (2000), Hoffmann and Lepski (2002) and Cai and Low (2004a,b). Here, we only focus on the inhomogeneous aspect of the confidence band. Adaptation with respect to the smoothness is beyond the scope of this study, and we would encounter the same limitations.

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 (2004) 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 space dependent convergence rate (whenever it exists) to be different from the classical minimax rate ψ_n not only up to the constants but in order.

3. CONSTRUCTION OF AN ESTIMATOR

3.1. Main idea. The estimator \widehat{f}_n described below is using both kernel and local polynomial methods. Its construction is divided in two parts: first, at the discretisation points x_j defined by (1.11), we use a Nadaraya-Watson estimator with a design data driven bandwidth. This part of the estimator is used to attain the minimax constant. Between the discretisation points, the estimator is defined by a Taylor expansion where the derivatives estimates are done by local polynomial estimation.

3.2. The estimator at points x_j . We consider the bandwidth $H_n(x)$ defined by (1.10) and we define

$$H_n^M = \max_{j \in \mathcal{J}_n} H_n(x_j),$$

where x_j and \mathcal{J}_n are defined in section 1.4. From Leonov (1997, 1999) we know that the function φ_s defined by (1.6) is even and compactly supported. We denote by $[-T_s, T_s]$ its support and $\tau_n \triangleq \min(2c_s T_s H_n^M, \delta_n)$ where $\delta_n = (\log n)^{-1}$ and

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

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 (1.6), we consider the kernel

$$K_s = \frac{\varphi_s}{\int_{\mathbb{R}} \varphi_s}. \quad (3.2)$$

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

$$\hat{f}_n(x_j) \triangleq \begin{cases} \frac{1}{nH_n(x_j)} \sum_{i=1}^n Y_i K_s\left(\frac{X_i - x_j}{c_s H_n(x_j)}\right) & \text{if } j \in \mathcal{J}_{2,n}, \\ \max\left[\delta_n, \frac{1}{nH_n(x_j)} \sum_{i=1}^n K_s\left(\frac{X_i - x_j}{c_s H_n(x_j)}\right)\right] & \\ \bar{f}_n(x_j) & \text{if } j \in \mathcal{J}_{1,n} \cup \mathcal{J}_{3,n}. \end{cases} \quad (3.3)$$

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

3.3. Between the points x_j – local polynomial estimation. We recall that $k = \lfloor s \rfloor$ where s is the smoothness of the unknown signal f . For any interval $I \subset [0, 1]$, 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 (1.9) – for some $x \in [0, 1]$ and $h > 0$, 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 $0 \leq p, q \leq k$. Let us define

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

where $\Omega_{n,I} = \{\lambda(\mathbf{X}_I) \leq 1/\sqrt{n\bar{\mu}_n(I)}\}$ and $\lambda(M)$ is the smallest eigenvalue of a matrix M and \mathbf{I}_{k+1} is the identity matrix on \mathbb{R}^{k+1} . Note that the correction term in $\bar{\mathbf{X}}_I$ entails $\lambda(\bar{\mathbf{X}}_I) \geq 1/\sqrt{n\bar{\mu}_n(I)}$. 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

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

where

$$h_n = (\sigma/L)^{2/(2s+1)} (\log n/n)^{1/(2s+1)},$$

and the estimator at the boundaries of $[0, 1]$ is given by

$$\bar{f}_n(x_j) \triangleq (\hat{\theta}_{I(x_j, t_n)})_0,$$

where $t_n = (\sigma/L)^{2/(2s+1)} n^{-1/(2s+1)}$. Note that the boundary estimator is a local polynomial estimator with the pointwise bandwidth of estimation t_n . If we define

$$\Gamma_{n,I} = \left\{ \min_{1 \leq m \leq k} \|\phi_{I,m}\|_I \geq \frac{1}{\sqrt{n}} \right\}, \quad (3.4)$$

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

$$\hat{f}_n(x) \triangleq \hat{f}_n(x_j) + \left(\sum_{m=1}^k \frac{\tilde{f}_n^{(m)}(x_j)}{m!} (x - x_j)^m \right) \mathbf{1}_{\Gamma_{n,I}(x_j, h_n)}. \quad (3.5)$$

4. PROOF OF THEOREM 1 AND PROPOSITION 1

The proof of theorem 1 needs several preliminary results. In section 4.1 we state the most important lemmas while section 4.2 is devoted to useful results concerning local polynomial estimation. We delay the proofs of these lemmas until section 4.4, since they can be skipped in a first reading. The proofs of theorem 1 and proposition 1 are given in section 4.3. We define the risk

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

and the discretised risk $\mathcal{E}_{n,f}^\Delta = \sup_{j \in \mathcal{J}_n} r_{n,\mu}(x_j)^{-1} |\hat{f}_n(x_j) - f(x_j)|$.

In the following, the notation $o(1)$ stands for a deterministic and positive quantity going to 0 as $n \rightarrow +\infty$ independent of f while $O(1)$ stands for a quantity bounded by a positive quantity independent of f . If A is non negative, we also define $O(A) = O(1) \times A$. We denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. We consider the norms $\|g\|_\infty = \sup_{x \in [0,1]} |g(x)|$, $\|g\|_2 = (\int_0^1 g^2(x) dx)^{1/2}$, and $\|x\|_\infty = \max_{0 \leq m \leq k} |x_m|$, $\|x\|_2 = (\sum_{0 \leq m \leq k} x_m^2)^{1/2}$ when $x \in \mathbb{R}^{k+1}$.

Since $\bar{\mu}_n(I(x, h))/h$ is close to $\mu(x)$ in probability, we have that $H_n(x)$ is close to

$$h_{n,\mu}(x) \triangleq \left(\frac{\log n}{n\mu(x)} \right)^{1/(2s+1)}.$$

To avoid overloaded notations, it is convenient to write K instead of K_s and to introduce for $j \in \mathcal{J}_n$,

$$H_j = H_n(x_j), \quad h_j = h_{n,\mu}(x_j), \quad \mu_j = \mu(x_j), \quad r_j = r_{n,\mu}(x_j),$$

$$K_{i,j} = K\left(\frac{X_i - x_j}{c_s h_j}\right), \quad \bar{K}_{i,j} = K\left(\frac{X_i - x_j}{c_s H_j}\right), \quad W_{i,j} = \frac{\bar{K}_{i,j}}{\sum_{i=1}^n \bar{K}_{i,j}},$$

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 (3.1). We denote by \mathfrak{X}_n the sigma algebra generated by the observations X_i , $1 \leq i \leq n$.

4.1. Preparatory results. We define

$$A_{n,j} \triangleq \left\{ \left| \left(\sum_{i=1}^n \bar{K}_{i,j} \right) / \bar{q}_j - 1 \right| \leq L_1 \delta_n^{s \wedge 1} \right\},$$

where L_1 is a positive constant, and

$$B_{n,j} \triangleq \left\{ \left| \left(\sum_{i=1}^n K_{i,j} \right) / q_j - 1 \right| \leq \delta_n \right\}, \quad C_{n,j} \triangleq \left\{ |H_j / h_j - 1| \leq \delta_n \right\},$$

$$E_{n,j} \triangleq \left\{ \left| \left(\sum_{i=1}^n \bar{K}_{i,j}^2 \right) / q_j - \|K\|_2^2 \right| \leq L_2 \delta_n^{s \wedge 1} \right\},$$

where L_2 is a fixed positive constant and

$$\mathcal{B}_n = \bigcap_{j \in \mathcal{J}_{2,n}} (A_{n,j} \cap B_{n,j} \cap E_{n,j}) \cap \bigcap_{j \in \mathcal{J}_n} C_{n,j}. \quad (4.1)$$

A control over the probability of this event is given in lemma 7 below. Let us denote $Z_n = \max_{j \in \mathcal{J}_{2,n}} |Z_{n,j}|$ where $Z_{n,j} = r_j^{-1} \sum_{i=1}^n \xi_i W_{i,j}$. Informally, the variable Z_n corresponds to the variance term of $\mathcal{E}_{n,f}^\Delta$. We recall that M_n is equal to the cardinal of \mathcal{J}_n .

Lemma 1 (variance term). *For any $\varepsilon > 0$,*

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

Proof. Conditionally on \mathfrak{X}_n , $Z_{n,j}$ is centered Gaussian with variance

$$v_j^2 = \sigma^2 r_j^{-2} \sum_{i=1}^n W_{i,j}^2.$$

On \mathcal{B}_n , we have for any $j \in \mathcal{J}_{2,n}$ and n large enough

$$\sum_{i=1}^n W_{i,j}^2 = \frac{\sum_{i=1}^n \bar{K}_{i,j}^2}{\left(\sum_{i=1}^n \bar{K}_{i,j} \right)^2} \leq (1 + o(1)) \frac{\|K\|_2^2}{q_j} = (1 + o(1)) \frac{\|K\|_2^2 r_j^2}{c_s \log n},$$

where we used the definition of $h_n(x)$, thus $v_j^2 \leq (1 + \varepsilon)\sigma^2\|K\|_2^2/(c_s \log n)$. Using the standard Gaussian deviation, we obtain

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

and bounding from above the probability of $\cup_{j \in \mathcal{J}_{2,n}} \{ |Z_{n,j}| \mathbf{1}_{\mathcal{B}_n} > (1 + \varepsilon)Lc_s^s \|K\|_2 \}$ by the sum of the probabilities, and since $|\mathcal{J}_{2,n}| \leq M_n \leq (\log n)^{2s/(2s+1)} n^{1/(2s+1)}$, the lemma follows. \square

For any $j \in \mathcal{J}_{n,2}$, we define

$$b_{n,f} = \max_{j \in \mathcal{J}_{2,n}} |b_{n,f,j}| \quad \text{and} \quad U_{n,f} = \max_{j \in \mathcal{J}_{2,n}} |U_{n,f,j}|,$$

where $b_{n,f,j} = \mathbb{E}_{f,\mu}^n \{ B_{n,f,j} \mathbf{1}_{\mathcal{B}_n} \}$, $U_{n,f,j} = B_{n,f,j} - b_{n,f,j}$ and

$$B_{n,f,j} = r_j^{-1} \sum_{i=1}^n (f(X_i) - f(x_j)) W_{i,j}.$$

The quantities $b_{n,f}$ and $U_{n,f}$ correspond to bias terms of the risk $\mathcal{E}_{n,f}^\Delta$.

Lemma 2 (first bias term). *We have*

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

where $\mathcal{B}(s, L)$ is defined by (A.2).

Lemma 3 (second bias term). *There is a constant $D_U > 0$ such that for any $\varepsilon > 0$,*

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

The proofs of these lemmas are delayed until section 4.4.

4.2. Local polynomial estimation. In this section we give results concerning local polynomial estimation. This well known estimation procedure provides an efficient method for recovering both a function and its derivatives. The lemma 4 below is one version of the bias variance decomposition of the local polynomial estimator, which is classical: see Korostelev and Tsybakov (1993), Fan and Gijbels (1995, 1996), Spokoiny (1998) and Tsybakov (2003), among many others. To a vector $\theta \in \mathbb{R}^{k+1}$ we associate the polynomial

$$P_\theta(y) = \theta_0 + \theta_1 y + \cdots + \theta_k y^k.$$

If $\hat{\theta}_I$ is the solution of the system $\bar{\mathbf{X}}_I \theta = \mathbf{Y}_I$ (see section 3.3) for $I = I(x, h)$, we define $\hat{f}_I(y) = P_{\hat{\theta}_I}(y - x)$. If $V_{I,k} = \text{Span}\{\phi_{I,m}; 0 \leq m \leq k\}$, we note that on $\Omega_{n,I}$, \hat{f}_I satisfies

$$\langle \hat{f}_I, \phi \rangle_I = \langle Y, \phi \rangle_I, \quad \forall \phi \in V_{I,k}. \quad (4.2)$$

By definition, we have $\tilde{f}_n^{(m)}(x_j) = \hat{f}_{I(x_j, h_n)}^{(m)}(x_j)$, where $\hat{f}_I^{(m)}$ is the derivative of order m of \hat{f}_I , and $\bar{f}_n(x_j) = \hat{f}_{I(x_j, t_n)}(x_j)$, see section 3.3. We introduce the diagonal matrix $\mathbf{\Lambda}_I$ with entries

$$(\mathbf{\Lambda}_I)_{m,m} = \|\phi_{I,m}\|_I^{-1},$$

for $0 \leq m \leq k$, where $\|\cdot\|_I^2 \triangleq \langle \cdot, \cdot \rangle_I$, the symmetrical matrix

$$\mathcal{G}_I \triangleq \mathbf{\Lambda}_I \bar{\mathbf{X}}_I \mathbf{\Lambda}_I,$$

where $\bar{\mathbf{X}}_I$ is introduced in section 3.3 and \mathcal{G} the matrix with entries

$$(\mathcal{G})_{p,q} = \frac{\chi_{p+q}}{\sqrt{\chi_{2p} \chi_{2q}}},$$

for $0 \leq p, q \leq k$, where $\chi_m = (1 + (-1)^m)/(2(m+1))$. It is easy to see that $\lambda(\mathcal{G}) > 0$ (we recall that $\lambda(M)$ is the smallest eigenvalue of a matrix M). We define the event

$$\Omega_n = \bigcap_{j \in \mathcal{J}_n} \Omega_{n,I(x_j, h_n)} \cap \bigcap_{j \in \mathcal{J}_n} \Omega_{n,I(x_j, t_n)},$$

where $\Omega_{n,I}$ is defined in section 3.3 and

$$\mathcal{L}_n = \bigcap_{j \in \mathcal{J}_n} \mathcal{L}_{n,I(x_j, h_n)} \cap \bigcap_{j \in \mathcal{J}_n} \mathcal{L}_{n,I(x_j, t_n)},$$

where if $I = I(x, h)$ for some $x \in [0, 1]$, $h > 0$,

$$\mathcal{L}_{n,I} = \{|\lambda(\mathcal{G}_I) - \lambda(\mathcal{G})| \leq \delta_n\}.$$

For $0 \leq m \leq 2k$ an interval $I \subset [0, 1]$ and $\delta > 0$, we define

$$\bar{\mathcal{D}}_{n,m,I,\delta} \triangleq \left\{ \left| \frac{1}{|\bar{\mu}_n(I)|I^m} \int_I \phi_{j,m} d\bar{\mu}_n - \chi_m \right| \leq \delta \right\},$$

and

$$\mathcal{D}_n = \bigcap_{m=0}^{2k} \left(\bigcap_{j \in \mathcal{J}_n} \bar{\mathcal{D}}_{n,m,I(x_j, h_n), \delta_n} \cap \bigcap_{j \in \mathcal{J}_n} \bar{\mathcal{D}}_{n,m,I(x_j, t_n), \delta_n} \right).$$

We define

$$\mathcal{N}_n = \bigcap_{j \in \mathcal{J}_n} \mathcal{N}_{n,I(x_j, h_n)} \cap \bigcap_{j \in \mathcal{J}_n} \mathcal{N}_{n,I(x_j, t_n)},$$

where

$$\mathcal{N}_{n,I(x,h)} = \left\{ \left| \frac{\bar{\mu}_n(I(x, h))}{\mu(x)h} - 1 \right| \leq \delta_n \right\}.$$

Finally, we introduce

$$\mathcal{C}_n = \Omega_n \cap \mathcal{L}_n \cap \mathcal{D}_n \cap \mathcal{N}_n. \quad (4.3)$$

A control on the probability of this event is given in lemma 7 below. We recall that M_n is the cardinal of \mathcal{J}_n .

Lemma 4. *There exists a centered Gaussian vector $W \in \mathbb{R}^{(k+1)M_n}$ with*

$$\mathbb{E}_{f,\mu}^n \{W_p^2\} = 1, \quad 0 \leq p \leq (k+1)M_n,$$

such that on \mathcal{C}_n , one has for any $0 \leq m \leq k$ and $f \in \Sigma(s, L)$:

$$\max_{j \in \mathcal{J}_n} |\tilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \leq (1 + o(1))CLh_n^{s-m}(1 + (\log n)^{-1/2}W^M), \quad (4.4)$$

where

$$W^M \triangleq \max_{0 \leq p \leq (k+1)M_n} |W_p|,$$

and $C = C_{\lambda,m,q,k}$ where $C_{\lambda,m,q,k} = \lambda^{-1}(\mathcal{G})(k+1)m!\sqrt{2m+1}(1 \vee q^{-1/2})$. For the estimator near the boundaries, we have for $a = 1$ and $a = 3$:

$$\max_{j \in \mathcal{J}_{a,n}} |\bar{f}_n(x_j) - f(x_j)| \leq (1 + o(1))\bar{C}Lt_n^s(1 + W^{(a)}), \quad (4.5)$$

where

$$\begin{aligned} W^{(1)} &= \max_{0 \leq p \leq (k+1)|\mathcal{J}_{1,n}|} |W_p| \\ W^{(3)} &= \max_{(k+1)(|\mathcal{J}_{1,n}| + |\mathcal{J}_{2,n}|) + 1 \leq p \leq (k+1)M_n} |W_p|, \end{aligned}$$

and $\bar{C} = C_{\lambda,0,q,k}$.

Lemma 5. *For any interval $I \subset [0, 1]$ and $p > 0$ we have*

$$\mathbb{E}_{f,\mu}^n \{|\widehat{\theta}_I|_0^p | \mathfrak{X}_n\} = O(n^{p/2}).$$

Moreover, for any $1 \leq m \leq k$, we have on $\Gamma_{n,I}$ (see section 3.3)

$$\mathbb{E}_{f,\mu}^n \{|\widehat{\theta}_I|_m^p | \mathfrak{X}_n\} = O(n^p).$$

The proofs of these lemmas are delayed until section 4.4. The following two lemmas are needed for the proof of theorem 1.

Lemma 6. *If $w(x) \leq A(1 + |x|)^b$ for some $A, b > 0$, we have*

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n \{w^2(\mathcal{E}_{n,f})\} = O(n^{2b(1+s/(2s+1))}). \quad (4.6)$$

We define $\Gamma_n = \cap_{j \in \mathcal{J}_n} \Gamma_{n,I(x_j, h_n)}$ where $\Gamma_{n,I}$ is defined by (3.4). The probability \mathbb{P}_μ^n stands for the joint law of the X_1, \dots, X_n .

Lemma 7. *There exists an event $\mathcal{A}_n \in \mathfrak{X}_n$ such that for n large enough, under assumption D*

$$\mathbb{P}_\mu^n \{\mathcal{A}_n^c\} \leq \exp(-D_{\mathcal{A}}n^{s/(2s+1)}), \quad (4.7)$$

where $D_{\mathcal{A}} > 0$ and

$$\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n \cap \Gamma_n, \quad (4.8)$$

where \mathcal{B}_n is defined by (4.1) and \mathcal{C}_n is defined by (4.3).

4.3. Proofs of the main results. The next proposition is a deviation inequality for the discretised risk $\mathcal{E}_{n,f}^\Delta$. This proposition is of special importance in the proof of theorem 1 and proposition 1.

Proposition 2. *There is $D_\mathcal{E} > 0$ such that for any $\varepsilon > 0$, we have*

$$\sup_{f \in \Sigma^{\mathcal{Q}}(s,L)} \mathbb{P}_{f,\mu}^n \{ \mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n} > (1 + \varepsilon)P \} \leq \exp \left(- D_\mathcal{E} \varepsilon (1 \wedge \varepsilon) (\log n)^{2s/(2s+1)} \right), \quad (4.9)$$

for n large enough. Moreover,

$$\sup_{f \in \Sigma^{\mathcal{Q}}(s,L)} \mathbb{E}_{f,\mu}^n \{ w^2(\mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n}) \} = O(1). \quad (4.10)$$

Proof. We decompose the risk into three parts

$$\mathcal{E}_{n,f}^\Delta = \mathcal{E}_{n,f}^{\Delta,1} + \mathcal{E}_{n,f}^{\Delta,2} + \mathcal{E}_{n,f}^{\Delta,3}, \quad (4.11)$$

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 $a = 3$, the quantity $\mathcal{E}_{n,f}^{\Delta,a}$ is the risk at the boundaries of $[0, 1]$. Note that on \mathcal{B}_n , we have $\sum_{i=1}^n \bar{K}_{i,j}/(nH_j) > c_s \mu_j (1 - L_1 \delta_n^{s \wedge 1}) > c_s q (1 - L_1 \delta_n^{s \wedge 1}) > \delta_n$ for n large enough. Hence, since $\mathcal{A}_n \subset \mathcal{B}_n$ (see lemma 7) we can decompose on \mathcal{A}_n the middle risk into bias and variance terms as follows:

$$\mathcal{E}_{n,f}^{\Delta,2} \leq b_{n,f} + U_{n,f} + Z_n. \quad (4.12)$$

In view of lemma 2 we have for n large enough $b_{n,f} \leq (1 + 2\varepsilon) L c_s^s \mathcal{B}(s, 1)$ and using equation (A.3) we obtain

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

Then, in view of the lemmas 1 and 3, it is easy to find $D_2 > 0$ such that for any $f \in \Sigma^{\mathcal{Q}}(s, L)$ and n large enough,

$$\mathbb{P}_{f,\mu}^n \{ \mathcal{E}_{n,f}^{\Delta,2} \mathbf{1}_{\mathcal{A}_n} > (1 + 2\varepsilon)P \} \leq \exp \left(- D_2 \varepsilon (1 \wedge \varepsilon) \log n \right). \quad (4.13)$$

Using lemma 4, we obtain

$$\mathcal{E}_{n,f}^{\Delta,1} \mathbf{1}_{\mathcal{A}_n} \leq L_3 \delta_n^{s/(2s+1)} (1 + W^{(1)}), \quad (4.14)$$

where $W^{(1)} = \max_{0 \leq p \leq (k+1) \times |\mathcal{J}_{1,n}|} |W_p|$ and $L_3 = \bar{C} \|\mu\|_\infty^{s/(2s+1)}$. 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

$$\mathbb{E}_{f,\mu}^n \{ W^{(1)} \} \leq \sqrt{2 \log((k+1) |\mathcal{J}_{n,1}|)} = O(\sqrt{\log \log n}),$$

since $|\mathcal{J}_{1,n}| = O(\log n)$, and that 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, when n is large enough,

$$\begin{aligned} \mathbb{P}_{f,\mu}^n \{ \mathcal{E}_{n,f}^{\Delta,1} \mathbf{1}_{\mathcal{A}_n} > 2\varepsilon P \} &\leq \mathbb{P}_{f,\mu}^n \{ W^{(1)} - \mathbb{E}_{f,\mu}^n \{ W^{(1)} \} > \varepsilon P \delta_n^{-s/(2s+1)} / L_3 \} \\ &\leq 2 \exp \left(- \varepsilon^2 P^2 \delta_n^{-2s/(2s+1)} / (2L_3^2) \right). \end{aligned}$$

The same result holds for $\mathcal{E}_{n,f}^{\Delta,3}$. Hence, together with (4.13), for a good choice of $D_{\mathcal{E}}$ we obtain (4.9). It is easy to prove (4.10) from (4.9). For any $f \in \Sigma^Q(s, L)$ and $p > 0$, when n is large enough,

$$\begin{aligned} \mathbb{E}_{f,\mu}^n \{ (\mathcal{E}_{n,f}^{\Delta})^p \mathbf{1}_{\mathcal{A}_n} \} &= p \int_0^{+\infty} t^{p-1} \mathbb{P}_{f,\mu}^n \{ \mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{A}_n} > t \} dt \\ &\leq (2P)^p + p e^{D_{\mathcal{E}}} \int_{2P}^{+\infty} t^{p-1} \exp(-D_{\mathcal{E}} t / P) dt = O(1), \end{aligned}$$

thus (4.10), since $w(x) \leq A(1 + |x|^b)$. \square

Proof of theorem 1. Let $x \in [x_j, x_{j+1})$. Since $\mu \in \Sigma(\nu, \varrho)$ with $0 < \nu \leq 1$ we have clearly $\mu^{s/(2s+1)} \in \Sigma(s\nu/(2s+1), \varrho^{s/(2s+1)})$ and using assumption D,

$$\sup_{x \in [x_j, x_{j+1})} |r_{n,\mu}(x)^{-1} - r_j^{-1}| \leq r_j^{-1} \left(\frac{\varrho}{q} \right)^{s/(2s+1)} \Delta_n^{s\nu/(2s+1)} = o(1) r_j^{-1}. \quad (4.15)$$

Since $f \in \Sigma^Q(s, L)$, writing the Taylor expansion of f at $x \in [x_j, x_{j+1})$ we obtain:

$$\begin{aligned} |\widehat{f}_n(x) - f(x)| &\leq |\widehat{f}_n(x_j) - f(x_j)| \\ &\quad + \sum_{m=1}^k (\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)) \frac{(x - x_j)^m}{m!} + L \Delta_n^s, \end{aligned}$$

and in view of (4.15),

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

We consider the event \mathcal{A}_n from lemma 7. Since $\mathcal{A}_n \subset \mathcal{C}_n$ we have that on \mathcal{A}_n , in view of lemma 4 and for any $1 \leq m \leq k$,

$$\begin{aligned} \max_{j \in \mathcal{J}_n} r_j^{-1} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \frac{\Delta_n^m}{m!} \\ \leq (1 + o(1)) \delta_n^m \|\mu\|_{\infty}^{s/(2s+1)} C(1 + (\log n)^{-1/2} W^M), \end{aligned}$$

and then

$$\mathcal{E}_{n,f} \mathbf{1}_{\mathcal{A}_n} \leq (1 + o(1)) \mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{A}_n} + O(1) \delta_n (1 + \delta_n^{1/2} W^M) + o(1).$$

We define $\mathcal{W}_n \triangleq \{|W^M - \mathbb{E}_{f,\mu}^n \{ W^M \}| \leq \delta_n^{-1}\}$. Since $W^M = \max_{0 \leq p \leq (k+1)M_n} |W_p|$, we know in the same way as in the proof of proposition 2 that $\mathbb{E}_{f,\mu}^n \{ W^M \} \leq \sqrt{2 \log((k+1)M_n)} = O(\delta_n^{-1/2})$ and

$$\mathbb{P}_{f,\mu}^n \{ \mathcal{W}_n^c \} \leq 2 \exp(-\delta_n^{-2}/2). \quad (4.16)$$

Thus

$$\mathcal{E}_{n,f} \mathbf{1}_{\mathcal{A}_n \cap \mathcal{W}_n} \leq (1 + o(1)) \mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n} + o(1), \quad (4.17)$$

and since w is non-decreasing, 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 proposition 2, lemmas 6, 7 and the fact that w 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 proposition 1. We consider the event \mathcal{W}_n defined in the proof of theorem 1. Since $\mathcal{A}_n \subset \mathcal{B}_n \subset \mathcal{C}_{n,j}$ for any $j \in \mathcal{J}_n$ we have

$$(1 - o(1))r_j \leq R_n(x_j) \leq (1 + o(1))r_j \quad (4.18)$$

on \mathcal{A}_n . In view of (4.15) and (4.17) we have for any $j \in \mathcal{J}_n$, $x \in [x_j, x_{j+1})$ on $\mathcal{A}_n \cap \mathcal{W}_n$

$$\begin{aligned} R_n(x)^{-1} |\widehat{f}_n(x) - f(x)| &= \frac{r_{n,\mu}(x)}{R_n(x_j)} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \\ &\leq (1 + o(1)) \mathcal{E}_{n,f} \leq (1 + o(1)) \mathcal{E}_{n,f}^\Delta + o(1). \end{aligned}$$

Thus, if $\mathcal{F}_{n,f,\beta} = \{\sup_{x \in [0,1]} R_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \leq (1 + \beta)P\}$ lemma 7, proposition 2 and (4.16) entail for any $f \in \Sigma^Q(s,L)$,

$$\begin{aligned} \mathbb{P}_{f,\mu}^n \{\mathcal{F}_{n,f,\beta}^c\} &\leq \mathbb{P}_{f,\mu}^n \{\mathcal{F}_{n,f,\beta}^c \cap \mathcal{A}_n \cap \mathcal{W}_n\} + \mathbb{P}_{f,\mu}^n \{\mathcal{A}_n^c \cup \mathcal{W}_n^c\} \\ &\leq \mathbb{P}_{f,\mu}^n \{\mathcal{E}_{n,f}^\Delta \mathbf{1}_{\mathcal{A}_n} > (1 + \beta/2)P\} + \mathbb{P}_{f,\mu}^n \{\mathcal{A}_n^c \cup \mathcal{W}_n^c\} \\ &\leq \exp(-D_c \beta(2 \wedge \beta)(\log n)^{2s/(2s+1)}), \end{aligned}$$

for a good choice of D_c . When n is large enough, the choice $\beta = \beta(n, \alpha)$ makes the last part of the above inequality equal to α , hence (1.13). Using again (4.18), lemma 7 and (4.15) it is easy to obtain (1.14). \square

4.4. Proof of lemmas 2, 3, 4, 5, 6 and 7. Since $b_{n,f}$ and $U_{n,f}$ only depend on f via its values in $[0, 1]$, we have

$$\sup_{f \in \Sigma(s,L)} b_{n,f} = \sup_{f \in \Sigma(s,L;\mathbb{R})} b_{n,f}, \quad \sup_{f \in \Sigma(s,L)} U_{n,f} = \sup_{f \in \Sigma(s,L;\mathbb{R})} U_{n,f}. \quad (4.19)$$

Here, it is convenient to introduce $P_j \triangleq \sum_{i=1}^n (f(X_i) - f(x_j)) \bar{K}_{i,j}$ and $Q_j \triangleq \sum_{i=1}^n \bar{K}_{i,j}$.

Proof of lemma 2. On $A_{n,j} \cap C_{n,j}$ we have $(1 - o(1))q_j \leq Q_j \leq (1 + o(1))q_j$ and since $\mathcal{B}_n \subset A_{n,j} \cap C_{n,j}$ for any $j \in \mathcal{J}_{2,n}$, we have

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

Recalling that $K = \varphi_s / \int \varphi_s$ with $\varphi_s \in \Sigma(s, 1; \mathbb{R})$ we have for any $x, y \in \mathbb{R}$

$$|K(x) - K(y)| \leq \kappa |x - y|^{s_1},$$

where $s_1 = s \wedge 1$ and $\kappa = (\int \varphi_s)^{-1}$ when $s \in (0, 1]$ and $\kappa = \|K'\|_\infty$ when $s > 1$. Since $\text{Supp } K = [-T_s, T_s]$, we have for n large enough on \mathcal{B}_n :

$$\begin{aligned} |\bar{K}_{i,j} - K_{i,j}| &\leq \kappa \left| \frac{X_i - x_j}{c_s H_j} \right|^{s_1} \left| \frac{H_j}{h_j} - 1 \right|^{s_1} \mathbf{1}_{|X_i - x_j| \leq c_s T_s (H_j \vee h_j)} \\ &\leq \kappa T_s^{s_1} \left(\frac{\delta_n}{1 - \delta_n} \right)^{s_1} \mathbf{1}_{|X_i - x_j| \leq c_s T_s (1 + \delta_n) h_j} = o(1) \mathbf{1}_{M_{i,j}}, \end{aligned} \quad (4.20)$$

where $M_{i,j} \triangleq \{|X_i - x_j| \leq c_s T_s (1 + \delta_n) h_j\}$. We introduce $\nu_{f,j}(x) = \mathbf{1}_{f(x) \geq f(x_j)} - \mathbf{1}_{f(x) < f(x_j)}$, $R_{i,j} = (f(X_i) - f(x_j))K_{i,j}$, $S_{i,j} = \nu_{f,j}(X_i)(f(X_i) - f(x_j))\mathbf{1}_{M_{i,j}}$, $R_j = \sum_{i=1}^n R_{i,j}$ and $S_j = \sum_{i=1}^n S_{i,j}$. Then,

$$\begin{aligned} &\frac{1}{r_j q_j} |\mathbb{E}_{f,\mu}^n \{ P_j \mathbf{1}_{\mathcal{B}_n} \}| \\ &\leq \frac{1}{r_j q_j} (|\mathbb{E}_{f,\mu}^n \{ R_j \}| + o(1) |\mathbb{E}_{f,\mu}^n \{ S_j \}|) \\ &\leq \frac{1}{r_j \mu_j} \left(\left| \int (f(x_j + y c_s h_j) - f(x_j)) K(y) \mu(x_j + y c_s h_j) dy \right| \right. \\ &\quad \left. + o(1) \left| \int_{|y| \leq (1 + \delta_n) T_s} (f(x_j + y c_s h_j) - f(x_j)) \nu_{f,j}(x_j + c_s y h_j) \mu(x_j + y c_s h_j) dy \right| \right), \end{aligned}$$

and since $\mu \in \Sigma_q(\nu, \varrho)$ we have

$$\begin{aligned} b_{n,f,j} &\leq \frac{1 + o(1)}{r_j} \left| \int (f(x_j + y c_s h_j) - f(x_j)) K(y) dy \right| \\ &\quad + \frac{o(1)}{r_j q} \int_{|y| \leq 2T_s} |f(x_j + y c_s h_j) - f(x_j)| dy. \end{aligned}$$

Using (4.19) and the fact that $\Sigma(s, L; \mathbb{R})$ is invariant by translation,

$$\begin{aligned} \sup_{f \in \Sigma(s, L; \mathbb{R})} b_{n,f,j} &\leq (1 + o(1)) \sup_{f \in \Sigma(s, L; \mathbb{R})} \max_{j \in \mathcal{J}_{2,n}} \frac{1}{r_j} \left(\left| \int (f(c_s h_j y) - f(0)) K(y) dy \right| \right. \\ &\quad \left. + o(1) \int_{|y| \leq 2T} |f(c_s h_j y) - f(0)| dy \right). \end{aligned} \quad (4.21)$$

Now we use an argument which is known as *renormalisation*, see Donoho and Low (1992). We introduce the functional operator $\mathcal{U}_{a,b} f(\cdot) = a f(b \cdot)$. 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 $a = (L c_s^s h_j^s)^{-1}$

and $b = c_s h_j$ entails

$$\sup_{f \in \Sigma(s, L; \mathbb{R})} b_{n, f} \leq (1 + o(1)) L c_s^s \mathcal{B}(s, 1) + o(1) \sup_{f \in \Sigma(s, 1; \mathbb{R})} \int_{|y| \leq 2T} |f(y) - f(0)| dy,$$

where $\mathcal{B}(s, 1)$ is given by (A.2) and where we recall that $r_j = h_j^s$. We define $f_k(y) = f(0) + f'(0)y + \dots + f^{(k)}(0)y^k/k!$. Since $f \in \Sigma(s, L; \mathbb{R})$, we have $f - f_k \in \Sigma(s, L; \mathbb{R})$ and finally

$$\sup_{f \in \Sigma(s, L; \mathbb{R})} b_{n, f} \leq (1 + o(1)) L c_s^s \mathcal{B}(s, 1) + o(1) \int_{|y| \leq 2T} |y|^s dy. \quad \square$$

Proof of lemma 3. We recall that $U_{n, f, j} \triangleq r_j^{-1} (B_j - \mathbb{E}_{f, \mu}^n \{B_j \mathbf{1}_{\mathcal{B}_n}\})$. We use the same notations as in the proof of lemma 2. On \mathcal{B}_n we have $(1 - o(1))q_j \leq Q_j \leq (1 + o(1))q_j$, and since $\mathbb{E}_{f, \mu}^n \{P_j^2\} \leq 4Q^2 \|K\|_\infty^2 n^2$ we obtain in view of lemma 7:

$$\frac{1}{r_j q_j} |\mathbb{E}_{f, \mu}^n \{P_j \mathbf{1}_{\mathcal{B}_n^c}\}| \leq \frac{1}{r_j q_j} \sqrt{\mathbb{E}_{f, \mu}^n \{P_j^2\}} \sqrt{\mathbb{P}_\mu^n \{\mathcal{B}_n^c\}} = o(1).$$

Then, it is easy to see that on \mathcal{B}_n ,

$$|U_{n, f, j}| \leq \frac{1}{r_j q_j} \left((1 + o(1)) |P_j - \mathbb{E}_{f, \mu}^n \{P_j\}| + o(1) |\mathbb{E}_{f, \mu}^n \{P_j \mathbf{1}_{\mathcal{B}_n}\}| \right) + o(1),$$

and we know from the proof of lemma 2 that

$$\frac{1}{r_j q_j} |\mathbb{E}_{f, \mu}^n \{P_j \mathbf{1}_{\mathcal{B}_n}\}| \leq \sup_{f \in \Sigma(s, L)} \max_{j \in \mathcal{J}_{2, n}} \frac{1}{r_j q_j} |\mathbb{E}_{f, \mu}^n \{P_j \mathbf{1}_{\mathcal{B}_n}\}| \leq (1 + o(1)) L c_s^s \mathcal{B}(s, 1),$$

thus $|U_{n, f, j}| \leq (1 + o(1))(r_j q_j)^{-1} |P_j - \mathbb{E}_{f, \mu}^n \{P_j\}| + o(1)$ on \mathcal{B}_n . From the proof of lemma 2, we know that $(r_j q_j)^{-1} |\mathbb{E}_{f, \mu}^n \{S_j\}| = O(1)$, and using (4.20) it is an easy computation to obtain that on \mathcal{B}_n ,

$$|P_j - \mathbb{E}_{f, \mu}^n \{P_j\}| \leq |R_j - \mathbb{E}_{f, \mu}^n \{R_j\}| + o(1) |S_j - \mathbb{E}_{f, \mu}^n \{S_j\}| + o(1) |\mathbb{E}_{f, \mu}^n \{S_j\}|.$$

Then we have for n large enough

$$\begin{aligned} \mathbb{P}_{f, \mu}^n \{ |U_{n, f, j}| \mathbf{1}_{\mathcal{B}_n} > \varepsilon \} &\leq \mathbb{P}_{f, \mu}^n \left\{ |R_j - \mathbb{E}_{f, \mu}^n \{R_j\}| > \frac{\varepsilon r_j q_j}{3} \right\} \\ &\quad + \mathbb{P}_{f, \mu}^n \left\{ |S_j - \mathbb{E}_{f, \mu}^n \{S_j\}| > \frac{\varepsilon r_j q_j}{3} \right\}. \end{aligned}$$

We use Bernstein inequality to the sum of variables $\bar{R}_{i, j} \triangleq R_{i, j} - \mathbb{E}_{f, \mu}^n \{R_{i, j}\}$ and $\bar{S}_{i, j} \triangleq S_{i, j} - \mathbb{E}_{f, \mu}^n \{S_{i, j}\}$, $1 \leq i \leq n$. The variables $(\bar{R}_{i, j})_{1 \leq i \leq n}$ are clearly independent, centered and satisfy $|\bar{R}_{i, j}| \leq 4QK_\infty$. In view of (4.19) and since $\mu \in \Sigma_q(\nu, \varrho)$, it is

easy to prove with the same arguments as in the end of the proof of lemma 2 that

$$\begin{aligned}
\mathbb{E}_{f,\mu}^n \{\bar{R}_{i,j}^2\} &\leq \mathbb{E}_{f,\mu}^n \{R_{i,j}^2\} \\
&\leq (1 + o(1))c_s h_j \mu_j \int (f(x_j + c_s h_j y) - f(x_j))^2 K^2(y) dy \\
&\leq (1 + o(1))c_s h_j \mu_j \sup_{f \in \Sigma(s,L;\mathbb{R})} \int (f(x_j + c_s h_j y) - f(x_j))^2 K^2(y) dy \\
&\leq (1 + o(1))L^2 (c_s h_j)^{2s+1} \mu_j \sup_{f \in \Sigma(s,L;\mathbb{R})} \int (f(y) - f(0))^2 K^2(y) dy \\
&\leq (1 + o(1))L^2 (c_s h_j)^{2s+1} \mu_j \int y^{2s} K^2(y) dy / (k!)^2.
\end{aligned}$$

Then $\sum_{i=1}^n \mathbb{E}_{f,\mu}^n \{\bar{R}_{i,j}^2\} = O(r_j^2 q_j)$ and the Bernstein inequality entails that for n large enough, there is a constant $D_4 > 0$ such that

$$\mathbb{P}_{f,\mu}^n \{|R_j - \mathbb{E}_{f,\mu}^n \{R_j\}| > \varepsilon r_j q_j / 3\} \leq 2 \exp(-D_4 \varepsilon (1 \wedge \varepsilon) n^{s/(2s+1)}).$$

The variables $(\bar{S}_{i,j})_{1 \leq i \leq n}$ are independent, centered and such that $|\bar{S}_{i,j}| \leq 4Q$, and in the same way as previously we can prove $\sum_{i=1}^n \mathbb{E}_{f,\mu}^n \{\bar{S}_{i,j}^2\} = O(r_j^2 q_j)$. Using again Bernstein inequality, it is easy to find D_5 such that

$$\mathbb{P}_{f,\mu}^n \{|S_j - \mathbb{E}_{f,\mu}^n \{S_j\}| > \varepsilon r_j q_j / 3\} \leq 2 \exp(-D_5 \varepsilon (1 \wedge \varepsilon) n^{s/(2s+1)}),$$

and since $|\mathcal{J}_{2,n}| \leq M_n$, we have for any $f \in \Sigma^Q(s, L)$,

$$\begin{aligned}
\mathbb{P}_{f,\mu}^n \{|U_{n,f}| \mathbf{1}_{\mathcal{B}_n} > \varepsilon\} &\leq \sum_{j \in \mathcal{J}_{2,n}} \mathbb{P}_{f,\mu}^n \{|U_{n,f,j}| \mathbf{1}_{\mathcal{B}_n} > \varepsilon\} \\
&\leq 4M_n \exp(- (D_4 \wedge D_5) \varepsilon (1 \wedge \varepsilon) n^{s/(2s+1)}).
\end{aligned}$$

Since $4M_n \exp(- (D_4 \wedge D_5) \varepsilon (1 \wedge \varepsilon) n^{s/(2s+1)} / 2)$ goes to 0 as n goes to $+\infty$, the lemma follows with $D_U = (D_4 \wedge D_5) / 2$. \square

Proof of lemma 4. We take $I = I(x, h)$ for some $x \in [0, 1]$, $h > 0$ and define the vector θ_I with coordinates $(\theta_I)_m = f^{(m)}(x) / m!$ for $0 \leq m \leq k$. Since $\bar{\mathbf{X}}_I = \mathbf{X}_I$ on $\Omega_{n,I}$, we have $\mathbf{\Lambda}_I^{-1}(\hat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1} \mathbf{\Lambda}_I \mathbf{X}_I(\hat{\theta}_I - \theta_I)$. If $f_I(y) = P_{\theta_I}(y - x)$, we have in view of (4.2) for any $0 \leq m \leq k$:

$$\begin{aligned}
(\mathbf{X}_I(\hat{\theta}_I - \theta_I))_m &= \langle \hat{f}_I - f_I, \phi_{I,m} \rangle_I = \langle Y - f_I, \phi_{I,m} \rangle_I \\
&= \langle f - f_I, \phi_{I,m} \rangle_I + \langle \xi, \phi_{I,m} \rangle_I,
\end{aligned}$$

thus $\mathbf{X}_I(\hat{\theta}_I - \theta_I) \triangleq \mathbf{B}_I + \mathbf{V}_I$. Since $f \in \Sigma(s, L)$,

$$(\mathbf{\Lambda}_I \mathbf{B}_I)_m \leq \|\phi_{I,m}\|_I^{-1} |\langle f - f_I, \phi_{I,m} \rangle_I| \leq \|f - f_I\|_I \leq Lh^s / k!,$$

then we can write

$$\mathbf{\Lambda}_I^{-1}(\hat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1} \frac{Lh^s}{k!} u + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}} \mathcal{G}_I^{-1/2} \gamma_I,$$

where $u \in \mathbb{R}^{k+1}$ is such that $\|u\|_\infty \leq 1$ and $\gamma_I = (\sigma \sqrt{n\bar{\mu}_n(I)})^{-1} \mathcal{G}_I^{-1/2} \mathbf{\Lambda}_I \mathbf{D}_I \xi \triangleq \mathbf{T}_I \xi$, where \mathbf{D}_I is the matrix of size $n\bar{\mu}_n(I) \times (k+1)$ with entries $(\mathbf{D}_I)_{i,m} = (X_i - x)^m$,

so that $\mathbf{X}_I = (n\bar{\mu}_n(I))^{-1}\mathbf{D}'_I\mathbf{D}_I$. Since $\mathbf{T}'_I\mathbf{T}_I = \sigma^{-1}\mathbf{I}_{k+1}$, we obtain that γ_I is, conditionally on \mathfrak{X}_n , centered Gaussian with covariance equal to \mathbf{I}_{k+1} .

Consider $I = I(x_j, h)$ for some $j \in \mathcal{J}_n$, $h > 0$. From the inequality $\|\cdot\|_\infty \leq \|\cdot\| \leq \sqrt{k+1}\|\cdot\|_\infty$ and since $\|\mathcal{G}_I^{-1/2}\| \leq \sqrt{k+1}\|\mathcal{G}_I^{-1}\|$ (\mathcal{G}_I is symmetrical with entries smaller than 1 in absolute value) we get

$$\begin{aligned} \|\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I)\|_\infty &\leq \|\mathcal{G}_I^{-1} \frac{Lh^s}{k!} u\|_\infty + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}} \|\mathcal{G}_I^{-1/2} \gamma_I\|_\infty \\ &\leq \|\mathcal{G}_I^{-1}\| (k+1) (Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}} \|\gamma_I\|_\infty) \\ &= \lambda^{-1}(\mathcal{G}_I) (k+1) (Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}} \max_{0 \leq m \leq k} |W_{(k+1)j+m}|), \end{aligned}$$

where $W \triangleq (\gamma_{I(x_0, h)}, \dots, \gamma_{I(x_{M_n}, h)})'$. If $\mathbf{T} \triangleq (\mathbf{T}_{I(x_0, h)}, \dots, \mathbf{T}_{I(x_{M_n}, h)})'$ we have $W = \mathbf{T}\xi$, thus W is a centered Gaussian vector and for any $(k+1)j \leq m \leq (k+1)j+k$, $j \in \mathcal{J}_n$ we have

$$\mathbb{E}_{f, \mu}^n \{W_m^2\} = (\text{Var}\{W\})_{m,m} = (\text{Var}\{\gamma_{I(x_j, h)}\})_{m-(k+1)j, m-(k+1)j} = 1,$$

since $\text{Var}\{\gamma_{I(x_j, h)}\} = \mathbf{I}_{k+1}$. Then, we have proved that on $\cap_{j \in \mathcal{J}_n} \Omega_{n, I(x_j, h)}$,

$$\begin{aligned} \max_{j \in \mathcal{J}_n} \|\mathbf{\Lambda}_{I(x_j, h)}^{-1}(\widehat{\theta}_{I(x_j, h)} - \theta_{I(x_j, h)})\|_\infty \\ \leq \lambda^{-1}(\mathcal{G}_{I(x_j, h)}) (k+1) (Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I(x_j, h))}} W^M), \end{aligned}$$

where $W^M = \max_{0 \leq m \leq (k+1)|\mathcal{J}_n|} |W_m|$. Since $\mathcal{C}_n \subset \mathbb{N}_n \cap \Omega_n \cap \mathcal{L}_n$, we have on \mathcal{C}_n for $h = h_n$ or $h = t_n$,

$$\begin{aligned} \max_{j \in \mathcal{J}_n} \|\mathbf{\Lambda}_{I(x_j, h)}^{-1}(\widehat{\theta}_{I(x_j, h)} - \theta_{I(x_j, h)})\|_\infty \\ \leq (1 + o(1)) \lambda^{-1}(\mathcal{G})(k+1) (Lh^s + \frac{\sigma}{\sqrt{nh\mu_j}} W^M). \end{aligned}$$

Since $\mathcal{C}_n \subset \mathcal{D}_n$, we have for any $j \in \mathcal{J}_n$, $0 \leq m \leq k$,

$$\mathcal{C}_n \subset \bar{\mathcal{D}}_{n, 2m, I(x_j, h_n), \delta_n} \cap \bar{\mathcal{D}}_{n, 2m, I(x_j, t_n), \delta_n},$$

thus on \mathcal{C}_n , when $h = h_n$ or $h = t_n$, we clearly have

$$(\mathbf{\Lambda}_{I(x_j, h)})_{m,m} = \|\phi_{I(x_j, h), m}\|_{I(x_j, h)}^{-1} \leq (1 + o(1)) h^{-m} \sqrt{2m+1}.$$

Since $\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j) = m!((\widehat{\theta}_{I(x_j, h_n)})_m - (\theta_{I(x_j, h_n)})_m)$, it follows that on \mathcal{C}_n :

$$\begin{aligned} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \\ \leq (1 + o(1)) \lambda^{-1}(\mathcal{G}) m! \sqrt{2m+1} (k+1) h_n^{-m} (Lh_n^s + \frac{\sigma}{\sqrt{nh_n\mu_j}} W^M) \\ \leq (1 + o(1)) C L h_n^{s-m} (1 + (\log n)^{-1/2} W^M), \end{aligned}$$

thus (4.4). Inequality (4.5) is obtained similarly. \square

Proof of lemma 5. If $\bar{\mu}_n(I) = 0$ we have $\hat{\theta}_I = 0$ and the result is obvious, thus we assume $\bar{\mu}_n(I) > 0$. In this case, $\mathbf{\Lambda}_I$, $\bar{\mathbf{X}}_I$ and \mathcal{G}_I are invertible, and by definition of $\hat{\theta}_I$,

$$\hat{\theta}_I = \mathbf{\Lambda}_I \mathbf{\Lambda}_I^{-1} \hat{\theta}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{\Lambda}_I \bar{\mathbf{X}}_I \hat{\theta}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{\Lambda}_I \mathbf{Y}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} (\mathbf{B}_I + \mathbf{V}_I),$$

where $(\mathbf{B}_I)_m = \|\phi_{I,m}\|_I^{-1} \langle f, \phi_{I,m} \rangle_I$ and $(\mathbf{V}_I)_m = \|\phi_{I,m}\|_I^{-1} \langle \xi, \phi_{I,m} \rangle_I$. Since $\|f\|_\infty \leq Q$ we have $|(\mathbf{B}_I)_m| = \|\phi_{I,m}\|_I^{-1} |\langle f, \phi_{I,m} \rangle_I| \leq \|f\|_I \leq Q$, thus $\|\mathbf{B}_I\|_\infty \leq Q$.

Conditionally on \mathfrak{X}_n , \mathbf{V}_I is centered Gaussian and it is an easy computation to see that its covariance matrix is equal to $\sigma^2 (n\bar{\mu}_n(I))^{-1} \mathbf{\Lambda}_I \mathbf{X}_I \mathbf{\Lambda}_I$. Then $\mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{V}_I$ is conditionally on \mathfrak{X}_n centered Gaussian with covariance matrix $\sigma^2 (n\bar{\mu}_n(I))^{-1} \bar{\mathbf{X}}_I^{-1} \mathbf{X}_I \bar{\mathbf{X}}_I^{-1}$. If e_m is the canonical vector with coordinates $(e_m)_p = \mathbf{1}_{p=m}$, we have

$$|(\hat{\theta}_I)_m| = |\langle \hat{\theta}_I, e_m \rangle| = |\langle \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{B}_I, e_m \rangle| + \sigma \sqrt{k+1} \gamma,$$

where $\gamma = (\sigma \sqrt{k+1})^{-1} \langle \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{V}_I, e_m \rangle$. By definition, we have $\|\bar{\mathbf{X}}_I^{-1}\| = \lambda^{-1} (\bar{\mathbf{X}}_I) \leq \sqrt{n\bar{\mu}_n(I)}$, and clearly $\|\mathbf{X}_I\| \leq k+1$ and $\|\mathbf{\Lambda}_I^{-1}\| \leq 1$. Then, conditional on \mathfrak{X}_n , γ is centered Gaussian with variance

$$\frac{\langle e_m, \bar{\mathbf{X}}_I^{-1} \mathbf{X}_I \bar{\mathbf{X}}_I^{-1} e_m \rangle}{(k+1)n\bar{\mu}_n(I)} \leq \frac{\|\bar{\mathbf{X}}_I^{-1}\|^2 \|\mathbf{X}_I\|}{(k+1)n\bar{\mu}_n(I)} \leq 1.$$

Since $\|\mathcal{G}_I^{-1}\| \leq \|\mathbf{\Lambda}_I^{-1}\| \|\bar{\mathbf{X}}_I^{-1}\| \|\mathbf{\Lambda}_I^{-1}\| \leq \sqrt{n\bar{\mu}_n(I)} \leq \sqrt{n}$ and $(\mathbf{\Lambda}_I)_{0,0} = 1$, we have

$$\mathbb{E}_{f,\mu}^n \{ |(\hat{\theta}_I)_0|^p | \mathfrak{X}_n \} \leq (k+1)^{p/2} n^{p/2} (Q \vee 1)^p \mathbb{E}_{f,\mu}^n \{ (1 + \sigma |\gamma|)^p | \mathfrak{X}_n \} = O(n^{p/2}),$$

for any $I \subset [0, 1]$, and since $\|\mathbf{\Lambda}_I\| \leq \sqrt{n}$ on $\Gamma_{n,I}$, it follows that

$$\mathbb{E}_{f,\mu}^n \{ |(\hat{\theta}_I)_m|^p | \mathfrak{X}_n \} \leq (k+1)^{p/2} n^p (Q \vee 1)^p \mathbb{E}_{f,\mu}^n \{ (1 + \sigma |\gamma|)^p | \mathfrak{X}_n \} = O(n^p),$$

for any $1 \leq m \leq k$. \square

Proof of lemma 6. We show that for any $p > 0$,

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n \{ \mathcal{E}_{n,f}^p \} = O(n^{p(1+s/(2s+1))}), \quad (4.22)$$

which entails (4.6). By definition of $H_n(x)$, we have $H_n(x) \geq (\log n/n)^{1/(2s)}$ for any $x \in [0, 1]$. Since $\|f\|_\infty \leq Q$, we have for any $j \in \mathcal{J}_{2,n}$,

$$|\hat{f}_n(x_j)| \leq \delta_n^{-1} (n/\log n)^{1/(2s)} (Q + |\bar{\xi}_n|/\sqrt{n}) \|K_s\|_\infty,$$

where $\bar{\xi}_n = \sum_{i=1}^n \xi_i/\sqrt{n}$ is standard Gaussian. Then,

$$\begin{aligned} \mathbb{E}_{f,\mu}^n \{ |\hat{f}_n(x_j)|^p | \mathfrak{X}_n \} &\leq \delta_n^{-p} ((n/\log n)^{p/(2s)} (Q \vee 1))^p \mathbb{E}_{f,\mu}^n \{ (1 + |\bar{\xi}_n|)^p | \mathfrak{X}_n \} \|K_s\|_\infty^p \\ &= O(n^{p/(2s)} (\log n)^{p(1-1/(2s))}). \end{aligned}$$

When $j \in \mathcal{J}_{n,1} \cup \mathcal{J}_{n,3}$, we have $\hat{f}_n(x_j) = \hat{\theta}_{I(x_j, t_n)}$ and in view of lemma 5,

$$\mathbb{E}_{f,\mu}^n \{ |\hat{f}_n(x_j)|^p | \mathfrak{X}_n \} = O(n^{p/2}).$$

For any $j \in \mathcal{J}_n$, since $\tilde{f}_n^{(m)}(x_j) = m! (\hat{\theta}_{I(x_j, h_n)})_m$, we have in view of lemma 5 that on $\Gamma_{n,I(x_j, h_n)}$,

$$\mathbb{E}_{f,\mu}^n \{ |\tilde{f}_n^{(m)}(x_j)|^p | \mathfrak{X}_n \} = O(n^p),$$

for any $1 \leq m \leq k$. Then, we obtain that for any $\|f\|_\infty \leq Q$,

$$\mathcal{E}_{n,f} = O((n/\log n)^{s/(2s+1)}) \left(\sup_{x \in [0,1]} |\widehat{f}_n(x)| + Q \right),$$

and since

$$\sup_{x \in [0,1]} |\widehat{f}_n(x)| \leq \max_{j \in \mathcal{J}_n} \left(|\widehat{f}_n(x_j)| + \left(\sum_{m=1}^k \frac{|\widetilde{f}_n^{(m)}(x_j)|}{m!} \right) \mathbf{1}_{\Gamma_{n,I}(x_j, h_n)} \right) = O(n^p),$$

thus (4.22) and (4.6). \square

Proof of lemma 7. The proof is divided in several steps. We recall that $q_j = nc_s h_j \mu_j$ and $\bar{q}_j = nc_s H_j \mu_j$.

Step 1. We prove that for any $j \in \mathcal{J}_{2,n}$ and n large enough,

$$\mathbb{P}_\mu^n \{B_{n,j}^c\} \leq 2 \exp(-D_1 \delta_n^2 n^{2s/(2s+1)}), \quad (4.23)$$

where D_1 is a positive constant. Consider the sequence of i.i.d variables $\zeta_{i,j} \triangleq K_{i,j} - \mathbb{E}_\mu^n \{K_{i,j}\}$, $1 \leq i \leq n$. Since $\mu \in \Sigma_q(\nu, \varrho)$ and $\int K = 1$, we have for n large enough $|\mathbb{E}_\mu^n \{K_{1,j}\}/q_j - 1| \leq \delta_n/2$, thus $B_{n,j}^c \subset \{|\sum_{i=1}^n \zeta_{i,j}|/q_j \leq \delta_n/2\}$. Since $|\zeta_{i,j}| \leq 2\|K\|_\infty$ and for n large enough $\sum_{i=1}^n \mathbb{E}_\mu^n \{\zeta_{i,j}^2\} \leq (1 + \delta_n)q_j \int K^2$, the Bernstein inequality entails (4.23).

Step 2. We prove that for any $j \in \mathcal{J}_{n,2}$,

$$\mathbb{P}_\mu^n \{A_{n,j}^c \cap C_{n,j}\} \leq 2 \exp(-D_2 \delta_{2,n}^2 n^{2s/(2s+1)}), \quad (4.24)$$

where D_2 is a positive constant and $\delta_{2,n} \triangleq \delta_n^{s_1}$, $s_1 = s \wedge 1$. In view of (4.20), we have on $C_{n,j}$

$$|\bar{K}_{i,j} - K_{i,j}| \leq \kappa T_s^{s_1} \left(\frac{\delta_n}{1 - \delta_n} \right)^{s_1} \mathbf{1}_{M_{i,j}} \quad (4.25)$$

where we recall that $M_{i,j} = \{|X_i - x_j| \leq c_s T_s (1 + \delta_n) h_j\}$. We define $\eta_{i,j} \triangleq \mathbf{1}_{M_{i,j}} - \mathbb{P}_\mu^n \{M_{i,j}\}$. On $C_{n,j}$ we have for n large enough $2c_s T_s H_n^M \leq \delta_n$, and since $x_j \in [\tau_n, 1 - \tau_n]$,

$$\begin{aligned} x_j \leq 1 - \tau_n &= 1 - 2c_s T_s H_n^M \leq 1 - 2c_s T_s H_j \\ &\leq 1 - 2c_s T_s (1 - \delta_n) h_j \leq 1 - c_s T_s (1 + \delta_n) h_j \end{aligned}$$

for n large enough. On the other hand we have similarly $x_j \geq c_s T_s (1 + \delta_n) h_j$. Thus, since $\mu \in \Sigma_q(\nu, \varrho)$ we have

$$\left| \frac{\mathbb{P}_\mu^n \{M_{i,j}\}}{(1 + \delta_n) c_s h_j \mu_j} - 2T_s \right| \leq \frac{1}{q} \int_{|y| \leq T} |\mu(x_j + c_s y (1 + \delta_n) h_j) - \mu_j| dy = O(h_n^\nu). \quad (4.26)$$

Since $x_j \in [c_s T_s(1 + \delta_n)h_j, 1 - (1 + \delta_n)c_s T_s h_j] \subset [c_s T_s h_j, 1 - c_s T_s h_j]$, we have for n large enough on $C_{n,j}$,

$$\begin{aligned} \left| \frac{\mathbb{E}_{f,\mu}^n \{K_{1,j}\}}{c_s H_j \mu_j} - 1 \right| &\leq \frac{h_j}{H_j \mu_j} \int |K(y)| |\mu(x_j + y c_s h_j) - \mu_j| dy + \left| \frac{h_j}{H_j} - 1 \right| \\ &\leq O(h_n^\nu) + \frac{\delta_n}{1 - \delta_n}. \end{aligned} \quad (4.27)$$

Then, combining (4.25), (4.26) and (4.27) we obtain that on $C_{n,j}$ and for n large enough,

$$\begin{aligned} \left| \frac{\sum_{i=1}^n \bar{K}_{i,j}}{q_j} - 1 \right| &\leq \frac{o(1)}{q_j} \left| \sum_{i=1}^n \eta_{i,j} \right| + \frac{\kappa T^{s_1} \delta_n^{s_1}}{(1 - \delta_n)^{s_1}} \frac{\mathbb{P}_\mu^n \{M_{1,j}\}}{c_s H_j \mu_j} \\ &\quad + \frac{1}{\bar{q}_j} \left| \sum_{i=1}^n \zeta_{i,j} \right| + O(h_n^\nu) + \frac{\delta_n}{1 - \delta_n} \\ &\leq \frac{o(1)}{q_j} \left| \sum_{i=1}^n \eta_{i,j} \right| + \frac{1 + o(1)}{q_j} \left| \sum_{i=1}^n \zeta_{i,j} \right| + 2(2\kappa T^{s_1+1} + 1)\delta_n^{s_1}, \end{aligned}$$

and taking $L_1 \triangleq 4(\kappa T^{s_1+1} + 1)$, we obtain

$$\mathbb{P}_{f,\mu}^n \{A_{n,j}^c \cap C_{n,j}\} \leq \mathbb{P}_\mu^n \left\{ \left| \sum_{i=1}^n \eta_{i,j} \right| > \delta_n^{s_1} q_j \right\} + \mathbb{P}_\mu^n \left\{ \left| \sum_{i=1}^n \zeta_{i,j} \right| > \delta_n^{s_1} q_j / 2 \right\}.$$

Then, applying Bernstein inequality to the sum of variables $\eta_{i,j}$ and $\zeta_{i,j}$, $1 \leq i \leq n$, we obtain (4.24). We can prove

$$\mathbb{P}_\mu^n \{E_{n,j}^c \cap C_{n,j}\} \leq 2 \exp(-D_3 \delta_{2,n}^2 n^{2s/(2s+1)}), \quad (4.28)$$

where D_3 is a positive constant in the same way as for the proof of (4.24) with a good choice for L_2 .

Step 3. We define the event

$$D_{n,m,I(x,h),\delta} \triangleq \left\{ \left| \frac{1}{\mu(x)h^{m+1}} \int_{I(x,h)} \phi_{I(x,h),m} d\bar{\mu}_n - \chi_m \right| \leq \delta \right\},$$

and we prove that if $\delta_{1,n} \triangleq 1 - (1 + \delta_n)^{-(2s+1)}$,

$$D_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{1,n}} \cap D_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{1,n}} \subset C_{n,j}. \quad (4.29)$$

From the definitions of H_j and h_j (see section 1.4) 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

$$D_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{1,n}} \subset \{(1 - \delta_n)h_j < H_j\}.$$

We can prove in the same way that on the other hand,

$$D_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{1,n}} \subset \{(1+\delta_n)h_j \geq H_j\},$$

hence (4.29).

Step 4. We prove (4.8). If $\delta_{3,n} = \delta_n/(2 - \delta_n)$, we clearly have for any interval I ,

$$D_{n,m,I,\delta_{3,n}} \cap D_{n,0,I,\delta_{3,n}} \subset \bar{D}_{n,m,I,\delta_n}.$$

Using the fact that $\lambda(M) = \inf_{\|x\|=1} \langle x, Mx \rangle$ for any symmetrical matrix M and since $\mathcal{G}_I, \mathcal{G}, \mathbf{X}_I$ are symmetrical, it is easy to see that

$$\bigcap_{0 \leq p, q \leq k} \left\{ |(\mathcal{G}_I - \mathcal{G})_{p,q}| \leq \frac{\delta_n}{(k+1)^2} \right\} \subset \mathcal{L}_{n,I}, \quad (4.30)$$

and that

$$\begin{aligned} \bigcap_{m=0}^{2k} \bar{D}_{n,m,I,\frac{\delta_n}{(k+1)^2}} &\subset \bigcap_{0 \leq p, q \leq k} \left\{ |(\mathbf{X}_I - \mathbf{X})_{p,q}| \leq \frac{\delta_n}{(k+1)^2} \right\} \\ &\subset \{|\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leq \delta_n\}. \end{aligned}$$

Recalling that if $I = I(x_j, h)$,

$$(\mathcal{G}_I)_{p,q} = \frac{\langle \phi_{I,p}, \phi_{I,q} \rangle_I}{\|\phi_{I,p}\|_I \|\phi_{I,q}\|_I} = \frac{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,p+q} d\bar{\mu}_n}{\sqrt{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2p} d\bar{\mu}_n} \sqrt{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2q} d\bar{\mu}_n}},$$

it is easy to see that if $\delta_{4,n} = \delta_n/((2 - \delta_n)(2k+1)(k+1)^2)$,

$$D_{n,2p,I,\delta_{4,n}} \cap D_{n,2q,I,\delta_{4,n}} \cap D_{n,p+q,I,\delta_{4,n}} \subset \left\{ |(\mathcal{G}_I - \mathcal{G})_{p,q}| \leq \frac{\delta_n}{(k+1)^2} \right\},$$

thus

$$\bigcap_{m=0}^{2k} D_{n,m,I,\delta_{4,n}} \subset \mathcal{L}_{n,I},$$

and clearly for n large enough, if $I = I(x_j, h_n)$ or $I = I(x_j, t_n)$,

$$\bigcap_{m=0}^{2k} D_{n,m,I,\delta_{4,n}} \subset \{|\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leq \delta_n\} \cap \left\{ \left| \frac{\bar{\mu}_n(I)}{|I|\mu_j} - 1 \right| \leq \delta_n \right\} \subset \Omega_{n,I}. \quad (4.31)$$

Moreover, if $I = I(x_j, h_n)$, we have on $\bar{D}_{n,2m,I,\delta_n}$ for any $1 \leq m \leq k$ and n large enough,

$$\|\phi_{I,m}\|_I \geq (1 - o(1))h_n^m \sqrt{2m+1} \geq 1/\sqrt{n}. \quad (4.32)$$

We define

$$\begin{aligned} D_{n,m} \triangleq \bigcap_{j \in \mathcal{J}_n} \left(D_{n,m,I(x_j,h_n),\delta_{5,n}} \cap D_{n,m,I(x_j,t_n),\delta_{5,n}} \right. \\ \left. \cap D_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{5,n}} \cap D_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{5,n}} \right), \end{aligned}$$

where $\delta_{5,n} = \delta_{4,n} \wedge \delta_{3,n} \wedge \delta_{1,n}$, $D_n = \bigcap_{m=0}^{2k} D_{n,m}$ and we choose

$$\mathcal{A}_n \triangleq D_n \cap A_n \cap B_n \cap E_n.$$

In view of (4.29), (4.30), (4.31), (4.32) we have $\mathcal{A}_n \subset C_n \cap \Omega_n \cap \mathcal{L}_n \cap \Gamma_n$ and since $D_{n,0,I,\delta} = N_{n,I}$ we obtain (4.8).

Step 5. We prove (4.7). Using Bernstein inequality, it is easy to show that for n large enough, if $h = h_n$, $h = t_n$, $h = (1 - \delta_n)h_j$ or $h = (1 + \delta_n)h_j$,

$$\mathbb{P}_\mu^n \{D_{n,m,I(x_j,h),\delta_{5,n}}^c\} \leq 2 \exp(-D_4 \delta_{5,n}^2 n h) \leq 2 \exp(-D_5 n^{s/(2s+1)}),$$

with D_4, D_5 positive constants, where we used the fact that $\delta_{5,n}^2 n^{s/(2s+1)} > 1$ for n large enough and $nh \geq D_6 n^{2s/(2s+1)}$. In view of (4.29) we have $D_n \subset C_n$, hence

$$\begin{aligned} \mathbb{P}_\mu^n \{\mathcal{A}_n^c\} &\leq \mathbb{P}_{f,\mu}^n \{D_n^c\} + \mathbb{P}_{f,\mu}^n \{A_n^c \cap C_n\} + \mathbb{P}_{f,\mu}^n \{B_n^c \cap C_n\} \\ &\quad + \mathbb{P}_{f,\mu}^n \{E_n^c \cap C_n\} + 3\mathbb{P}_{f,\mu}^n \{C_n^c\} \\ &\leq 4\mathbb{P}_{f,\mu}^n \{D_n^c\} + \mathbb{P}_{f,\mu}^n \{A_n^c \cap C_n\} + \mathbb{P}_{f,\mu}^n \{B_n^c \cap C_n\} + \mathbb{P}_{f,\mu}^n \{E_n^c \cap C_n\} \\ &\leq 2(8k + 7)M_n \exp(-2D_{\mathcal{A}} n^{s/(2s+1)}) \leq \exp(-D_{\mathcal{A}} n^{s/(2s+1)}), \end{aligned}$$

for n large enough, where $D_{\mathcal{A}} \triangleq (D_1 \vee D_2 \vee D_3 \vee D_5)/2$, where we used (4.23), (4.24) and (4.28). \square

5. PROOF OF THEOREM 2

The proof of the lower bound is heavily based on arguments found in Korostelev (1993), Donoho (1994), Korostelev and Nussbaum (1999) and Bertin (2004c). It is mainly a modification of the former proof in Bertin (2004c). It consists in a classical reduction to the Bayesian risk over an hardest cubical subfamily of functions, see for instance Donoho (1994). The main difference with the former proofs is that the subfamily of functions depends on the design via the bandwidth $h_{n,\mu}(x)$, which is adapted to the local amount of data.

5.1. Preparatory results. We begin with some definitions. We recall that φ_s is defined by (1.6) and that it has a compact support $[-T_s, T_s]$. Let $h_n^I \triangleq \max_{x \in I_n} h_{n,\mu}(x)$ 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

$$x_j = a_n + j \Xi_n, \quad j \in \mathcal{J}_n \triangleq \{1, \dots, M_n\}. \quad (5.1)$$

In order to unload the notations, we denote again $\mu_j = \mu(x_j)$, $h_j = h_{n,\mu}(x_j)$.

Lemma 8. *Let define the event*

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

and $\mathbb{H}_n \triangleq \bigcap_{j \in \mathcal{J}_n} \mathbb{H}_{n,j}$. We have

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\mu^n \{\mathbb{H}_n\} = 1.$$

Proof. We use Bernstein inequality to the sum of variables $\varphi_s^2((X_i - x_j)/(c_s h_j))$, for $1 \leq i \leq n$, where we use the fact that $\|\varphi_s\|_2 = 1$ (see section A) and we derive a deviation inequality for the events $H_{n,j}^c$. Then, bounding from above the probability of $\cup_{j \in \mathcal{J}_n} H_{n,j}^c$ by the probabilities sum, the result follows easily. \square

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 the functions

$$f(x; \theta) = \sum_{j \in \mathcal{J}_n} \theta_j f_j(x), \quad f_j(x) = L c_s^s h_j^s \varphi_s\left(\frac{x - x_j}{c_s h_j}\right).$$

Clearly, $f_j \in \Sigma(s, L)$. Let us show that $f(\cdot; \theta) \in \Sigma(s, L)$. We note that

$$\text{Supp}\left(\varphi_s\left(\frac{\cdot - x_j}{c_s h_j}\right)\right) = [x_j - c_s T_s h_j, x_j + c_s T_s h_j] \triangleq I_j.$$

If $x, y \in I_j$ then $f(x; \theta) = \theta_j f_j(x)$, $f(y; \theta) = \theta_j f_j(y)$ and the result is obvious. To complete the proof, it suffices to consider the case $x \in I_j$ and $y \in I_{j+1}$. In this case, we have

$$\begin{aligned} & |f^{(k)}(x; \theta) - f^{(k)}(y; \theta)| \\ &= |\theta_j f_j^{(k)}(x) - \theta_{j+1} f_{j+1}^{(k)}(y)| \\ &\leq |f_j^{(k)}(x) - f_j^{(k)}(x_j + c_s T_s h_j)| + |f_{j+1}^{(k)}(x_{j+1} - c_s T_s h_{j+1}) - f_{j+1}^{(k)}(y)| \\ &\leq L(|x - x_j - c_s T_s h_j|^{s-k} + |x_{j+1} - c_s T_s h_{j+1} - y|^{s-k}) \\ &\leq L((2c_s T_s h_j)^{s-k} + (2c_s T_s h_{j+1})^{s-k}) \leq 2L(2c_s T_s h_n^I)^{s-k}. \end{aligned}$$

Moreover, since $x \in I_j$ and $y \in I_{j+1}$ we have

$$|x - y| \geq x_{j+1} - x_j - c_s T_s (h_j + h_{j+1}) \geq \Xi_n - 2c_s T_s h_n^I = 2^{1/(s-k)} (2c_s T_s h_n^I),$$

and finally

$$|f^{(k)}(x; \theta) - f^{(k)}(y; \theta)| \leq L|x - y|^{s-k}, \quad (5.2)$$

thus $f(\cdot; \theta) \in \Sigma(s, L)$. For any $j \in \mathcal{J}_n$, we define the statistics

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

Lemma 9. *Conditionally on \mathfrak{X}_n , the y_j are Gaussian and independent. Moreover, if $v_j^2 = \mathbb{E}_{f, \mu}^n \{y_j^2 | \mathfrak{X}_n\}$, we have on $H_{n,j}$*

$$\mathbb{E}_{f, \mu}^n \{y_j | \mathfrak{X}_n\} = \theta_j, \quad \frac{2s+1}{2(1+\varepsilon)\log n} \leq v_j^2 \leq \frac{2s+1}{2(1-\varepsilon)\log n}. \quad (5.3)$$

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 H_n in the form

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

where g_v is the density of $\mathcal{N}(0, v^2)$, and λ^n is the Lebesgue measure over \mathbb{R}^n .

Proof. By construction the f_j have disjoint supports, thus it is easy to see that conditionally on \mathfrak{X}_n the y_j are Gaussian independent with conditional mean θ_j . Using the definition of H_n and since

$$\mathbb{E}_{f,\mu}^n \{y_j^2 | \mathfrak{X}_n\} = \frac{\sigma^2}{\sum_{i=1}^n f_j^2(X_i)},$$

it is an easy computation to see that on H_n , we have (5.3). The last part of the lemma follows from the following computation:

$$\begin{aligned} & \prod_{i=1}^n g_\sigma(Y_i) \prod_{j \in \mathcal{J}_n} \frac{g_{v_j}(y_j - \theta_j)}{g_{v_j}(y_j)} \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \prod_{i=1}^n \exp(-Y_i^2 / (2\sigma^2)) \prod_{j \in \mathcal{J}_n} \exp((2\theta_j y_j - \theta_j) / (2v_j^2)) \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \prod_{i=1}^n \left[\exp\left(\frac{-Y_i^2 + \sum_{j \in \mathcal{J}_n} (2Y_j \theta_j f_j(X_i) - \theta_j^2 f_j(X_i)^2)}{2\sigma^2}\right) \right] \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \prod_{i=1}^n \exp\left(-\frac{(Y_i - f(X_i; \theta))^2}{2\sigma^2}\right) = \frac{d\mathbb{P}_{f,\mu}^n}{d\lambda^n} |_{\mathfrak{X}_n}(Y_1, \dots, Y_n). \quad \square \end{aligned}$$

5.2. Proof of theorem 2. We denote in the following $\Sigma = \Sigma(s, L)$ and $\mathcal{E}_{n,f,T}^I = \sup_{x \in I} r_{n,\mu}(x)^{-1} |T(x) - f(x)|$. Since w is nondecreasing and $f(\cdot; \theta) \in \Sigma$ for any $\theta \in \Theta$, we have for any distribution \mathcal{B} on Θ by a minoration of the minimax risk by the Bayesian risk,

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

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

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

where $\inf_{\hat{\theta}}$ is taken among any measurable vector (with respect to the observations (1.1)) in \mathbb{R}^{M_n} . Then, theorem 2 follows from lemma 8 if we prove that on H_n ,

$$\inf_{\hat{\theta}} \int_{\Theta} \mathbb{P}_{\theta}^n \left\{ \max_{j \in \mathcal{J}_n} |\hat{\theta}_j - \theta_j| \geq 1 - \varepsilon | \mathfrak{X}_n \right\} \mathcal{B}(d\theta) \geq 1 - o(1),$$

or equivalently, that on H_n

$$\sup_{\hat{\theta}} \int_{\Theta} \mathbb{P}_{\hat{\theta}}^n \{ \max_{j \in \mathcal{J}_n} |\hat{\theta}_j - \theta_j| < 1 - \varepsilon | \mathfrak{X}_n \} \mathcal{B}(d\theta) = o(1). \quad (5.4)$$

To prove (5.4), we choose

$$\Theta = \Theta_{\varepsilon}^{M_n}, \quad \Theta_{\varepsilon} = \{-(1 - \varepsilon), 1 - \varepsilon\}, \quad \mathcal{B} = \bigotimes_{j \in \mathcal{J}_n} b_{\varepsilon}, \quad b_{\varepsilon} = \frac{1}{2}(\delta_{-(1-\varepsilon)} + \delta_{1-\varepsilon}),$$

where δ stands for the Dirac mass. Note that using lemma 9, the left hand side of (5.4) 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 an easy argument shows that

$$\hat{\theta}_j = (1 - \varepsilon) \mathbf{1}_{y_j \geq 0} - (1 - \varepsilon) \mathbf{1}_{y_j < 0}$$

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 distribution density $g_{v_j}(\cdot - \theta_j)$, the left hand side of (5.4) is smaller than

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

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

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

where we used lemma 9 and the fact that for $x > 0$, $\Phi(-x) = \frac{(1+o(1)) \exp(-x^2/2)}{x\sqrt{2\pi}}$. It follows that the left hand side of (5.4) is smaller than

$$\begin{aligned} \left(1 - \frac{D_1}{\sqrt{\log n}} n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)} \right)^{M_n} \\ \leq \exp\left(|I_n| \Xi_n^{-1} \log\left(1 - D_1 n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)} (\log n)^{-1/2}\right)\right), \end{aligned}$$

and if D_2 is a positive constant,

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

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

APPENDIX A. WELL KNOWN FACTS ON OPTIMAL RECOVERY

A.1. Explicit values. To our knowledge, the function φ_s is only known for $s \in (0, 1] \cup \{2\}$. We recall that the optimal recovery kernel is defined by

$$K_s = \frac{\varphi_s}{\int_{\mathbb{R}} \varphi_s},$$

where φ_s is given by (1.6). The kernel K_s for $s \in (0, 1]$ was found by Korostelev (1993) and Fuller (1961) for $s = 2$. See also Leonov (1997, 1999), Lepski and Tsybakov (2000) and Bertin (2004b). When $s \in (0, 1]$,

$$K_s(t) = \frac{s+1}{2s} \varphi_s^{-1/s}(0) (1 - \varphi_s^{-1}(0)|t|^s)_+,$$

where $x_+ = \max(0, x)$, and

$$\varphi_s(0) = \left(\frac{(2s+1)(s+1)}{4s^2} \right)^{s/(2s+1)}.$$

When $s = 2$, we have

$$\varphi_s(t) = \theta^{-2/5} g_2(\theta^{-2/5} t),$$

where for $t \geq 0$

$$\begin{aligned} g_2(t) &= \sum_{j \geq 0} \left((-1)^j q^j + \frac{1}{2} (-1)^{j+1} (t - t_{2j})^2 \right) \mathbf{1}_{t \in [t_{2j-1}, t_{2j+1}]}, \\ q &= \frac{1}{16} \left(3 + \sqrt{33} - \sqrt{26 + 6\sqrt{33}} \right)^2, \\ \theta &= \frac{2(23q^2 - 14q + 23)\sqrt{1+q}}{30(1 - q^{5/2})}, \end{aligned}$$

and $t_{-1} = t_0 = 0$, $t_1 = \sqrt{1+q}$ and for any $j \in \mathbb{N} - \{0\}$, $t_{2j} = 2\sqrt{1+q} \sum_{i=0}^{j-1} q^{i/2}$, $t_{2j+1} = t_{2j} + q^{j/2} \sqrt{1+q}$. Note that φ_2 is piecewise quadratic and infinitely oscillating around 0 at the boundaries of its support. For these values of s ,

$$P = P_s = \begin{cases} \left(\frac{s+1}{2s^2} \right)^{s/(2s+1)} & \text{when } s \in (0, 1], \\ \left(\frac{2}{5} \right)^{2/5} \theta^{-2/5} & \text{when } s = 2. \end{cases}$$

In figure 3 we give an illustration of the kernel K_s for $s = 1/2$, $s = 1$ and $s = 2$.

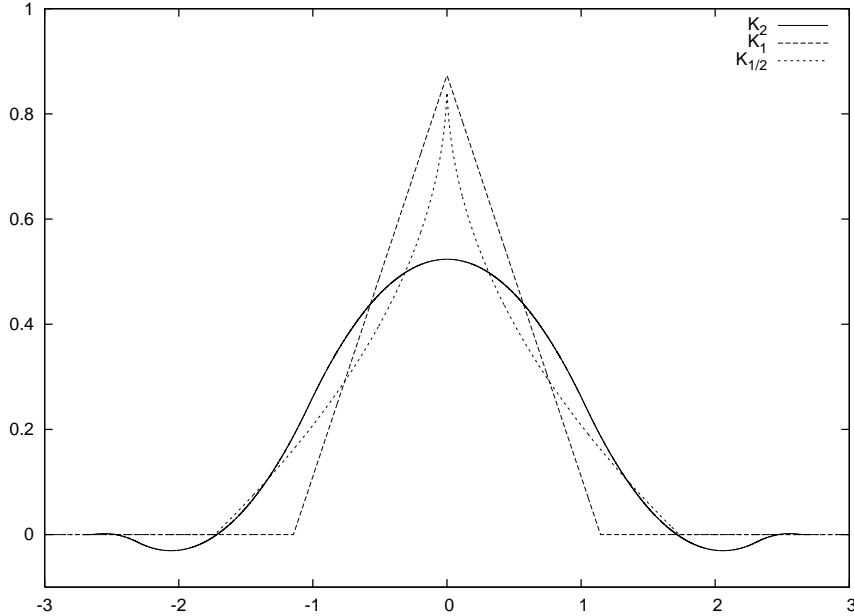


FIGURE 3. Optimal recovery kernels K_s for $s = 1/2$, $s = 1$ and $s = 2$.

A.2. Optimal recovery. The next results are well known and can be found in Donoho (1994), Leonov (1997, 1999), Lepski and Tsybakov (2000) and Bertin (2004b). The problem consists in recovering f from

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}, \quad (\text{A.1})$$

where $\varepsilon > 0$, z is an unknown deterministic function such that $\|z\|_2 \leq 1$ and $f \in C(s, L; \mathbb{R}) \triangleq \Sigma(s, L; \mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. This problem is well known, and the link between this problem and the statistical estimation in sup norm in the white noise model

$$dY_t^\varepsilon = f(t)dt + \varepsilon dW_t, \quad t \in \mathbb{R},$$

was made by Donoho (1994), see also Leonov (1999). The minimax error for the problem of optimal recovery of f at 0 in the model (A.1) is defined by

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

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

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

Note that φ_s satisfies $\varphi_s(0) = E_s(1, 1)$. 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

$$\mathcal{B}(s, L) \triangleq \sup_{f \in C(s, L; \mathbb{R})} \left| \int K_s(t)(f(t) - f(0)) \right|, \quad (\text{A.2})$$

we have the decomposition

$$E_s(1, 1) = \mathcal{B}(s, 1) + \|K\|_2,$$

and in particular if P is given by (1.5) and c_s by (3.1) we have

$$P = Lc_s^s(\mathcal{B}(s, 1) + \|K\|_2). \quad (\text{A.3})$$

REFERENCES

- ANTONIADIS, A., GREGOIRE, G. and VIAL, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35** 225–232.
- 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 hölder classes. *Bernoulli*, **10** 873–888.
- BERTIN, K. (2004b). *Estimation asymptotiquement exacte en norme sup de fonctions multidimensionnelles*. Ph.D. thesis, Université Paris 6.
- BERTIN, K. (2004c). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, **123** 225–242.
- BROWN, L. and CAI, T. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, **26** 1783–1799.
- 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.
- CAI, T. T. and LOW, M. G. (2004a). An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics*, **32** 1805–1840.
- CAI, T. T. and LOW, M. G. (2004b). Adaptive confidence balls. *The Annals of Statistics*. To appear.
- 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.
- FAN, J. and GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B. Methodological*, **57** 371–394.
- FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability, Chapman & Hall, London.
- FULLER, A. T. (1961). Relay control systems optimized for various performance criteria,. *Automatic and remote control*, **1**.
- GAÏFFAS, S. (2004). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*. To appear, available at <http://hal.ccsd.cnrs.fr/ccsd-00003086/en/>.

- HOFFMANN, M. and LEPSKI, O. V. (2002). Random rates in anisotropic regression. *The Annals of Statistics*, **30** 325–396.
- 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.
- KOROSTELEV, V. and TSYBAKOV, A. (1993). *Minimax theory of image reconstruction*. Springer-Verlag, New York.
- 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. 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.
- LOW, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics*, **25** 2547–2554.
- MAXIM, V. (2003). *Restauration de signaux bruités sur des plans d'expérience aléatoires*. Ph.D. thesis, Université Joseph Fourier, Grenoble 1.
- MICCHELLI, C. A. and RIVLIN, T. J. (1977). A survey of optimal recovery. *Optimal estimation in approximation theory* 1 – 54.
- PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics*, **28** 298–335.
- PINSKER, M. S. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Problems of Information Transmission*, **16** 52–68.
- SPOKOINY, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.
- STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10** 1040–1053.
- TSYBAKOV, A. (2003). *Introduction à l'estimation non-paramétrique*. Springer.
- WONG, M.-Y. and ZHENG, Z. (2002). Wavelet threshold estimation of a regression function with random design. **80** 256–284.

LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES, UNIVERSITÉ PARIS 7, 175 RUE DU CHEVALERET, 75013 PARIS,
E-mail address: gaiffas@math.jussieu.fr