UNIFORM ESTIMATION OF A SIGNAL BASED ON INHOMOGENEOUS DATA

STÉPHANE GAÏFFAS

Modal'X, Université Paris X – Nanterre
email: stephane.gaiffas@u-paris10.fr

ABSTRACT. We want to reconstruct a signal based on inhomogeneous data (the amount of data can vary strongly), using the model of regression with a random design. Our aim is to understand the consequences of inhomogeneity on the accuracy of estimation within the minimax framework. Using the uniform metric weighted by a spatially-dependent rate as a benchmark for an estimator accuracy, we are able to capture the deformation of the usual minimax rate in situations with local lacks of data (modelled by a design density with vanishing points). In particular, we construct an estimator both design and smoothness adaptive, and a new criterion is developed to prove the optimality of these deformed rates.

1. INTRODUCTION

Motivations. A problem particularly prominent in statistical literature is the adaptive reconstruction of a function based on irregularly sampled noisy data. In several practical situations, the statistician cannot obtain "nice" regularly sampled observations, because of various constraints linked with the source of the data, or the way the data is obtained. For instance, in signal or image processing, the irregular sampling can be due to the process of motion or disparity compensation (used in advanced video processing), while in topography, measurement constraints are linked with the properties of the ground. See Feichtinger and Gröchenig (1994) for a survey on irregular sampling, Almansa et al. (2003), Vàzquez et al. (2000) for applications concerning respectively satellite image and stereo imaging, and Jansen et al. (2004) for examples of geographical constraints.

Such constraints can result in potentially strong local lacks of data. Consequently, the accuracy of a procedure based on such data can become locally very poor. The aim of the paper is to study from a theoretical point of view the consequences of data *inhomogeneity*

Date: October 3, 2006.
2000 *Mathematics Subject Classification.* 62G05, 62G08.
Key words and phrases. nonparametric regression, adaptive estimation, minimax theory, random design.

on the reconstruction of a univariate signal. Natural questions arise: how does the inhomogeneity impact on the accuracy of estimation? What does the optimal convergence rate become in such situations? Can the rate vary strongly from place to place, and how?

The model. The widest spread way to model such observations is as follows. We model the available data $[(X_i, Y_i); 1 \leqslant i \leqslant n]$ by

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

where ξ_i are i.i.d. Gaussian standard and independent of the X_i's and $\sigma > 0$ is the noise level. The design variables X_i are i.i.d. with unknown density μ on $[0, 1]$. The more the density μ is "far" from the uniform law, the more the data drawn from (1.1) is inhomogeneous. A simple way to include situations with local lacks of data within the model (1.1) is to allow the density μ to be arbitrarily small at some points, and to vanish. This kind of behaviour is not commonly used in literature, since most papers assume μ to be uniformly bounded away from zero. We give references handling this kind of design below.

In practice, we don't know μ, since it requires to know in a precise way the constraints making the observation irregularly sampled, neither do we know the smoothness of f. Therefore, a convenient procedure shall adapt both to the design and to the smoothness of f. Such a procedure (that is proved to be optimal) is constructed here.

Methodology. We want to reconstruct f globally, with sup norm loss. The reason for choosing this metric is that it is exacting: roughly, it forces an estimator to behave well at every point simultaneously. This property is convenient here, since it allows to capture in a very simple way the consequences of inhomogeneity directly on the convergence rate.

In what follows, $a_n \lesssim b_n$ means $a_n \leqslant C b_n$ for any n, where $C > 0$. We say that a sequence of curves $v_n(\cdot) \geqslant 0$ is an upper bound over some class F if there is an estimator $\widehat{f}_n$ such that

$$\sup_{f \in F} \mathbf{E}_{f\mu} \big[w \big(\sup_{x \in [0,1]} v_n(x)^{-1} | \widehat{f}_n(x) - f(x) | \big) \big] \lesssim 1 \tag{1.2}$$

as $n \to +\infty$, where $\mathbf{E}_{f\mu}$ denotes the expectation with respect to the joint law $\mathbf{P}_{f\mu}$ of the $[(X_i, Y_i); 1 \leqslant i \leqslant n]$, and where $w(\cdot)$ is a loss function, that is a non-negative and non-decreasing function such that $w(0) = 0$ and $w(x) \leqslant A(1 + |x|^b)$ for some $A, b > 0$.

Literature. Pointwise estimation at a point where the design vanishes is studied in Hall et al. (1997), with the use of a local linear procedure. This design behaviour is given as an example in Guerre (1999), where a more general setting for the design is considered, with a Lipschitz regression function. In Gaïffas (2005a), pointwise minimax rates over Hölder classes are computed for several design behaviours, and an adaptive estimator for pointwise risk is constructed in Gaïffas (2005b). In these papers, it appears that, depending on the design behaviour at the estimation point, the range of minimax rates is very wide: from very slow (logarithmic) rates to very fast quasi-parametric rates.

Many adaptive techniques have been developed in literature for handling irregularly sampled data. Among wavelet methods, see Hall et al. (1997) for interpolation; Antoniadis et al. (1997), Antoniadis and Pham (1998), Brown and Cai (1998), Hall et al. (1998) and Wong and Zheng (2002) for tranformation and binning; Antoniadis and Fan (2001) for a penalization approach; Delouille et al. (2001) and Delouille et al. (2004) for the construction of design-adapted wavelet via lifting; Pensky and Wiens (2001) for projection-based techniques; Kerkyacharian and Picard (2004) for warped wavelets. For model selection, see Baraud (2002). See also the PhD manuscripts from Maxim (2003) and Delouille (2002).

2. Results

To measure the smoothness of f, we consider the standard Hölder class $H(s, L)$ where $s, L > 0$, defined as the set of all the functions $f : [0, 1] \to \mathbb{R}$ such that

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

where $\lfloor s\rfloor$ is the largest integer smaller than s. Minimax theory over such classes is standard: we know from Stone (1982) that within the model (1.1), the minimax rate is equal to $(\log n/n)^{s/(2s+1)}$ over such classes, when μ is continuous and uniformly bounded away from zero. If $Q > 0$, we define $H^Q(s, L) := H(s, L) \cap \{f \mid \|f\|_\infty \leqslant Q\}$ (the constant Q needs not to be known).

We use the notation $\mu(I) := \int_I \mu(t)dt$. If $F = H(s, L)$ is fixed, we consider the sequence of positive curves $h_n(\cdot) = h_n(\cdot; F, \mu)$ satisfying

$$Lh_n(x)^s = \sigma\Big(\frac{\log n}{n\mu([x - h, x + h])}\Big)^{1/2} \tag{2.1}$$

for any $x \in [0, 1]$, and we define

$$r_n(x; F, \mu) := Lh_n(x; F, \mu)^s.$$

Since $h \mapsto h^{2s}\mu([x - h, x + h])$ is increasing for any x, these curves are well-defined (for n large enough) and unique. In Theorem 1 below, we show that $r_n(\cdot)$ is an upper bound over Hölder classes, and the optimality of this rate is proved in Theorem 2.

Example. When $s = 1$, $\sigma = L = 1$ and $\mu(x) = 4|x - 1/2|\mathbf{1}_{[0,1]}(x)$, solving (2.1) leads to

$$r_n(x) = (\log n/n)^{\alpha_n(x)},$$

where the exponent $\alpha_n(\cdot)$ is given by

$$\alpha_n(x) = \begin{cases} \frac{1}{3}\big(1 - \frac{\log(1-2x)}{\log(\log n/n)}\big) & \text{when } x \in \big[0, \frac{1}{2} - (\frac{\log n}{2n})^{1/4}\big], \\ \frac{\log\big(((x-1/2)^4 + 4\log n/n)^{1/2} - (x-1/2)^2\big) - \log 2}{2\log(\log n/n)} & \\ & \text{when } x \in \big[\frac{1}{2} - (\frac{\log n}{2n})^{1/4}, \frac{1}{2} + (\frac{\log n}{2n})^{1/4}\big], \\ \frac{1}{3}\big(1 - \frac{\log(2x-1)}{\log(\log n/n)}\big) & \text{when } x \in \big[\frac{1}{2} + (\frac{\log n}{2n})^{1/4}, 1\big]. \end{cases}$$

Within this example, $r_n(\cdot)$ switches from one "regime" to another. Indeed, in this example there is a lack of data in the middle of the unit interval. The consequence is that $r_n(1/2) = (\log n/n)^{1/4}$ is slower than the rate at the boundaries $r_n(0) = r_n(1) = (\log n/n)^{1/3}$, which comes from the standard minimax rate $(\log n/n)^{s/(2s+1)}$ with $s = 1$. We show the shape of this deformed rate for several sample sizes in Figure 1.

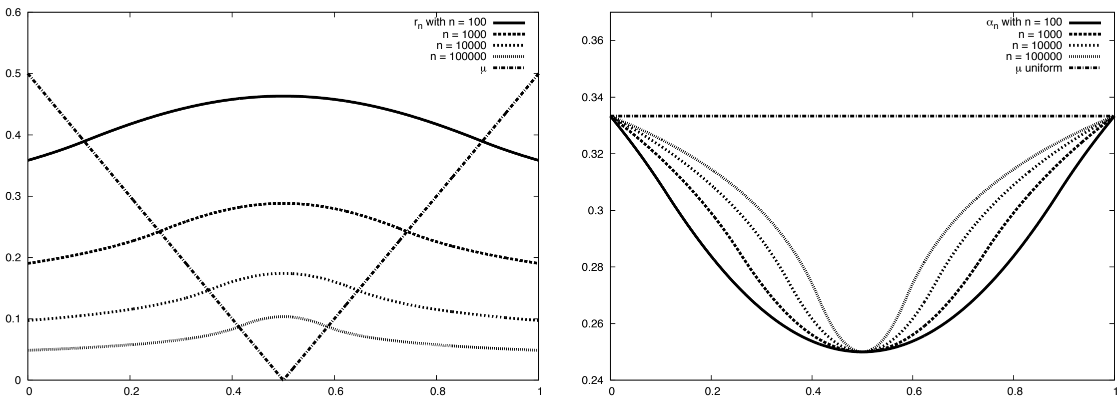

FIGURE 1. $r_n(\cdot)$ and $\alpha_n(\cdot)$ for several sample sizes

Upper bound. In this section, we show that the spatially-dependent rate $r_n(\cdot)$ defined by (2.1) is an upper bound in the sense of (1.2) over Hölder classes. The estimator used in this upper bound is both smoothness and design adaptive (it does not depend on the design density within its construction). This estimator is constructed in Section 3 below. Let R be a fixed natural integer.

Assumption D. We assume that μ is continuous, and that whether $\mu(x) > 0$ for any x, or $\mu(x) = 0$ for a finite number of x. Moreover, for any x such that $\mu(x) = 0$ we assume that $\mu(y) = |y - x|^{\beta(x)}$ for any y in a neighbourhood of x (where $\beta(x) \geqslant 0$).

Theorem 1. *Let $s \in (0, R+1]$ and assumption D holds. The estimator $\widehat{f}_n$ defined by (3.2) satisfies*

$$\sup_{f \in F} \mathbf{E}_{f\mu}\big[w(\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|)\big] \lesssim 1 \tag{2.2}$$

as $n \to +\infty$ for any $F = H^Q(s, L)$, where $r_n(\cdot) = r_n(\cdot; F, \mu)$ is given by (2.1).

This theorem assesses the adaptive estimator constructed in Section 3 below. The estimator $\widehat{f}_n$ is based on a precise estimation of the scaling coefficients (within a multiresolution analysis) of f. This method relies on a Lepski-type method (see for instance Lepski et al. (1997)) that we adapt for random designs.

Remark. Within Theorem 1, there are mainly two situations.

- $\mu(x) > 0$ for any x: we have $r_n(x) \asymp (\log n/n)^{s/(2s+1)}$ for any x, where $a_n \asymp b_n$ means $a_n \lesssim b_n$ and $b_n \lesssim a_n$. Hence, we find back the standard minimax rate in this situation. Note that this result is new since adaptive estimators over Hölder balls in regression with random design were not previously constructed.
- $\mu(x) = 0$ for one or several x: the rate $r_n(\cdot)$ can vary strongly from place to place, depending on the behaviour of μ. Indeed, the rate changes *in order* from one point to another, see the example above.

Remark. Implicitly, we assumed in Theorem 1 that $s \in (0, R + 1]$, where R is a tuning parameter of the procedure. Indeed, in the minimax framework considered here, the fact of knowing an upper bound for s is usual in the study of adaptive methods, and somehow, unavoidable. For instance, when considering adaptive wavelet methods, the "maximum smoothness" corresponds to the number of moments of the mother wavelet.

Optimality of $r_n(\cdot)$. We have seen that the rate $r_n(\cdot)$ defined by (2.1) is an upper bound over Hölder classes, see Theorem 1. In Theorem 2 below, we prove that this rate is indeed optimal. In order to show that $r_n(\cdot)$ is optimal in the minimax sense over some class F, the classical criterion consists in showing that

$$\inf_{\widehat{f}_n} \sup_{f\in F} \mathbf{E}_{f\mu}\big[w(\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|)\big] \gtrsim 1, \tag{2.3}$$

where the infimum is taken among all estimators based on the observations (1.1). However, this criterion does not exclude the existence of another normalisation $\rho_n(\cdot)$ that can improve $r_n(\cdot)$ in some regions of $[0, 1]$. Indeed, (2.3) roughly consists in a minoration of the uniform risk over the whole unit interval and then, only over some particular points. Therefore, we need a new criterion that strengthens the usual minimax one to prove the optimality of $r_n(\cdot)$. The idea is simple: we localize (2.3) by replacing the supremum over $[0,1]$ by a supremum over any (small) inverval $I_n \subset [0,1]$, that is

$$\inf_{\widehat{f}_n} \sup_{f \in F} \mathbf{E}_{f\mu}\big[w\big(\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)|\big)\big] \gtrsim 1, \quad \forall I_n. \tag{2.4}$$

It is noteworthy that in (2.4), the length of the intervals cannot be arbitrarily small. Actually, if an interval I_n has a length smaller than a given limit, (2.4) does not hold anymore. Indeed, beyond this limit, we can improve $r_n(\cdot)$ for the risk localized over I_n: we can construct an estimator $\widehat{f}_n$ such that

$$\sup_{f \in F} \mathbf{E}_{f\mu}\big[w\big(\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)|\big)\big] = o(1), \tag{2.5}$$

see Proposition 1 below. The phenomenon described in this section, which concerns the uniform risk, is linked with the results from Cai and Low (2005) for shrunk $\mathbb{L}^2$ risks. In what follows, $|I|$ stands for the length of an interval I.

Theorem 2. *Suppose that*

$$\mu(I) \gtrsim |I|^{\beta+1} \tag{2.6}$$

uniformly for any interval $I \subset [0,1]$*, where* $\beta \geqslant 0$ *and let* $F = H(s,L)$*. Then, for any interval* $I_n \subset [0,1]$ *such that*

$$|I_n| \sim n^{-\alpha} \tag{2.7}$$

with $\alpha \in (0, (1+2s+\beta)^{-1})$*, we have*

$$\inf_{\widehat{f}_n} \sup_{f \in F} \mathbf{E}_{f\mu}\big[w\big(\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)|\big)\big] \gtrsim 1 \tag{2.8}$$

as $n \to +\infty$*, where* $r_n(\cdot) = r_n(\cdot\,; F, \mu)$ *is given by* (2.1).

Corollary 1. *If* $v_n(\cdot)$ *is an upper bound over* $F = H(s,L)$ *in the sense of* (1.2)*, we have*

$$\sup_{x \in I_n} v_n(x)/r_n(x) \gtrsim 1$$

for any interval I_n *as in Theorem 2. Hence,* $r_n(\cdot)$ *cannot be improved uniformly over an interval with length* $n^{\varepsilon - 1/(1+2s+\beta)}$*, for any arbitrarily small* $\varepsilon > 0$.

Proposition 1. *Let* $F = H(s,L)$ *and* ℓ_n *be a positive sequence satisfying*

$$\log \ell_n = o(\log n).$$

a) Let μ *be such that* $0 < \mu(x) < +\infty$ *for any* $x \in [0,1]$*. Note that in this case,* $r_n(x) \asymp (\log n/n)^{s/(2s+1)}$ *for any* $x \in [0,1]$ *and that* (2.6) *holds with* $\beta = 0$*. If* I_n *is an interval* *satisfying*

$$|I_n| \sim (\ell_n/n)^{1/(1+2s)},$$

we can contruct an estimator $\widehat{f}_n$ *such that*

$$\sup_{f\in F} \mathbf{E}_{f\mu}\Big[w\Big(\Big(\frac{n}{\log n}\Big)^{s/(2s+1)} \sup_{x\in I_n} |\widehat{f}_n(x) - f(x)|\Big)\Big] = o(1).$$

b) Let $\mu(x_0) = 0$ *for some* $x_0 \in [0,1]$ *and* $\mu([x_0 - h, x_0 + h]) = h^{\beta+1}$ *where* $\beta \geqslant 0$ *for any* h *in a neighbourhood of* 0*. If*

$$I_n = [x_0 - (\ell_n/n)^{1/(1+2s+\beta)}, x_0 + (\ell_n/n)^{1/(1+2s+\beta)}],$$

we can contruct an estimator $\widehat{f}_n$ *such that*

$$\sup_{f\in F} \mathbf{E}_{f\mu}\big[w(\sup_{x\in I_n} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|)\big] = o(1).$$

This proposition entails that $r_n(\cdot)$ can be improved for localized risks (2.5) over intervals I_n with size $(\ell_n/n)^{1/(1+2s+\beta)}$ where ℓ_n can be a slow term such has $(\log n)^\gamma$ for any $\gamma \geqslant 0$. A consequence is that the lower bound in Theorem 2 cannot be improved, since (2.8) does not hold anymore when I_n has a length smaller than (2.7). This phenomenon is linked both to the choice of the uniform metric for measuring the error of estimation, and to the nature of the noise within the model (1.1). It is also a consequence of the minimax paradigm: it is well-known that the minimax risk actually concentrates on some critical functions of the considered class (that we rescale and place within I_n here, hence the critical length for I_n), which is a property allowing to prove lower bounds such as the one in Theorem 2.

3. Construction of an adaptive estimator

The adaptive method proposed here differs from the techniques mentioned in Introduction. Indeed, it is not appropriate here to apply a wavelet decomposition of the scaling coefficients at the finest scale since it is a $\mathbb{L}^2$-transform, while the criterion (1.2) considered here uses the uniform metric. This is the reason why we focus the analysis on a precise estimation of the scaling coefficients. The technique consists in a local polynomial approximation of f within adaptively selected bandwidths for each scaling coefficient.

Let $(V_j)_{j\geqslant 0}$ be a multiresolution analysis of $\mathbf{L}^2([0,1])$ with scaling function ϕ compactly supported and R-regular (the parameter R comes from Theorem 1), which ensures that

$$\|f - P_j f\|_\infty \lesssim 2^{-js} \tag{3.1}$$

for any $f \in H(s, L)$ with $s \in (0, R+1]$, where P_j denotes the projection onto V_j. We use P_j as an interpolation transform. Interpolation transforms in the unit interval are constructed in Donoho (1992) and Cohen et al. (1993). We have $P_j f = \sum_{k=0}^{2^j-1} \alpha_{jk}\phi_{jk}$, where $\phi_{jk}(\cdot) = 2^{j/2}\phi(2^j \cdot -k)$ and $\alpha_{jk} = \int f\phi_{jk}$. We consider the largest integer J such that $N := 2^J \leqslant n$, and we estimate the scaling coefficients at the high resolution J. For appropriate estimators $\widehat{\alpha}_{Jk}$ of α_{Jk}, we simply consider

$$\widehat{f}_n := \sum_{k=0}^{2^J-1} \widehat{\alpha}_{Jk}\phi_{Jk}. \tag{3.2}$$

Let us denote by Pol_R the set of all real polynomials with degree at most R. If $\bar{f}_k \in \mathrm{Pol}_R$ is close to f over the support of ϕ_{Jk}, then

$$\alpha_{Jk} = \int f\phi_{Jk} \approx \int \bar{f}_k\phi_{Jk}.$$

When the scaling function ϕ has R moments, that is

$$\int \phi(t)t^p dt = \mathbf{1}_{p=0}, \quad p \in \{0, \ldots, R\}, \tag{3.3}$$

and when f is s-Hölder for $s \in (0, R+1]$, accurate estimators of $\widehat{\alpha}_{Jk}$ are given by

$$\widehat{\alpha}_{Jk} := 2^{-J/2}\bar{f}_k(k2^{-J}). \tag{3.4}$$

If ϕ does not satisfies (3.3), $\int \bar{f}\phi_{Jk}$ can be computed exactly using a quadrature formula, in the same way as in Delyon and Juditsky (1995). Indeed, there is a matrix Q_J (characterized by ϕ) with entries (q_{Jkm}) for $(k, m) \in \{0, \ldots, 2^J - 1\}^2$ such that

$$\int P\phi_{Jk} = 2^{-J/2} \sum_{m \in \Gamma_{Jk}} q_{Jkm} P(m/2^J) \tag{3.5}$$

for any $P \in \mathrm{Pol}_R$. Within this equation, the entries of the quadrature matrix Q_J satisfy

$$q_{Jkm} \neq 0 \to |k - m| \leqslant L_\phi \text{ and } m \in \Gamma_{Jk}, \tag{3.6}$$

where $L_\phi > 0$ is the support length of ϕ. Therefore, the matrix Q_J is band-limited. For instance, if we consider the Coiflets basis, which satisfies the moment condition (3.3), we have $q_{Jkm} = \mathbf{1}_{k=m}$, and we can use directly (3.4). If the $(\phi(\cdot - k))_k$ are orthogonal, then $q_{Jkm} = \phi(m - k)$, see Delyon and Juditsky (1995).

For the sake of simplicity, we assume in what follows that ϕ satisfies the moment condition (3.3), thus α_{Jk} is estimated by (3.4). Each polynomial $\bar{f}_k$ in (3.4) is defined via a least square minimization which is localized within a data-driven bandwidth $\widehat{\Delta}_k$, hence

$$\bar{f}_k = \bar{f}_k^{(\widehat{\Delta}_k)}.$$

Below, we describe the computation of these polynomials and then, we define the selection rule for the $\widehat{\Delta}_k$.

Local polynomials. The polynomials used to estimate each scaling coefficients are defined via a slightly modified version of the local polynomial estimator (LPE). This linear method of estimation is standard, see for instance Fan and Gijbels (1995, 1996), among many others. For any interval $\delta \subset [0,1]$, we define the empirical sample measure

$$\bar{\mu}_n(\delta) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_\delta(X_i),$$

where $\mathbf{1}_\delta$ is the indicator of δ, and if $\bar{\mu}_n(\delta) > 0$, we introduce the pseudo-inner product

$$\langle f \, , \, g \rangle_\delta := \frac{1}{\bar{\mu}_n(\delta)} \int_\delta f g \, d\bar{\mu}_n, \tag{3.7}$$

and $\|g\|_\delta := \langle g \, , \, g \rangle_\delta^{1/2}$ the corresponding pseudo-norm. The LPE consists in looking for the polynomial $\bar{f}^{(\delta)}$ of degree R which is the closest to the data in the least square sense, with respect to the localized design-adapted norm $\| \cdot \|_\delta$:

$$\bar{f}^{(\delta)} := \underset{g \in \mathrm{Pol}_R}{\operatorname{argmin}} \|Y - g\|_\delta^2, \tag{3.8}$$

where we recall that Pol_R is the set of all real polynomials with degree at most R. We can rewrite (3.8) in a variational form, in which we look for $\bar{f}^{(\delta)} \in \mathrm{Pol}_R$ such that for any $\varphi \in \mathrm{Pol}_R$,

$$\langle \bar{f}^{(\delta)} \, , \, \varphi \rangle_\delta = \langle Y \, , \, \varphi \rangle_\delta, \tag{3.9}$$

where it suffices to consider only power functions $\varphi_{kp}(\cdot) = (\cdot - k/2^J)^p$, $0 \leqslant p \leqslant R$ when estimating in a neighbourhood of the regular sampling point $k/2^J$. The coefficients vector $\bar{\theta}_k^{(\delta)} \in \mathbb{R}^{R+1}$ of the polynomial $\bar{f}_k^{(\delta)}$ is therefore solution, when it makes sense, of the linear system

$$\mathbf{X}_k^{(\delta)} \theta = \mathbf{Y}_k^{(\delta)},$$

where for $0 \leqslant p, q \leqslant R$:

$$(\mathbf{X}_k^{(\delta)})_{p,q} := \langle \varphi_{kp} \, , \, \varphi_{kq} \rangle_\delta \quad \text{and} \quad (\mathbf{Y}_k^{(\delta)})_p := \langle Y \, , \, \varphi_{kp} \rangle_\delta. \tag{3.10}$$

We modify this system as follows: when the smallest eigenvalue of $\mathbf{X}_k^{(\delta)}$ (which is non-negative) is too small, we add a correcting term allowing to bound it from below. We introduce

$$\bar{\mathbf{X}}_k^{(\delta)} := \mathbf{X}_k^{(\delta)} + (n\bar{\mu}_n(\delta))^{-1/2}\mathbf{Id}_{R+1}\mathbf{1}_{\Omega_k(\delta)^\complement},$$

where $\mathbf{Id}_{R+1}$ is the identity matrix in $\mathbb{R}^{R+1}$ and

$$\Omega_k(\delta) := \big\{\lambda(\mathbf{X}_k^{(\delta)}) > (n\bar{\mu}_n(\delta))^{-1/2}\big\}, \tag{3.11}$$

where $\lambda(M)$ stands for the smallest eigenvalue of a matrix M. The quantity $(n\bar{\mu}_n(\delta))^{-1/2}$ comes from the variance of $\bar{f}_k^{(\delta)}$, and this particular choice preserves the convergence rate of the method. This modification of the classical LPE is convenient in situations with little data.

Definition 1. When $\bar{\mu}_n(\delta) > 0$, we consider the solution $\bar{\theta}_k^{(\delta)}$ of the linear system

$$\bar{\mathbf{X}}_k^{(\delta)}\theta = \mathbf{Y}_k^{(\delta)}, \tag{3.12}$$

and introduce $\bar{f}_k^{(\delta)}(x) := (\bar{\theta}_k^{(\delta)})_0 + (\bar{\theta}_k^{(\delta)})_1(x - k/2^J) + \cdots + (\bar{\theta}_k^{(\delta)})_R(x - k/2^J)^R$. When $\bar{\mu}_n(\delta) = 0$, we take simply $\bar{f}_k^{(\delta)} := 0$.

Adaptive bandwidth selection. The adaptive procedure selecting the intervals $\widehat{\Delta}_k$ is based on a method introduced by Lepski (1990), see also Lepski et al. (1997), and Lepski and Spokoiny (1997). If a family of linear estimators can be "well-sorted" by their respective variances (e.g. kernel estimators in the white noise model, see Lepski and Spokoiny (1997)), the Lepski procedure selects the largest bandwidth such that the corresponding estimator does not differ "significantly" from estimators with a smaller bandwidth. Following this principle, we construct a method which adapts to the unknown smoothness, and additionally to the original Lepski method, to the distribution of the data (the design density is unknown). Bandwidth selection procedures in local polynomial estimation can be found in Fan and Gijbels (1995), Goldenshluger and Nemirovski (1997) or Spokoiny (1998).

The idea of the adaptive procedure is the following: when $\bar{f}^{(\delta)}$ is close to f (that is, when δ is well-chosen), we have in view of (3.9)

$$\langle \bar{f}^{(\delta')} - \bar{f}^{(\delta)}\,,\, \varphi\rangle_{\delta'} = \langle Y - \bar{f}^{(\delta)}\,,\, \varphi\rangle_{\delta'} \approx \langle Y - f\,,\, \varphi\rangle_{\delta'} = \langle \xi\,,\, \varphi\rangle_{\delta'}$$

for any $\delta' \subset \delta$, $\varphi \in \mathrm{Pol}_R$, where the right-hand side is a noise term. Then, in order to "remove" this noise, we select the largest δ such that this noise term remains smaller than an appropriate threshold, for any $\delta' \subset \delta$ and $\varphi = \varphi_{kp}$, $p \in \{0, \dots, R\}$. The bandwidth $\widehat{\Delta}_k$

is selected in a fixed set of intervals G_k called *grid* (which is defined below) as follows:

$$\widehat{\Delta}_k := \operatorname*{argmax}_{\delta \in G_k} \Big\{ \bar{\mu}_n(\delta) \mid \forall \delta' \in G_k, \delta' \subset \delta, \ \forall p \in \{0, \ldots, R\}, \\ |\langle \bar{f}_k^{(\delta')} - \bar{f}_k^{(\delta)} \, , \varphi_{kp} \rangle_{\delta'}| \leqslant \|\varphi_{kp}\|_{\delta'} T_n(\delta, \delta') \Big\}, \tag{3.13}$$

where

$$T_n(\delta, \delta') := \sigma \Big[\Big(\frac{\log n}{n \bar{\mu}_n(\delta)} \Big)^{1/2} + D C_R \Big(\frac{\log(n \bar{\mu}_n(\delta))}{n \bar{\mu}_n(\delta')} \Big)^{1/2} \Big], \tag{3.14}$$

with $C_R := 1 + (R+1)^{1/2}$ and $D > (2(b+1))^{1/2}$, if we want to prove Theorem 1 with a loss function satisfying $w(x) \lesssim (1 + |x|^b)$. The threshold choice (3.14) can be understood in the following way: since the variance of $\bar{f}_k^{(\delta)}$ is of order $(n\bar{\mu}_n(\delta))^{-1/2}$, we see that the two terms in $T_n(\delta, \delta')$ are ratios between a penalizing log term and the variance of the estimators compared by the rule (3.13). The penalization term is linked with the number of comparisons necessary to select the bandwidth. To prove Theorem 1, we use the grid

$$G_k := \bigcup_{1 \leqslant i \leqslant n} \Big\{ \big[k2^{-J} - |X_i - k2^{-J}|, k2^{-J} + |X_i - k2^{J}| \big] \Big\}, \tag{3.15}$$

and we recall that the scaling coefficients are estimated by

$$\widehat{\alpha}_{Jk} := 2^{-J/2} \bar{f}_k^{(\widehat{\Delta}_k)}(k2^{-J}).$$

Remark. In this form, the adaptive estimator has a complexity $O(n^2)$. This can be decreased using a smaller grid. An example of such a grid is the following: first, we sort the (X_i, Y_i) into $(X_{(i)}, Y_{(i)})$ such that $X_{(i)} < X_{(i+1)}$. Then, we consider $i(k)$ such that $k/2^J \in [X_{(i(k))}, X_{(i(k)+1)}]$ (if necessary, we take $X_{(0)} = 0$ and $X_{(n+1)} = 1$) and for some $a > 1$ (to be chosen by the statistician) we introduce

$$G_k := \bigcup_{p=0}^{[\log_a(i(k)+1)]} \ \bigcup_{q=0}^{[\log_a(n-i(k))]} \Big\{ \big[X_{(i(k)+1-[a^p])}, X_{(i(k)+[a^q])} \big] \Big\}. \tag{3.16}$$

With this grid, the selection of the bandwidth is fast, and the complexity of the procedure is $O(n(\log n)^2)$. We can use this grid in practice, but we need extra assumptions on the design if we want to prove Theorem 1 with this grid choice.

4. Proofs

We recall that the weight function $w(\cdot)$ is non-negative, non-decreasing and such that $w(x) \leqslant A(1 + |x|)^b$ for some $A, b > 0$. We denote by μ^n the joint law of $X_1, \ldots, X_n$ and $\mathfrak{X}_n$ the sigma-field generated by $X_1, \ldots, X_n$. $|A|$ denotes both the length of an interval A and the cardinality of a finite set A. $M^\top$ is the transpose of M, and $\xi = (\xi_1, \ldots, \xi_n)^\top$.

Proof of Theorem 1. To prove the upper bound, we use the estimator defined by (3.2) where ϕ is a scaling function satisfying (3.3) (for instance the Coiflets basis), and where the scaling coefficients are estimated by (3.4). Using together (3.1) and the fact that $r_n(x) \gtrsim (\log n/n)^{s/(1+2s)}$ for any x, we have $\sup_{x\in[0,1]} r_n(x)^{-1}\|f - P_J f\|_\infty = o(1)$. Hence,

$$\begin{aligned}\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)| &\lesssim \sup_{x\in[0,1]} r_n(x)^{-1}\Big|\sum_{k=0}^{2^J-1}(\widehat{\alpha}_{Jk} - \alpha_{Jk})\phi_{Jk}(x)\Big| \\ &\lesssim \max_{0\leqslant k\leqslant 2^J-1} \sup_{x\in S_k} r_n(x)^{-1}2^{J/2}|\widehat{\alpha}_{Jk} - \alpha_{Jk}|,\end{aligned}$$

where S_k denotes the support of ϕ_{Jk}. Then, expanding f up to the degree $\lfloor s\rfloor \leqslant R$ and using (3.3), we obtain

$$\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)| \lesssim \max_{0\leqslant k\leqslant 2^J-1} \sup_{x\in S_k} r_n(x)^{-1}|\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - f(x_k)|. \tag{4.1}$$

Since $|S_k| = 2^{-J} \asymp n^{-1}$, we have

$$\sup_{x\in S_k} r_n(x)^{-1} \lesssim r_n(x_k)^{-1}. \tag{4.2}$$

Indeed, since μ is continuous, $r_n(\cdot)$ is continuously differentiable and we have $\sup_{x\in S_k}|r_n(x)^{-1} - r_n(x_k)^{-1}| \leqslant 2^{-J}\|(r_n^{-1})'\|_\infty$, where g' stands for the derivative of g. Moreover, $|(r_n(x)^{-1})'| \lesssim h_n'(x)h_n(x)^{-(s+1)} \lesssim n^{-1}$, since $h_n'(x) \lesssim 1$ and $h_n(x) \gtrsim (\log n/n)^{1/(2s+1)}$, thus (4.2).

In what follows, $\|\cdot\|_\infty$ denotes the supremum norm in $\mathbb{R}^{R+1}$. The following lemma is a version of the bias-variance decomposition of the local polynomial estimator, which is classical: see for instance Fan and Gijbels (1995, 1996), Goldenshluger and Nemirovski (1997), Spokoiny (1998), among others. We define the matrix

$$\mathbf{E}_k^{(\delta)} := \mathbf{\Lambda}_k^{(\delta)}\bar{\mathbf{X}}_k^{(\delta)}\mathbf{\Lambda}_k^{(\delta)},$$

where $\bar{\mathbf{X}}_k$ is given by (3.10) and $\mathbf{\Lambda}_k^{(\delta)} := \mathrm{diag}[\|\varphi_{k0}\|_\delta^{-1}, \ldots, \|\varphi_{kR}\|_\delta^{-1}]$.

Lemma 1. *Conditionally on $\mathfrak{X}_n$, for any $f \in H(s,L)$ and $\delta \in G_k$, we have*

$$|\bar{f}_k^{(\delta)}(x_k) - f(x_k)| \lesssim \lambda(\mathbf{E}_k^{(\delta)})^{-1}\big(L|\delta|^s + \sigma(n\bar{\mu}_n(\delta))^{-1/2}\|\mathbf{U}_k^{(\delta)}\xi\|_\infty\big)$$

on $\Omega_k(\delta)$, where $\mathbf{U}_k^{(\delta)}$ is a $\mathfrak{X}_n$-measurable matrix of size $(R+1)\times(n\bar{\mu}_n(\delta))$ satisfying $\mathbf{U}_k^{(\delta)}(\mathbf{U}_k^{(\delta)})^\top = \mathbf{Id}_{R+1}$.

Note that within Lemma 1, the bandwidth δ can change from one point x_k to another. We denote shortly $\mathbf{U}_k := \mathbf{U}_k^{(\delta_k)}$. Let us define $W := \mathbf{U}\xi$ where $\mathbf{U} := (\mathbf{U}_0^\top, \ldots, \mathbf{U}_{2^J}^\top)^\top$. In view of Lemma 1, W is conditionally on $\mathfrak{X}_n$ a centered Gaussian vector such that $\mathbf{E}_{f\mu}[W_k^2|\mathfrak{X}_n] = 1$

for any $k \in \{0, \dots, (R+1)2^J\}$. We introduce $W^N := \max_{0 \leqslant k \leqslant (R+1)2^J} |W_k|$ and the event $\mathcal{W}_N := \{|W^N - \mathbf{E}[W^N|\mathfrak{X}_n]| \leqslant L_W(\log n)^{1/2}\}$, where $L_W > 0$. We recall the following classical results about the supremum of a Gaussian vector (see for instance in Ledoux and Talagrand (1991)):

$$\mathbf{E}_{f\mu}\big[W^N|\mathfrak{X}_n\big] \lesssim (\log N)^{1/2} \lesssim (\log n)^{1/2},$$

and

$$\mathbf{P}_{f\mu}\big[\mathcal{W}_N^{\complement}|\mathfrak{X}_n\big] \lesssim \exp(-L_W^2(\log n)/2) = n^{-L_W^2/2}. \tag{4.3}$$

Let us define the event

$$\mathrm{T}_k := \{\bar{\mu}_n(\Delta_k) \leqslant \bar{\mu}_n(\widehat{\Delta}_k)\}$$

and $R_k := \sigma\big(\frac{\log n}{n\bar{\mu}_n(\Delta_k)}\big)^{1/2}$ where the intervals Δ_k are given by

$$\Delta_k := \operatorname*{argmax}_{\delta \in G_k} \Big\{\bar{\mu}_n(\delta) \mid L|\delta|^s \leqslant \sigma\Big(\frac{\log n}{n\bar{\mu}_n(\delta)}\Big)^{1/2}\Big\}.$$

There is an event $\mathrm{S}_n \in \mathfrak{X}_n$ such that $\mu^n[\mathrm{S}_n^{\complement}] = o(1)$ faster than any power of n, and such that $R_k \asymp r_n(x_k)$ and $\lambda(\mathbf{E}_k^{(\Delta_k)}) \gtrsim 1$, uniformly for any $k \in \{0, \dots, 2^J - 1\}$. This event is constructed below. We decompose

$$|\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - f(x_k)| \leqslant A_k + B_k + C_k + D_k,$$

where

$$A_k := |\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - f(x_k)|\mathbf{1}_{\mathcal{W}_N^{\complement} \cup \mathrm{S}_n^{\complement}},$$

$$B_k := |\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - f(x_k)|\mathbf{1}_{\mathrm{T}_k^{\complement} \cap \mathcal{W}_N \cap \mathrm{S}_n},$$

$$C_k := |\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - \bar{f}_k^{(\Delta_k)}(x_k)|\mathbf{1}_{\mathrm{T}_k \cap \mathrm{S}_n},$$

$$D_k := |\bar{f}_k^{(\Delta_k)}(x_k) - f(x_k)|\mathbf{1}_{\mathcal{W}_N \cap \mathrm{S}_n}.$$

Term A_k. For any $\delta \in G_k$, we have

$$|\bar{f}_k^{(\delta)}(x_k)| \lesssim (n\bar{\mu}_n(\delta))^{1/2}\|f\|_\infty(1 + W^N). \tag{4.4}$$

This inequality is proved below. Using (4.4), we can bound

$$\mathbf{E}_{f\mu}\big[w\big(\max_{0 \leqslant k \leqslant 2^J} r_n(x_k)^{-1}|\bar{f}_k^{(\widehat{\Delta}_k)}(x_k)|\big)|\mathfrak{X}_n\big]$$

by some power of n. Using $\|f\|_\infty \leqslant Q$ together with the fact that L_W can be arbitrarily large in (4.3) and since $\mu^n[S_n^{\complement}] = o(1)$ faster than any power of n, we obtain

$$\mathbf{E}_{f\mu}\big[w(\max_{0\leqslant k\leqslant 2^J} r_n(x_k)^{-1}A_k)\big] = o(1).$$

Term D_k. Using together Lemma 1, the definition of Δ_k and the fact that $W^N \lesssim (\log n)^{1/2}$ on $\mathcal{W}_N$, we have

$$|\bar f_k^{(\Delta_k)}(x_k) - f(x_k)| \leqslant \lambda(\mathbf{E}_k^{(\Delta_k)})^{-1} R_k (1 + (\log n)^{-1/2} W^N) \lesssim \lambda(\mathbf{E}_k^{(\Delta_k)})^{-1} r_n(x_k)$$

on $\mathcal{W}_N \cap \mathrm{S}_n$, thus

$$\mathbf{E}_{f\mu}\big[w(\max_{0\leqslant k\leqslant 2^J} r_n(x_k)^{-1}D_k)\big] \lesssim 1.$$

Term C_k. We introduce $G_k(\delta) := \{\delta' \in G_k | \delta' \subset \delta\}$ and the following events:

$$\mathcal{T}_k(\delta,\delta',p) := \big\{|\langle \bar f_k^{(\delta)} - \bar f_k^{(\delta')}, \varphi_{kp}\rangle_{\delta'}| \leqslant \sigma \|\varphi_{kp}\|_{\delta'} T_n(\delta,\delta')\big\},$$
$$\mathcal{T}_k(\delta,\delta') := \cap_{0\leqslant p\leqslant R} \mathcal{T}_k(\delta,\delta'),$$
$$\mathcal{T}_k(\delta) := \cap_{\delta'\in G_k(\delta)} \mathcal{T}_k(\delta,\delta').$$

By the definition (3.13) of the selection rule, we have $\mathrm{T}_k \subset \mathcal{T}_k(\widehat\Delta_k, \Delta_k)$. Let $\delta \in G_k, \delta' \in G_k(\delta)$. On $\mathcal{T}_k(\delta,\delta') \cap \Omega_k(\delta')$ we have (see below)

$$|\bar f_k^{(\delta)}(x_k) - \bar f_k^{(\delta')}(x_k)| \lesssim \lambda(\mathbf{E}_k^{(\delta')})^{-1} \Big(\frac{\log n}{n\bar\mu_n(\delta')}\Big)^{1/2}. \tag{4.5}$$

Thus, using (4.5), we obtain

$$\mathbf{E}_{f\mu}\big[w(\max_{0\leqslant k\leqslant 2^J} r_n(x_k)^{-1}C_k)\big] \lesssim 1.$$

Term B_k. By the definition (3.13) of the selection rule, we have $\mathrm{T}_k^{\complement} \subset \mathcal{T}_k(\Delta_k)^{\complement}$. We need the following lemma.

Lemma 2. *If* $\delta \in G_k$ *satisfies*

$$L|\delta|^s \leqslant \sigma \Big(\frac{\log n}{n\bar\mu_n(\delta)}\Big)^{1/2} \tag{4.6}$$

and $f \in H(s,L)$, *we have*

$$\mathbf{P}_{f\mu}\big[\mathcal{T}_k(\delta)^{\complement} | \mathfrak{X}_n\big] \leqslant (R+1)(n\bar\mu_n(\delta))^{1-D^2/2}$$

on $\Omega_k(\delta)$, *where* D *is the constant from the threshlod* (3.14).

Using together Lemma 2, $\|f\|_\infty \leqslant Q$ and (4.4), we obtain

$$\mathbf{E}_{f\mu}\big[w\big(\max_{0\leqslant k\leqslant 2^J} R_k^{-1}|\bar{f}_k^{(\widehat{\Delta}_k)}(x_k) - f(x_k)|\mathbf{1}_{\mathrm{T}_k^{\complement}\cap\mathcal{W}_N}\big)|\mathfrak{X}_n\big] \lesssim 1,$$

thus

$$\mathbf{E}_{f\mu}\big[w\big(\max_{0\leqslant k\leqslant 2^J} r_n(x_k)^{-1}B_k\big)\big] \lesssim 1,$$

and Theorem 1 follows. □

Proof of Lemma 1. On $\Omega_k(\delta)$, we have $\bar{\mathbf{X}}_k^{(\delta)} = \mathbf{X}_k^{\delta}$, and $\lambda(\mathbf{X}_k^{(\delta)}) > (n\bar{\mu}_n(\delta))^{-1/2} > 0$, thus $\mathbf{X}_k^{(\delta)}$ and $\mathbf{E}_k^{(\delta)}$ are invertible. Let f_k be the Taylor polynomial of f at x_k up to the order $\lfloor s\rfloor$ and $\theta_k \in \mathbb{R}^{R+1}$ be the coefficient vector of f_k. Using $f \in H(s,L)$, we obtain

$$\begin{aligned}|\bar{f}_k^{(\delta)}(x_k) - f(x_k)| &\lesssim |\langle(\mathbf{\Lambda}_k^{(\delta)})^{-1}(\bar{\theta}_k^{(\delta)} - \theta_k)\,,\, e_1\rangle| + |\delta|^s\\ &= |\langle(\mathbf{E}_k^{(\delta)})^{-1}\mathbf{\Lambda}_k^{(\delta)}\mathbf{X}_k^{(\delta)}(\bar{\theta}_k^{(\delta)} - \theta_k)\,,\, e_1\rangle| + |\delta|^s.\end{aligned}$$

In view of (3.9), we have on $\Omega_k(\delta)$ for any $p \in \{0,\ldots,R\}$:

$$\begin{aligned}(\mathbf{X}_k^{(\delta)}(\bar{\theta}_k^{(\delta)} - \theta_k))_p &= \langle\bar{f}_k^{(\delta)} - f_k\,,\, \varphi_{kp}\rangle_\delta\\ &= \langle Y - f_k\,,\, \varphi_{kp}\rangle_\delta\end{aligned}$$

thus, $\mathbf{X}_k^{(\delta)}(\bar{\theta}_k^{(\delta)} - \theta_k) = B_k^{(\delta)} + V_k^{(\delta)}$ where $(B_k^{(\delta)})_p := \langle f - f_k\,,\, \varphi_{kp}\rangle_\delta$ and $(V_k^{(\delta)})_p := \langle\xi\,,\, \varphi_{kp}\rangle_\delta$, which correspond respectively to bias and variance terms. Since $f \in H(s,L)$ and $\lambda(M)^{-1} = \|M^{-1}\|$ for any symmetrical and positive matrix M, we have

$$|\langle(\mathbf{E}_k^{(\delta)})^{-1}\mathbf{\Lambda}_k^{(\delta)}B_k^{(\delta)}\,,\, e_1\rangle| \lesssim \lambda(\mathbf{E}_k^{(\delta)})^{-1}L|\delta|^s.$$

Since $(V_k^{(\delta)})_p = (n\bar{\mu}_n(\delta))^{-1}\mathbf{D}_k^{(\delta)}\xi$ where $\mathbf{D}_k^{(\delta)}$ is the $(R+1)\times(n\bar{\mu}_n(\delta))$ matrix with entries $(\mathbf{D}_k^{(\delta)})_{i,p} := (X_i - x_k)^p$, $X_i \in \delta$, we can write

$$|\langle(\mathbf{E}_k^{(\delta)})^{-1}\mathbf{\Lambda}_k^{(\delta)}V_k^{(\delta)}\,,\, e_1\rangle_\delta| \lesssim \sigma(n\bar{\mu}_n(\delta))^{-1/2}\|(\mathbf{E}_k^{(\delta)})^{-1/2}\|\|\mathbf{U}_k^{(\delta)}\xi\|_\infty,$$

where $\mathbf{U}_k^{(\delta)} := (n\bar{\mu}_n(\delta))^{-1/2}(\mathbf{E}_k^{(\delta)})^{-1/2}\mathbf{\Lambda}_k^{(\delta)}\mathbf{D}_k^{(\delta)}$ satisfies $\mathbf{U}_k^{(\delta)}(\mathbf{U}_k^{(\delta)})^\top = \mathbf{Id}_{R+1}$ since $\mathbf{E}_k^{(\delta)} = \mathbf{\Lambda}_k^{(\delta)}\mathbf{X}_k^{(\delta)}\mathbf{\Lambda}_k^{(\delta)}$ and $\mathbf{X}_k^{(\delta)} = (n\bar{\mu}_n(\delta))^{-1}\mathbf{D}_k^{(\delta)}(\mathbf{D}_k^{(\delta)})^\top$, thus the lemma. □

Proof of (4.4). If $\bar{\mu}_n(\delta) = 0$, we have $\bar{f}_k^{(\delta)} = 0$ by definition and the result is obvious, thus we assume $\bar{\mu}_n(\delta) > 0$. Since $\lambda(\bar{\mathbf{X}}_k^{(\delta)}) \geqslant (n\bar{\mu}_n(\delta))^{-1/2} > 0$, $\bar{\mathbf{X}}_k^{(\delta)}$ and $\mathbf{\Lambda}_k^{(\delta)}$ are invertible and $\mathbf{E}_k^{(\delta)}$ also is. The proof of (4.4) is then similar to that of Lemma 1, where the bias is bounded by $\|f\|_\infty$ and where we use the fact that $\lambda(\bar{\mathbf{X}}_k^{(\delta)}) \geqslant (n\bar{\mu}_n(\delta))^{-1/2}$ to control the variance term. □

Proof of (4.5)**.** Let us define $\mathbf{H}_k^{(\delta)} := \mathbf{\Lambda}_k^{(\delta)} \mathbf{X}_k^{(\delta)}$. On $\Omega_k(\delta)$, we have:

$$|\bar{f}_k^{(\delta)}(x_k) - \bar{f}_k^{(\delta')}(x_k)| = |(\bar{\theta}_k^{(\delta)} - \bar{\theta}_k^{(\delta')})_0| \lesssim \lambda(\mathbf{E}_k^{(\delta')})^{-1} \|\mathbf{H}_k^{(\delta')}(\bar{\theta}_k^{(\delta)} - \bar{\theta}_k^{(\delta')})\|_\infty.$$

Since on $\Omega_k(\delta')$, $(\mathbf{H}_k^{(\delta')}(\bar{\theta}_k^{(\delta)} - \bar{\theta}_k^{(\delta')}))_p = \langle \bar{f}_k^{(\delta)} - \bar{f}_k^{(\delta')}, \varphi_{kp}\rangle_{\delta'} / \|\varphi_{kp}\|_{\delta'}$, and since $\delta' \subset \delta$, we obtain (4.5) on $\mathcal{T}_k(\delta, \delta')$. □

Proof of Lemma 2. We denote by $\mathbf{P}_k^{(\delta)}$ the projection onto $\mathrm{Span}\{\varphi_{k0}, \ldots, \varphi_{kR}\}$ with respect to the inner product $\langle \cdot\,, \cdot \rangle_\delta$. Note that on $\Omega_k(\delta)$, we have $\bar{f}_k^{(\delta)} = \mathbf{P}_k^{(\delta)} Y$. Let $\delta \in G_k$ and $\delta' \in G_k(\delta)$. In view of (3.9), we have on $\Omega_k(\delta)$ for any $\varphi = \varphi_{kp}$, $p \in \{0, \ldots, R\}$:

$$\begin{aligned} \langle \bar{f}_k^{(\delta')} - \bar{f}_k^{(\delta)}, \varphi \rangle_{\delta'} &= \langle Y - \bar{f}_k^{(\delta)}, \varphi \rangle_{\delta'} \\ &= \langle f - \mathbf{P}_k^{(\delta)} Y, \varphi \rangle_{\delta'} + \langle \xi, \varphi \rangle_{\delta'} \\ &= A_k - B_k + C_k, \end{aligned}$$

where $A_k := \langle f - \mathbf{P}_k^{(\delta)} f, \varphi \rangle_{\delta'}$, $B_k := \sigma \langle \mathbf{P}_k^{(\delta)} \xi, \varphi \rangle_{\delta'}$ and $C_k := \sigma \langle \xi, \varphi \rangle_{\delta'}$. If f_k is the Taylor polynomial of f at x_k up to the order $\lfloor s \rfloor$, since $\delta' \subset \delta$ and $f \in H(s, L)$ we have:

$$|A_k| \leqslant \|\varphi\|_{\delta'} \|f - f_k + \mathbf{P}_k^{(\delta)}(f_k - f)\|_\delta \leqslant \|\varphi\|_{\delta'} \|f - f_k\|_\delta \lesssim \|\varphi\|_{\delta'} L |\delta|^s,$$

and using (4.6), we obtain $|A_k| \lesssim \|\varphi\|_{\delta'} \sigma \big(\frac{\log n}{n \bar{\mu}_n(\delta)}\big)^{1/2}$. Since $\mathbf{P}_k^{(\delta)}$ is an orthogonal projection, the variance of B_k is equal to

$$\begin{aligned} \sigma^2 \mathbf{E}_{f\mu}\big[\langle \mathbf{P}_k^{(\delta)} \xi, \varphi \rangle_{\delta'}^2 | \mathfrak{X}_n\big] &\leqslant \sigma^2 \|\varphi\|_{\delta'}^2 \mathbf{E}_{f\mu}\big[\|\mathbf{P}_k^{(\delta)} \xi\|_{\delta'}^2 | \mathfrak{X}_n\big] \\ &= \sigma^2 \|\varphi\|_{\delta'}^2 \,\mathrm{Tr}(\mathbf{P}_k^{(\delta)}) / (n \bar{\mu}_n(\delta')), \end{aligned}$$

where $\mathrm{Tr}(M)$ stands for the trace of a matrix M. Since $\mathbf{P}_k^{(\delta)}$ is the projection onto Pol_R, $\mathrm{Tr}(\mathbf{P}_k^{(\delta)}) \leqslant R + 1$, and the variance of B_k is smaller than $\sigma^2 \|\varphi\|_{\delta'}^2 (R+1)/(n \bar{\mu}_n(\delta'))$. Then,

$$\mathbf{E}_{f\mu}[(B + C)^2 | \mathfrak{X}_n] \leqslant \sigma^2 \|\varphi\|_{\delta'}^2 C_R^2 / (n \bar{\mu}_n(\delta')). \tag{4.7}$$

In view of the threshold choice (3.14), we have

$$\begin{aligned} &\big\{ |\langle \bar{f}_k^{(\delta)} - \bar{f}_k^{(\delta')}, \varphi \rangle_{\delta'}| > \|\varphi\|_{\delta'} T_n(\delta, \delta') \big\} \\ &\qquad \subset \Big\{ \frac{\|\varphi\|_{\delta'}^{-1} |B_k + C_k|}{\sigma (n \bar{\mu}_n(\delta'))^{-1/2} C_R} > D \big(\log(n \bar{\mu}_n(\delta))\big)^{1/2} \Big\}, \end{aligned}$$

and using (4.7) together with $\mathbf{P}[|N(0,1)| > x] \leqslant \exp(-x^2/2)$ and $|G_k(\delta)| \leqslant (n\bar{\mu}_n(\delta))$, we obtain

$$\mathbf{P}_{f\mu}[\mathcal{T}(\delta)^{\complement}|\mathfrak{X}_n] \leqslant \sum_{\delta' \in G_k(\delta)} \sum_{p=0}^{R} \exp\big(- D^2 \log(n\bar{\mu}_n(\delta))/2\big)$$
$$\leqslant (R+1)(n\bar{\mu}_n(\delta))^{1-D^2/2},$$

which concludes the proof. □

Construction of S_n**.** We construct an event $\mathrm{S}_n \in \mathfrak{X}_n$ such that $\mu^n\big[\mathrm{S}_n^{\complement}\big] = o(1)$ faster than any power of n, and such that on this event, $R_k \asymp r_n(x_k)$ and $\lambda(\mathbf{E}_k^{(\Delta_k)}) \gtrsim 1$ uniformly for any $k \in \{0, \ldots, 2^J\}$. We need preliminary approximation results, linked with the approximation of μ by $\bar{\mu}_n$. The following deviation inequalities use Berstein inequality for the sum of independent random variables, which is standard. We have

$$\mu^n\Big[\Big|\frac{\bar{\mu}_n(\delta)}{\mu(\delta)} - 1\Big|\Big] \lesssim \exp\big(- \varepsilon^2 n\mu(\delta)\big) \tag{4.8}$$

for any interval $\delta \subset [0,1]$ and $\varepsilon \in (0,1)$. Let us define the events

$$\mathrm{D}_{n,a}^{(\delta)}(x,\varepsilon) := \Big\{ \Big| \frac{1}{\mu(\delta)} \int_\delta \Big(\frac{\cdot - x}{|\delta|}\Big)^a d\bar{\mu}_n - e_a(x,\mu)\Big| \leqslant \varepsilon \Big\}$$

where $e_a(x,\mu) := (1+(-1)^a)(\beta(x)+1)/(a+\beta(x)+1)$ (a is a natural integer) where we recall that $\beta(x)$ comes from assumption D (if x is such that $\mu(x) > 0$ then $\beta(x) = 0$). Using together Bernstein inequality and the fact that

$$\frac{1}{\mu(\delta)} \int_\delta \Big(\frac{t-x}{|\delta|}\Big)^a \mu(t)dt \to e_a(x,\mu)$$

as $|\delta| \to 0$, we obtain

$$\mu^n\big[(\mathrm{D}_{n,a}^{(\delta)}(x,\varepsilon))^{\complement}\big] \lesssim \exp\big(- \varepsilon^2 n\mu(\delta)\big). \tag{4.9}$$

By definition (3.15) of G_k, we have $\Delta_k = [x_k - H_n(x_k), x_k + H_n(x_k)]$ where

$$H_n(x) := \operatorname*{argmin}_{h\in[0,1]} \Big\{ Lh^s \geqslant \sigma\Big(\frac{\log n}{n\bar{\mu}_n([x-h,x+h])}\Big)^{1/2}\Big\} \tag{4.10}$$

is an approximation of $h_n(x)$ (see (2.1)). Since $\bar{\mu}_n$ is "close" to μ, these quantities are close to each other for any x. Indeed, if $\delta_n(x) := [x - h_n(x), x + h_n(x)]$ and $\Delta_n(x) := [x - H_n(x), x + H_n(x)]$ we have using together (4.10) and (2.1):

$$\big\{H_n(x) \leqslant (1+\varepsilon)h_n(x)\big\} = \Big\{\frac{\bar{\mu}_n[(1+\varepsilon)\delta_n(x)]}{\mu[\delta_n(x)]} \geqslant (1-\varepsilon)^{-2}\Big\} \tag{4.11}$$

for any $\varepsilon \in (0,1)$, where $(1+\varepsilon)\delta_n(x) := [x - (1+\varepsilon)h_n(x), x + (1+\varepsilon)h_n(x)]$. Hence, for each $x = x_k$, the left hand side event of (4.11) has a probability that can be controlled under assumption D by (4.8), and the same argument holds for $\{H_n(x) > (1-\varepsilon)h_n(x)\}$. Combining (4.8), (4.9) and (4.11), we obtain that the event

$$\mathrm{B}_{n,a}(x,\varepsilon) := \Big\{ \Big| \frac{1}{\bar\mu_n(\Delta_n(x))} \int_{\Delta_n(x)} \Big(\frac{\cdot - x}{|\delta_n(x)|} \Big)^a d\bar\mu_n - e_a(x,\mu) \Big| \leqslant \varepsilon \Big\}$$

satisfies also (4.9) for n large enough. This proves that $(\mathbf{X}_k^{(\Delta_k)})_{p,q}$ and $(\mathbf{\Lambda}_k^{(\Delta_k)})_p$ are close to $e_{p+q}(x_k,\mu)$ and $e_{2p}(x_k,\mu)^{-1/2}$ respectively on the event

$$\mathrm{S}_n := \bigcap_{a \in \{0,\dots,2R\}} \bigcap_{k \in \{0,\dots,2^J-1\}} \mathrm{B}_{n,a}(x_k,\varepsilon).$$

Using the fact that $\lambda(M) = \inf_{\|x\|=1} x^\top M x$ for a symmetrical matrix M, where $\lambda(M)$ denotes the smallest eigenvalue of M, we can conclude that for n large enough,

$$\lambda(\mathbf{\Lambda}_k^{(\Delta_k)} \mathbf{X}_k^{(\Delta_k)} \mathbf{\Lambda}_k^{(\Delta_k)}) \gtrsim \min_{x\in[0,1]} \lambda(\mathbf{E}(x,\mu)),$$

where $\mathbf{E}(x,\mu)$ has entries $(\mathbf{E}(x,\mu))_{p,q} = e_{p+q}(x,\mu)/(e_{2p}(x,\mu)e_{2q}(x,\mu))^{1/2}$. Since $\mathbf{E}(x,\mu)$ is definite positive for any $x \in [0,1]$, we obtain that on S_n, $\lambda(\mathbf{X}_k^{(\Delta_k)}) \gtrsim 1$, thus $\mathrm{S}_n \subset \Omega_n(\Delta_k)$ and $\lambda(\mathbf{E}_k^{(\Delta_k)}) \gtrsim 1$ uniformly for any $k \in \{0,\dots,2^J-1\}$, since $\mathbf{E}_k^{(\Delta_k)} = \mathbf{\Lambda}_k^{(\Delta_k)} \mathbf{X}_k^{(\Delta_k)} \mathbf{\Lambda}_k^{(\Delta_k)}$ on $\Omega_n(\Delta_k)$. Moreover, since $R_k = LH_n(x_k)^s$, using together (4.8) and (4.11), we obtain $R_k \asymp r_n(x_k)$ uniformly for $k \in \{0,\dots,2^J-1\}$. □

Proof of Theorem 2. The main features of the proof are first, a reduction to the Bayesian risk over an hardest cubical subfamily of functions for the $\mathbb{L}^\infty$ metrics, which is standard: see Korostelev (1993), Donoho (1994), Korostelev and Nussbaum (1999) and Bertin (2004), and the choice of rescaled hypothesis with design-adapted bandwidth $h_n(\cdot)$, necessary to achieve the rate $r_n(\cdot)$.

Let us consider $\varphi \in H(s,L;\mathbb{R})$ (the extension of $H(s,L)$ to the whole real line) with support $[-1,1]$ and such that $\varphi(0) > 0$. We define

$$a := \min\Big[1, \Big(\frac{2}{\|\varphi\|_\infty^2} \Big(\frac{1}{1+2s+\beta} - \alpha \Big) \Big)^{1/(2s)} \Big]$$

and

$$\Xi_n := 2a(1 + 2^{1/(s-\lfloor s \rfloor)}) \sup_{x\in[0,1]} h_n(x),$$

where we recall that $\lfloor s \rfloor$ is the largest integer smaller than s. Note that (2.6) entails

$$\Xi_n \lesssim (\log n/n)^{1/(1+2s+\beta)}. \tag{4.12}$$

If $I_n = [c_n, d_n]$, we introduce $x_k := c_n + k\,\Xi_n$ for $k \in K_n := \{1, \ldots, [|I_n|\,\Xi_n^{-1}]\}$, and denote for the sake of simplicity $h_k := h_n(x_k)$. We consider the family of functions

$$f(\cdot;\theta) := \sum_{k \in K_n} \theta_k f_k(\cdot), \quad f_k(\cdot) := L a^s h_k^s \varphi\Big(\frac{\cdot - x_k}{h_k}\Big),$$

which belongs to $H(s, L)$ for any $\theta \in [-1, 1]^{|K_n|}$. Using Bernstein inequality, we can see that

$$\mathrm{H}_n := \bigcap_{k \in K_n} \Big\{ \frac{\bar{\mu}_n([x_k - h_k, x_k + h_k])}{\mu([x_k - h_k, x_k + h_k])} \geqslant 1/2 \Big\}$$

satisfies

$$\mu^n[\mathrm{H}_n] = 1 - o(1). \tag{4.13}$$

Let us introduce $b := c^s \varphi(0)$. For any distribution $\mathbf{B}$ on $\Theta_n \subset [-1, 1]^{|K_n|}$, by a minoration of the minimax risk by the Bayesian risk, and since w is non-decreasing, the left hand side of (2.8) is smaller than

$$\begin{aligned} w(b) \inf_{\widehat{\theta}} \int_{\Theta_n} \mathbf{P}_\theta^n\big[\max_{k \in K_n} |\widehat{\theta}_k - \theta_k| \geqslant 1\big]\mathbf{B}(d\theta) \\ \geqslant w(b) \int_{\mathrm{H}_n} \inf_{\widehat{\theta}} \int_{\Theta_n} \mathbf{P}_\theta^n\big[\max_{k \in K_n} |\widehat{\theta}_k - \theta_k| \geqslant 1 | \mathfrak{X}_n\big]\mathbf{B}(d\theta) d\mu^n. \end{aligned}$$

Hence, together with (4.13), Theorem 2 follows if we show that on H_n

$$\sup_{\widehat{\theta}} \int_{\Theta_n} \mathbf{P}_\theta^n\big[\max_{k \in K_n} |\widehat{\theta}_k - \theta_k| < 1 | \mathfrak{X}_n\big]\mathbf{B}(d\theta) = o(1). \tag{4.14}$$

We denote by $L(\theta; Y_1, \ldots, Y_n)$ the conditional on $\mathfrak{X}_n$ likelihood function of the observations Y_i from (1.1) when $f(\cdot) = f(\cdot;\theta)$. Conditionally on $\mathfrak{X}_n$, we have

$$L(\theta; Y_1, \ldots, Y_n) = \prod_{1 \leqslant i \leqslant n} g_\sigma(Y_i) \prod_{k \in K_n} \frac{g_{v_k}(y_k - \theta_k)}{g_{v_k}(y_k)},$$

where g_v is the density of $N(0, v^2)$, $v_k^2 := \mathbf{E}\{y_k^2 | \mathfrak{X}_n\}$ and

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

Thus, choosing

$$\mathbf{B} := \bigotimes_{k \in K_n} \mathbf{b}, \quad \mathbf{b} := (\delta_{-1} + \delta_1)/2, \quad \Theta_n := \{-1, 1\}^{|K_n|},$$

the left hand side of (4.14) is smaller than

$$\int \frac{\prod_{1 \leqslant i \leqslant n} g_\sigma(Y_i)}{\prod_{k \in K_n} g_{v_k}(y_k)} \Big(\prod_{k \in K_n} \sup_{\widehat{\theta}_k} \int_{\{-1,1\}} \mathbf{1}_{|\widehat{\theta}_k - \theta_k| < 1}\, g_{v_k}(y_k - \theta_k) \mathbf{b}(d\theta_k) \Big) dY_1 \times \cdots \times dY_n,$$

and $\widehat{\theta}_k = \mathbf{1}_{y_k \geqslant 0} - \mathbf{1}_{y_k < 0}$ are strategies reaching the supremum. Then, in (4.14), it suffices to take the supremum over estimators $\widehat{\theta}$ with coordinates $\widehat{\theta}_k \in \{-1, 1\}$ measurable with respect to y_k only. Since conditionally on $\mathfrak{X}_n$, y_k is in law $N(\theta_k, v_k^2)$, the left hand side of (4.14) is smaller than

$$\prod_{k \in K_n} \Big(1 - \inf_{\widehat{\theta}_k \in \{-1,1\}} \int_{\{-1,1\}} \int \mathbf{1}_{|\widehat{\theta}_k(u) - \theta_k| \geqslant 1} g_{v_k}(u - \theta_k) du \, \mathbf{b}(d\theta_k) \Big).$$

Moreover, if $\Phi(x) := \int_{-\infty}^{x} g_1(t) dt$

$$\inf_{\widehat{\theta}_k \in \{-1,1\}} \int_{\{-1,1\}} \int \mathbf{1}_{|\widehat{\theta}_k(u) - \theta_k| \geqslant 1} g_{v_k}(u - \theta_k) du \, \mathbf{b}(d\theta_k)$$
$$\geqslant \frac{1}{2} \int \min\big(g_{v_k}(u-1), g_{v_k}(u+1) \big) du = \Phi(-1/v_k).$$

On H_n, we have in view of (2.1)

$$v_k^2 = \frac{\sigma^2}{\sum_{i=1}^n f_k^2(X_i)} \geqslant \frac{2}{(1-\delta) \|\varphi\|_\infty^2 c^{2s} \log n},$$

and since $\Phi(-x) \geqslant \exp(-x^2/2)(x\sqrt{2\pi})$ for any $x > 0$, we obtain

$$\Phi(-1/v_k) \gtrsim (\log n)^{-1/2} n^{\{\alpha - 1/(1+2s+\beta)\}/2} =: L_n.$$

Thus, the left hand side of (4.14) is smaller than $(1 - L_n)^{|K_n|}$, and since

$$|I_n| \Xi_n^{-1} L_n \gtrsim n^{\{1/(1+2s+\beta) - \alpha\}/2} (\log n)^{1/2 - 1/(1+2s+\beta)} \to +\infty$$

as $n \to +\infty$, Theorem 2 follows. □

Proof of Corollary 1. Let us consider the loss function $w(\cdot) = |\cdot|$, and let $\widehat{f}_n^v$ be an estimator converging with rate $v_n(\cdot)$ over F in the sense of (2.2). Hence,

$$1 \lesssim \sup_{f \in F} \mathbf{E}_{f\mu} \big[\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n^v(x) - f(x)| \big]$$
$$\leqslant \sup_{x \in I_n} \frac{v_n(x)}{r_n(x)} \sup_{f \in F} \mathbf{E}_{f\mu} \big[\sup_{x \in I_n} v_n(x)^{-1} |\widehat{f}_n^v(x) - f(x)| \big] \lesssim \sup_{x \in I_n} \frac{v_n(x)}{r_n(x)},$$

where we used Theorem 2. □

Proof of Proposition 1. Without loss of generality, we consider the loss $w(\cdot) = |\cdot|$. For proving Proposition 1, we use the linear LPE. If we denote by $\partial^m f$ the m-th derivative of f, a slight modification of the proof of Lemma 1 gives for $f \in H(s, L)$ with $s > m$,

$$|\partial^m \bar{f}_k^{(\delta)}(x_k) - \partial^m f(x_k)| \lesssim \lambda(\mathbf{E}_k^{(\delta)})^{-1} |\delta|^{-m} \big(L|\delta|^s + \sigma(n\bar{\mu}_n(\delta))^{-1/2} W^N \big),$$

where in the same way as in the proof of Theorem 1, W^N satisfies

$$\mathbf{E}_{f\mu}[W^N|\mathfrak{X}_n] \lesssim (\log N)^{1/2}, \tag{4.15}$$

with N depending on the size of the supremum, to be specified below. First, we prove a). Since $|I_n| \sim (\ell_n/n)^{1/(2s+1)}$, if $I_n = [a_n, b_n]$, the points

$$x_k := a_n + (k/n)^{1/(2s+1)}, \quad k \in \{0, \dots, N\},$$

where $N := [\ell_n]$ belongs to I_n. We consider the bandwidth

$$h_n = \Big(\frac{\log \ell_n}{n}\Big)^{1/(2s+1)}, \tag{4.16}$$

and we take $\delta_k := [x_k - h_n, x_k + h_n]$. Note that since $\mu(x) > 0$ for any x, $\bar{\mu}_n(\delta) \asymp |\delta|$ as $|\delta| \to 0$ with probability going to 1 faster than any power of n (using Berstein inequality, for instance). We consider the estimator defined by

$$\widehat{f}_n(x) := \sum_{m=0}^{r} \partial^m \bar{f}_k^{(\delta_k)}(x_k)(x - x_k)^m/m! \quad \text{for } x \in [x_k, x_{k+1}), \quad k \in \{0, \dots, [\ell_n]\}, \tag{4.17}$$

where $r := \lfloor s \rfloor$. Using a Taylor expansion of f up to the degree r together with (4.16) gives

$$(n/\log n)^{s/(1+2s)} \sup_{x \in I_n} |\widehat{f}_n(x) - f(x)| \lesssim \Big(\frac{\log \ell_n}{\log n}\Big)^{s/(1+2s)} (1 + (\log \ell_n)^{-1/2} W^N).$$

Then, integrating with respect to $\mathbf{P}_{f\mu}(\cdot|\mathfrak{X}_n)$ and using (4.15) where $N = [\ell_n]$ entails a), since $\log \ell_n = o(\log n)$.

The proof of b) is similar to that of a). In this setting, the rate $r_n(\cdot)$ (see (2.1)) can be written as $r_n(x) = (\log n/n)^{\alpha_n(x)}$ for x in I_n (for n large enough) where $\alpha_n(x_0) = s/(1 + 2s + \beta)$ and $\alpha_n(x) > s/(1 + 2s + \beta)$ for $x \in I_n - \{x_0\}$. We define

$$x_{k+1} = \begin{cases} x_k + n^{-\alpha_n(x_k)/s} & \text{for } k \in \{-N, \dots, -1\} \\ x_k + n^{-\alpha_n(x_{k+1})/s} & \text{for } k \in \{0, \dots, N\}, \end{cases}$$

where $N := [\ell_n]$. All the points fit in I_n, since $|x_{-N} - x_N| \leqslant \sum_{-N \leqslant k \leqslant N} n^{-\min(\alpha_n(x_k), \alpha_n(x_{k+1}))/s} \leqslant 2(\ell_n/n)^{1/(1+2s+\beta)}$. We consider the bandwidths

$$h_k := (\log \ell_n/n)^{\alpha_n(x_k)/s},$$

and the intervals $\delta_k = [x_k - h_k, x_k + h_k]$. We keep the same definition (4.17) for $\widehat{f}_n$. Since x_0 is a local extremum of $r_n(\cdot)$, we have in the same way as in the proof of a) that

$$\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \lesssim \Big[\max_{-N \leqslant k \leqslant -1} \Big(\frac{\log \ell_n}{\log n} \Big)^{\alpha_n(x_k)} + \max_{0 \leqslant k \leqslant N-1} \Big(\frac{\log \ell_n}{\log n} \Big)^{\alpha_n(x_{k+1})} \Big] (1 + (\log \ell_n)^{-1/2} W^N),$$

hence

$$\mathbf{E}_{f\mu} \big[\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \big] \lesssim \Big(\frac{\log \ell_n}{\log n} \Big)^{s/(1+2s+\beta)} = o(1),$$

which concludes the proof of Proposition 1. □

References

Almansa, A., Rouge, B. and Jaffard, S. (2003). Irregular sampling in satellite images and reconstruction algorithms. In *CANUM 2003*. CANUM 2003, http://www.math.univ-montp2.fr/canum03/communications/ms/andres.almansa.pdf.

Antoniadis, A. and Fan, J. Q. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, **96** 939–967.

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35** 225–232.

Antoniadis, A. and Pham, D. T. (1998). Wavelet regression for random or irregular design. *Comput. Statist. Data Anal.*, **28** 353–369.

Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, **6** 127–146 (electronic).

Bertin, K. (2004). 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.

Cai, T. T. and Low, M. G. (2005). Nonparametric estimation over shrinking neighborhoods: superefficiency and adaptation. *Ann. Statist.*, **33** 184–213.

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelets transforms. *Appl. Comput. Harmon. Anal.*, **1** 54–81.

Delouille, V. (2002). *Nonparametric stochastic regression using design-adapted wavelets*. Ph.D. thesis, Université catholique de Louvain.

Delouille, V., Franke, J. and von Sachs, R. (2001). Nonparametric stochastic regression with design-adapted wavelets. *Sankhyā Ser. A*, **63** 328–366. Special issue on wavelets.

Delouille, V., Simoens, J. and Von Sachs, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *Journal of the American Statistical Society*, **99** 643–658.

Delyon, B. and Juditsky, A. (1995). Estimating wavelet coefficients. In *Lecture notes in Statistics* (A. Antoniadis and G. Oppenheim, eds.), vol. 103. Springer-Verlag, New York, 151–168.

Donoho, D. (1992). Interpolating wavelet tranforms. Tech. rep., Department of Statistics, Stanford University, http://www-stat.stanford.edu/ donoho/Reports/1992/interpol.ps.Z.

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probability Theory and Related Fields*, **99** 145–170.

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.

Feichtinger, H. G. and Gröchenig, K. (1994). Theory and practice of irregular sampling. In *Wavelets: mathematics and applications*. Stud. Adv. Math., CRC, Boca Raton, FL, 305–363.

Gaïffas, 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/ .

Gaïffas, S. (2005b). On pointwise adaptive curve estimation based on inhomogeneous data. Preprint LPMA no 974 available at http://hal.ccsd.cnrs.fr/ccsd-00004605/en/.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics*, **6** 135–170.

Guerre, E. (1999). Efficient random rates for nonparametric regression under arbitrary designs. Personal communication.

Hall, P., Marron, J. S., Neumann, M. H. and Tetterington, D. M. (1997). Curve estimation when the design density is low. *The Annals of Statistics*, **25** 756–770.

Hall, P., Park, B. U. and Turlach, B. A. (1998). A note on design transformation and binning in nonparametric curve estimation. *Biometrika*, **85** 469–476.

Jansen, M., Nason, P. G. and Silverman, B. W. (2004). Multivariate nonparametric regression using lifting. Tech. rep., University of Bristol, UK, http://www.stats.ox.ac.uk/˜silverma/pdf/jansennasonsilverman.pdf.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, **10** 1053–1105.

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 contant. *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.

LEPSKI, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, **35** 454–466.

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 SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, **25** 2512–2546.

MAXIM, V. (2003). *Restauration de signaux bruités sur des plans d'experience aléatoires*. Ph.D. thesis, Université Joseph Fourier, Grenoble 1.

PENSKY, M. and WIENS, D. P. (2001). On non-equally spaced wavelet regression. *Advances in Soviet Mathematics*, **53** 681–690.

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.

VÀZQUEZ, C., KONRAD, J. and DUBOIS, E. (2000). Wavelet-based reconstruction of irregularly-sampled images: application to stereo imaging. In *Proc. Int. Conf. on Image Processing, ICIP-2000*. IEEE, http://iss.bu.edu/jkonrad/Publications/local/cpapers/Vazq00icip.pdf.

WONG, M.-Y. and ZHENG, Z. (2002). Wavelet threshold estimation of a regression function with random design. *Journal of Multivariate Analysis*, **80** 256–284.

MODAL'X, UNIVERSITÉ PARIS X – NANTERRE, BÂTIMENT G, 200 AVENUE DE LA RÉPUBLIQUE, 92000 NANTERRE

E-mail address: stephane.gaiffas@u-paris10.fr