Pointwise Adaptive Estimation of a Multivariate Function

Nicolas Klutchnikoff [*]

June 1, 2014

Abstract

In this paper, we address the problem of pointwise estimation in the Gaussian white noise model. We propose a new data-driven procedure that achieves (up to a multiplicative logarithmic term) the minimax rate of convergence over a scale of anisotropic Hölder spaces. Moreover we present a general criterion in order to define what should be an "optimal" procedure of estimation and we prove that our procedure satisfies this criterion. The extra logarithmic term can thus be viewed as an unavoidable price to pay for adaptation.

Keywords and phrases : Gaussian white noise, adaptive estimation, anisotropic Hölder spaces, kernel estimators, pointwise risk.

AMS 2000 subject classification : 62G05, 62G20.

1 Introduction

The aim of this paper is to construct a data-driven procedure of estimation in order to reconstruct a noisy signal in the multivariate Gaussian white noise model and to prove that this procedure is, in some sense, *optimal.*

Many nonparametric procedures of estimation can be used in order to reconstruct a noisy signal : projection on finite dimensional vector spaces (Fourier

[*]Crest (Ensai) and Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS

or wavelet reconstruction), local smoothing (kernel estimators, local polynomials), etc. Each method usually depends on a tuning parameter which could be difficult to calibrate from both practical and theoretical viewpoints. Nevertheless data-driven methods have been developed to select these parameters. Among others let us mention the following popular methods: cross-validation (Duin, 1976, Stone, 1984), coefficient thresholding (Donoho, Johnstone, Kerkyacharian, and Picard, 1995), Lepski's method (Lepski, 1990). The question that arises is the following: are the resulting estimators optimal in some sense? A classical way to answer this question is to prove that the constructed procedure of estimation is adaptive with respect to some nuisance parameter, that is, it achieves the minimax rate of convergence simultaneously over a scale of spaces indexed by these parameters. Unfortunately obtaining such a property is not obvious and in many situations the proposed estimator achieves the minimax rate of convergence only up to an extra multiplicative logarithmic factor. Lepski (1990) and Tsybakov (1998) proved that in particular statistical models this extra term can be viewed as an unavoidable price to pay for adaptation. Nevertheless their definitions, designed for univariate settings, suffer from some major defects in a multivariate setup. The first part of this paper is devoted to the introduction (in a general statistical model) of a new criterion of optimality in order to overcome the main drawbacks of the previous notions.

On the second part of this paper we focus on our main problem: the construction of an optimal procedure of estimation to reconstruct the unknown signal at a given fixed point. This or similar problems received attention of many authors. Among others let us emphasize the following works: Lepski (1990), Lepski and Spokoiny (1997), Goldenshluger and Nemirovski (1997), Tsybakov (1998), Klemelä and Tsybakov (2001), Chichignoud (2012). Recently, Goldenshluger and Lepski (2008) proposed for this problem a universal pointwise selection rule. Considering different collection of kernels, the associated procedures of estimation can adapt to the smoothness or the structure of the underlying function. In particular it can be proved that there exists an adaptive procedure of estimation that achieves the minimax rate of convergence (up to a multiplicative logarithmic factor) simultaneously over a scale of anisotropic Hölder spaces. Unfortunately the Goldenshluger-Lepski procedure does not take into account the "end point effect" (see § 4.1, third point) that appears in Lepski (1990), Tsybakov (1998), Chichignoud (2012) among others. In words, this phenomenon implies that, for the most regular space in the considered scale the extra log-term disappears. Hence a theoretical question arises: does there exist an estimator that outperforms the Goldenshluger-Lepski procedure at the final point of regularity?

We propose to answer this question by constructing a specific procedure of estimation that achieves this goal. The main idea is to follow the principle of Lepski's procedure : the estimators are compared pairwise using a criterion in order to select a bandwidth that realizes a bias-variance trade-off. In order to make these comparisons possible in an anisotropic setup, two different estimators are compared by introducing an auxiliary estimator. The same methodology was successfully used in Kerkyacharian, Lepski, and Picard (2001), Bertin (2005), Goldenshluger and Lepski (2008) among others. We finally prove that our procedure is optimal with respect to our criterion.

Our paper is organized as follows. In Section 2 we present our new definition of what should be an "optimal" procedure of estimation. Section 3 is devoted to the presentation of the statistical framework as well as the description of our procedure of estimation. The main results are stated in Section 4 and their proofs are given in Section 5. The proofs of technical or auxiliary results are gathered in the appendix.

2 Adaptive rates of convergence

2.1 Statistical framework

In this section, we consider a general statistical experiment $(\mathcal{Y}^\varepsilon, \mathcal{B}^\varepsilon, \{\mathbf{P}_f^\varepsilon\}_{f\in\Sigma})$ generated by an observation $\mathbf{Y}^\varepsilon$ where $\varepsilon > 0$ can be understood as the noise level. We address the problem of estimation of a functional G of f where G acts from Σ to some Banach space $(E, \|\cdot\|)$.

To measure the performance of an arbitrary estimator $\tilde{f}_\varepsilon$, with respect to a function $f \in \Sigma$, we consider the normalized risk

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, f, \psi_\varepsilon) = \left(\mathbf{E}_f^\varepsilon\left[\left(\psi_\varepsilon^{-1}\|G(\tilde{f}_\varepsilon) - G(f)\|\right)^q\right]\right)^{1/q},$$

where $q > 0$ is fixed. Here $\psi_\varepsilon > 0$ is a normalization and $\mathbf{E}_f^\varepsilon$ denotes the expectation with respect to $\mathbf{P}_f^\varepsilon$. We consider a collection of subspaces Σ_κ of Σ where the "nuisance" parameters κ belongs to some known set $\mathcal{K} \subset \mathbb{R}^m$. We then define, for any $\kappa \in \mathcal{K}$, the maximal normalized risk over Σ_κ by

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, \Sigma_\kappa, \psi_\varepsilon) = \sup_{f\in\Sigma_\kappa} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, f, \psi_\varepsilon).$$

For an arbitrary estimator $\tilde{f}_\varepsilon$ one can consider the family of normalizations $\{R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, \Sigma_\kappa, 1)\}_{\kappa\in\mathcal{K},\varepsilon>0}$ which measure the quality of this estimator with respect to the collection of spaces considered. In words, the quality of $\tilde{f}_\varepsilon$ is measured simultaneously over all Σ_κ, $\kappa \in \mathcal{K}$. The question we would like to answer in this section is the following: how to choose the "best" estimator or, equivalently, how to define the optimal family of normalizations?

An obvious answer can be done if there exists a single procedure of estimation f_ε^* such that

$$\limsup_{\varepsilon \to 0} R_\varepsilon^{(q)}(f_\varepsilon^*, \Sigma_\kappa, N_\varepsilon(\kappa)) < +\infty, \quad \forall \kappa \in \mathcal{K}$$

where $N_\varepsilon(\kappa) = \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, \Sigma_\kappa, 1)$ is the minimax rate of convergence over Σ_κ (the infimum is taken over all possible estimators). In what follows, the family of normalizations $\{N_\varepsilon(\kappa)\}_{\kappa \in \mathcal{K}}$ will be denoted by N.

Things turn out to be more difficult if no estimator can achieve the minimax rate of convergence simultaneously over each Σ_κ: we thus have to compare families of normalizations that can be viewed as functions of κ. Since, in general, no notion of "natural" order can be used in such a situation, this leads to introduce a criterion of optimality. Lepski (1990) and Tsybakov (1998) proposed two different options to solve this problem. Since both of them are not satisfactory in a general framework (see § 2.5 below) we propose a new definition in order to overcome the main drawbacks of these criteria.

2.2 Achievable families of normalizations

The first idea is to compare families of normalizations that can be "achieved" by a procedure of estimation.

Definition 1. *A family of normalizations $\Psi = \{\psi_\varepsilon(\kappa)\}_{\kappa \in \mathcal{K}}$ is called* achievable *if there exists a procedure of estimation $\hat{f}_\varepsilon^\Psi$ which satisfies the following* adaptive upper bound

$$\limsup_{\varepsilon \to 0} R_\varepsilon^{(q)}(\hat{f}_\varepsilon^\Psi, \Sigma_\kappa, \psi_\varepsilon(\kappa)) < +\infty, \quad \forall \kappa \in \mathcal{K}. \tag{AUB}$$

In words, $\hat{f}_\varepsilon^\Psi$ achieves the normalization $\psi_\varepsilon(\kappa)$ simultaneously over each space $\Sigma(\kappa)$ in the collection. Let us notice that, for such a family, we have the inequality $\psi_\varepsilon(\kappa) \gtrsim N_\varepsilon(\kappa)$ for all $\kappa \in \mathcal{K}$. (here $u_\varepsilon \gtrsim v_\varepsilon$ stands for $\liminf_{\varepsilon \to 0} u_\varepsilon(\kappa)/v_\varepsilon(\kappa) > 0$). Moreover, if the minimax rates of convergence is not achievable (see Lepski (1990), Tsybakov (1998) or corollary 2 for instance) there exists a nuisance parameter κ_0 such that

$$N_\varepsilon^{-1}(\kappa_0)\psi_\varepsilon(\kappa_0) \xrightarrow[\varepsilon\to 0]{} +\infty.$$

In such a situation, the family Ψ can be "improved" at least at point κ_0 using the minimax on $\Sigma(\kappa_0)$ estimator for all values of the nuisance parameter κ.

In this in mind we will use the following principle in order to give our notion of optimality : *the "optimal" achievable family of normalizations should have "small number" of points where it can be improved.*

2.3 Adaptive rate of convergence

In order to compare two achievable families of normalizations, namely Ψ and Φ, we introduce two subsets of $\mathcal{K}$, $\mathcal{K}(\Psi \ll \Phi)$ and $\mathcal{K}(\Psi \ggg \Phi)$ defined by

$$\mathcal{K}(\Psi \ll \Phi) = \left\{ \kappa \in \mathcal{K} : \quad \frac{\Psi_\varepsilon(\kappa)}{\Phi_\varepsilon(\kappa)} \xrightarrow[\varepsilon\to 0]{} 0 \right\}$$
$$\mathcal{K}(\Psi \ggg \Phi) = \left\{ \kappa \in \mathcal{K} : \quad \frac{\Psi_\varepsilon(\kappa)}{\Phi_\varepsilon(\kappa)} \frac{\Psi_\varepsilon(\ell)}{\Phi_\varepsilon(\ell)} \xrightarrow[\varepsilon\to 0]{} +\infty, \quad \forall \ell \in K(\Psi \ll \Phi) \right\}.$$

In other words, $\mathcal{K}(\Psi \ll \Phi)$ consists of all nuisance parameters where Φ can be outperformed by Ψ whereas $\mathcal{K}(\Psi \ggg \Phi)$ consists of all points κ where Φ is better than Ψ and, moreover, where the loss of Φ with respect to Ψ on $K(\Psi \ll \Phi)$ is "compensated".

Our principle of the choice between two achievable families is to compare the "massiveness" of these two sets. This leads to the following definition.

Definition 2. *Assume that $\mathcal{K} \subseteq \mathbb{R}^m$ contains an open set. A family of normalizations Φ is called* adaptive *(or the* adaptive rate of convergence*) if :*

- *It is an achievable family.*
- *If Ψ is another achievable family of normalizations then $\mathcal{K}(\Psi \ll \Phi)$ is contained in a $(m-1)$ manifold and $\mathcal{K}(\Psi \ggg \Phi)$ contains an open set of $\mathbb{R}^m$.*

2.4 Remarks

First, let us mention that this definition guarantees the uniqueness (in order) of the adaptive rate of convergence. Indeed, let us consider two optimal families of normalizations Φ and $\tilde{\Phi}$ and let us assume that there exists $\kappa_0 \in \mathcal{K}(\Phi \lll \tilde{\Phi})$. Since $\tilde{\Phi}$ is optimal, the set $\mathcal{K}(\Phi \ggg \tilde{\Phi})$ contains an open set of $\mathbb{R}^m$. Moreover, it is easily seen that $\mathcal{K}(\Phi \ggg \tilde{\Phi}) \subseteq \mathcal{K}(\tilde{\Phi} \lll \Phi)$. This is in contraction with the optimality of Φ which implies that $\mathcal{K}(\tilde{\Phi} \lll \Phi)$ is contained in a $(m-1)$-manifold. Thus $\mathcal{K}(\Phi \lll \tilde{\Phi})$ and $\mathcal{K}(\tilde{\Phi} \lll \Phi)$ (using a similar argument) are empty. This implies that for all $\kappa \in \mathcal{K}$ the rates $\phi_\varepsilon(\kappa)$ and $\tilde{\phi}_\varepsilon(\kappa)$ are of the same order.

Second, notice that if N is an achievable family, then it is the optimal one. Indeed for any achievable family Ψ the set $\mathcal{K}(\Psi \lll N)$ is empty.

Third, assume that the estimator $\hat{f}_\varepsilon^\Phi$ achieves the optimal family Φ. Assume also that Ψ is another achievable family of normalizations such that $\hat{f}_\varepsilon^\Psi$ achieves this family. Then $\hat{f}_\varepsilon^\Phi$ outperforms $\hat{f}_\varepsilon^\Psi$ in the following sense : the loss of $\hat{f}_\varepsilon^\Psi$ with respect to $\hat{f}_\varepsilon^\Phi$ is much bigger on the large set $\mathcal{K}(\Psi \ggg \Phi)$ than its gain on the small set $\mathcal{K}(\Phi \lll \Psi)$.

2.5 Comparison with previous notions

Lepski (1990) and Tsybakov (1998) already proposed optimality criteria. Nevertheless these criteria are not well-adapted to general statistical frameworks. Let us mention some drawbacks for both definitions.

In his paper, Lepski compares each achievable family of normalizations Ψ with respect to the family of the minimax rates through the following quantity

$$\Lambda_\varepsilon(\Psi) = \sup_{\kappa \in \mathcal{K}} \frac{\psi_\varepsilon(\kappa)}{N_\varepsilon(\kappa)}. \tag{1}$$

The optimal family of normalizations Φ is essentially defined as follows : for all achievable family Ψ the following property holds : $\Lambda_\varepsilon(\Phi) \lesssim \Lambda_\varepsilon(\Psi)$ (another technical condition implies the uniqueness in order of Φ).

Assume that Φ is such that for almost all $\kappa \in \mathcal{K}$ the ratio $\phi_\varepsilon(\kappa)/N_\varepsilon(\kappa)$ tends to infinity as ε goes to 0 and let us consider the family $\tilde{N}$ defined, for fixed α_0, by

$$\tilde{N}_\varepsilon(\alpha) = \begin{cases} N_\varepsilon(\alpha) & \text{if } \alpha \neq \alpha_0 \\ N_\varepsilon(\alpha_0)\Lambda_\varepsilon^2(\Phi) & \text{otherwise} \end{cases}$$

Using the criterion proposed by Lepski, it is impossible to prove that $\tilde{N}$ is not achievable. Nevertheless, one can think that a reasonable criterion should choose this family as the optimal one if it is achievable. This drawback comes from the supremum in (1) : this definition of optimality is in some sense too "global".

Tsybakov introduced another criterion in order to overcome this drawback when the set $\mathcal{K}$ is finite. The optimal family Φ is compared to all other achievable families Ψ and should satisfy the following property : if there exists a nuisance parameter κ_0 such that $\phi_\varepsilon^{-1}(\kappa_0)\psi_\varepsilon(\kappa_0) \to 0$ (that is $\hat{f}_\varepsilon^\Psi$ is better than $\hat{f}_\varepsilon^\Phi$ on $\Sigma(\kappa_0)$), then there exists a nuisance parameter κ_1 such that $(\phi_\varepsilon^{-1}(\kappa_0)\psi_\varepsilon(\kappa_0)) \cdot (\phi_\varepsilon^{-1}(\kappa_1)\psi_\varepsilon(\kappa_1)) \to +\infty$ (and thus $\hat{f}_\varepsilon^\Psi$ is worse than $\hat{f}_\varepsilon^\Phi$ even if a compensation by the previous gain is made). Unfortunately, the same problem can arise if $\mathcal{K}$ is not finite : if $\tilde{N}$ is achievable, for all $\alpha \neq \alpha_0$ we have $\phi_\varepsilon^{-1}(\alpha)\tilde{N}_\varepsilon(\alpha) \to 0$ on the one hand, and $(\phi_\varepsilon^{-1}(\alpha)\tilde{N}_\varepsilon(\alpha)) \cdot (\phi_\varepsilon^{-1}(\alpha_0)\tilde{N}_\varepsilon(\alpha_0)) \to +\infty$ on the other hand which is not in contradiction with the optimality of Φ. In this case the drawback comes from the fact that the same point $\kappa_1 = \alpha_0$ corresponds to different points $\kappa_0 = \alpha$: this definition is too "local".

Notice that our definition is more local than the Lepski's definition and more global than the Tsybakov's one. We will use this criterion of optimality in Theorem 1 below.

3 Statistical model and estimation procedure

3.1 Multivariate Gaussian white noise model

We consider a real valued multivariate function f defined on $\mathbb{R}^d$ (with $d \in \mathbb{N}^*$) that is observed in presence of an additive random noise. More precisely, we assume that the observation $\mathbf{Y}^\varepsilon = (Y_\varepsilon(u), u \in \mathcal{D})$ satisfies the equation

$$Y_\varepsilon(\mathrm{d}u) = f(u)\mathrm{d}u + \varepsilon W(\mathrm{d}u) \tag{2}$$

where $u = (u_1, \ldots, u_d)$ belongs to the open set $\mathcal{D} = (-1, 1)^d$, W is the standard Wiener process in $\mathbb{R}^d$ and $0 < \varepsilon < 1$ is the noise level.

3.2 Anisotropic Hölder spaces

Such a functional space $\mathbb{H}_d(\beta, L)$ is defined using two parameters: a vector $\beta = (\beta_1, \dots, \beta_d)$ that describes the smoothness of f in each direction of $\mathbb{R}^d$ and a real number $L > 0$ that corresponds to the radius of a ball in a functional vector space.

In what follows, since we have to measure the smoothness of f in the different directions of the ambient space $\mathbb{R}^d$ we consider for all function $g : \mathbb{R}^d \to \mathbb{R}$ the following notation: for $u \in \mathbb{R}^d$ and $i = 1, \dots, d$, we consider the function $g_i(\cdot|u)$ defined for all $\eta \in \mathbb{R}$ by $g_i(\eta|u) = g(u + \eta e_i)$ where e_i denotes the i^{th} vector of the canonical basis of $\mathbb{R}^d$. Let us now define formally $\mathbb{H}_d(\beta, L)$.

Definition 3. *Consider for all $i = 1, \dots, d$, the integer $\lfloor \beta_i \rfloor = \sup\{\ell \in \mathbb{N} : \ell < \beta_i\}$ and let us define $\mathbb{H}_d(\beta, L)$ as the set of all functions $g : \mathbb{R}^d \to \mathbb{R}$ such that, for all $i = 1, \dots, d$, the two following inequalities hold:*

$$\left\| g_i^{(k)}(\cdot|u) \right\|_\infty \leq L, \qquad k = 0 \dots, \lfloor \beta_i \rfloor, \quad u \in \mathbb{R}^d,$$

and

$$|g_i^{(\lfloor \beta_i \rfloor)}(\eta|u) - g_i^{(\lfloor \beta_i \rfloor)}(0|u)| \leq L|\eta|^{\beta_i - \lfloor \beta_i \rfloor}, \qquad u \in \mathbb{R}^d, \quad \eta \in \mathbb{R},$$

where $g_i^{(k)}(\cdot|u)$ denotes the k^{th} order derivative of $g_i(\cdot|u)$.

3.3 Risks and minimax rates

Since we are interested in pointwise estimation, we consider a fixed point $t \in \mathcal{D}$. Using the notations introduced in § 2.1 we consider $\Sigma = \mathcal{C}^0(\mathcal{D}, \mathbb{R})$, $G(f) = f(t)$ and $(E, \|\cdot\|) = (\mathbb{R}, |\cdot|)$. In words, the risk can be written as

$$R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, \mathbb{H}_d(\beta, L)) = \inf_{\tilde{f}_\varepsilon} \sup_{f \in \mathbb{H}_d(\beta, L)} \left(\mathbf{E}_f^\varepsilon \left| \tilde{f}_n(t) - f(t) \right|^q \right)^{1/q}.$$

Let us recall that, in this framework, the minimax rate of convergence over $\mathbb{H}_d(\beta, L)$ is well-known. Indeed, it can be proved that

$$\inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(q)}(\tilde{f}_\varepsilon, \mathbb{H}_d(\beta, L)) \asymp L^{1/(2\bar{\beta}+1)} \varepsilon^{2\bar{\beta}/(2\bar{\beta}+1)} =: N_\varepsilon(\beta, L)$$

where: the infimum is taken over all possible estimators; $\mathbf{E}_f^\varepsilon$ denotes the expectation with respect to the distribution $\mathbf{P}_f^\varepsilon$ of the observation $\mathbf{Y}^\varepsilon$ that satisfies (2); $\bar{\beta}$ is the "effective smoothness" of β which is defined by $\bar{\beta} = 1/(\sum \beta_i^{-1})$; the notation $u_\varepsilon \asymp v_\varepsilon$ stands for both $u_\varepsilon \lesssim v_\varepsilon$ and $u_\varepsilon \gtrsim u_\varepsilon$.

Remark also that this rate is achieved, for example, using a kernel estimator with properly chosen bandwidth $h^{(N)}$ defined by:

$$h_i^{(N)}(\beta, L) = \left(\frac{\|K\|_2 \, \varepsilon}{L} \right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i}}, \qquad i = 1, \ldots, d. \tag{3}$$

3.4 Procedure of estimation

In this section we fix $b = (b_1, \ldots, b_d) \in (\mathbb{R}_+^*)^d$ and $0 < L_* < L^*$. The procedure of estimation constructed in this paper depends on these parameters at least through the choice of a kernel of "order b". In the sequel of this section we present a finite collection of kernel estimators and our selection rule that consists in selecting (in a data-driven way) an estimator among all the estimators in this collection.

Collection of kernel estimators

Firstly we consider, for any $h = (h_1, \ldots, h_d) \in (\mathbb{R}_+^*)^d$, the multivariate kernel

$$K_h(u) = V_h^{-1} \prod_{i=1}^{d} \omega_i \left(\frac{u_i}{h_i} \right), \quad u \in \mathbb{R}^d$$

where $V_h = \prod_{i=1}^d h_i$ and $\omega_i : \mathbb{R} \to \mathbb{R}$ is a univariate kernel that satisfies the following classical assumptions:

(K1) The kernel ω_i belongs to $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$;

(K2) The following property holds:

$$\lambda_i^* = \int_{\mathbb{R}} |\omega_i(v)|(1 + |v|)^{b_i} \mathrm{d}v < +\infty.$$

(K3) The kernel ω_i is such that:

$$\int_{\mathbb{R}} \omega_i(v) v^\ell \mathrm{d}v = \delta_{0,\ell}, \qquad \text{for any } \ell = 0, \ldots, \lfloor b_i \rfloor.$$

For the sake of readability we denote $K = K_{(1,\dots,1)}$. Secondly, we construct a dyadic grid on the set of multivariate bandwidth. To do so, we introduce the following set of indices :

$$\mathcal{Z}_\varepsilon = \bigcup_{n=0}^{n_\varepsilon} \mathcal{Z}(n)$$

where n_ε is a positive integer that tends slowly to infinity as ε goes to 0 and

$$\mathcal{Z}(n) = \left\{ k \in \mathbb{Z}^d : \sum_{i=1}^d (k_i + 1) = n \text{ and } |k_i| \leq C_1 n + 1, \quad i = 1, \dots, d \right\}.$$

The explicit expressions for $n_\varepsilon = n_\varepsilon(b, L_*, L^*)$ and $C_1 = C_1(b)$ are given in § 5.1. We finally consider the collection $\{\hat{f}_k\}_{k \in \mathcal{Z}_\varepsilon}$ of kernel estimators defined by :

$$\hat{f}_k(t) = \int_{\mathbb{R}^d} K_{h^{(k)}}(t-u) Y_\varepsilon(\mathrm{d}u), \qquad k \in \mathcal{Z}_\varepsilon$$

where the bandwidth $h^{(k)}$ satisfies

$$h_i^{(k)} = h_i^* 2^{-k_i} \quad \text{with} \quad h_i^* = \left(\frac{\|K\|_2 \, \varepsilon}{L_*} \right)^{\frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i}}, \quad i = 1, \dots, d.$$

Bandwidth selection

For $k, \ell \in \mathcal{Z}_\varepsilon$ we define

$$|k| = \sum_{i=1}^d (k_i + 1) \quad \text{and} \quad k \wedge \ell = (k_1 \wedge \ell_1, \dots, k_d \wedge \ell_d).$$

We also define

$$\sigma_\varepsilon(k) = 2^{-d/2} \|K\|_2 \, \varepsilon \sqrt{\frac{2^{|k|}}{\prod_{i=1}^d h_i^*}} \quad \text{and} \quad S_\varepsilon(k) = \sigma_\varepsilon(k) \sqrt{1 + |k| \log 2}.$$

Equipped with these definitions we introduce the random set of “admissible” multivariate indices

$$\mathcal{A} = \left\{ k \in \mathcal{Z}_\varepsilon : |\hat{f}_{k \wedge \ell}(t) - \hat{f}_\ell(t)| \leq C^* S_\varepsilon(\ell), \quad \forall \ell \in \mathcal{Z}_\varepsilon \text{ s.t. } \sigma_\varepsilon(\ell) \geq \sigma_\varepsilon(k) \right\}$$

where the constant C^* is defined in § 5.1. We finally define $\hat{f}$ by :

$$\hat{f}(t) = \begin{cases} \hat{f}_{\hat{k}}(t) & \text{if } \mathcal{A} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \quad \text{where} \quad \hat{k} = \arg\min_{k \in \mathcal{A}} \sigma_\varepsilon(k).$$

4 Main result

Set $\mathcal{B} = \prod_{i=1}^{d}(0, b_i]$ and $\mathcal{L} = [L_*, L^*]$ and let us define $\mathcal{K} = \mathcal{B} \times \mathcal{L}$. We consider the family of normalizations Φ defined by :

$$\phi_\varepsilon(\beta, L) = L^{\frac{1}{2\bar{\beta}+1}} \left(\varepsilon \sqrt{1 + 2 \log \frac{N_\varepsilon(\beta, L)}{N_\varepsilon(b, L_*)}} \right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1}}, \quad (\beta, L) \in \mathcal{K}.$$

We also define, for $\gamma > 0$ the following set :

$$\mathcal{L}(\gamma) = \{(\beta, L) \in \mathcal{K} : \bar{\beta} = \gamma\}.$$

Theorem 1. *Let Ψ be an achievable family of normalisations such that there exists $(\beta_0, L_0) \in \mathcal{K}(\Psi \ll \Phi)$. Then*

$$\mathcal{K}(\Psi \ll \Phi) \subseteq \mathcal{L}(\bar{\beta}_0) \quad \text{and} \quad \bigcup_{\gamma > \bar{\beta}_0} \mathcal{L}(\gamma) \subseteq \mathcal{K}(\Psi \ggg \Phi).$$

Theorem 2. *The estimator $\hat{f}$ constructed in § 3.4 satisfies the following property : for all $0 < \varepsilon < L_* / \|K\|_2$,*

$$\sup_{(\beta,L) \in \mathcal{B} \times \mathcal{L}} R_\varepsilon^{(q)}(\hat{f}, \mathbb{H}_d(\beta, L), \phi_\varepsilon(\beta, L)) \leq M_q$$

where $M_q = M_q(b, L_*, L^*)$ is an absolute constant.

As direct consequences of Theorems 1 and 2, let us mention the following results :

Corollary 1. *The family of normalizations Φ is the adaptive rate of convergence.*

Corollary 2. *The family of normalizations N as well as the family $\tilde{N}$ defined in § 2.5 are not achievable.*

4.1 Remarks

Firstly, remark that the result stated in Theorem 1 can not be improved. Indeed, Goldenshluger and Lepski (2008) proved that for all $\gamma \in (0, \bar{b}]$, there exists an estimator that achieves a rate of order

$$\left(\varepsilon \sqrt{\log |\log \varepsilon|} \right)^{2\gamma/(2\gamma+1)}$$

over the union of all $\mathbb{H}_d(\beta, L)$ with $\beta \in \mathcal{L}(\gamma)$. This implies that their estimator achieves a family of normalizations Ψ which satisfies $\mathcal{K}(\Psi \ll \Phi) = \mathcal{L}(\gamma)$.

Secondly, remark that the result stated in Theorem 2 is stronger than (AUB) since the constant M_q does not depend on the nuisance parameter (β, L). Indeed, this implies that

$$\limsup_{\varepsilon \to 0} \sup_{(\beta,L)\in\mathcal{K}} R_\varepsilon^{(q)}(\hat{f}, \mathbb{H}_d(\beta, L), \phi_\varepsilon(\beta, L)) < +\infty.$$

Thirdly, let us remark that the procedure constructed by Goldenshluger and Lepski (2008) achieves the rate Ψ defined by

$$\psi_\varepsilon(\beta, L) = L^{\frac{1}{2\bar{\beta}+1}} \left(\varepsilon \sqrt{1 + \log \prod_{i=1}^{d} \frac{h_i^{\max}}{h_i^{(N)}(\beta, L)}} \right)^{\frac{2\bar{\beta}}{2\bar{\beta}+1}},$$

where $h_i^{(N)}(\beta, L) \in (0, h_i^{\max}]$ is defined in (3) and the bandwidth $h^{\max}$ is a tuning parameter of the Goldenshluger-Lepski procedure. Thus in order to obtain the end-point effect we have to choose $h^{\max} = h^*$. However, if $d = 2$, the choices $b = (1, 1)$ and $\beta = (a, 1)$ with $a < 1$ lead to $h_2(\beta, L_*) > h_2^*$. This implies that the procedure is not optimal over the whole scale of anisotropic Hölder spaces $\mathbb{H}_2(\beta, L)$ for $\beta \in (0, 1]^2$. In our procedure $h^{\max}$ and h^* can be different (recall that $k_i \in \mathbb{Z}$). This allows us to remove this limitation.

Fourthly, let us briefly explain the idea behind the construction of $\mathcal{A}$. Assume that the unknown function f belongs to $\mathbb{H}_d(\beta, L)$ and that there is no noise (that is, we put formally $\varepsilon = 0$). Then for all $k \in \mathcal{Z}_\varepsilon$, we have $\hat{f}_k = \mathbf{E}_f^\varepsilon \hat{f}_k$. This leads, using Lemma 3, to

$$|\hat{f}_{k\wedge\ell}(t) - \hat{f}_\ell(t)| = |\mathbf{E}_f^\varepsilon \hat{f}_{k\wedge\ell} - \mathbf{E}_f^\varepsilon \hat{f}_\ell| \leq 2\mathfrak{B}_{\beta,L}(h^{(k)}).$$

Using Lemma 4 and the notations introduced in § 5.1 it is easily seen that there exists a bandwidth $h^{(k(\beta,L))}$ on the geometrical grid such that :

$$\mathfrak{B}_{\beta,L}(h^{(k(\beta,L))}) \leq 2C_2 S_\varepsilon(k(\beta, L)).$$

Assuming that $\sigma_\varepsilon(\ell) \geq \sigma_\varepsilon(k(\beta, L))$ (or in an equivalent way that $|\ell| \geq |k(\beta, L)|$), we obtain that $S_\varepsilon(\ell) \geq S_\varepsilon(k(\beta, L))$. We then deduce that, under this assumption,

$$\mathfrak{B}_{\beta,L}(h^{(k(\beta,L))}) \leq 2C_2 S_\varepsilon(\ell),$$

that is, $k(\beta, L) \in \mathcal{A}$ as soon as $C^* \geq 2C_2$. Remark also that the relation $\sigma_\varepsilon(\ell) \geq \sigma_\varepsilon(k(\beta, L))$ clearly induces a partial ordering on the set $\mathcal{Z}_\varepsilon$.

Let us finally comment the introduction of the set of indices $\mathcal{Z}_\varepsilon$. To obtain the "end-point effect" the construction of a geometric grid of bandwidth is based on $h^* = (h_1^*, \dots, h_d^*)$ which allows our estimator to achieve the minimax rate of convergence over $\mathbb{H}_d(b, L_*)$. This implies that some $k(\beta, L)$, that satisfied (4), are not in $\mathbb{N}^d$. Moreover, $\mathcal{Z}_\varepsilon$ is constructed such that the quantities that appear in (6) and (11) are finite. This requires that $\mathcal{Z}_\varepsilon$ is "small enough".

5 Proof

5.1 Basic notations

In this section we introduce some notations and define some constants that will be used throughout the rest of the paper.

First, we consider the partial ordering on $\mathcal{Z}_\varepsilon$ defined by :

$$k \prec \ell \iff \sigma_\varepsilon(k) \leq \sigma_\varepsilon(\ell).$$

Next, for $(\beta, L) \in \mathcal{K}$ we define the following quantities :

$$h_i(\beta, L) = \left(\frac{\|K\|_2\, \varepsilon}{L} \sqrt{1 + 2\log \frac{N_\varepsilon(\beta, L)}{N_\varepsilon(b, L_*)}} \right)^{\frac{2\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}, \qquad i = 1, \dots, d$$

and $k_i(\beta, L)$ be such that :

$$2^{k_i(\beta,L)} \leq h_i^*/h_i(\beta, L) < 2^{k_i(\beta,L)+1}. \tag{4}$$

we finally consider the following quantites :

$$n_\varepsilon = \left\lfloor \frac{2}{\log 2} \left\{ \frac{2\bar b}{2\bar b+1} \log\left(\frac{L_*}{\|K\|_2\, \varepsilon} \right) + \log\left(\frac{L^* \vee 1}{L_*} \right) \right\} \right\rfloor + d,$$

$$C_1 = C_1(b) = \left(\frac{2\bar b+1}{2\bar b} \times \frac{\log(2) + \sqrt{2\log(2)}}{\log(2)} \right),$$

$$\lambda^* = \lambda^*(K) = \sum_{i=1}^{d} 2^{b_i} \lambda_i^* \quad \text{and} \quad C_2 = 2^{3/2}\lambda^*.$$

and

$$C^* = 2(C_2 + \sqrt{6q+4}).$$

5.2 Auxiliary results

Lemma 1. *Set* (α, L_α) *and* (β, L_β) *in* $\mathcal{K}$. *Assume that* $\bar{\alpha} < \bar{\beta}$ *and let* ν *be a positive constant such that* $\nu < 2(\bar{\beta} - \bar{\alpha})/[(2\bar{\alpha}+1)(2\bar{\beta}+1)]$. *Set* $\tilde{f}$ *be an arbitrary estimator. Then*

$$\mathcal{R}_\varepsilon^{(q)}(\tilde{f}) = R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\alpha, L_\alpha), \phi_\varepsilon(\alpha, L_\alpha)) + R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\beta, L_\beta), \varepsilon^{-\nu}\phi_\varepsilon(\beta, L_\beta))$$

is such that $\liminf_{\varepsilon \to 0} \mathcal{R}_\varepsilon^{(q)}(\tilde{f}) > 0$.

Lemma 2. *For any* $(\beta, L) \in \mathcal{K}$, *the multi-index* $k(\beta, L)$ *belongs to* $\mathcal{Z}_\varepsilon$.

Lemma 3. *Set* $(\beta, L) \in \mathcal{B} \times \mathcal{L}$ *and* $f \in \mathbb{H}_d(\beta, L)$. *Let us denote*

$$\lambda_i(\beta) = \int_{\mathbb{R}^d} |K(u)| \frac{|u_i|^{\beta_i}}{\lfloor \beta_i \rfloor !} \mathrm{d}u,$$

and

$$\mathfrak{B}_{\beta,L}(h) = L \sum_{i=1}^{d} \lambda_i(\beta) h^{\beta_i}.$$

The following properties hold, for all $(k, \ell) \in \mathcal{Z}_\varepsilon^2$ *:*

$$\begin{cases} |\mathbf{E}_f^\varepsilon \hat{f}_k - f(t)| \leq \mathfrak{B}_{\beta,L}(h^{(k)}) \\ |\mathbf{E}_f^\varepsilon \hat{f}_{k \wedge \ell} - \mathbf{E}_f^\varepsilon \hat{f}_\ell| \leq 2\mathfrak{B}_{\beta,L}(h^{(k)}) \end{cases}$$

Lemma 4. *Set* $(\beta, L) \in \mathcal{B} \times \mathcal{L}$. *There exists a positive constant* C_2, *that depends only on* K *and* b, *such that the following inequalities hold :*

$$\mathfrak{B}_{\beta,L}(h^{(k(\beta,L))}) \leq C_2 S_\varepsilon(k(\beta, L)).$$

Lemma 5. *Set* $(\beta, L) \in \mathcal{B} \times \mathcal{L}$ *and* $f \in \mathbb{H}_d(\beta, L)$. *Let* κ, *and* $\kappa(s)$, *for* $s \in \mathbb{N}$, *be defined as in Section 5.4. Then*

$$\mathbf{P}_f^\varepsilon(\kappa(s) \notin \mathcal{A}) \leq 2 \sum_{\substack{\ell \in \mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s)}} 2^{-\delta|\ell|},$$

where $\delta = 3q + 2$.

Lemma 6. *Set* $(\beta, L) \in \mathcal{B} \times \mathcal{L}$ *and* $f \in \mathbb{H}_d(\beta, L)$. *For* $s \in \mathbb{N}$, *let* $B(s)$ *and* $\kappa(s)$ *be defined as in Section 5.4. For all* $p > 0$ *there exists an absolute constant* A_p *such that*

$$\mathbf{E}_f^\varepsilon \left(\left| \hat{f}_{\hat{k}}(t) - f(t) \right|^p \mathbf{1}_{\hat{k} \in B(s)} \right) \leq A_p S_\varepsilon^p(\kappa(s)).$$

5.3 Proof of Theorem 1

First, remark that Lemma 1 implies the following property: set (α, L_α) and (β, L_β) in $\mathcal{K}$ be such that $\bar{\alpha} < \bar{\beta}$ and assume that Ψ is an achievable family of normalizations that satisfies

$$\frac{\psi_\varepsilon(\alpha, L_\alpha)}{\Phi_\varepsilon(\alpha, L_\alpha)} \underset{\varepsilon \to 0}{\longrightarrow} 0.$$

Then

$$\frac{\psi_\varepsilon(\alpha, L_\alpha)}{\Phi_\varepsilon(\alpha, L_\alpha)} \frac{\psi_\varepsilon(\beta, L_\beta)}{\Phi_\varepsilon(\beta, L_\beta)} \underset{\varepsilon \to 0}{\longrightarrow} +\infty. \tag{5}$$

Indeed, let us consider an estimator $\hat{f}$ that achieves Ψ (that is, it satisfies (AUB) with Ψ). Then:

$$\begin{aligned}\mathcal{R}_\varepsilon^{(q)}(\hat{f}) = &\frac{\psi_\varepsilon(\alpha, L_\alpha)}{\phi_\varepsilon(\alpha, L_\alpha)} R_\varepsilon^{(q)}(\hat{f}, \mathbb{H}_d(\alpha, L_\alpha), \psi_\varepsilon(\alpha, L_\alpha)) \\ &+ \varepsilon^\nu \frac{\phi_\varepsilon(\alpha, L_\alpha)}{\psi_\varepsilon(\alpha, L_\alpha)} \times \frac{\psi_\varepsilon(\alpha, L_\alpha)}{\phi_\varepsilon(\alpha, L_\alpha)} \frac{\psi_\varepsilon(\beta, L_\beta)}{\phi_\varepsilon(\beta, L_\beta)} R_\varepsilon^{(q)}(\hat{f}, \mathbb{H}_d(\beta, L_\beta), \psi_\varepsilon(\beta, L_\beta)).\end{aligned}$$

Since $\hat{f}$ achieves Ψ, the first term of the right hand side of this inequality tends to 0 as ε goes to 0. This implies that:

$$\liminf_{\varepsilon \to 0} \varepsilon^\nu \frac{\phi_\varepsilon(\alpha, L_\alpha)}{\psi_\varepsilon(\alpha, L_\alpha)} \times \frac{\psi_\varepsilon(\alpha, L_\alpha)}{\phi_\varepsilon(\alpha, L_\alpha)} \frac{\psi_\varepsilon(\beta, L_\beta)}{\phi_\varepsilon(\beta, L_\beta)} > 0.$$

This property combined with the fact that

$$\frac{\phi_\varepsilon(\alpha, L_\alpha)}{\psi_\varepsilon(\alpha, L_\alpha)} \lesssim \frac{\phi_\varepsilon(\alpha, L_\alpha)}{N_\varepsilon(\alpha, L_\alpha)} = |\log \varepsilon|^{1/(2\bar{\alpha}+1)}$$

leads immediately to (5). Note that this remark readily implies that (β, L_β) belongs to $\mathcal{K}(\Psi \ggg \Phi)$ as soon as (β_0, L_0) belongs to $\mathcal{K}(\Psi \ll \Phi)$ and $\bar{\beta} > \bar{\beta}_0$ (here β_0 plays the role of α).

Next, let us complete the proof of Theorem 1. Assume that (β_0, L_0) belongs to $\mathcal{K}(\Psi \ll \Phi)$ and set (α, L_α) be such that:

$$\frac{\psi_\varepsilon(\alpha, L_\alpha)}{\phi_\varepsilon(\alpha, L_\alpha)} \underset{\varepsilon \to 0}{\longrightarrow} 0.$$

Using our remark we deduce that $\bar{\alpha} \geq \bar{\beta}_0$. Indeed otherwise we obtain (5) with β_0 playing the role of β which is in contradiction with $(\beta_0, L_0) \in \mathcal{K}(\Psi \ll \Phi)$.

Combining these results Theorem 1 follows easily.

5.4 Proof of Theorem 2

Set $0 < \varepsilon < \|K\|_2/L_*$, $(\beta, L) \in \mathcal{K}$ and $f \in \mathbb{H}_d(\beta, L)$. For the sake of readability we denote $\kappa = k(\beta, L)$.

First, we assume that $\mathcal{A}$ is a non-empty set and we consider the following decomposition of $\mathcal{Z}_\varepsilon$:

$$\mathcal{Z}_\varepsilon = B(0) \uplus \left(\biguplus_{s\geq 1} B(s) \cap \bar{B}(s-1) \right)$$

where $B(s) = \{k \in \mathcal{Z}_\varepsilon : |k| \leq |\kappa| + sd\}$ and $\bar{B}(s)$ is its complementary into $\mathcal{Z}_\varepsilon$. This leads to the decomposition of the risk that follows :

$$\mathbf{E}_f^\varepsilon \left(\left| \hat{f}(t) - f(t) \right|^q \mathbf{1}_{\mathcal{A}\neq\emptyset} \right) \leq R(0, q) + \sum_{s\geq 1} \sqrt{R(s, 2q)\tilde{D}(s)},$$

where for all $p > 0$ we define $R(s, p) = \mathbf{E}_f^\varepsilon[|\hat{f}_{\hat{k}}(t) - f(t)|^p \mathbf{1}_{\hat{k}\in B(s)}]$ and $\tilde{D}(s) = \mathbf{P}_f^\varepsilon(\hat{k} \notin B(s-1))$.

Let us now consider the point $\kappa(s) \in B(s) \cap \bar{B}(s-1)$ defined by $\kappa_i(s) = \kappa_i + s$. This point is closely related to κ since it is linked with κ in an *isotropic* way. Moreover note that, thanks to the definition of $\hat{k}$, if $\hat{k} \notin B(s-1)$, then $\kappa(s-1) \notin \mathcal{A}$. This leads to

$$\mathbf{E}_f^\varepsilon \left(\left| \hat{f}(t) - f(t) \right|^q \mathbf{1}_{\mathcal{A}\neq\emptyset} \right) \leq R(0, q) + \sum_{s\geq 1} \sqrt{R(s, 2q)D(s)},$$

where $D(s) = \mathbf{P}_f^\varepsilon(\kappa(s-1) \notin \mathcal{A})$. Lemmas 5 and 6 provide upper bound for both $R(s, p)$ and $D(s)$. Using these results we obtain :

$$\begin{aligned}
(*) &= \mathbf{E}_f^\varepsilon \left(\left| \hat{f}(t) - f(t) \right|^q \mathbf{1}_{\mathcal{A}\neq\emptyset} \right) \\
&\leq A_q S_\varepsilon^q(\kappa) + (2A_{2q})^{1/2} \sum_{s\geq 1} S_\varepsilon^q(\kappa(s)) \sqrt{\sum_{\substack{\ell\in\mathcal{Z}_\varepsilon \\ \ell\succ\kappa(s-1)}} 2^{-\delta|\ell|}} \\
&\leq A_q S_\varepsilon^q(\kappa) + (2A_{2q})^{1/2} \sum_{s\geq 1} S_\varepsilon^q(\kappa(s)) \sum_{\substack{\ell\in\mathcal{Z}_\varepsilon \\ \ell\succ\kappa(s-1)}} 2^{-\delta|\ell|/2},
\end{aligned}$$

where the quantities A_p and δ are defined in these lemmas. Moreover, it is easily seen that

$$S_\varepsilon^q(\kappa(s)) = S_\varepsilon^q(\kappa) \left(\frac{1 + |\kappa(s)| \log 2}{1 + |\kappa| \log 2} \right)^{q/2} 2^{q(|\kappa(s)|-|\kappa|)}.$$

This implies, using that $\delta = 3q+2$, the following inequality

$$(*) \leq A_q S_\varepsilon^q(\kappa) + (2A_{2q})^{1/2} S_\varepsilon^q(\kappa) \sum_{s\geq 1} \sum_{\substack{\ell\in\mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} \left(\frac{1+|\kappa(s)|\log 2}{2^{|\ell|}(1+|\kappa|\log 2)} \right)^{q/2} 2^{-|\ell|}$$

$$\leq S_\varepsilon^q(\kappa) \left(A_q + (2^{dq+1}A_{2q})^{1/2} \sum_{s\geq 1} \sum_{\substack{\ell\in\mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} 2^{-|\ell|} \right).$$

Now, let us remark that

$$\sum_{s\geq 1} \sum_{\substack{\ell\in\mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} 2^{-|\ell|} \leq \sum_{s\geq 0} \sum_{n\geq s} \#\mathcal{Z}(n) 2^{-n}$$

$$= \sum_{n\geq 0} (n+1) \#\mathcal{Z}(n) 2^{-n}. \tag{6}$$

Thanks to our definition of $\mathcal{Z}(n)$ the last quantity is finite and does not depend on ε. Thus, there exists an absolute constant $M_{1,q}$ (that depends also on b, L_* and L^*) such that

$$\mathbf{E}_f^\varepsilon \left(\left| \hat{f}(t) - f(t) \right|^q \mathbf{1}_{\mathcal{A}\neq\emptyset} \right) \leq M_{1,q} S_\varepsilon^q(\kappa) \leq (\sqrt{d+1}\, \|K\|_2^{\frac{2\bar{b}}{2\bar{b}+1}})^q M_{1,q} \phi_\varepsilon^q(\beta, L).$$

Last inequality follows from Lemma 4.

Next, we assume that $\mathcal{A}$ is empty. Let us denote $s^* = \max\{s \in \mathbb{N} : |\kappa(s)| \leq n_\varepsilon\}$ and remark that $\mathcal{A}$ is empty implies that $\kappa(s^*) \notin \mathcal{A}$. Thus:

$$\mathbf{P}_f^\varepsilon(\mathcal{A} = \emptyset) \leq \mathbf{P}_f^\varepsilon(\kappa(s^*) \notin \mathcal{A}).$$

Using Lemma 5 and similar arguments as previously we obtain that there exists a constant $M_{2,q}$ (that depends also on b, L_* and L^*) such that

$$\mathbf{P}_f^\varepsilon(\mathcal{A} = \emptyset) \leq 2 \sum_{\ell \succ \kappa(s^*)} 2^{-\delta|\ell|}$$

$$\leq M_{2,q} 2^{-n_\varepsilon}.$$

Using the definition of n_ε, it can be easily proved that:

$$\mathbf{P}_f^\varepsilon(\mathcal{A} = \emptyset) \leq M_{2,q}\sigma_\varepsilon(0) \leq M_{2,q} S_\varepsilon(\kappa).$$

This leads to the following inequality:

$$\begin{aligned}
\mathbf{E}_f^\varepsilon \left(\left| \hat{f}(t) - f(t) \right|^q \mathbf{1}_{\mathcal{A}=\emptyset} \right) &= \mathbf{E}_f^\varepsilon \left(|f(t)|^q \mathbf{1}_{\mathcal{A}=\emptyset} \right) \\
&\leq L^q \mathbf{P}_f^\varepsilon(\mathcal{A} = \emptyset) \\
&\leq L^q M_{2,q} S_\varepsilon^q(\kappa) \leq (\sqrt{d+1}\, \|K\|_2^{\frac{2\bar{b}}{2\bar{b}+1}}\, L^*)^q M_{2,q} \phi_\varepsilon^q(\beta, L).
\end{aligned}$$

A Proof of lemma 1

First, remark that we can assume that $L_\alpha = L_\beta = L_*$. Indeed, since $\mathbb{H}_d(\alpha, L_*) \subset \mathbb{H}_d(\alpha, L_\alpha)$, if $\tilde{f}$ is an arbitrary estimator:

$$\begin{aligned}\mathcal{R}_\varepsilon^{(q)}(\tilde{f}) &\geq \frac{\phi_\varepsilon(\alpha, L_*)}{\phi_\varepsilon(\alpha, L_\alpha)} R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\alpha, L_*), \phi_\varepsilon(\alpha, L_*)) \\ &\quad + \frac{\phi_\varepsilon(\beta, L_*)}{\phi_\varepsilon(\beta, L_\beta)} R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\beta, L_*), \phi_\varepsilon(\beta, L_*)) \\ &\geq \frac{L_*}{L^*} \tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}),\end{aligned}$$

where

$$\tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}) = R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\alpha, L_*), \phi_\varepsilon(\alpha, L_*)) + R_\varepsilon^{(q)}(\tilde{f}, \mathbb{H}_d(\beta, L_*), \phi_\varepsilon(\beta, L_*)).$$

In what follows, for simplicity we will denote $L = L_*$. Set $\varkappa$ a positive parameter to be chosen. We consider $h_i = h_i(\varepsilon)$ defined by the formula

$$h_i = \left(\varkappa \varepsilon \sqrt{|\log \varepsilon|}\right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1} \frac{1}{\alpha_i}},$$

and we consider the two following functions:

$$\begin{cases} f_0 \equiv 0 \\ f_1(x) = L \varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} \varphi_\varepsilon(\alpha) f\left(\frac{x_1 - t_1}{h_1}, \ldots, \frac{x_d - t_d}{h_d}\right) \end{cases}$$

where f belongs to $\mathbb{H}_d(\alpha, 1)$. Hence f_1 belongs to $\mathbb{H}_d(\alpha, L)$ and, if $\mathbf{E}_0$ and $\mathbf{E}_1$ denote respectively $\mathbf{E}^n_{f_0}$ and $\mathbf{E}^n_{f_1}$, we have:

$$\begin{aligned}\tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_0 \left|\varepsilon^\nu \varphi_\varepsilon^{-1}(\beta) \tilde{f}_\varepsilon(t)\right|^q + \mathbf{E}_1 \left|\varphi_\varepsilon^{-1}(\alpha)(\tilde{f}_\varepsilon(t) - f_1(t))\right|^q \\ &\geq \mathbf{E}_0 \left|\varepsilon^\nu \varphi_\varepsilon^{-1}(\beta) \tilde{f}_\varepsilon(t)\right|^q + \mathbf{E}_1 \left|\varphi_\varepsilon^{-1}(\alpha) \tilde{f}_\varepsilon(t) - z\right|^q,\end{aligned}$$

where z denotes $\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} f(0)$ and $\varphi_\varepsilon(\delta)$ stands for $\phi_\varepsilon(\delta, L)$ $(\delta = \alpha, \beta)$. For ease of exposition, we consider the following notations:

$$\lambda_\varepsilon = \varepsilon^\nu \frac{\varphi_\varepsilon(\alpha)}{\varphi_\varepsilon(\beta)} = \varepsilon^\nu \left(\varepsilon \sqrt{|\log \varepsilon|}\right)^{-\varrho} \text{ where } \varrho = \frac{2(\bar{\beta} - \bar{\alpha})}{(2\bar{\beta}+1)(2\bar{\alpha}+1)}.$$

and

$$\tilde{\theta} = \varphi_\varepsilon^{-1}(\alpha) |\tilde{f}_\varepsilon(t)|.$$

Using these notations, we have :

$$\tilde{\mathcal{R}}_{\varepsilon}^{(q)}(\tilde{f}_{\varepsilon}) \geq \mathbf{E}_0 \left|\lambda_{\varepsilon}\tilde{\theta}\right|^q + \mathbf{E}_1 \left|\tilde{\theta} - z\right|^q$$

By changing the probability measure, we obtain :

$$\tilde{\mathcal{R}}_{\varepsilon}^{(q)}(\tilde{f}_{\varepsilon}) \geq \mathbf{E}_1 \left[\left|\lambda_{\varepsilon}\tilde{\theta}\right|^q Z_{\varepsilon} + \left|\tilde{\theta} - z\right|^q\right]$$

where Z_{ε} denotes the classical likelyhood ratio :

$$\begin{aligned} Z_{\varepsilon} &= \frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}_1}(\mathbf{Y}^{\varepsilon}) \\ &= \exp\left(-\frac{1}{\varepsilon}\int_{\mathbb{R}^d} f_1(u)dW(u) - \frac{1}{2\varepsilon^2}\|f_1\|_2^2\right) \\ &= \exp\left(-\frac{\|f_1\|_2}{\varepsilon}\xi - \frac{1}{2}\left(\frac{\|f_1\|_2}{\varepsilon}\right)^2\right). \end{aligned}$$

In the last line, $\xi \sim \mathcal{N}(0,1)$. Now we consider, for $\delta > 0$ and $a > 0$ the events

$$\tilde{\Lambda}_{\delta} = \left\{|\tilde{\theta}| > \delta\right\} \quad \text{and} \quad \Lambda_a = \{|\xi| \leq a\}.$$

Clearly, if δ is small enough, denoting by $\mathbf{1}$ the characteristic function of a set, we have :

$$\tilde{\mathcal{R}}_{\varepsilon}^{(q)}(\tilde{f}_{\varepsilon}) \geq \mathbf{E}_1 \left[(\delta\lambda_{\varepsilon})^q Z_{\varepsilon}\mathbf{1}_{\tilde{\Lambda}_{\delta}} + (z-\delta)^q \mathbf{1}_{\tilde{\Lambda}_{\delta}^c}\right].$$

Moreover, it can be easily seen that

$$\begin{aligned} Z_{\varepsilon}\mathbf{1}_{\Lambda_a} &\geq \exp\left(-\frac{\|f_1\|}{\varepsilon}a - \frac{1}{2}\frac{\|f_1\|^2}{\varepsilon^2}\right) \\ &\geq \exp\left(-\frac{1}{2}\left(\frac{\|f_1\|}{\varepsilon} + a\right)^2\right) \end{aligned}$$

that implies the following inequalities

$$\begin{aligned} \tilde{\mathcal{R}}_{\varepsilon}^{(q)}(\tilde{f}_{\varepsilon}) &\geq \mathbf{E}_1 \left[(\delta\lambda_{\varepsilon})^q Z_{\varepsilon}\mathbf{1}_{\tilde{\Lambda}_{\delta}\cap\Lambda_a} + (z-\delta)^q \mathbf{1}_{\tilde{\Lambda}_{\delta}^c}\right] \\ &\geq \mathbf{E}_1 \left[(\delta\lambda_{\varepsilon})^q \exp\left(-\frac{1}{2}\left(\frac{\|f_1\|}{\varepsilon} + a\right)^2\right)\mathbf{1}_{\tilde{\Lambda}_{\delta}\cap\Lambda_a} + (z-\delta)^q \mathbf{1}_{\tilde{\Lambda}_{\delta}^c}\right]. \end{aligned}$$

Let us point out that $\|f_1\|/\varepsilon = L\varkappa\|f\|\sqrt{|\log\varepsilon|}$. Thus, choosing $a = L\varkappa\|f\|\sqrt{|\log\varepsilon|}\wedge 1$, we obtain :

$$
\begin{aligned}
\tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_1\left[(\delta\lambda_\varepsilon)^q \exp\left(-\left(\frac{\|f_1\|}{\varepsilon}\right)^2\right)\mathbf{1}_{\tilde{\Lambda}_\delta\cap\Lambda_a} + (z-\delta)^q\mathbf{1}_{\tilde{\Lambda}_\delta^c}\right] \\
&\geq \mathbf{E}_1\left[(\delta\lambda_\varepsilon)^q\varepsilon^{(\varkappa\|f\|)^2}\mathbf{1}_{\tilde{\Lambda}_\delta\cap\Lambda_a} + (z-\delta)^q\mathbf{1}_{\tilde{\Lambda}_\delta^c}\right] \\
&\geq \mathbf{E}_1\left[(\delta\eta_\varepsilon)^q\varepsilon^{q(\nu-\varrho)+(\varkappa\|f\|)^2}\mathbf{1}_{\tilde{\Lambda}_\delta\cap\Lambda_a} + (z-\delta)^q\mathbf{1}_{\tilde{\Lambda}_\delta^c}\right],
\end{aligned}
$$

where $\eta_\varepsilon = (|\log\varepsilon|)^{-\varrho/2}$. Now, let us introduce

$$
t_\varepsilon = \frac{q}{|\log\varepsilon|}\left(|\log(\delta\eta_\varepsilon)| + \log ALf(0)\right) \to 0,
$$

whith

$$
A = \left(\frac{\sqrt{q(\varrho-\nu)}}{L\|f\|}\right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}.
$$

Let us choose :

$$
\varkappa = \frac{\sqrt{q(\varrho-\nu)-t_\varepsilon}}{L\|f\|}.
$$

Using these notations, we obtain

$$
(L\varkappa\|f\|)^2 = q\varrho - t_\varepsilon \qquad \text{and} \qquad (\delta\eta_\varepsilon)^q\varepsilon^{-q\varrho+L\varkappa\|f\|} = (ALf(0))^q
$$

This implies that

$$
\begin{aligned}
\tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}_\varepsilon) &\geq \mathbf{E}_1\left[(ALf(0))^q\mathbf{1}_{\tilde{\Lambda}_\delta\cap\Lambda_a} + (L\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}f(0)-\delta)^q\mathbf{1}_{\tilde{\Lambda}_\delta^c}\right] \\
&\geq \mathbf{E}_1\left[(L\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}f(0)-\delta)^q\left(\mathbf{1}_{\tilde{\Lambda}_\delta+\tilde{\Lambda}_\delta^c}\right)\mathbf{1}_{\Lambda_a}\right] \\
&\geq (L\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}f(0)-\delta)^q\mathbf{P}_1\left[\Lambda_a\right] \\
&\geq (L\varkappa^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}f(0)-\delta)^q\mathbf{P}\left[|\xi|\geq 1\right].
\end{aligned}
$$

And, then

$$
\liminf_{\varepsilon\to 0}\tilde{\mathcal{R}}_\varepsilon^{(q)}(\tilde{f}_\varepsilon) \geq \left(\frac{L^{\frac{1}{2\bar{\alpha}+1}}f(0)}{\|f\|^{\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}}\left(q(\varrho-\nu)\right)^{\frac{\bar{\alpha}}{2\bar{\alpha}+1}}\right)^q\mathbf{P}\left[|\xi|\geq 1\right] > 0.
$$

The lemma follows.

B Proof of Lemma 2

The proof of this lemma is divided into several steps. Firstly, we prove that the following equation holds :

$$0 \le \sum_{i=1}^{d} (k_i(\beta, L) + 1) \le n_\varepsilon.$$

Indeed, let us remark that, thanks to the definition of $k_i(\beta, L)$, we have for all $i = 1, \dots, d$

$$2^{k_i(\beta,L)} \le h_i^*/h_i(\beta, L) \le 2^{k_i(\beta,L)+1}. \tag{7}$$

Now, remark that

$$\log \prod_{i=1}^{d} \frac{h_i^*}{h_i(\beta, L)} = x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta}+1} \log(1 + x_\varepsilon(\beta, L)).$$

where

$$\begin{aligned} x_\varepsilon(\beta, L) &= \frac{4(\bar{b} - \bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \log \frac{l_*}{\|K\|\varepsilon} + \frac{2}{2\bar{\beta}+1} \log \frac{L}{l_*} \\ &= \rho_\varepsilon^2(\beta, L) - 1 \ge 0. \end{aligned}$$

Using the first inequality in (7), taking the product over i and using basic inequalities, the first step follows easily. Secondly, we will prove that, if $\sum_i (k_i + 1) = n$, then :

$$|k_i(\beta, L)| \le C_1 n + 1.$$

In this situation, using the second inequality in (7) and taking the product over i we obtain :

$$x_\varepsilon(\beta, L) \le n \log 2 + \log(1 + x_\varepsilon(\beta, L))$$

which leads to the following inequalities :

$$x_\varepsilon(\beta, L) \le n \log(2) + \sqrt{2n \log(2)} \le n \left(\log(2) + \sqrt{2 \log(2)} \right). \tag{8}$$

Using Equation (7) it is sufficient to provide both upper and lower bounds for the quantity $H_i = (\log h_i^*/h_i(\beta, L))$ that can be written as

$$\begin{aligned} H_i = &\left(\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \right) \log \frac{\ell_*}{\|K\|_2\, \varepsilon} \\ &+ \frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} \log \frac{L}{\ell_*} - \frac{\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} \log(1 + x_\varepsilon(\beta, L)). \end{aligned}$$

Upper bound for H_i

Using that $\bar{\beta} \leq \beta_i$ we obtain :

$$H_i \leq \left(\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \right) \log \frac{\ell_*}{\|K\|_2 \varepsilon} + \frac{2}{2\bar{\beta}+1} \log \frac{L}{l_*}$$

Simple calculations prove that

$$\begin{aligned} H_i &\leq \frac{2\bar{b}+1}{2\bar{b}} \frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)} \log \frac{\ell_*}{\|K\|_2 \varepsilon} + \frac{2}{2\bar{\beta}+1} \log \frac{L}{l_*} \\ &\leq \frac{2\bar{b}+1}{2\bar{b}} x_\varepsilon(\beta, L). \end{aligned}$$

Combining this equation with (7) and (8) the upper bound follows.

Lower bound for H_i

First, remark that :

$$\frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i} - \frac{2\bar{b}}{2\bar{b}+1} \frac{1}{b_i} \geq -\frac{1}{2\bar{b}} \frac{4(\bar{b}-\bar{\beta})}{(2\bar{b}+1)(2\bar{\beta}+1)}.$$

This leads to

$$\begin{aligned} H_i &\geq -\frac{1}{2\bar{b}} \left(x_\varepsilon(\beta, L) + \frac{2\bar{\beta}}{2\bar{\beta}+1} \frac{\bar{b}}{\beta_i} \ln(1 + x_\varepsilon(\beta, L)) \right) \\ &\geq -\frac{2\bar{b}+1}{2\bar{b}} x_\varepsilon(\beta, L) \end{aligned}$$

which, using (7) and (8), implies the lower bound.

C Proof of Lemma 3

Remark first that, in order to prove this lemma, it is sufficient to prove the following property : if $h = (h_1, \ldots, h_d)$ and $\eta = (\eta_1, \ldots, \eta_d)$ are two multivariate bandwidth such that, for all i, η_i is either h_i or 0, then

$$\left| \int_{\mathbb{R}^d} K(u) \left(f(t - h \cdot u) - f(t - \eta \cdot u) \right) \mathrm{d}u \right| \leq L \sum_{i \in J} \lambda_i(\beta) h_i^{\beta_i}, \tag{9}$$

where $x \cdot y$ stands for the component-wise product of vectors of $\mathbb{R}^d$ and $J = \{i = 1, \ldots, d : \eta_i = 0\}$. Indeed, taking $h = h^{(k)}$ and $\eta = (0, \ldots, 0)$ the first inequality follows. Let us prove that Equation (9) implies the second inequality as well. Remark that :

$$\left|\mathbf{E}_f^\varepsilon \hat{f}_{k\wedge\ell}(t) - \mathbf{E}_f^\varepsilon \hat{f}_\ell(t)\right| = \left|\int_{\mathbb{R}^d} K(u) \left(f(t - h^{(k\wedge\ell)} \cdot u) - f(t - h^{(\ell)} \cdot u)\right) \mathrm{d}u\right|.$$

Consider $J = \{i = 1, \ldots, d : h_i^{(k)} > h_i^{(\ell)}\}$ be such that if $i \notin J$ then $h_i^{(k\wedge\ell)} = h_i^{(\ell)}$ and let us define η in the following way : if $i \in J$ then $\eta_i = 0$ otherwise $\eta_i = h_i^{(k\wedge\ell)} = h_i^{(\ell)}$. We then obtain :

$$\begin{aligned}\left|\mathbf{E}_f^\varepsilon \hat{f}_{k\wedge\ell}(t) - \mathbf{E}_f^\varepsilon \hat{f}_\ell(t)\right| \leq &\left|\int_{\mathbb{R}^d} K(u) \left(f(t - h^{(k\wedge\ell)} \cdot u) - f(t - \eta \cdot u)\right) \mathrm{d}u\right| \\ &+ \left|\int_{\mathbb{R}^d} K(u) \left(f(t - \eta \cdot u) - f(t - h^{(\ell)} \cdot u)\right) \mathrm{d}u\right|\end{aligned}$$

The result then follows from Equation (9) applied with $h = h^{(k\wedge\ell)}$ or $h = h^{(\ell)}$ since in both cases $i \in J$ implies $h_i \leq h_i^{(k)}$. To conclude, let us prove that Equation (9) holds. If $x, y \in \mathbb{R}^d$ we consider the vector $[x, y]_i = (x_1, \ldots, x_{i-1}, 0, y_{i+1}, \ldots, y_d)$. Using this notation, we obtain the telescopic formula :

$$\begin{aligned}\mathcal{C} &= f(t - h \cdot u) - f(t - \eta \cdot u) \\ &= \sum_{i=1}^{d} f_i(-h_i u_i | t - [h, \eta]_i \cdot u) - f_i(-\eta_i u_i | t - [h, \eta]_i \cdot u) \\ &= \sum_{i \in J} f_i(-h_i u_i | t - [h, \eta]_i \cdot u) - f_i(-\eta_i u_i | t - [h, \eta]_i \cdot u).\end{aligned}$$

Last inequality comes from the fact that $i \notin J$ implies $h_i = \eta_i$. Using a Taylor expansion we obtain :

$$\begin{aligned}\mathcal{C} = &\sum_{i \in J} \sum_{j=1}^{\lfloor \beta_i \rfloor} f_i^{(j)}(0 | t - [h, \eta]_i \cdot u \frac{(-h_i u_i)^j}{j!} \\ &+ \sum_{i \in J} \frac{(-h_i u_i)^{\lfloor \beta_i \rfloor}}{\lfloor \beta_i \rfloor !} \left(f_i^{(\lfloor \beta_i \rfloor)}(\theta_i | t - [h, \eta]_i \cdot u) - f_i^{(\lfloor \beta_i \rfloor)}(0 | t - [h, \eta]_i \cdot u)\right),\end{aligned}$$

where $|\theta_i| \leq h_i |u_i|$. Using the properties of our kernel combined with the fact that $t - [h, \eta]_i \cdot u$ does not depend on u_i and Fubini's theorem, we obtain in an usual way :

$$\begin{aligned}&\left|\int_{\mathbb{R}^d} K(u) \left(f(t - h \cdot u) - f(t - \eta \cdot u)\right) \mathrm{d}u\right| \\ &\qquad \leq \sum_{i \in J} \int_{\mathbb{R}^d} |K(u)| \frac{|h_i u_i|^{\lfloor \beta_i \rfloor}}{\lfloor \beta_i \rfloor !} L |\theta_i|^{\beta_i - \lfloor \beta_i \rfloor}.\end{aligned}$$

Result follows.

D Proof of Lemma 4

Thanks to the definition of $k(\beta, L)$ we obtain that, for all i, we have the following inequalities :

$$h_i(\beta, L) \le h_i^{(k(\beta,L))} \le 2h_i(\beta, L).$$

This implies that

$$\begin{aligned}\mathfrak{B}_{\beta,L}(h^{k(\beta,L)}) &\le L\sum_{i=1}^{d} \lambda_i(\beta)(2h_i(\beta, L))^{\beta_i} \\ &\le L\sum_{i=1}^{d} 2^{b_i}\lambda_i(\beta)(h_i(\beta, L))^{\beta_i} \\ &\le L\sum_{i=1}^{d} \lambda_i^* 2^{b_i}(h_i(\beta, L))^{\beta_i}\end{aligned}$$

Now, a direct calculation, using the definition of $h(\beta, L)$ shows that

$$\mathfrak{B}_{\beta,L}(h^{k(\beta,L)}) \le \lambda^* \frac{\|K\|_2\, \varepsilon\rho_\varepsilon(\beta, L)}{\left(\prod_{i=1}^{d}, h_i(\beta, L)\right)^{1/2}}$$

which implies immediately that

$$\begin{aligned}\mathfrak{B}_{\beta,L}(h^{k(\beta,L)}) &\le \sqrt{2}\lambda^* \frac{\|K\|_2\, \varepsilon\rho_\varepsilon(\beta, L)}{\left(\prod_{i=1}^{d} h_i^{k(\beta,L)}\right)^{1/2}} \\ &= \sqrt{2}\lambda^* S_\varepsilon(k(\beta, L)) \frac{\rho_\varepsilon(\beta, L)}{\sqrt{1 + |k(\beta, L)| \log 2}}\end{aligned}$$

Using the notations introduced in the proof of Lemma 2, we obtain :

$$\frac{\rho_\varepsilon^2(\beta, L)}{1 + |k(\beta, L)| \log 2} = \frac{1 + x_\varepsilon(\beta, L)}{1 + |k(\beta, L)| \log 2}$$

Moreover it is easily seen that

$$|k(\beta, L)| \log 2 \ge x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta} + 1} \log(1 + x_\varepsilon(\beta, L)),$$

which implies:

$$x_\varepsilon(\beta, L) \le |k(\beta, L)| \log 2 + \sqrt{2|k(\beta, L)| \log 2},$$

and, consequently:

$$\frac{\rho_\varepsilon^2(\beta, L)}{1 + |k(\beta, L)| \log 2} = \frac{1 + x_\varepsilon(\beta, L)}{1 + |k(\beta, L)| \log 2} \le 2.$$

This proves that:

$$\mathfrak{B}_{\beta,L}(h^{k(\beta,L)}) \le 2^{3/2} \lambda^* S_\varepsilon(k(\beta, L)).$$

It is easily seen that:

$$|k(\beta, L)| \log 2 \le x_\varepsilon(\beta, L) - \frac{1}{2\bar{\beta} + 1} \log(1 + x_\varepsilon(\beta, L)) + d \log 2$$

Thanks to our choice of ε small enough, $x_\varepsilon \ge 0$. This leads to

$$1 + |k(\beta, L)| \log 2 \le (1 + d)(1 + x_\varepsilon(\beta, L))$$

We finally obtain:

$$\begin{aligned}
\mathfrak{B}_{\beta,L}(h^{k(\beta,L)}) &\le (2^{3/2}\lambda^*) \frac{\|K\|_2\, \varepsilon \rho_\varepsilon(\beta, L)}{\left(\prod_{i=1}^d h_i(\beta, L)\right)^{1/2}} \frac{\sqrt{1 + |k(\beta, L)| \log 2}}{\rho_\varepsilon(\beta, L)} \\
&\le \left(2^{3/2} \lambda^* \sqrt{1 + d}\right) \|K\|_2^{\frac{2\bar{\beta}}{2\bar{\beta}+1}} \phi_\varepsilon(\beta, L) \\
&\le \left(2^{3/2} \lambda^* \sqrt{1 + d}\right) (\|K\|_2 \vee 1)^{2\bar{b}/(2\bar{b}+1)} \phi_\varepsilon(\beta, L).
\end{aligned}$$

This ends the proof.

E Proof of Lemma 5

In this proof, we use the notations introduced in Section 5.4. If $\kappa(s-1) \notin \mathcal{A}$, there exists $\ell \in \mathcal{Z}_\varepsilon$ such that $\ell \succ \kappa(s-1)$ and:

$$|\hat{f}_{\kappa(s-1)\wedge\ell}(t) - \hat{f}_\ell(t)| > C^* S_\varepsilon(\ell).$$

In what follows, for the sake of readability, we denote $\ell[s-1] = \kappa(s-1) \wedge \ell$. We have:

$$\mathbf{P}_f^\varepsilon(\kappa(s-1) \notin \mathcal{A}) \le \sum_{\substack{\ell \in \mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} \mathbf{P}_f^\varepsilon\left(|\hat{f}_{\ell[s-1]}(t) - \hat{f}_\ell(t)| > C^* S_\varepsilon(\ell).\right).$$

Note that for all k we can decompose of $\hat{f}_k(t) - f(t)$ into a bias term and a centered stochastic gaussian term in the following way :

$$\hat{f}_k(t) - f(t) = \mathbf{E}_f^\varepsilon \hat{f}_k(t) - f(t) + \sigma_\varepsilon(k)\xi_k$$

where $\xi_k \sim \mathcal{N}(0,1)$. This leads to

$$|\hat{f}_{\ell[s-1]}(t) - \hat{f}_\ell(t)| \leq |\mathbf{E}_f^\varepsilon \hat{f}_{\ell[s-1]}(t) - \mathbf{E}_f^\varepsilon \hat{f}_\ell(t)| + \sigma_\varepsilon(\ell[s-1])\xi_{\ell[s-1]} + \sigma_\varepsilon(\ell)\xi_\ell. \quad (10)$$

Now, Lemmas 3 implies

$$|\mathbf{E}_f^\varepsilon \hat{f}_{\ell[s-1]}(t) - \mathbf{E}_f^\varepsilon \hat{f}_\ell(t)| \leq 2\mathfrak{B}_{\beta,L}(h^{(\kappa(s-1))})$$

Moreover, using Lemma 4

$$\mathfrak{B}_{\beta,L}(h^{(\kappa(s-1))}) \leq \mathfrak{B}_{\beta,L}(h^{(\kappa)}) \leq C_* S_\varepsilon(\kappa) \leq C_* S_\varepsilon(\kappa(s-1)) \leq 2C_* S_\varepsilon(\ell).$$

Last inequality holds since $\kappa(s-1) \prec \ell$. Note also that $\sigma_\varepsilon(\ell[s-1]) \leq \sigma_\varepsilon(\ell)$. This implies that

$$\begin{aligned}
\mathbf{P}_f^\varepsilon(\kappa(s-1) \notin \mathcal{A}) &\leq \sum_{\substack{\ell \in \mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} \mathbf{P}_f^\varepsilon\left(\xi_{\ell[s-1]} + \xi_\ell > (C^* - 2C_*)\sqrt{1 + |\ell| \log 2}\right) \\
&\leq 2 \sum_{\substack{\ell \in \mathcal{Z}_\varepsilon \\ \ell \succ \kappa(s-1)}} 2^{-\delta|\ell|},
\end{aligned}$$

where $\delta = (C^* - 2C_*)^2/8$. This ends the proof of our lemma.

F Proof of Lemma 6

Throughout this proof, C is an absolute constant (that can depend on p), the value of which may change from line to line. Let us decompose $R(s,p)$ in the following way :

$$\begin{aligned}
R(s,p) &= \mathbf{E}_f^\varepsilon \left(|\hat{f}(t) - f(t)|^p \mathbf{1}_{\hat{k} \in B(s)} \right) \\
&\leq C \left(\mathbf{E}_f^\varepsilon(I_1^p) + \mathbf{E}_f^\varepsilon(I_2^p) + \mathbf{E}_f^\varepsilon(I_3^p) \right)
\end{aligned}$$

where, using the notation $\hat{k}[s] = \hat{k} \wedge \kappa(s)$, we define :

$$\begin{cases}
I_1 = |\hat{f}_{\hat{k}}(t) - \hat{f}_{\hat{k}[s]}(t)|^p \mathbf{1}_{\hat{k} \in B(s)} \\
I_2 = |\hat{f}_{\hat{k}[s]}(t) - \hat{f}_{\kappa(s)}(t)|^p \mathbf{1}_{\hat{k} \in B(s)} \\
I_3 = |\hat{f}_{\kappa(s)}(t) - f(t)|^p \mathbf{1}_{\hat{k} \in B(s)}.
\end{cases}$$

Now, we provide upper bounds for these expectations. Firstly, let us consider the third (which is the simplest) expectation. Using similar arguments as in Equation (10) and Lemmas 4, we obtain:

$$\begin{aligned}
\mathbf{E}_f^\varepsilon(I_3^p) &\leq \mathbf{E}_f^\varepsilon\left((C_* S_\varepsilon(\kappa(s)) + \sigma_\varepsilon(\kappa(s))\xi_{\kappa(s)})^p \mathbf{1}_{\hat{k}\in B(s)}\right) \\
&\leq C\left(S_\varepsilon^p(\kappa(s)) + \mathbf{E}_f^\varepsilon\left((\sigma_\varepsilon(\kappa(s))\xi_{\kappa(s)})^p \mathbf{1}_{\hat{k}\in B(s)}\right)\right) \\
&\leq C\left(S_\varepsilon^p(\kappa(s)) + \sigma_\varepsilon^p(\kappa(s)) m_p\right),
\end{aligned}$$

where $m_p = \mathbf{E}(|\mathcal{N}(0,1)|^p)$. Secondly, let us upper bound the second expectation. Note that under the event $\{\hat{k} \in B(s)\}$, then $\hat{k}$ belongs to $\mathcal{A}$ (this is always true) and satisfies $|\hat{k}| \leq |\kappa(s)|$. Since $\kappa(s)$ belongs to $\mathcal{Z}_\varepsilon$, the construction of the procedure implies that

$$\mathbf{E}_f^\varepsilon(I_2^p) \leq (C^*)^p S_\varepsilon^p(\kappa(s)).$$

We finally have to bound the first expectation. Using similar arguments as above we obtain:

$$\mathbf{E}_f^\varepsilon(I_2^p) \leq C\left(S_\varepsilon^p(\kappa(s)) + \mathbf{E}_f^\varepsilon\left((\sigma_\varepsilon(\hat{k})\xi_{\hat{k}})^p \mathbf{1}_{\hat{k}\in B(s)}\right) + \mathbf{E}_f^\varepsilon\left((\sigma_\varepsilon(\hat{k}[s])\xi_{\hat{k}[s]})^p \mathbf{1}_{\hat{k}\in B(s)}\right)\right).$$

Now, let us control $J = \mathbf{E}_f^\varepsilon((\sigma_\varepsilon(\hat{k}[s])\xi_{\hat{k}[s]})^p \mathbf{1}_{\hat{k}\in B(s)})$. The second expectation on the right hand side on the previous inequality can be upper bounded in the same way. Let us consider, for all ℓ, the event

$$\Lambda_\ell = \left\{|\xi_{\ell[s]}| \leq 2\sqrt{1 + |\ell| \log 2}\right\}.$$

Using this event, J can be written as

$$\begin{aligned}
J &= \mathbf{E}_f^\varepsilon\left((\sigma_\varepsilon(\hat{k}[s])\xi_{\hat{k}[s]})^p \mathbf{1}_{\hat{k}\in B(s)}(\mathbf{1}_{\Lambda_{\hat{k}}} + \mathbf{1}_{\bar{\Lambda}_{\hat{k}}})\right) \\
&\leq (2S_\varepsilon(\kappa(s)))^p + \sum_{k\in B(s)} \sigma_\varepsilon^p(k[s]) \mathbf{E}_f^\varepsilon(\xi_{k[s]}^p \mathbf{1}_{\bar{\Lambda}_k}) \\
&\leq (2S_\varepsilon(\kappa(s)))^p + \sum_{k\in B(s)} \sigma_\varepsilon^p(\kappa(s)) \sqrt{m_{2p} \mathbf{P}(|\mathcal{N}(0,1)| > 2\sqrt{1 + |k| \log 2})} \\
&\leq (2S_\varepsilon(\kappa(s)))^p + m_{2p}^{1/2} \sigma_\varepsilon^p(\kappa(s)) \sum_{k\in B(s)} 2^{-|k|}.
\end{aligned}$$

Moreover

$$\begin{aligned}
\sum_{k\in B(s)} 2^{-|k|} &\leq \sum_{n\geq 0} \sum_{k\in\mathcal{Z}(n)d} 2^{-n} \\
&\leq \sum_{n\geq 0} \#\mathcal{Z}(n) 2^{-n} < +\infty. \qquad (11)
\end{aligned}$$

Taking all together, this ends our proof.

References

K. Bertin. Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes. *Math. Methods Statist.*, 14(3):267–298, 2005. ISSN 1066-5307.

M. Chichignoud. Minimax and minimax adaptive estimation in multiplicative regression: locally Bayesian approach. *Probab. Theory Related Fields*, 153(3-4):543–586, 2012. ISSN 0178-8051. doi: 10.1007/s00440-011-0354-7. URL http://dx.doi.org/10.1007/s00440-011-0354-7.

David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57 (2):301–369, 1995. ISSN 0035-9246. URL http://links.jstor.org/sici?sici=0035-9246(1995)57:2<301:WSA>2.0.CO;2-S&origin=MSN. With discussion and a reply by the authors.

R. P. W. Duin. On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans. on Computers*, C-25:1175–1179, 1976.

A. Goldenshluger and A. Nemirovski. On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.*, 6(2):135–170, 1997. ISSN 1066-5307.

Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008. ISSN 1350-7265. doi: 10.3150/08-BEJ144. URL http://dx.doi.org/10.3150/08-BEJ144.

Gérard Kerkyacharian, Oleg Lepski, and Dominique Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2): 137–170, 2001. ISSN 0178-8051. doi: 10.1007/PL00008800. URL http://dx.doi.org/10.1007/PL00008800.

Jussi Klemelä and Alexandre B. Tsybakov. Sharp adaptive estimation of linear functionals. *Ann. Statist.*, 29(6):1567–1600, 2001. ISSN 0090-5364. doi: 10.1214/aos/1015345955. URL http://dx.doi.org/10.1214/aos/1015345955.

O. V. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990. ISSN 0040-361X. doi: 10.1137/1135065. URL http://dx.doi.org/10.1137/1135065.

O. V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6):2512–2546, 1997. ISSN 0090-5364. doi: 10.1214/aos/1030741083. URL http://dx.doi.org/10.1214/aos/1030741083.

Charles J. Stone. An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, 12(4):1285–1297, 1984. ISSN 0090-5364. doi: 10.1214/aos/1176346792. URL http://dx.doi.org/10.1214/aos/1176346792.

A. B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469, 1998. ISSN 0090-5364. doi: 10.1214/aos/1024691478. URL http://dx.doi.org/10.1214/aos/1024691478.