Bernoulli **21**(4), 2015, 1984–2023
DOI: 10.3150/14-BEJ633

Pointwise adaptive estimation of a multivariate density under independence hypothesis

GILLES REBELLES

Institut de Mathématique de Marseille, Aix-Marseille Université, 39, rue F. Joliot-Curie, 13453 Marseille, France. E-mail: rebelles.gilles@neuf.fr

In this paper, we study the problem of pointwise estimation of a multivariate density. We provide a data-driven selection rule from the family of kernel estimators and derive for it a pointwise oracle inequality. Using the latter bound, we show that the proposed estimator is minimax and minimax adaptive over the scale of anisotropic Nikolskii classes. It is important to emphasize that our estimation method adjusts automatically to eventual independence structure of the underlying density. This, in its turn, allows to reduce significantly the influence of the dimension on the accuracy of estimation (curse of dimensionality). The main technical tools used in our considerations are pointwise uniform bounds of empirical processes developed recently in Lepski [*Math. Methods Statist.* **22** (2013) 83–99].

Keywords: adaptation; density estimation; independence structure; oracle inequality; upper function

1. Introduction

Let $X_i = (X_{i,1}, \dots, X_{i,d}), i \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^d$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ and having the density f with respect to the Lebesgue measure. Furthermore, $\mathbb{P}_f^{(n)}$ denotes the probability law of $X^{(n)} = (X_1, \dots, X_n), n \in \mathbb{N}^*$, and $\mathbb{E}_f^{(n)}$ is the mathematical expectation with respect to $\mathbb{P}_f^{(n)}$.

Our goal is to estimate the density f at a given point $x_0 \in \mathbb{R}^d$ using the observation $X^{(n)} = (X_1, \dots, X_n), n \in \mathbb{N}^*$. As an estimator, we mean any $X^{(n)}$-measurable mapping $\widehat{f}: \mathbb{R}^n \to \mathbb{R}$ and the accuracy of an estimator is measured by the *pointwise risk*:

$$\mathcal{R}_n^{(q)}[\widehat{f}, f] := \left(\mathbb{E}_f^{(n)} \left|\widehat{f}(x_0) - f(x_0)\right|^q\right)^{1/q}, \qquad q \geq 1.$$

The discussion of traditional methods and a part of the vast literature on the theory and application of the density estimation is given by Devroye and Györfi [7], Silverman [40] and Scott [39]. We do not pretend here to provide with a detailed overview and mention only the results which are relevant for considered problems. The minimax and adaptive minimax multivariate density estimation with $\mathbb{L}_p$-loss on particular functional classes was studied in Bretagnolle and Huber [2], Ibragimov and Khasminskii [21,22], Devroye and Lugosi [8–10], Efroimovich [13,14], Hasminskii and Ibragimov [20], Golubev [19], Donoho *et al.* [11], Kerkyacharian, Picard and Tribouley [26], Giné and Guillou [15], Juditsky and Lambert-Lacroix [23], Rigollet [36], Massart [33] (Chapter 7), Samarov and Tsybakov [38], Birgé [1], Mason [32], Giné and Nickl [16],

1350-7265

Chacón and Duong [5] and Goldenshluger and Lepski [18]. In Comte and Lacour [6], the pointwise setting was first considered in the context of multidimensional deconvolution model. More recently, in Goldenshluger and Lepski [17], adaptive minimax upper bounds were proved for multivariate density estimation with $\mathbb{L}_p$-risks on anisotropic Nikolskii classes using a local (pointwise) procedure. The use of Nikolskii classes allows to consider the estimation of anisotropic and inhomogeneous densities; see Ibragimov and Khasminskii [22], Goldenshluger and Lepski [18] and Lepski [29].

In this paper, we focus on the problem of the minimax and adaptive minimax pointwise multivariate density estimation over the scale of anisotropic Nikolskii classes.

Minimax estimation. In the framework of the minimax estimation, it is assumed that f belongs to a certain set of functions Σ, and then the accuracy of an estimator $\widehat{f}$ is measured by its *maximal risk* over Σ:

$$\mathcal{R}_n^{(q)}[\widehat{f}, \Sigma] := \sup_{f\in\Sigma} \big(\mathbb{E}_f^{(n)} \big|\widehat{f}(x_0) - f(x_0)\big|^q\big)^{1/q}, \qquad q \geq 1. \tag{1}$$

The objective here is to construct an estimator $\widehat{f_*}$ which achieves the asymptotic of *the minimax risk* (minimax rate of convergence):

$$\mathcal{R}_n^{(q)}[\widehat{f_*}, \Sigma] \asymp \inf_{\widehat{f}} \mathcal{R}_n^{(q)}[\widehat{f}, \Sigma] := \varphi_n(\Sigma).$$

Here, infimum is taken over all possible estimators.

Smoothness assumption. Let Σ be either Hölder classes $\mathbb{H}(\beta, L)$ or $\mathbb{L}_p$-Sobolev classes $\mathbb{W}(\beta, p, L)$ of univariate functions. Here, β represents the smoothness of the underlying density and p is the index of the norm where the smoothness is measured. Then

$$\begin{aligned} \varphi_n\big(\mathbb{H}(\beta, L)\big) &= n^{-\beta/(2\beta+1)}, \\ \varphi_n\big(\mathbb{W}(\beta, p, L)\big) &= n^{-(\beta-1/p)/(2(\beta-1/p)+1)}, \qquad \beta > 0, 1 < p < \infty. \end{aligned} \tag{2}$$

These minimax rates can be obtained from the results developed by Donoho and Low [12]; see also Ibragimov and Khasminskii [21,22], and Hasminskii and Ibragimov [20].

Let now $\Sigma = \mathbb{H}_d(\beta, L)$ where $\mathbb{H}_d(\beta, L)$ is an anisotropic Hölder class determined by the smoothness parameter $\beta = (\beta_1, \dots, \beta_d)$. In this case,

$$\varphi_n\big(\mathbb{H}_d(\beta, L)\big) = n^{-\overline{\beta}/(2\overline{\beta}+1)}, \qquad \overline{\beta} := \left[\sum_{i=1}^d 1/\beta_i\right]^{-1}, \qquad \beta_i > 0, i = \overline{1, d}. \tag{3}$$

The latter result can be obtained from Kerkyacharian, Lepski and Picard [24], Proposition 1, in the framework of the Gaussian white noise model. The similar minimax results will be established for pointwise multivariate density estimation in Section 3.2; see Theorems 2 and 3.

It is important to emphasize that minimax rates depend heavily on the dimension d. Let us briefly discuss how to reduce the influence of the dimension on the accuracy of estimation (curse of dimensionality). The approach which have been recently proposed in Lepski [29] is to take into account the eventual independence structure of the underlying density.

Structural assumption. Note $\mathcal{I}_d$ the set of all subsets of $\{1,\dots,d\}$ and $\mathfrak{P}$ the set of all partitions of $\{1,\dots,d\}$ completed by the empty set $\varnothing$. For all $I\in\mathcal{I}_d$ and $x\in\mathbb{R}^d$ note also $x_I=(x_i)_{i\in I}$, $\overline{I}=\{1,\dots,d\}\setminus I$, $|I|=\mathrm{card}(I)$ and put

$$f_I(x_I):=\int_{\mathbb{R}^{|\overline{I}|}} f(x)\,\mathrm{d}x_{\overline{I}}.$$

Obviously, f_I is the marginal density of $X_{1,I}$ and, to take into account the independence structure of the density f, we consider the following set:

$$\mathfrak{P}(f):=\Big\{\mathcal{P}\in\mathfrak{P}:\ f(x)=\prod_{I\in\mathcal{P}} f_I(x_I),\forall x\in\mathbb{R}^d\Big\}.$$

In this paper, we focus on the problem of pointwise multivariate density estimation on anisotropic Nikolskii classes. In particular, we will prove that the minimax rate on the class $N^*_{p,d}(\beta,L,\mathcal{P})$ (introduced in Lepski [29], see the definition in Section 3.1) for fixed $\beta\in(0,+\infty)^d$, $p\in[1,+\infty]^d$, $L\in(0,+\infty)^d$, $\mathcal{P}\in\mathfrak{P}(f)$, are given by

$$\varphi_n\big(N^*_{p,d}(\beta,L,\mathcal{P})\big)=n^{-r/(2r+1)},\qquad r:=\inf_{I\in\mathcal{P}}\Big[\frac{1-\sum_{i\in I}1/(\beta_i p_i)}{\sum_{i\in I}1/\beta_i}\Big].$$

If $d=1$, then the structural assumption does not exist, that means formally $\mathcal{P}=\overline{\varnothing}$, and we come to the rates given in (2). Note that $N^*_{\infty,1}(\beta,L,\overline{\varnothing})$ coincides with the set of densities belonging to $\mathbb{H}(\beta,L)$ and that $N^*_{p,1}(\beta,L,\overline{\varnothing})$ contains the set of densities belonging to $\mathbb{W}(\beta,p,L)$.

If $d\geq 2$, $p_i=\infty$, $i=\overline{1,d}$, and $\mathcal{P}=\overline{\varnothing}$ we find again the rates given in (3), and $N^*_{\infty,d}(\beta,L,\overline{\varnothing})$ coincides with a set of densities belonging to $\mathbb{H}_d(\beta,L)$. Note however that if $\mathcal{P}\neq\overline{\varnothing}$ the latter rates can be essentially improved. Indeed, if, for instance, $\beta=(\boldsymbol{\beta},\dots,\boldsymbol{\beta})$ and $\mathcal{P}^*=\{\{1\},\dots,\{d\}\}$, then $r=\boldsymbol{\beta}$ and

$$n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+d)}=\varphi_n\big(\mathbb{H}_d(\beta,L)\big)\gg\varphi_n\big(N^*_{\infty,d}\big(\beta,L,\mathcal{P}^*\big)\big)=n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+1)}.\tag{4}$$

Moreover, $\varphi_n(N^*_{\infty,d}(\beta,L,\mathcal{P}^*))$ does not depend on the dimension d.

We remark that minimax rates (accuracy of estimation) depend heavily on the parameters β, p and $\mathcal{P}$. Their knowledge cannot be often supposed in particular practice. It makes necessary to find an estimator whose construction would be parameter's free.

Adaptive minimax estimation. In the framework of the adaptive minimax estimation the underlying density f is supposed to belong to the given scale of functional classes $\{\Sigma_\alpha,\alpha\in\mathcal{A}\}$. For instance, if $\Sigma_\alpha=\mathbb{H}(\beta,L)$, $\alpha=(\beta,L)$, or if $\Sigma_\alpha=\mathbb{W}(\beta,p,L)$, $\alpha=(\beta,p,L)$.

The first question arising in the framework of the adaptive approach consists in the following: does there exists an estimator $\widehat{f_*}$ such that

$$\limsup_{n\to+\infty}\big\{\varphi_n^{-1}(\alpha)\mathcal{R}_n^{(q)}[\widehat{f_*},\Sigma_\alpha]\big\}<+\infty\qquad\forall\alpha\in\mathcal{A},\tag{5}$$

where $\varphi_n(\alpha)$ is the minimax rate of convergence over Σ_α.

As it was shown in Lepski [31] for the Gaussian white noise model, the answer of this question is negative if $\Sigma_\alpha = \mathbb{H}(\beta, L)$, $\alpha = (\beta, L)$. Brown and Low [3] extended this result to the pointwise density estimation. Further Butucea [4] extended the results of Brown and Low [3] over the scale of $\mathbb{L}_p$-Sobolev classes $\mathbb{W}(\beta, p, L)$. In Section 3.3.2, we will prove that the answer is also negative for multivariate density estimation at a given point over the scale of anisotropic Nikolskii classes $N^*_{p,d}(\beta, L, \mathcal{P})$.

Thus, for problems in which (5) does not hold we need first to find a family of normalizations $\Psi = \{\Psi_n(\Sigma_\alpha), \alpha \in \mathcal{A}\}$ and an estimator $\widehat{f}_\Psi$ such that

$$\limsup_{n\to+\infty} \bigl\{\Psi_n^{-1}(\alpha)\mathcal{R}_n^{(q)}[\widehat{f}_\Psi, \Sigma_\alpha]\bigr\} < +\infty \qquad \forall \alpha \in \mathcal{A}. \tag{6}$$

Any family of normalizations satisfying (6) is called admissible and the estimator $\widehat{f}_\Psi$ is called Ψ-adaptive. Next, we have to provide with the criterion of optimality allowing to select "the best" admissible family of normalizations, usually called adaptive rate of convergence. The first criterion was proposed in Lepski [31] and it was improved later in Tsybakov [41] and in Klutchnikoff [27].

In particular, in Lepski [31] and in Butucea [4], it was shown that the adaptive rate of convergence for the considered problem is

$$\Psi_n\bigl(\mathbb{H}(\beta, L)\bigr) = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{\beta/(2\beta+1)}, & \beta \in (0, \beta_{\max}), \\ \left(\dfrac{1}{n}\right)^{\beta/(2\beta+1)}, & \beta = \beta_{\max}, \end{cases}$$

$$\Psi_n\bigl(\mathbb{W}(\beta, p, L)\bigr) = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{(\beta-1/p)/(2(\beta-1/p)+1)}, & \beta \in (0, \beta_{\max}), \\ \left(\dfrac{1}{n}\right)^{(\beta-1/p)/(2(\beta-1/p)+1)}, & \beta = \beta_{\max}, \end{cases}$$

with respect to the criterion in Lepski [31] and Tsybakov [41], respectively. Here, $\beta_{\max}$ is an arbitrary positive number.

Later Klutchnikoff [27] studied the pointwise adaptive minimax estimation over anisotropic Hölder classes, in the Gaussian white noise model. The consideration of anisotropic functional classes required to develop a new criterion of optimality. Following this criterion, Klutchnikoff [27] proved that the adaptive rate of convergence is

$$\Psi_n\bigl(\mathbb{H}_d(\beta, L)\bigr) = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{\overline{\beta}(2\overline{\beta}+1)}, & \beta \in \prod_{i=1}^{d}\bigl(0, \beta_i^{(\max)}\bigr)^d, \\ \left(\dfrac{1}{n}\right)^{\overline{\beta}^{(\max)}/(2\overline{\beta}^{(\max)}+1)}, & \beta = \beta^{(\max)}. \end{cases}$$

Recently, Comte and Lacour [6] found a similar form of admissible sequence for pointwise adaptive minimax estimation in the deconvolution model.

In Section 3.3, we provide with minimax adaptive estimator in pointwise multivariate density estimation over the scale of anisotropic Nikolskii classes. We will take into account not only the approximation properties of the underlying density but the eventual independence structure as well. To analyze the accuracy of the proposed estimator, we establish so-called pointwise oracle inequality proved in Section 5.3. We will also show that the adaptive rate of convergence is given by

$$\Psi_n\big(N^*_{p,d}(\beta, L, \mathcal{P})\big) = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{r/(2r+1)}, & 0 < r < r_{\max}, \\ \left(\dfrac{1}{n}\right)^{r/(2r+1)}, & r = r_{\max}, \end{cases}$$

$$r := \inf_{I \in \mathcal{P}} \left[\frac{1 - \sum_{i\in I} 1/(\beta_i p_i)}{\sum_{i\in I} 1/\beta_i}\right].$$

To assert the optimality of this family of normalizations, we generalize the criterion proposed in Klutchnikoff [27]; see Section 3.3.2.

Organization of the paper. In Section 2, we provide a measurable data-driven selection rule based on bandwidth selection of kernel estimators and we derive an oracle-type inequality for the selected estimator at a given point. In Section 3, we treat the complete problem of minimax and adaptive minimax pointwise multivariate density estimation on a scale of anisotropic Nikolskii classes taking into account the independence structure of the underlying density. In Section 4, we briefly compare our local method with the global one developed in Lepski [29]. Proofs of all main results are given in Section 5. Proofs of technical lemmas are postponed to the Appendix.

2. Selection rule and pointwise oracle-type inequality

2.1. Kernel estimators related to independence structure

Let $\mathbf{K}:\mathbb{R} \to \mathbb{R}$ be a fixed symmetric kernel satisfying $\int \mathbf{K} = 1$, $\operatorname{supp}(\mathbf{K}) \subseteq [-1/2, 1/2]$, $\|\mathbf{K}\|_\infty < \infty$,

$$\exists L_{\mathbf{K}} > 0: \qquad \big|\mathbf{K}(x) - \mathbf{K}(y)\big| \le L_{\mathbf{K}}|x - y| \qquad \forall x, y \in \mathbb{R}. \tag{7}$$

For all $I \in \mathcal{I}_d$, $h \in (0, 1]^d$ and $x \in \mathbb{R}^d$ put also

$$K^{(I)}(x_I) := \prod_{i\in I} \mathbf{K}(x_i), \qquad V_{h_I} := \prod_{i\in I} h_i, \qquad K^{(I)}_{h_I}(x_I) := V^{-1}_{h_I} \prod_{i\in I} \mathbf{K}(x_i/h_i);$$

$$\widehat{f}^{(n)}_{h_I}(x_{0,I}) := n^{-1} \sum_{i=1}^{n} K^{(I)}_{h_I}(X_{i,I} - x_{0,I}).$$

Then introduce the family of estimators

$$\mathfrak{F}[\mathfrak{P}] := \left\{ \widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0) = \prod_{I\in\mathcal{P}} \widehat{f}^{(n)}_{h_I}(x_{0,I}), (h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P} \right\}.$$

Note first that $\widehat{f}^{(n)}_{(h,\overline{\varnothing})}(x_0) = \widehat{f}^{(n)}_{h}(x_0)$ is the Parzen–Rosenblatt estimator (see, e.g., Rosenblatt [37], Parzen [35]) with kernel $K^{(\overline{\varnothing})}$ and multibandwidth h.

Next, the introduction of the estimator $\widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0)$ is based on the following simple observation. If there exists $\mathcal{P} \in \mathfrak{P}(f)$, the idea is to estimate separately each marginal density corresponding to $I \in \mathcal{P}$. Since the estimated density possesses the product structure, we seek its estimator in the same form.

Below we propose a data driven selection from the family $\mathfrak{F}[\mathfrak{P}]$.

2.2. Auxiliary estimators and extra parameters

To define our selection rule, we need to introduce some notation and quantities.

Auxiliary estimators. For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$ put

$$\widetilde{G}_{h_I}(x_{0,I}) := 1 \vee \left[n^{-1} \sum_{i=1}^{n} \left| K^{(I)}_{h_I}(X_{i,I} - x_{0,I}) \right| \right].$$

Introduce for $I \in \mathcal{I}_d$ and $h, \eta \in (0,1]^d$ auxiliary estimators

$$\widehat{f}^{(n)}_{h_I,\eta_I}(x_{0,I}) := n^{-1} \sum_{i=1}^{n} K^{(I)}_{h_I \vee \eta_I}(X_{i,I} - x_{0,I}), \qquad h_I \vee \eta_I := (h_i \vee \eta_i)_{i\in I}.$$

Note that the idea to use such auxiliary estimators, defined with the multibandwidth $h \vee \eta$, appeared for the first time in Kerkyacharian, Lepski and Picard [24], in the framework of the Gaussian white noise model.

We endow the set $\mathfrak{P}$ with the operation "$\circ$" introduced in Lepski [29]: for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \circ \mathcal{P}' := \left\{ I \cap I' \neq \varnothing, I \in \mathcal{P}, I' \in \mathcal{P}' \right\} \in \mathfrak{P}.$$

Then we define for $h, \eta \in (0,1]^d$ and $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\widehat{f}^{(n)}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x_0) := \prod_{I\in\mathcal{P}\circ\mathcal{P}'} \widehat{f}^{(n)}_{h_I,\eta_I}(x_{0,I}). \tag{8}$$

Set of parameters. Our selection rule consists in choosing an estimator $\widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0)$ when the parameter $(h, \mathcal{P})$ belongs at most to the set $\mathfrak{H}[\mathfrak{P}]$ defined as follows.

Let $\mathfrak{z} > 0$, $\tau(s) \in (0,1]$, $s = 1, \dots, d$, be fixed numbers and let $\mathfrak{h}^{(I)}_I \in (0,1]^{|I|}$, $I \in \mathcal{I}_d$, be fixed multibandwidths. All these parameters will be chosen in accordance with our procedure.

Set also $\lambda := \sup_{I\in\mathcal{I}_d}\{1 \vee \lambda^{(2q)}_{|I|}[\mathbf{K}, \mathfrak{z}]\}$ and $a := \{2\lambda\sqrt{1+2q}\}^{-2}$, where constants $\lambda^{(q)}_s[\mathbf{K}, \mathfrak{z}]$, $s \in \mathbb{N}^*$, $q \geq 1$, are given in Section 5.1. The explicit expressions of $\lambda^{(q)}_s[\mathbf{K}, \mathfrak{z}]$ are too cumbersome and it is not convenient for us to present them right now.

For all $I \in \mathcal{I}_d$ and all integer $m > 0$ introduce

$$\mathfrak{H}^{(I)}_{m,1} := \big\{h_I \in (0,1]^{|I|}:\ v^{(I)}_m V_{\mathfrak{h}^{(I)}_I} \le V_{h_I} \le v^{(I)}_{m-1} V_{\mathfrak{h}^{(I)}_I}\big\} \cap \prod_{i\in I}\Big[\frac{1}{n}, \big(v^{(I)}_m\big)^{-\mathfrak{z}} \mathfrak{h}^{(I)}_i\Big],$$

$$\mathfrak{H}^{(I)}_{m,2} := \big\{h_I \in (0,1]^{|I|}:\ v^{(I)}_m V_{\max} \le V_{h_I} \le v^{(I)}_{m-1} V_{\max}\big\} \cap \prod_{i\in I}\Big[\frac{1}{n}, \big(v^{(I)}_m\big)^{-\mathfrak{z}} \mathfrak{h}^{(I)}_i\Big],$$

$$\mathfrak{H}^{(I)} := \Bigg(\bigcup_{m=1}^{M_n(I)} \mathfrak{H}^{(I)}_{m,1}\Bigg) \bigcup \Bigg(\bigcup_{m=1}^{M_n(I)} \mathfrak{H}^{(I)}_{m,2}\Bigg),$$

where $v^{(I)}_m := 2^{-m\tau(|I|)}$, $M_n(I)$ is the largest integer satisfying $v^{(I)}_{M_n(I)}[V_{\mathfrak{h}^{(I)}_I} \wedge V_{\max}] \ge \frac{\ln(n)}{an}$ and $M_n(I) \le \log_2(n)$, and $V_{\max}$ is defined below.

Define finally

$$\mathfrak{H}[\mathfrak{P}] := \big\{(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}:\ h_I \in \mathfrak{H}^{(I)}, \forall I \in \mathcal{P}\big\}.$$

Extra parameters. Let $\overline{\mathfrak{H}}$ and $\overline{\mathfrak{P}}$ be arbitrary subsets of $(0,1]^d$ and $\mathfrak{P}$, respectively. The selection rule (9)–(10) below run over $\overline{\mathfrak{H}}[\overline{\mathfrak{P}}] := (\overline{\mathfrak{H}} \times \overline{\mathfrak{P}}) \cap \mathfrak{H}[\mathfrak{P}]$ and the reasons for introducing these extra parameters are discussed in Remark 1. In particular, for measurability reasons, we will always suppose that $\overline{\mathfrak{H}}$ is either a compact or a finite subset of $(0,1]^d$.

Set $\Lambda_n(x_0) := 3\lambda d^2[2\overline{G}_n(x_0)]^{d^2-1}$, where

$$\overline{G}_n(x_0) := \sup_{(h,\mathcal{P})\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]}\ \sup_{(\eta,\mathcal{P}')\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]}\ \sup_{I\in\mathcal{P}\circ\mathcal{P}'} \big[2\widetilde{G}_{h_I\vee\eta_I}(x_{0,I})\big].$$

Put also $V_{\max} := \sup_{\mathcal{P}\in\overline{\mathfrak{P}}} \inf_{I\in\mathcal{P}} V_{\mathfrak{h}^{(I)}_I}$ and, for $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$,

$$\delta(h,\mathcal{P}) := \sup_{\mathcal{P}'\in\overline{\mathfrak{P}}}\ \sup_{I\cap I'\in\mathcal{P}\circ\mathcal{P}'} \Bigg[\frac{V_{\mathfrak{h}^{(I)}_{I\cap I'}\vee\mathfrak{h}^{(I')}_{I\cap I'}}}{V_{h_{I\cap I'}}}\Bigg] \vee \Bigg[\frac{V_{\max}}{\inf_{I\in\mathcal{P}} V_{\mathfrak{h}^{(I)}_I}}\Bigg].$$

Define finally, for $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$,

$$\widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0) := \sqrt{\frac{[\overline{G}_n(x_0)]^2\{1 \vee \ln\delta(h,\mathcal{P})\}}{nV(h,\mathcal{P})}}, \qquad V(h,\mathcal{P}) := \inf_{I\in\mathcal{P}} V_{h_I}.$$

2.3. Selection rule

For $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$ introduce

$$\widehat{\Delta}_{(h,\mathcal{P})}(x_0) := \sup_{(\eta,\mathcal{P}')\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]} \big[\big|\widehat{f}^{(n)}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x_0) - \widehat{f}^{(n)}_{(\eta,\mathcal{P}')}(x_0)\big| - \Lambda_n(x_0)\big\{\widehat{\mathcal{U}}_{(\eta,\mathcal{P}')}(x_0) + \widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\big\}\big]_+. \quad (9)$$

Define finally $(\widehat{h}, \widehat{\mathcal{P}})$ satisfying

$$\widehat{\Delta}_{(\widehat{h},\widehat{\mathcal{P}})}(x_0) + 2\Lambda_n(x_0)\widehat{\mathcal{U}}_{(\widehat{h},\widehat{\mathcal{P}})}(x_0) = \inf_{(h,\mathcal{P})\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]} \big[\widehat{\Delta}_{(h,\mathcal{P})}(x_0) + 2\Lambda_n(x_0)\widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\big]. \qquad (10)$$

The selected estimator is $\widehat{f}_n(x_0) := \widehat{f}^{(n)}_{(\widehat{h},\widehat{\mathcal{P}})}(x_0)$.

Similarly to Section 2.1 in Lepski [29] it is easy to show that $(\widehat{h}, \widehat{\mathcal{P}})$ is $X^{(n)}$-measurable and that $(\widehat{h}, \widehat{\mathcal{P}}) \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$. It follows that $\widehat{f}_n(x_0)$ is also a $X^{(n)}$-measurable random variable.

Remark 1. The necessity to introduce the extra parameters $\overline{\mathfrak{H}}$ and $\overline{\mathfrak{P}}$ is dictated by several reasons. The first one is computational namely the computation of $\widehat{\Delta}_{(h,\mathcal{P})}(x_0)$ and $(\widehat{h}, \widehat{\mathcal{P}})$. However, the computational aspects of the choice of $\overline{\mathfrak{P}}$ and $\overline{\mathfrak{H}}$ are quite different. Typically, $\overline{\mathfrak{H}}$ can be chosen as an appropriate grid in $(0, 1]^d$, for instance, dyadic one, that is sufficient for proving adaptive properties of the proposed estimator. The choice of $\overline{\mathfrak{P}}$ is much more delicate. The reason of considering $\overline{\mathfrak{P}}$ instead of $\mathfrak{P}$ is explained by the fact that the cardinality of $\mathfrak{P}$ grows exponentially with the dimension d. Therefore, if $\overline{\mathfrak{P}} = \mathfrak{P}$, for large values of d our procedure is not practically feasible in view of huge amount of comparisons to be done. In the latter case, the interest of our result is theoretical. Note also that the best attainable trade-off between approximation and stochastic errors depends heavily on both the number of observations and the effective dimension $d(f) = \inf_{\mathcal{P}\in\mathfrak{P}(f)} \sup_{I\in\mathcal{P}} |I|$. Thus, if $d(f)$ is big the corresponding independence structure does not bring a real improvement of the estimation accuracy. So, in practice, $\overline{\mathfrak{P}}$ is chosen to satisfy $\sup_{I\in\mathcal{P}} |I| \le d_0$, $\forall \mathcal{P} \in \overline{\mathfrak{P}} \setminus \{\varnothing\}$. The choice of the parameter d_0 (made by a statistician) is based on the compromised between the sample size n, the desirable quality of estimation and the number of computations. For instance, one can consider $d_0 = 1$, that means that $\overline{\mathfrak{P}}$ contains two elements, $\{\{1, \dots, d\}\}$ and $\{\{1\}, \dots, \{d\}\}$. The latter case corresponds to the observations having independent components and it can be illustrated in Example 1 below. On the other hand, in the case of low dimension d, one can always take $\overline{\mathfrak{P}} = \mathfrak{P}$, since if $d = 2$, $|\mathfrak{P}| = 2$, $d = 3$, $|\mathfrak{P}| = 5$, $d = 4$, $|\mathfrak{P}| = 12$, etc.

Other reasons are related to the possibility to consider various problems arising in the framework of minimax and minimax adaptive estimation and they will be discussed in detail in Sections 3.2 and 3.3.2. Here, we only mention that the choice $\overline{\mathfrak{P}} = \{\varnothing\}$ allows to study the adaptive estimation of a multivariate density on $\mathbb{R}^d$ without taking into account eventual independence structure. We would like to emphasize that the latter problem was not studied in the literature.

At last the introduction of $\overline{\mathfrak{P}}$ allows to minimize the assumptions imposed on the density to be estimated. In particular, the oracle inequality corresponding to $\overline{\mathfrak{P}} = \{\varnothing\}$ is proved over the set of bounded densities; see Corollary 1.

In spite of the fact that the construction of the proposed procedure does not require any condition on the density f, the following assumption will be used for computing its risk:

$$f \in \mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}] := \Big\{ f \colon \sup_{\mathcal{P},\mathcal{P}'\in\overline{\mathfrak{P}}} \sup_{I\in\mathcal{P}\circ\mathcal{P}'} \|f_I\|_\infty \le \mathbf{f}, \exists \mathcal{P} \in \mathfrak{P}(f) \cap \overline{\mathfrak{P}} \Big\}, \qquad 0 < \mathbf{f} < +\infty. \qquad (11)$$

Note that the considered class of densities is determined by $\overline{\mathfrak{P}}$ and in particular, if $\overline{\varnothing} \in \overline{\mathfrak{P}}$,

$$\mathbb{F}_d[\mathbf{f}, \mathfrak{P}] = \Big\{ f\colon \sup_{I \in \mathcal{I}_d} \|f_I\|_\infty \le \mathbf{f} \Big\} \subseteq \mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}],$$

$$\mathbb{F}_d\big[\mathbf{f}, \{\overline{\varnothing}\}\big] = \big\{ f\colon \|f\|_\infty \le \mathbf{f} \big\}, \qquad \mathbb{F}_d\big[\mathbf{f}, \{\mathcal{P}\}\big] = \Big\{ f\colon \sup_{I \in \mathcal{P}} \|f_I\|_\infty \le \mathbf{f}, \mathcal{P} \in \mathfrak{P}(f) \Big\}.$$

2.4. Oracle-type inequality

For $I \in \mathcal{I}_d$ and $(h, \eta) \in (0,1]^d \times [0,1]^d$ introduce

$$\mathcal{B}_{h_I,\eta_I}(x_{0,I}) := \int_{\mathbb{R}^{|I|}} K^{(I)}(u)\big[f_I\big(x_{0,I} + (h_I \vee \eta_I)u\big) - f_I(x_{0,I} + \eta_I u)\big]\,\mathrm{d}u,$$

where here and later $y_I x_I$ denotes the coordinate-vise product of $y_I, x_I \in \mathbb{R}^{|I|}$.

For $(h, \mathcal{P}) \in (0,1]^d \times \mathfrak{P}$ define $\mathcal{B}_{(h,\mathcal{P})}(x_0) := \sup_{\mathcal{P}' \in \overline{\mathfrak{P}}} \sup_{I \in \mathcal{P} \circ \mathcal{P}'} \sup_{\eta \in [0,1]^d} |\mathcal{B}_{h_I,\eta_I}(x_{0,I})|$.

Introduce finally, if exists $\mathcal{P} \in \mathfrak{P}(f) \cap \overline{\mathfrak{P}}$,

$$\mathfrak{R}_n(f) := \inf_{(h,\mathcal{P}) \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]\colon \mathcal{P} \in \mathfrak{P}(f)} \Bigg[\mathcal{B}_{(h,\mathcal{P})}(x_0) + \sqrt{\frac{1 \vee \ln \delta(h,\mathcal{P})}{nV(h,\mathcal{P})}} \Bigg].$$

The quantity $\mathfrak{R}_n(f)$ can be viewed as the optimal trade-off between approximation and stochastic errors provided by estimators involved in the selection rule.

Theorem 1. *Let $\overline{\mathfrak{H}} \subseteq (0,1]^d$ and $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$ be arbitrary subsets such that $\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$ is non-empty.*

Then for any $0 < \mathbf{f} < +\infty$, any $q \ge 1$ and any integer $n \ge 3$:

$$\mathcal{R}_n^{(q)}[\widehat{f}_n, f] \le \alpha_1 \mathfrak{R}_n(f) + \alpha_2 [nV_{\max}]^{-1/2} \qquad \forall f \in \mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}], \tag{12}$$

where $\alpha_1 := \alpha_1(q, d, \mathbf{K}, \mathbf{f})$ and $\alpha_2 := \alpha_2(q, d, \mathbf{K}, \mathbf{f})$ are given in the proof of the theorem.

Considering the case $\overline{\mathfrak{P}} = \{\overline{\varnothing}\}$ and noting $\mathfrak{H} = \mathfrak{H}^{(\overline{\varnothing})}$ we come to the following consequence of Theorem 1.

Corollary 1. *Let assumptions of Theorem 1 be fulfilled. Then, for all densities f such that $\|f\|_\infty \le \mathbf{f}$,*

$$\mathcal{R}_n^{(q)}[\widehat{f}_n, f] \le \alpha_1 \inf_{h \in \overline{\mathfrak{H}} \cap \mathfrak{H}} \Bigg[\sup_{\eta \in [0,1]^d} \big|\mathcal{B}_{h,\eta}(x_0)\big| + \sqrt{\frac{1 \vee \ln(V_{\mathfrak{h}}/V_h)}{nV_h}} \Bigg] + \alpha_2 [nV_{\mathfrak{h}}]^{-1/2}. \tag{13}$$

Looking at the assertion of Theorem 1 and its Corollary 1 it is not clear what can be gained by taking into account eventual independence structure. This issue will be scrutinized in Section 3, but some conclusions can be deduced directly from the latter results. Consider the following example.

Example 1. For any $t \in \mathbb{R}$, put

$$f(t) = \tfrac{64}{15}\left\{4t\mathbf{1}_{[0,1/8)}(t) + \left(\tfrac{3}{4} - 2t\right)\mathbf{1}_{(1/8,1/4]}(t) + \tfrac{1}{4}\mathbf{1}_{(1/4,3/4]}(t) + (1-t)\mathbf{1}_{(3/4,1]}(t)\right\},$$

and define $f_d(x) = \prod_{i=1}^{d} f(x_i)$, $x \in \mathbb{R}^d$. It is easily seen that f_d is a probability density and the goal is to estimate $f(x_0)$, $x_0 \in (3/8, 7/8)^d$.

Choose $\mathfrak{h} = (1, \dots, 1)$, $h = (1/4, \dots, 1/4)$ and let $\overline{\mathfrak{H}} = \{\mathfrak{h}, h\}$. Put $\mathcal{P}_1 = \{\{1, \dots, d\}\}$, $\mathcal{P}_2 = \{\{1\}, \dots, \{d\}\}$ and let $\overline{\mathfrak{P}} = \{\mathcal{P}_1, \mathcal{P}_2\}$. Since, in this case, $\overline{\mathfrak{H}} \times \overline{\mathfrak{P}}$ contains 4 elements, our estimator can be computed in a reasonable time.

Moreover, in accordance with the oracle-type inequality proved in Theorem 1, the accuracy provided by the selected estimator is proportional to $\sqrt{[4\ln(4)]/n}$. On the other hand, the pointwise risk of the kernel estimator with optimally chosen bandwidth and kernel is proportional to $\sqrt{[d4^d \ln(4)]/n}$ if the independence structure is not taken into account. As we see, the adaptation to eventual independence structure can lead to significant improvement of the constant. This shows that the proposed methodology has an interest beyond derivation of minimax rates, which is the subject of the next section.

3. Minimax and adaptive minimax pointwise estimation

In this section, we provide with minimax and adaptive minimax estimation over a scale of anisotropic Nikolskii classes.

3.1. Anisotropic Nikolskii densities classes with independence structure

Let $\{e_1, \dots, e_s\}$ denote the canonical basis in $\mathbb{R}^s$, $s \in \mathbb{N}^*$.

Definition 1. *Let $p = (p_1, \dots, p_s)$, $p_i \in [1, \infty]$, $\beta = (\beta_1, \dots, \beta_s)$, $\beta_i > 0$ and $L = (L_1, \dots, L_s)$, $L_i > 0$. A function $f : \mathbb{R}^s \to \mathbb{R}$ belongs to the anisotropic Nikolskii class $\mathbb{N}_{p,s}(\beta, L)$ if*

$$\text{(i)} \qquad \left\| D_i^k f \right\|_{p_i} \le L_i \qquad \forall k = \overline{0, \lfloor \beta_i \rfloor}, \forall i = \overline{1, s};$$

$$\text{(ii)} \qquad \left\| D_i^{\lfloor \beta_i \rfloor} f(\cdot + t e_i) - D_i^{\lfloor \beta_i \rfloor} f(\cdot) \right\|_{p_i} \le L_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \qquad \forall t \in \mathbb{R}, \forall i = \overline{1, s}.$$

Here, $D_i^k f$ denotes the kth order partial derivate of f with respect to the variable t_i, and $\lfloor \beta_i \rfloor$ is the largest integer strictly less than β_i.

The following collection $\{N_{p,d}^*(\beta, L, \mathcal{P})\}_{\mathcal{P}}$ was introduced in Lepski [29] in order to take into account the smoothness of the underlying density and its eventual independence structure simultaneously.

$$N_{p,d}^*(\beta, L, \mathcal{P}) := \left\{ f \in \mathbb{N}_{p,d}^*(\beta, L) \colon\ f \ge 0, \int f = 1, f(x) = \prod_{I \in \mathcal{P}} f_I(x_I), \forall x \in \mathbb{R}^d \right\},$$

where $f \in \mathbb{N}^*_{p,d}(\beta, L)$ means that

$$f_I \in \mathbb{N}_{p_I,|I|}(\beta_I, L_I) \qquad \forall I \in \mathcal{I}_d. \tag{14}$$

We remark that this collection of functional classes was used in the case of adaptive estimation, that is, when the partition $\mathcal{P} \in \mathfrak{P}$ is unknown. However, when the minimax estimation is considered ($\mathcal{P}$ is fixed), we do not need that condition (14) holds for any $I \in \mathcal{I}_d$. It suffices to consider only I belonging to $\mathcal{P}$, and we come to the following definition.

Definition 2 (Minimax estimation). *Let $p = (p_1, \dots, p_d)$, $p_i \in [1, \infty]$, $\beta = (\beta_1, \dots, \beta_d)$, $\beta_i > 0$, $L = (L_1, \dots, L_d)$, $L_i > 0$ and $\mathcal{P} \in \mathfrak{P}$. A probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $N_{p,d}(\beta, L, \mathcal{P})$ if*

$$f(x) = \prod_{I \in \mathcal{P}} f_I(x_I) \qquad \forall x \in \mathbb{R}^d, \qquad f_I \in \mathbb{N}_{p_I,|I|}(\beta_I, L_I) \qquad \forall I \in \mathcal{P}. \tag{15}$$

Let us now come back to the adaptive estimation. As it was discussed in Remark 1, the adaptation is not necessarily considered with respect to $\mathfrak{P}$. If $\overline{\mathfrak{P}} \subset \mathfrak{P}$ is used instead of $\mathfrak{P}$, the assumption (14) is too restrictive and can be weakened in the following way.

Denote $\overline{\mathfrak{P}}^* := \{\mathcal{P} \circ \mathcal{P}' \colon \mathcal{P}, \mathcal{P}' \in \overline{\mathfrak{P}}\}$ and $\overline{\mathcal{I}}^*_d := \{I \in \mathcal{I}_d \colon \exists \mathcal{P} \in \overline{\mathfrak{P}}^*, I \in \mathcal{P}\}$.

Definition 3 (Adaptive estimation). *Let $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$ and $(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \overline{\mathfrak{P}}$ be fixed. A probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $\overline{N}_{p,d}(\beta, L, \mathcal{P})$ if*

$$f(x) = \prod_{I \in \mathcal{P}} f_I(x_I) \qquad \forall x \in \mathbb{R}^d; \qquad f_I \in \mathbb{N}_{p_I,|I|}(\beta_I, L_I) \qquad \forall I \in \overline{\mathcal{I}}^*_d. \tag{16}$$

Some remarks are in order.

(1) We note that if $\overline{\mathfrak{P}} = \mathfrak{P}$, then $\overline{N}_{p,d}(\beta, L, \mathcal{P}) = N^*_{p,d}(\beta, L, \mathcal{P})$, but for some $\overline{\mathfrak{P}} \subset \mathfrak{P}$, one has $N^*_{p,d}(\beta, L, \mathcal{P}) \subset \overline{N}_{p,d}(\beta, L, \mathcal{P})$. The latter inclusion shows that the condition (16) is weaker than $f \in N^*_{p,d}(\beta, L, \mathcal{P})$. In particular, if $\overline{\mathfrak{P}} = \{\overline{\varnothing}\}$, then $\overline{N}_{p,d}(\beta, L, \overline{\varnothing}) = \{f \in \mathbb{N}_{p,d}(\beta, L) \colon f \geq 0, \int f = 1\} \supset N^*_{p,d}(\beta, L, \overline{\varnothing})$.

(2) Note that if $\overline{\mathfrak{P}} = \{\mathcal{P}\}$, then $\overline{N}_{p,d}(\beta, L, \mathcal{P})$ coincides with the class $N_{p,d}(\beta, L, \mathcal{P})$ used for minimax estimation. But $\overline{N}_{p,d}(\beta, L, \mathcal{P}) \subset N_{p,d}(\beta, L, \mathcal{P})$ for all $\mathcal{P} \in \overline{\mathfrak{P}}$ for any other choices of $\overline{\mathfrak{P}}$.

3.2. Minimax results

For $(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \mathfrak{P}$ define

$$r := r(\beta, p, \mathcal{P}) = \inf_{I \in \mathcal{P}} \gamma_I(\beta, p), \qquad \gamma_I := \gamma_I(\beta, p) = \frac{1 - \sum_{i \in I} 1/(\beta_i p_i)}{\sum_{i \in I} 1/\beta_i}, \qquad I \in \mathcal{P};$$

$$\varphi_n(\beta, p, \mathcal{P}) := \left(\frac{1}{n}\right)^{r/(2r+1)}, \qquad \rho_n(\beta, p, \mathcal{P}) := \mathbf{1}_{\{r\le 0\}} + \varphi_n(\beta, p, \mathcal{P})\mathbf{1}_{\{r>0\}}. \tag{17}$$

As it will follow from Theorems 2 and 3 below $\varphi_n(\beta, p, \mathcal{P})$ is the minimax rate of convergence on $N_{p,d}(\beta, L, \mathcal{P})$. Hence, similarly to the standard representation of minimax rates, the parameter r can be interpreted as a smoothness index corresponding to the independence structure.

Theorem 2. $\forall(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \mathfrak{P}$, $\forall L \in (0, \infty)^d$, $\exists c > 0$:

$$\liminf_{n\to+\infty}\Big\{\rho_n^{-1}(\beta, p, \mathcal{P}) \inf_{\widetilde{f}_n} \mathcal{R}_n^{(q)}\big[\widetilde{f}_n, N_{p,d}(\beta, L, \mathcal{P})\big]\Big\} \ge c,$$

where infimum is taken over all possible estimators.

Note that the assertion of Theorem 2 will be deduced from more general result established in Proposition 1 below. It is also important to emphasize that if $r \le 0$ there is no uniformly consistent estimator for the considered problem and, to the best of our knowledge, this fact was not known before. Let us provide an example with a density for which $r < 0$.

Example 2. Suppose that $d = 1$ and, therefore, $\mathcal{P} = \overline{\varnothing}$ (no independence structure). For any $x \in \mathbb{R}$, put

$$g(x) = \mathbf{1}_{\{0\}}(x) + \frac{1}{2\sqrt{x}}\mathbf{1}_{(0,1]}(x).$$

Some straightforward computations allows us to assert that $g \notin N_{p,1}(\beta, L, \overline{\varnothing})$, $\forall L > 0$, if $p\beta \ge 1$ (i.e., $r \ge 0$), and that $g \in N_{1,1}(1/2, L, \overline{\varnothing})$ for some $L > 0$ ($p = 1$, $\beta = 1/2$). Thus, in this case, one has $r < 0$.

Our goal now is to show that $\varphi_n(\beta, p, \mathcal{P})$ is the minimax rate of convergence on $N_{p,d}(\beta, L, \mathcal{P})$ and that a minimax estimator belongs to the collection $\mathfrak{F}[\mathfrak{P}]$. In fact, we prove that the minimax estimator is $\widehat{f}^{(n)}_{(\mathbf{h},\mathcal{P})}$ with properly chosen kernel $\mathbf{K}$ and bandwidth $\mathbf{h}$.

For a given integer $l \ge 2$ and a given symmetric Lipschitz function $u : \mathbb{R} \to \mathbb{R}$ satisfying $\operatorname{supp}(u) \subseteq [-1/(2l), 1/(2l)]$ and $\int_{\mathbb{R}} u(y)\,\mathrm{d}y = 1$ set

$$u_l(z) := \sum_{i=1}^{l} \binom{l}{i} (-1)^{i+1} \frac{1}{i} u\left(\frac{z}{i}\right), \qquad z \in \mathbb{R}. \tag{18}$$

Furthermore, we use $\mathbf{K} \equiv u_l$ in the definition of estimators collection $\mathfrak{F}[\mathfrak{P}]$.

The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [26]. In particular, it was shown that

$$\int_{\mathbb{R}} \mathbf{K}(z)\,\mathrm{d}z = 1, \qquad \int_{\mathbb{R}} z^k \mathbf{K}(z)\,\mathrm{d}z = 0 \qquad \forall k = 1, \dots, l-1. \tag{19}$$

Choose finally $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_d)$, where

$$\mathbf{h}_i = n^{-(\gamma_I(\beta,p)/(2\gamma_I(\beta,p)+1))(1/\beta_i(I))}, \qquad i \in I, I \in \mathcal{P}.$$

Here,

$$\beta_i(I) := \varkappa(I)\beta_i \varkappa_i^{-1}(I), \qquad \varkappa(I) := 1 - \sum_{k\in I}(\beta_k p_k)^{-1}, \qquad \varkappa_i(I) := 1 - \sum_{k\in I}\big(p_k^{-1} - p_i^{-1}\big)\beta_k^{-1}.$$

Theorem 3. *For all $(\beta, p, \mathcal{P}) \in (0, l]^d \times [1, \infty]^d \times \mathfrak{P}$ such that $r(\beta, p, \mathcal{P}) > 0$ and all $L \in (0, \infty)^d$*

$$\limsup_{n\to+\infty}\big\{\varphi_n^{-1}(\beta, p, \mathcal{P})\mathcal{R}_n^{(q)}\big[\widehat{f}^{(n)}_{(\mathbf{h},\mathcal{P})}, N_{p,d}(\beta, L, \mathcal{P})\big]\big\} < \infty.$$

To get the statement of this theorem, we apply Theorem 1 with $\overline{\mathfrak{P}} = \{\mathcal{P}\}$ and $\overline{\mathfrak{H}} = \{\mathbf{h}\}$. In view of the embedding theorem for anisotropic Nikolskii classes (formulated in the proof of Lemma 3 and available when $r(\beta, p, \mathcal{P}) > 0$), there exists a number $\mathbf{f} := \mathbf{f}(\beta, p) > 0$ such that $N_{p,d}(\beta, L, \mathcal{P}) \subseteq \mathbb{F}_d[\mathbf{f}, \{\mathcal{P}\}]$. It makes possible the application of Theorem 1.

Let us briefly discuss several consequences of Theorems 2 and 3. First, if $\mathcal{P} = \overline{\varnothing}$, we obtain the minimax rate on the anisotropic Nikolskii class $\mathbb{N}_{p,d}(\beta, L)$. In particular, if $p_i = +\infty$, $i = \overline{1, d}$, we find the minimax rate on the anisotropic Hölder class $\mathbb{H}_d(\beta, L)$ given in (3). If $d = 1$, then our results coincide with those presented in (2).

Next, in view of Theorem 2 there is no consistent estimator for $f(x_0)$ on $\mathbb{N}_{p,d}(\beta, L)$ if $r(\beta, p, \overline{\varnothing}) \le 0$. On the other hand, if $f \in N_{p,d}(\beta, L, \mathcal{P})$ and $r(\beta, p, \mathcal{P}) > 0$, then such estimator for $f(x_0)$ does exist in view of Theorem 3 even if $r(\beta, p, \overline{\varnothing}) < 0$.

Note also that the condition $r(\beta, p, \overline{\varnothing}) > 0$ is sufficient to find a consistent estimator on each functional class $N_{p,d}(\beta, L, \mathcal{P})$, $\mathcal{P} \in \mathfrak{P}$, and that the same condition is necessary for the estimation over $N_{p,d}(\beta, L, \varnothing)$. It allows us to compare the influence of the independence structure on the accuracy of estimation. For example, we see that

$$\varphi_n\big(\mathbb{H}_d(\beta, L)\big) \gg \varphi_n(\beta, p, \mathcal{P}), \qquad p_i = \infty, i = \overline{1, d}.$$

We conclude that the existence of an independence structure improves significantly the accuracy of estimation.

We finish this section with the result being a refinement of Theorem 2.

Proposition 1. $\forall(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \mathfrak{P}$, $\forall L \in (0, \infty)^d$, $\exists c > 0$:

$$\liminf_{n\to+\infty}\Big\{\rho_n^{-1}(\beta, p, \mathcal{P})\inf_{\widetilde{f}_n}\mathcal{R}_n^{(q)}\big[\widetilde{f}_n, N^*_{p,d}(\beta, L, \mathcal{P})\big]\Big\} \ge c,$$

where infimum is taken over all possible estimators.

Remark 2. Recall (see Section 3.1) that $N^*_{p,d}(\beta, L, \mathcal{P}) \subseteq \overline{N}_{p,d}(\beta, L, \mathcal{P}) \subseteq N_{p,d}(\beta, L, \mathcal{P})$. Hence, the statement of Theorem 3 remains true if one replaces $N_{p,d}(\beta, L, \mathcal{P})$ by $\overline{N}_{p,d}(\beta, L, \mathcal{P})$,

$\overline{\mathfrak{P}} \subseteq \mathfrak{P}$. Thus, Proposition 1 together with Theorem 3 allows us to assert that $\rho_n(\beta, p, \mathcal{P})$ is the minimax rate of convergence on $\overline{N}_{p,d}(\beta, L, \mathcal{P})$.

3.3. Adaptive estimation

3.3.1. *Adaptive estimation. Upper bound*

Let $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$, such that $\overline{\varnothing} \in \overline{\mathfrak{P}}$, be fixed. Denote $d(\mathcal{P}) := \sup_{I\in\mathcal{P}} |I|$, $\mathcal{P} \in \overline{\mathfrak{P}}$, and $\overline{d} := \inf_{\mathcal{P}\in\overline{\mathfrak{P}}} \mathrm{d}(\mathcal{P})$.

Set $\beta_i^{(\max)} = \beta_{\max} > (d - \overline{d})/2$, $p_i^{(\max)} = +\infty$, $i = \overline{1, d}$, and suppose additionally that $l \geq 2 \vee \beta_{\max}$. Choose $\mathbf{K} \equiv u_l$, $\mathfrak{z} := \frac{1}{2\beta_{\max}}$ and $\tau(s)$, $s = 1, \dots, d$, satisfying

$$\tau(s) := 2\beta_{\max}/(2\beta_{\max} + \overline{d}).$$

Let $\overline{\mathfrak{H}}$ be the dyadic grid in $(0, 1]^d$ and let $\mathfrak{h}_I^{(I)}$, $I \in \mathcal{I}_d$, be the projection on the dyadic grid in $(0, 1]^{|I|}$ of the multibandwith $\mathrm{h}_I^{(I)}$ given by

$$\mathrm{h}_i^{(I)} := n^{-1/(2\beta_{\max}+\overline{d})}, \qquad i \in I. \tag{20}$$

Consider the estimator $\widehat{f}_n(x_0)$ defined by the selection rule (9)–(10), in Section 2.3.

For $(\beta, p, \mathcal{P}) \in (0, \beta_{\max}]^d \times [1, \infty]^d \times \overline{\mathfrak{P}}$ introduce

$$\psi_n(\beta, p, \mathcal{P}) := \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{r/(2r+1)}, & r := r(\beta, p, \mathcal{P}) < r_{\max}, \\ \left(\dfrac{1}{n}\right)^{r_{\max}/(2r_{\max}+1)}, & r := r(\beta, p, \mathcal{P}) = r_{\max}, \end{cases} \qquad r_{\max} := \frac{\beta_{\max}}{\overline{d}}. \tag{21}$$

Theorem 4. *For any $(\beta, p) \in (0, \beta_{\max}]^d \times [1, \infty]^d$ such that $r(\beta, p, \overline{\varnothing}) > 0$, any $\mathcal{P} \in \overline{\mathfrak{P}}$ and any $L \in (0, \infty)^d$*

$$\limsup_{n\to+\infty} \left\{ \psi_n^{-1}(\beta, p, \mathcal{P}) \mathcal{R}_n^{(q)}\left[\widehat{f}_n, \overline{N}_{p,d}(\beta, L, \mathcal{P})\right]\right\} < \infty.$$

Similarly to Theorem 3, the proof of Theorem 4 is mostly based on the result of Theorem 1. The application of Theorem 1 is possible because $\overline{N}_{p,d}(\beta, L, \mathcal{P}) \subseteq \mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}]$ for some $\mathbf{f} := \mathbf{f}(\beta, p) > 0$ that is guaranteed by the condition $r(\beta, p, \overline{\varnothing}) > 0$.

We would like to emphasize that the construction of $\widehat{f}_n(x_0)$ does not involved the knowledge of the parameters $(\beta, L, p, \mathcal{P})$. Using the modern statistical language, one can say that $\widehat{f}_n(x_0)$ is fully adaptative.

Note, however, that the precision $\psi_n(\beta, p, \mathcal{P})$ given by this estimator does not coincide with minimax rate of convergence $\varphi_n(\beta, p, \mathcal{P})$ whenever $r \neq r_{\max}$. In the next section, we prove that

$\psi_n(\beta, p, \mathcal{P})$ found in Theorem 4 is an optimal payment for adaptation.

3.3.2. *Adaptive estimation. Criterion of optimality*

Let $\{\Sigma_{(\alpha,b)}, (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ be the scale of functional classes where $\mathcal{A} \subset \mathbb{R}^m$ is a (m)-dimensional manifold and $\mathfrak{B}$ is a finite set. Recall that the family $\Psi = \{\Psi_n(\alpha, b), (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ of normalizations is called admissible if there exists an estimator $\widehat{f}_\Psi$ such that

$$\limsup_{n\to+\infty}\big\{\Psi_n^{-1}(\alpha, b)\mathcal{R}_n^{(q)}[\widehat{f}_\Psi, \Sigma_{(\alpha,b)}]\big\} < +\infty \qquad \forall(\alpha, b) \in \mathcal{A} \times \mathfrak{B}. \tag{22}$$

The estimator $\widehat{f}_\Psi$ is called Ψ-adaptive.

In the considered problem, $\alpha = (\beta, p)$, $b = \mathcal{P}$ and

$$\mathcal{A} = \big\{(\beta, p) \in (0, \beta_{\max}]^d \times [1, \infty]^d\colon\ r(\beta, p, \overline{\varnothing}) > 0\big\}, \qquad \mathfrak{B} = \overline{\mathfrak{P}}.$$

As it follows from Theorem 4 $\psi_n(\beta, p, \mathcal{P})$ is an admissible family of normalizations and the estimator $\widehat{f}_n$ is ψ_n-adaptive.

Let $\Psi = \{\Psi_n(\alpha, b) > 0, (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ and $\widetilde{\Psi} = \{\widetilde{\Psi}_n(\alpha, b) > 0, (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ be arbitrary families of normalizations and put

$$\Upsilon_n(\alpha, b) := \frac{\widetilde{\Psi}_n(\alpha, b)}{\Psi_n(\alpha, b)}, \qquad \Upsilon_n(\alpha) := \inf_{b\in\mathfrak{B}} \Upsilon_n(\alpha, b).$$

Define the set $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi] \subseteq \mathcal{A}$ as follows:

$$\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi] := \Big\{\alpha \in \mathcal{A}\colon\ \lim_{n\to\infty} \Upsilon_n(\alpha) = 0\Big\}.$$

The set $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]$ can be viewed as the set where the family $\widetilde{\Psi}$ "outperforms" the family Ψ. For any $b \in \mathfrak{B}$, introduce

$$\mathcal{A}_b^{(\infty)}[\widetilde{\Psi}/\Psi] := \Big\{\alpha \in \mathcal{A}\colon\ \lim_{n\to\infty} \Upsilon_n(\alpha_0)\Upsilon_n(\alpha, b) = \infty, \forall\alpha_0 \in \mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]\Big\}.$$

Remark first that the set $\mathcal{A}_b^{(\infty)}[\widetilde{\Psi}/\Psi]$ is the set where the family Ψ "outperforms" the family $\widetilde{\Psi}$. Moreover, the "gain" provided by Ψ with respect to $\widetilde{\Psi}$ on $\mathcal{A}_b^{(\infty)}[\widetilde{\Psi}/\Psi]$ is much larger than its "loss" on $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]$.

The idea led to the criterion of optimality formulated below is to say that Ψ is "better" than $\widetilde{\Psi}$ if there exists $b \in \mathfrak{B}$ for which the set $\mathcal{A}_b^{(\infty)}[\widetilde{\Psi}/\Psi]$ is much more "massive" than $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]$.

Definition 4. (I) *A family of normalizations Ψ is called adaptive rate of convergence if*

1. *Ψ is an admissible family of normalizations*;
2. *for any admissible family of normalizations $\widetilde{\Psi}$ satisfying $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi] \neq \varnothing$*

- $\mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]$ *is contained in a* $(m-1)$*-dimensional manifold*,
- *there exists* $b \in \mathfrak{B}$ *such that* $\mathcal{A}_b^{(\infty)}[\widetilde{\Psi}/\Psi]$ *contains an open set of* $\mathcal{A}$.

(II) *If* Ψ *is an adaptive rate of convergence*, *then* $\widehat{f}_\Psi$ *satisfying* (22) *is called rate adaptive estimator.*

The aforementioned definition is inspired by Klutchnikoff's criterion; see Klutchnikoff [27]. Indeed if $\mathrm{card}(\mathfrak{B}) = 1$ the both definitions coincide.

Theorem 5. (i) *We can find no optimal rate adaptive estimator* (*satisfying* (5) *in Section* 1) *over the scale*

$$\{\overline{N}_{p,d}(\beta, L, \mathcal{P}), (\beta, p, L, \mathcal{P}) \in \mathfrak{A}\},$$

whenever $\mathfrak{A} \subseteq \{(\beta, p, L, \mathcal{P}) \in (0, \beta_{\max}]^d \times [1,\infty]^d \times (0,\infty)^d \times \overline{\mathfrak{P}}\colon r(\beta, p, \overline{\varnothing}) > 0\}$ *contains at least two elements* $(\beta, p, L, \mathcal{P})$ *and* $(\beta', p', L', \mathcal{P}')$ *such that* $r(\beta, p, \mathcal{P}) \neq r(\beta', p', \mathcal{P}')$.

(ii) $\widehat{f}_n(x_0)$ *is rate adaptive estimator of* $f(x_0)$ *and* ψ_n *is the adaptive rate of convergence*, *in the sense of Definition* 4, *over the scale*

$$\{\overline{N}_{p,d}(\beta, L, \mathcal{P}), (\beta, p, L, \mathcal{P}) \in (0, \beta_{\max}]^d \times [1,\infty]^d \times (0,\infty)^d \times \overline{\mathfrak{P}}, r(\beta, p, \overline{\varnothing}) > 0\}.$$

It is important to emphasize that our results cover a large class of problems in the framework of pointwise density estimation.

In particular, if $\overline{\mathfrak{P}} = \{\overline{\varnothing}\}$, we deduce that $\widehat{f}_n(x_0)$ is rate adaptive estimator of $f(x_0)$ over

$$\{\overline{N}_{p,d}(\beta, L, \overline{\varnothing}), (\beta, p, L) \in (0, \beta_{\max}]^d \times [1,\infty]^d \times (0,\infty)^d, r(\beta, p, \overline{\varnothing}) > 0\}.$$

The adaptive rate of convergence for this problem is given by

$$\psi_n(\beta, p, \overline{\varnothing}) := \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{r/(2r+1)}, & (\beta, p) \neq (\beta^{(\max)}, p^{(\max)}), r := \dfrac{1 - \sum_{i=1}^d 1/(\beta_i p_i)}{\sum_{i=1}^d 1/\beta_i}, \\ \left(\dfrac{1}{n}\right)^{r_{\max}/(2r_{\max}+1)}, & (\beta, p) = (\beta^{(\max)}, p^{(\max)}), r_{\max} := \dfrac{\beta_{\max}}{d}. \end{cases}$$

To the best of our knowledge, the latter result is new. It is precise and generalizes the results of Butucea [4] ($d = 1$) and Comte and Lacour [6] for the deconvolution model when the noise variable is equal to zero.

Another interesting fact is related to the set of "nuisance" parameters where the adaptive rate of convergence $\psi_n(\beta, p, \mathcal{P})$ coincides with the minimax one. In all known for us problems of pointwise adaptive estimation this set contains a single element. However, as it follows from Theorem 5, this set may contain several elements. Indeed, if, for instance, $d = 4$ and $\overline{\mathfrak{P}} = \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3\}$ with $\mathcal{P}_1 = \{\{1\}, \{2\}, \{3, 4\}\}$, $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}\}$, $\mathcal{P}_3 = \{\{1, 2, 3, 4\}\}$, then $\widehat{f}_n(x_0)$ is rate adaptive estimator of $f(x_0)$ over

$$\{\overline{N}_{p,4}(\beta, L, \mathcal{P}), (\beta, p, L, \mathcal{P}) \in (0, \beta_{\max}]^4 \times [1,\infty]^4 \times (0,\infty)^4 \times \overline{\mathfrak{P}}, r(\beta, p, \overline{\varnothing}) > 0\}.$$

In this case, the adaptive rate of convergence satisfies

$$\psi_n(\beta, p, \mathcal{P}) := \left(\frac{1}{n}\right)^{r_{\max}/(2r_{\max}+1)}, \qquad (\beta, p, \mathcal{P}) \in \{\beta^{(\max)}\} \times \{p^{(\max)}\} \times \{\mathcal{P}_1, \mathcal{P}_2\},$$
$$r_{\max} := \frac{\beta_{\max}}{2}.$$

Thus, in the considered example the aforementioned set contains two elements.

Finally, let us note that there is a "ln-price" to pay for adaptation with respect to the structure of independence even if the smoothness parameters β, L and p are known. This result follows from the bound (41) established in the proof of Theorem 5.

4. Discussion: Comparison with the global method in Lepski [29]

The latter paper deals with the rate optimal adaptive estimation of a probability density under sup-norm loss. It is obvious that the estimator constructed in Lepski [29] is fully data-driven and can be also used in pointwise estimation. However, this estimator is neither minimax nor optimally minimax adaptive when pointwise estimation is considered. Below, we discuss this issue in detail.

Oracle approach. Obviously, the use of a local method allows to control better the error of approximation since $\mathcal{B}_{(h,\mathcal{P})}(x_0)$ is smaller than $\sup_{x\in\mathbb{R}^d} \mathcal{B}_{(h,\mathcal{P})}(x)$. Moreover, our local method controls better the stochastic error since $\ln\delta(h,\mathcal{P})$ is smaller than $\ln(n)$. The latter fact is explained by the use of different constructions of the selection rule. First, it concerns the choice of the regularization parameter h. Whereas Lepski [29] uses kernel convolution, we use the "operation" $\vee$ on the set of bandwidth parameters. Next, in pointwise estimation, we select the parameter $(h,\mathcal{P})$ from very special set whose construction is new. It is important to emphasize that the consideration of the parameter set used in Lepski [29] is too "rough" in order to bring an optimal pointwise adaptive estimator. Both reasons required the introduction of novel technical arguments for pointwise estimation with respect to those in Lepski [29] for estimation under sup-norm loss; see the definition of our selection rule in Section 2.3, and the proofs of Proposition 2, Lemma 1 and Theorem 1 in the next section. Note, however, that the adaptation to eventual independence structure in both papers has rest upon the same methodology.

The following example illustrates clearly how the quality of estimation provided by Lepski's estimator can be significantly improved by application of our local method.

Example 3. Considering the problem described in Example 1, we compare both methods.

- *Local method.* We obtain from our local oracle inequality that

$$\left(\mathbb{E}_f^{(n)}\left|\widehat{f}_n(x_0) - f_d(x_0)\right|^q\right)^{1/q} \le \left(\alpha_1\sqrt{4\ln(4)} + \alpha_2\right)n^{-1/2}, \qquad \alpha_1, \alpha_2 > 0.$$

- *Global method.* The best quality of estimation provided by Theorem 1 in Lepski [29] is

$$\left(\mathbb{E}_f^{(n)}\left|\widetilde{f}_n(x_0) - f_d(x_0)\right|^q\right)^{1/q} \le (2C_1 + C_2)\left(n/\ln(n)\right)^{-1/3}, \qquad C_1, C_2 > 0.$$

It is also important to emphasize that our Theorem 1 presents other advantages with respect to that in Lepski [29].

(a) We derive our oracle-type inequality over the functional class $\mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}]$ which contains the class $\mathbb{F}_d[\mathbf{f}]$ used in Lepski [29] that allows to obtain upper bounds under more general assumptions. For instance, if $\overline{\mathfrak{P}} = \{\{1, \dots, d\}\}$, we do not need that all marginals are uniformly bounded, that is not true when we use Theorem 1 in Lepski [29]; see our Corollary 1 above.

(b) The oracle-type inequality for sup-norm risk cannot be used in general for other type of loss functions. Contrary to this, the pointwise risk can be integrated that allows to obtain the results under $\mathbb{L}_p$-loss; see, for example, Lepski, Mammen and Spokoiny [30] and Goldenshluger and Lepski [17]. In this context, the establishing of local oracle inequality with the term $\ln \delta(h, \mathcal{P})$ instead of $\ln(n)$ is crucial.

Minimax adaptive estimation. Comparing the minimax rate of convergence defined by (17), we find a price to pay for adaptation in the pointwise setting. This does not exist in the estimation under sup-norm loss. Note nevertheless that this price to pay for adaptation is not unavoidable for all values of nuisance parameter $(\beta, p, L, \mathcal{P})$. This explains the necessity of the introduction of the optimality criterion presented in Section 3.3.2.

Let us also compare our results with those obtained in Lepski [29].

Example 4. Consider that $\overline{\mathfrak{P}}$ still contains the elements $\mathcal{P}_1$ and $\mathcal{P}_2$ defined in Example 1 and that $d = 2$. Put $\beta_{\max} = 1$.

- *Local method.* In view of our results, our estimator $\widehat{f}_n$ achieves the following minimax rate of convergence:

$$\inf_{\widehat{f}} \sup_{f \in \overline{N}_{\infty,2}(\beta^{(\max)}, L, \mathcal{P}_2)} \left(\mathbb{E}_f^{(n)} \left|\widehat{f}(x_0) - f(x_0)\right|^q\right)^{1/q} \asymp n^{-1/3},$$

  where infimum is taken over all possible estimators.
- *Global method.* In view of the results in Lepski [29], the estimator $\widetilde{f}_n$ proposed in the latter paper achieves the following minimax rate of convergence:

$$\inf_{\widehat{f}} \sup_{f \in \overline{N}_{\infty,2}(\beta^{(\max)}, L, \mathcal{P}_2)} \left(\mathbb{E}_f^{(n)} \|\widehat{f} - f\|_\infty^q\right)^{1/q} \asymp \left(n / \ln(n)\right)^{-1/3},$$

  where infimum is taken over all possible estimators.

Thus, the application of the procedure from Lepski [29] for pointwise adaptive estimation leads to the logarithmic loss of accuracy everywhere, while our estimator is rate optimal for some values of nuisance parameter.

5. Proofs of main results

The main technical tools used in the derivation of pointwise oracle inequality given in Theorem 1 are uniform bounds of empirical processes. We start this section with presenting of corresponding results those proof are postponed to the Appendix. In particular, we provide with the explicit expression of the constants $\lambda_s^{(q)}[\mathbf{K},\mathfrak{z}]$, $q\geq 1$, used in the selection rule (9)–(10). Our considerations here are mostly based on the results recently developed in Lepski [28].

5.1. Constants $\lambda_s^{(q)}[\mathbf{K},\mathfrak{z}]$

Set for any $s\in\mathbb{N}^*, q\geq 1$, $\lambda_s^{(q)}[\mathbf{K},\mathfrak{z}] := \{3q+sq[1\vee\mathfrak{z}](1+1/\underline{\tau})\}^{1/2}\lambda_s^{(q)}$, where $\underline{\tau} := \inf_{I\in\mathcal{I}_d}\tau(|I|)>0$,

$$\lambda_s^{(q)} := \lambda_s^{(q)}[\mathbf{K}] = \left\{\left(10se^s+\frac{10seL_{\mathbf{K}}}{\|\mathbf{K}\|_\infty}\right)\vee(48e)\right\}\left[\sqrt{7}+7\sqrt{(1+q)\|\mathbf{K}\|_\infty^s}\right]C_{s,1}^{(q)}\|\mathbf{K}\|_\infty^s$$

and $C_{s,1}^{(q)} := [144s\delta_*^{-2}+5q+3+36C_s]\vee 1$.

Here, δ_* is the smallest solution of the equation $8\pi^2\delta(1+[\ln\delta]^2)=1$ and

$$C_s := s\sup_{\delta>\delta_*}\frac{1}{\delta^2}\left[1+\ln\left(\frac{9216(s+1)\delta^2}{[s^*(\delta)]^2}\right)\right]_+ + s\sup_{\delta>\delta_*}\frac{1}{\delta^2}\left[1+\ln\left(\frac{9216(s+1)\delta}{s^*(\delta)}\right)\right]_+,$$

$$s^*(\delta) := \frac{(6/\pi^2)}{1+[\ln\delta]^2}.$$

5.2. Pointwise uniform bounds of kernel-type empirical processes

Let $s\in\mathbb{N}^*$, $s\leq d$, and let $Y_i=(Y_{i,1},\dots,Y_{i,s})$, $i\in\mathbb{N}^*$, be a sequence of $\mathbb{R}^s$-valued i.i.d. random vectors defined on a complete probability space $(\Omega,\mathfrak{A},\mathsf{P})$ and having the density g with respect to the Lebesgue measure. Later on $\mathbb{P}_g^{(n)}$ denotes the probability law of $Y^{(n)} := (Y_1,\dots,Y_n)$ and $\mathbb{E}_g^{(n)}$ is the mathematical expectation with respect to $\mathbb{P}_g^{(n)}$. Assume that $\|g\|_\infty\leq\mathbf{g}$ where $\mathbf{g}>0$ is a given number.

Set $a_s^{(q)} := (2\sqrt{1+q}[1\vee\lambda_s^{(q)}])^{-2}$ and

$$\mathcal{H}_n^{(s)} := \prod_{i=1}^s\left[h_i^{(\min)}(n),h_i^{(\max)}(n)\right]\subseteq\left[\frac{1}{n},1\right]^s, \qquad \mathfrak{H}_s^{(q)}(n) := \left\{h\in\mathcal{H}_n^{(s)}:\ nV_h\geq\left[a_s^{(q)}\right]^{-1}\ln(n)\right\}.$$

For any $h\in\mathcal{H}_n^{(s)}$, $y_0\in\mathbb{R}^s$ and $u\geq 1$ set also

$$K(y) := \prod_{i=1}^s\mathbf{K}(y_i), \qquad V_h := \prod_{i=1}^s h_i, \qquad K_h(y) := V_h^{-1}\prod_{i=1}^s\mathbf{K}(y_i/h_i) \qquad \forall y\in\mathbb{R}^s,$$

$$G_h(y_0) := 1\vee\left[\int_{\mathbb{R}^s}\left|K_h(y-y_0)\right|g(y)\,\mathrm{d}y\right], \qquad \widetilde{G}_h(y_0) := 1\vee\left[n^{-1}\sum_{i=1}^n\left|K_h(Y_i-y_0)\right|\right],$$

$$\mathcal{U}_h^{(u)}(y_0) := \sqrt{\frac{[G_h(y_0)]^2}{nV_h}\left\{1\vee\ln\left(\frac{V_{h^{(\max)}}}{V_h}\right)+u\right\}}.$$

For a given $y_0 \in \mathbb{R}^s$ consider the empirical processes

$$\xi_h^{(n)}(y_0) := n^{-1}\sum_{i=1}^{n}\big[K_h(Y_i - y_0) - \mathbb{E}_g^{(n)}\big\{K_h(Y_i - y_0)\big\}\big], \qquad h \in \mathcal{H}_n^{(s)},$$

$$\overline{\xi}_h^{(n)}(y_0) := n^{-1}\sum_{i=1}^{n}\big[\big|K_h(Y_i - y_0)\big| - \mathbb{E}_g^{(n)}\big\{\big|K_h(Y_i - y_0)\big|\big\}\big], \qquad h \in \mathcal{H}_n^{(s)}.$$

Proposition 2. *For all $q \ge 1$, all integer $n \ge 3$ and all number u satisfying $1 \le u \le q\ln(n)$*

(i) $$\mathbb{E}_g^{(n)}\Big\{\sup_{h\in\mathfrak{H}_s^{(q)}(n)}\big[\big|\xi_h^{(n)}(y_0)\big| - \lambda_s^{(q)}\mathcal{U}_h^{(u)}(y_0)\big]_+\Big\}^q \le C_s^{(q)}(\mathbf{K},\mathbf{g})[nV_{h^{(\max)}}]^{-q/2}\mathrm{e}^{-u};$$

(ii) $$\mathbb{E}_g^{(n)}\Big\{\sup_{h\in\mathfrak{H}_s^{(q)}(n)}\Big[\big|\overline{\xi}_h^{(n)}(y_0)\big| - \frac{1}{2}G_h(y_0)\Big]_+\Big\}^q \le C_s^{(q)}(\mathbf{K},\mathbf{g})[nV_{h^{(\max)}}]^{-q/2}\mathrm{e}^{-u};$$

(iii) $$\Big(\mathbb{E}_g^{(n)}\Big\{\sup_{h\in\mathfrak{H}_s^{(q)}(n)}\big[G_h(y_0) - 2\widetilde{G}_h(y_0)\big]_+\Big\}^q\Big)^{1/q} \le 2\big[C_s^{(q)}(\mathbf{K},\mathbf{g})\big]^{1/q}[nV_{h^{(\max)}}]^{-1/2}\mathrm{e}^{-u/q}.$$

The expression of the constant $C_s^{(q)}(\mathbf{K},\mathbf{g})$ is given in the proof of the proposition.

5.3. Oracle-type inequality

5.3.1. *Auxiliary result*

For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$ set

$$b_{h_I}(x_{0,I}) := \int_{\mathbb{R}^{|I|}} K_{h_I}^{(I)}(x_I - x_{0,I}) f_I(x_I)\,\mathrm{d}x_I, \qquad \xi_{h_I}^{(n)}(x_{0,I}) := \widehat{f}_{h_I}^{(n)}(x_{0,I}) - b_{h_I}(x_{0,I});$$

$$G_{h_I}(x_{0,I}) := 1 \vee \Big[\int_{\mathbb{R}^{|I|}}\big|K_{h_I}^{(I)}(x_I - x_{0,I})\big| f(x_I)\,\mathrm{d}x_I\Big],$$

$$G(x_0) := \sup_{(h,\mathcal{P})\in\overline{\mathfrak{H}}[\mathfrak{P}]}\ \sup_{(\eta,\mathcal{P}')\in\overline{\mathfrak{H}}[\mathfrak{P}]}\ \sup_{I\in\mathcal{P}\circ\mathcal{P}'} G_{h_I\vee\eta_I}(x_{0,I}).$$

For any $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$ put

$$\mathcal{U}_{(h,\mathcal{P})}(x_0) := \sqrt{\frac{[G(x_0)]^2\{1 \vee \ln\delta(h,\mathcal{P})\}}{nV(h,\mathcal{P})}}.$$

Define also $\bar{\mathbf{f}}_n(x_0) := 12\lambda d^3(2\max\{\overline{G}_n(x_0), 1 \vee \mathbf{f}\|\mathbf{K}\|_1^d\})^{d^2}$ and

$$\xi_n(x_0) := \sup_{(h,\mathcal{P})\in\overline{\mathfrak{H}}[\mathfrak{P}]}\ \sup_{(\eta,\mathcal{P}')\in\overline{\mathfrak{H}}[\mathfrak{P}]}\ \sup_{I\in\mathcal{P}\circ\mathcal{P}'}\big[\big|\xi_{h_I\vee\eta_I}^{(n)}(x_{0,I})\big| - \lambda\big\{\mathcal{U}_{(h,\mathcal{P})}(x_0) + \mathcal{U}_{(\eta,\mathcal{P}')}(x_0)\big\}\big]_+.$$

Lemma 1. *Set* $\mathbf{f} > 0$. *For any* $q \geq 1$ *there exist constants* $\mathbf{c}_i := \mathbf{c}_i(2q, d, \mathbf{K}, \mathbf{f}, \mathfrak{z})$, $i = 1, 2, 3, 4$, *such that* $\forall n \geq 3$, $\forall f \in \mathbb{F}_d[\mathbf{f}, \overline{\mathfrak{P}}]$, $\forall (h, \mathcal{P}) \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$, $\mathcal{P} \in \mathfrak{P}(f)$,

$$\text{(i)} \qquad \left(\mathbb{E}_f^{(n)} \left|\xi_n(x_0)\right|^{2q}\right)^{1/2q} \leq \mathbf{c}_1 [n V_{\max}]^{-1/2};$$

$$\text{(ii)} \qquad \left(\mathbb{E}_f^{(n)} \left[G(x_0) - \overline{G}_n(x_0)\right]_+^{2q}\right)^{1/2q} \leq \mathbf{c}_2 [n V_{\max}]^{-1/2};$$

$$\text{(iii)} \qquad \left(\mathbb{E}_f^{(n)} \left|\bar{\mathbf{f}}_n(x_0)\right|^{2q}\right)^{1/2q} \leq \mathbf{c}_3;$$

$$\text{(iv)} \qquad \left(\mathbb{E}_f^{(n)} \left|\widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\right|^{2q}\right)^{1/2q} \leq \mathbf{c}_4 \mathcal{U}_{(h,\mathcal{P})}(x_0).$$

5.3.2. *Proof of Theorem* 1

We divide the proof into several steps.

(1) Let $(h, \mathcal{P}) \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$, $\mathcal{P} \in \mathfrak{P}(f)$, be fixed. By the triangle inequality, we have

$$\begin{aligned} \left|\widehat{f}_n(x_0) - f(x_0)\right| &\leq \left|\widehat{f}^{(n)}_{(\widehat{h},\widehat{\mathcal{P}})}(x_0) - \widehat{f}^{(n)}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})}(x_0)\right| + \left|\widehat{f}^{(n)}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})}(x_0) - \widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0)\right| \\ &\quad + \left|\widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0) - f(x_0)\right| \\ &\leq 2\left[\widehat{\Delta}_{(h,\mathcal{P})}(x_0) + 2\Lambda_n(x_0)\widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\right] + \left|\widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0) - f(x_0)\right|. \end{aligned} \tag{23}$$

Here, we have used that $\widehat{f}^{(n)}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})}(x_0) = \widehat{f}^{(n)}_{(\widehat{h},\widehat{\mathcal{P}}),(h,\mathcal{P})}(x_0)$ and the definition of $(\widehat{h}, \widehat{\mathcal{P}})$.

In what follows, we will use the inequality: for $m \in \mathbb{N}^*$ and $a_i, b_i \in \mathbb{R}$, $i = \overline{1, m}$,

$$\left|\prod_{i=1}^m a_i - \prod_{i=1}^m b_i\right| \leq m\Big(\sup_{i=\overline{1,m}} \max\big\{|a_i|, |b_i|\big\}\Big)^{m-1} \sup_{i=\overline{1,m}} |a_i - b_i|. \tag{24}$$

Here and later, we assume that the product and the supremum over empty set are equal to one and zero, respectively.

(2) Since $\mathcal{P} \in \mathfrak{P}(f)$, using (24) we have

$$\begin{aligned} \left|\widehat{f}^{(n)}_{(h,\mathcal{P})}(x_0) - f(x_0)\right| &\leq d\Big(\sup_{I\in\mathcal{P}} \max\big\{\widehat{G}_{h_I}(x_{0,I}), \mathbf{f}\big\}\Big)^{d-1} \sup_{I\in\mathcal{P}} \left|\widehat{f}^{(n)}_{h_I}(x_{0,I}) - f_I(x_{0,I})\right| \\ &\leq d\big(\max\big\{\overline{G}_n(x_0), \mathbf{f}\big\}\big)^{d-1} \left[\mathcal{B}_{(h,\mathcal{P})}(x_0) + \xi_n(x_0) + 2\lambda \mathcal{U}_{(h,\mathcal{P})}(x_0)\right], \end{aligned} \tag{25}$$

since $\overline{G}_n(x_0) \geq \widehat{G}_{h_I}(x_{0,I}) \geq 1$ and $|\widehat{f}^{(n)}_{h_I}(x_{0,I}) - f_I(x_{0,I})| \leq |\xi^{(n)}_{h_I}(x_{0,I})| + |b_{h_I}(x_{0,I}) - f_I(x_{0,I})|$, $\forall I \in \mathcal{P}$.

(3) Set $\bar{\mathbf{f}}^{(1)}_n := d[\overline{G}_n(x_0)]^{d(d-1)}$. For any $(\eta, \mathcal{P}') \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$, we get from the inequality (24)

$$\left|\widehat{f}^{(n)}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x_0) - \widehat{f}^{(n)}_{(\eta,\mathcal{P}')}(x_0)\right| \leq \bar{\mathbf{f}}^{(1)}_n \sup_{I'\in\mathcal{P}'} \left|\prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} \widehat{f}^{(n)}_{h_{I\cap I'},\eta_{I\cap I'}}(x_{0,I\cap I'}) - \widehat{f}^{(n)}_{\eta_{I'}}(x_{0,I'})\right|.$$

Introduce, for all $I \in \mathcal{I}_d$ and all $\eta \in (0, 1]^d$, $b_{h_I,\eta_I}(x_{0,I}) := \int_{\mathbb{R}^{|I|}} K^{(I)}_{h_I\vee\eta_I}(u - x_{0,I}) f_I(u)\, \mathrm{d}u$.

Put also $\bar{\mathbf{f}}_n^{(2)} := d(\max\{\overline{G}_n(x_0), G(x_0)\})^{d-1}$. For any $(\eta, \mathcal{P}') \in \bar{\mathfrak{H}}[\overline{\mathfrak{P}}]$ and any $I' \in \mathcal{P}'$, in view of (24),

$$\begin{aligned}
&\Biggl|\prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} \widehat{f}^{(n)}_{h_{I\cap I'},\eta_{I\cap I'}}(x_{0,I\cap I'}) - \prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} b_{h_{I\cap I'},\eta_{I\cap I'}}(x_{0,I\cap I'})\Biggr| \\
&\quad\le \bar{\mathbf{f}}_n^{(2)} \sup_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} \bigl|\xi^{(n)}_{h_{I\cap I'}\vee\eta_{I\cap I'}}(x_{0,I\cap I'})\bigr|, \\
&\Biggl|\prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} b_{h_{I\cap I'},\eta_{I\cap I'}}(x_{0,I\cap I'}) - b_{\eta_{I'}}(x_{0,I'})\Biggr| \le \bar{\mathbf{f}}_n^{(2)} \mathcal{B}_{(h,\mathcal{P})}(x_0).
\end{aligned}$$

For the last inequality, we have used that $\mathcal{P} \in \mathfrak{P}(f)$ and, therefore, for any $\eta \in (0,1]^d$ and any $I' \in \mathcal{I}_d$

$$b_{\eta_{I'}}(x_{0,I'}) = \int_{\mathbb{R}^{|I'|}} K^{(I')}_{\eta_{I'}}(x_{I'} - x_{0,I'}) \prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} f_{I\cap I'}(x_{I\cap I'})\,\mathrm{d}x_{I'} = \prod_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} b_{\eta_{I\cap I'}}(x_{0,I\cap I'}).$$

(4) Applying the triangle inequality, we get since $\bar{\mathbf{f}}_n^{(2)} \ge 1$ and $\mathcal{U}_{(h,\mathcal{P})}(x_0) > 0$, for any $(\eta, \mathcal{P}') \in \bar{\mathfrak{H}}[\overline{\mathfrak{P}}]$,

$$\begin{aligned}
&\bigl|\widehat{f}^{(n)}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x_0) - \widehat{f}^{(n)}_{(\eta,\mathcal{P}')}(x_0)\bigr| \\
&\quad\le \bar{\mathbf{f}}_n^{(1)} \sup_{I'\in\mathcal{P}'} \Bigl\{\bar{\mathbf{f}}_n^{(2)} \sup_{I\in\mathcal{P}:\ I\cap I'\neq\varnothing} \bigl|\xi^{(n)}_{h_{I\cap I'}\vee\eta_{I\cap I'}}(x_{0,I\cap I'})\bigr| + \bar{\mathbf{f}}_n^{(2)}\mathcal{B}_{(h,\mathcal{P})}(x_0) + \bigl|\xi^{(n)}_{\eta_{I'}}(x_{0,I'})\bigr|\Bigr\} \\
&\quad\le \bar{\mathbf{f}}_n^{(1)}\bar{\mathbf{f}}_n^{(2)}\mathcal{B}_{(h,\mathcal{P})}(x_0) + 2\bar{\mathbf{f}}_n^{(1)}\bar{\mathbf{f}}_n^{(2)}\xi_n(x_0) + 3\lambda\bar{\mathbf{f}}_n^{(1)}\bar{\mathbf{f}}_n^{(2)}\bigl\{\mathcal{U}_{(\eta,\mathcal{P}')}(x_0) + \mathcal{U}_{(h,\mathcal{P})}(x_0)\bigr\}.
\end{aligned}$$

Put $\widetilde{\mathbf{f}}_n^{(2)} := d[2\overline{G}_n(x_0)]^{d-1}$ and $\mathcal{U}(x_0) := \sup_{(\eta,\mathcal{P}')\in\bar{\mathfrak{H}}[\overline{\mathfrak{P}}]}\mathcal{U}_{(\eta,\mathcal{P}')}(x_0)$. We obtain that

$$\begin{aligned}
&\widehat{\Delta}_{(h,\mathcal{P})}(x_0) \\
&\quad\le 2\bar{\mathbf{f}}_n^{(1)}\bar{\mathbf{f}}_n^{(2)}\bigl\{\mathcal{B}_{(h,\mathcal{P})}(x_0) + \xi_n(x_0)\bigr\} + 3\lambda\bar{\mathbf{f}}_n^{(1)}\bigl\{\mathcal{U}(x_0) + \mathcal{U}_{(h,\mathcal{P})}(x_0)\bigr\}\bigl[\bar{\mathbf{f}}_n^{(2)} - \widetilde{\mathbf{f}}_n^{(2)}\bigr]_+ \\
&\qquad + 3\lambda\bar{\mathbf{f}}_n^{(1)}\widetilde{\mathbf{f}}_n^{(2)}\Bigl\{\sup_{(\eta,\mathcal{P}')\in\bar{\mathfrak{H}}[\overline{\mathfrak{P}}]}\bigl[\mathcal{U}_{(\eta,\mathcal{P}')}(x_0) - \widehat{\mathcal{U}}_{(\eta,\mathcal{P}')}(x_0)\bigr]_+ \\
&\qquad\qquad + \bigl[\mathcal{U}_{(h,\mathcal{P})}(x_0) - \widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\bigr]_+\Bigr\};
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\widehat{\Delta}_{(h,\mathcal{P})}(x_0) \\
&\quad\le \bar{\mathbf{f}}_n(x_0)\bigl\{\mathcal{B}_{(h,\mathcal{P})}(x_0) + \xi_n(x_0) + \bigl[G(x_0) - \overline{G}_n(x_0)\bigr]_+\bigr\},
\end{aligned}$$

where $\bar{\mathbf{f}}_n(x_0) := 12\lambda d^3(2\max\{\overline{G}_n(x_0), 1\vee \mathbf{f}\|\mathbf{K}\|_1^d\})^{d^2}$, since $\lambda \wedge \|\mathbf{K}\|_1 \geq 1$,

$$\mathcal{U}_{(\eta,\mathcal{P}')}(x_0) \leq \big(1\vee \mathbf{f}\|\mathbf{K}\|_1^d\big)\sqrt{\frac{1\vee \ln\delta(h,\mathcal{P})}{nV(h,\mathcal{P})}} \leq 1\vee \mathbf{f}\|\mathbf{K}\|_1^d \qquad \forall\big(\eta,\mathcal{P}'\big)\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}],$$

and $[a^m - b^m]_+ \leq m(\max\{a,b\})^{m-1}[a-b]_+$, $\forall a,b>0$, $\forall m\in\mathbb{N}^*$.

(5) Finally, we deduce from (23), (25) and (26), using again $\lambda \wedge \|\mathbf{K}\|_1 \geq 1$, that

$$\begin{aligned}&\big|\widehat{f}_n(x_0) - f(x_0)\big| \\ &\quad\leq 3\bar{\mathbf{f}}_n(x_0)\big\{\mathcal{B}_{(h,\mathcal{P})}(x_0) + \mathcal{U}_{(h,\mathcal{P})}(x_0) + \widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0) + \xi_n(x_0) + \big[G(x_0) - \overline{G}_n(x_0)\big]_+\big\}.\end{aligned} \tag{27}$$

By the Cauchy–Schwarz inequality

$$\begin{aligned}&\big(\mathbb{E}_f^{(n)}\big|\widehat{f}_n(x_0) - f(x_0)\big|^q\big)^{1/q} \\ &\quad\leq 3\big(\mathbb{E}_f^{(n)}\big|\bar{\mathbf{f}}_n(x_0)\big|^{2q}\big)^{1/(2q)}\big[\mathcal{B}_{(h,\mathcal{P})}(x_0) + \mathcal{U}_{(h,\mathcal{P})}(x_0) + \big(\mathbb{E}_f^{(n)}\big|\widehat{\mathcal{U}}_{(h,\mathcal{P})}(x_0)\big|^{2q}\big)^{1/(2q)} \\ &\qquad\qquad + \big(\mathbb{E}_f^{(n)}\big|\xi_n(x_0)\big|^{2q}\big)^{1/(2q)} + \big(\mathbb{E}_f^{(n)}\big[G(x_0) - \overline{G}_n(x_0)\big]_+^{2q}\big)^{1/(2q)}\big].\end{aligned}$$

Applying Lemma 1,

$$\big(\mathbb{E}_f^{(n)}\big|\widehat{f}_n(x_0) - f(x_0)\big|^q\big)^{1/q} \leq 3\mathbf{c}_3\big[\mathcal{B}_{(h,\mathcal{P})}(x_0) + (1+\mathbf{c}_4)\mathcal{U}_{(h,\mathcal{P})}(x_0) + (\mathbf{c}_1+\mathbf{c}_2)[nV_{\max}]^{-1/2}\big],$$

and we come to the assertion of Theorem 1 with $\alpha_1 = 3\mathbf{c}_3(1+\mathbf{c}_4)(1\vee\mathbf{f}\|\mathbf{K}\|_1^d)$ and $\alpha_2 = 3\mathbf{c}_3(\mathbf{c}_1+\mathbf{c}_2)$.

5.4. Lower bound for minimax estimation

5.4.1. *Auxiliary result*

The result formulated in Lemma 2 below is a direct consequence of the general bound obtain in Kerkyacharian, Lepski and Picard [25], Proposition 7.

Let $(\beta, p, \mathcal{P}) \in (0,\infty)^d \times [1,\infty]^d \times \mathfrak{P}$ and $L\in(0,\infty)^d$ be fixed.

Lemma 2. *Suppose that there exists* $\{f_0, f_1\}\subset N^*_{p,d}(\beta, L, \mathcal{P})$ *such that* $\mathbb{P}_{f_1}^{(n)}$ *is absolutely continuous with respect to* $\mathbb{P}_{f_0}^{(n)}$ *and*

$$\big|f_1(x_0) - f_0(x_0)\big| \geq s_n(\beta, p, \mathcal{P}); \tag{28}$$

$$\limsup_{n\to+\infty}\mathbb{E}_{f_0}^{(n)}\bigg[\frac{\mathrm{d}\mathbb{P}_{f_1}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}}\big(X^{(n)}\big) - 1\bigg]^2 \leq C < \infty. \tag{29}$$

Then, for all $q \ge 1$,

$$\liminf_{n\to+\infty}\Big\{s_n^{-1}(\beta,p,\mathcal{P})\inf_{\widetilde{f}_n}\mathcal{R}_n^{(q)}\big[\widetilde{f}_n, N^*_{p,d}(\beta,L,\mathcal{P})\big]\Big\} \ge \frac{1}{2}\big(1-\sqrt{C/(C+4)}\big),$$

where infimum is taken over all possible estimators.

5.4.2. *Proof of Proposition* 1

Set $\mathcal{N}(x) := \prod_{i=1}^d \sqrt{2\pi}^{-1}\exp(-x_i^2/2)$ and let $f_0(x) := \sigma^{-1}\mathcal{N}(x/\sigma)$. It is easily seen that one can find $\sigma > 0$ such that

$$\begin{aligned} f_0 &\in N^*_{p,d}(\beta, \underline{L}/2, \mathcal{P}) \subseteq N^*_{p,d}(\beta, L, \mathcal{P}),\\ \underline{L}_i &:= 2\wedge L_i, \qquad i=\overline{1,d}. \end{aligned} \tag{30}$$

Let $I = \{i_1,\dots,i_m\}\in\mathcal{P}$ be such that $r := r(\beta,p,\mathcal{P}) = \gamma_I(\beta,p)$ and $g:\mathbb{R}\to\mathbb{R}$ such that $\operatorname{supp}(g)\subseteq(-1/2,1/2)$, $g\in\bigcap_{i\in I}N_{p_i,1}(\beta_i,1/2)$, $\int g = 0$, and $|g(0)| = \|g\|_\infty$. Define

$$G(x_I) = A_n\prod_{l=1}^m g\left(\frac{x_{i_l}-x_{0,i_l}}{\delta_{l,n}}\right),$$

where $A_n, \delta_{l,n}\to 0$, $l=\overline{1,m}$, if $n\to\infty$, will be chosen later. Note that $G\in\mathbb{N}_{p_I,|I|}(\beta_I,\underline{L}_I/2)$ if

$$A_n\delta_{l,n}^{-\beta_{i_l}}\left(\prod_{j=1}^m\delta_{j,n}\right)^{1/p_{i_l}} \le \frac{\underline{L}_{i_l}}{c_l}, \qquad l=\overline{1,m}, c_l = \|g\|_{p_{i_l}}^{m-1}. \tag{31}$$

Introduce

$$f_1(x) = \prod_{i\notin I}\big\{\big[2\pi\sigma^2\big]^{-1}\exp\big(-x_i^2/2\sigma^2\big)\big\}\left\{\prod_{i\in I}\big[2\pi\sigma^2\big]^{-1}\exp\big(-x_i^2/2\sigma^2\big)+G(x_I)\right\}. \tag{32}$$

It is obvious that there exists $A_0 > 0$ such that if $A_n \le A_0$ then $f_1(x) > 0$ for any $x\in\mathbb{R}^d$. Note also that the condition $\int g = 0$ implies that $\int f_1 = 1$. We conclude that f_1 is a probability density. Furthermore, assumptions (30)–(31) and the definition of f_0 allow us to assert that $f_1\in N^*_{p,d}(\beta,L,\mathcal{P})$. We remark that

$$\big|f_1(x_0)-f_0(x_0)\big| = c_1^*A_n, \qquad c_1^* := (\sigma\sqrt{2\pi})^{m-d}\big|g(0)\big|^m\prod_{i\notin I}\exp\big(-x_{0,i}^2/2\sigma^2\big).$$

Then Assumption (28) of Lemma 2 is fulfilled when $s_n(\beta,p,\mathcal{P})\le c_1^*A_n$.

Since $X_k, k=\overline{1,n}$, are i.i.d. random fields and $\int g=0$ it is easily check that

$$\mathbb{E}_{f_0}^{(n)}\left[\frac{\mathrm{d}\mathbb{P}_{f_1}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}}\big(X^{(n)}\big)\right]^2 \le \left[1+\frac{2}{f_{0,I}(x_{0,I})}A_n^2\left(\prod_{j=1}^m \delta_{j,n}\right)\|g\|_2^{2m}\right]^n$$
$$\le \exp\left[\frac{2\|g\|_2^{2m}}{f_{0,I}(x_{0,I})}nA_n^2\left(\prod_{j=1}^m \delta_{j,n}\right)\right],$$

for n large enough. Here, we have used that $\operatorname{supp}(G)\subseteq \Pi_n := \prod_{l=1}^m [x_{0,i_l}-\delta_{l,n}/2, x_{0,i_l}+\delta_{l,n}/2]$ and that $\inf_{x_I\in\Pi_n} f_{0,I}(x_I)\ge f_{0,I}(x_{0,I})/2$ for n large enough.

Since $\mathbb{E}_{f_0}^{(n)}[\frac{\mathrm{d}\mathbb{P}_{f_1}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}}(X^{(n)})]=1$, Assumption (29) of Lemma 2 is fulfilled if

$$\exp\left[\frac{2\|g\|_2^{2m}}{f_{0,I}(x_{0,I})}nA_n^2\left(\prod_{j=1}^m \delta_{j,n}\right)\right]-1\le C.$$

The latter inequality holds if

$$nA_n^2\left(\prod_{j=1}^m \delta_{j,n}\right)\le t^2, \qquad t:=\sqrt{\left[c_2^*\right]^{-1}\ln(C+1)}, \qquad c_2^*:=\frac{2\|g\|_2^{2m}}{f_{0,I}(x_{0,I})}. \tag{33}$$

To finalize our proof, we study separately two cases: $r>0$ and $r\le 0$. Note first that $r=(1-1/s_I)/(1/\beta_I)$, where

$$\frac{1}{s_I}:=\sum_{i\in I}\frac{1}{\beta_i p_i}, \qquad \frac{1}{\beta_I}:=\sum_{i\in I}\frac{1}{\beta_i},$$

(1) *Case* $r>0$. Solving the system

$$A_n\delta_{l,n}^{-\beta_{i_l}}\left(\prod_{j=1}^m \delta_{j,n}\right)^{1/p_{i_l}}=\frac{L_{i_l}}{c_l}, \qquad l=\overline{1,m}, \qquad nA_n^2\left(\prod_{j=1}^m \delta_{j,n}\right)=t^2,$$

we obtain

$$\delta_{l,n}=\left(\frac{c_l}{L_{i_l}}\right)^{1/\beta_{i_l}}\left(\frac{t^2}{n}\right)^{1/(\beta_{i_l}p_{i_l})}A_n^{1/\beta_{i_l}-2/(\beta_{i_l}p_{i_l})}, \qquad A_n=R\left(\frac{t^2}{n}\right)^{r/(2r+1)},$$
$$R=\left[\prod_{l=1}^m\left(\frac{L_{i_l}}{c_l}\right)^{1/(2\beta_{i_l})}\right]^{1/(1-1/s_I-1/2\beta_I)}.$$

It is easily seen that $A_n, \delta_{l,n}\to 0, l=\overline{1,m}$, if $n\to\infty$ and one can choose $C=1$.

We conclude that, if $r > 0$, Lemma 2 is applicable with $s_n(\beta, p, \mathcal{P}) = c_1^* R(\frac{t^2}{n})^{r/(2r+1)}$.

(2) *Case* $r \le 0$. We choose $A_n \equiv A$, where the constant A satisfies $0 < A < A_0$. Solving the system

$$A\delta_{l,n}^{-\beta_{i_l}} \left(\prod_{j=1}^{m} \delta_{l,n}\right)^{1/p_{i_l}} \le \frac{\underline{L}_{i_l}}{c_l}, \qquad l = \overline{1,m}, \qquad nA^2 \left(\prod_{j=1}^{m} \delta_{j,n}\right) \le t^2,$$

$$\delta_{l,n} \ge \left(\frac{Ac_l}{\underline{L}_{i_l}}\right)^{1/\beta_{i_l}} \left(\prod_{j=1}^{m} \delta_{j,n}\right)^{1/(p_{i_l}\beta_{i_l})}, \qquad \prod_{j=1}^{m} \delta_{j,n} \le R_2 n^{-1},$$

$$R_2 = \frac{\ln(C+1)}{c_2^* A^2}.$$

Note that one can choose A such that $\max_{l=\overline{1,m}} (\frac{Ac_l}{\underline{L}_{i_l}})^{1/\beta_{i_l}} \le 1$ and $C = 1$. Since $s_I \le 1$, we obtain the following solution:

$$\delta_{l,n} = \left(\frac{R_2}{n}\right)^{s_I/(p_{i_l}\beta_{i_l})} \to 0, \qquad l = \overline{1,m}, n \to \infty.$$

We conclude that, if $r \le 0$, Lemma 2 is applicable with $s_n(\beta, p, \mathcal{P}) = c_1^* A$.

This completes the proof of Proposition 1.

5.5. Upper bounds for minimax and adaptive minimax estimation

The proof of Theorems 3 and 4 is based on application of Theorem 1. Note that in view of the embedding theorem for anisotropic Nikolskii classes (formulated in the proof of Lemma 3), there exists a number $\mathbf{f} := \mathbf{f}(\beta, p) > 0$ such that $\sup_{I \in \mathcal{P}} \|f_I\|_\infty \le \mathbf{f}$ if $r(\beta, p, \mathcal{P}) > 0$ or such that $\sup_{\mathcal{P} \in \overline{\mathfrak{P}}^*} \sup_{I \in \mathcal{P}} \|f_I\|_\infty \le \mathbf{f}$ if $r(\beta, p, \overline{\emptyset}) > 0$. It makes possible the application of Theorem 1.

5.5.1. *Auxiliary result*

The result formulated in Lemma 3 below is a consequence of Theorem 6.9 in Nikolskii [34].

Let $l \ge 2$ be a fixed integer and $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$ be a fixed set of partitions of $\{1, \ldots, d\}$. Let $f \in \overline{N}_{p,d}(\beta, L, \mathcal{P})$, where $\beta \in (0, l]^d$, $\mathcal{P} \in \overline{\overline{\mathfrak{P}}}$, $p \in [0, \infty]^d$ satisfy $r(\beta, p, \mathcal{P}) > 0$ and $L \in (0, \infty)^d$.

Lemma 3. *There exists* $\mathbf{c} := \mathbf{c}(\mathbf{K}, d, p, l, \mathcal{P}) > 0$ *such that*

$$\mathcal{B}_{h_I,\eta_I}(x_{0,I}) \le \mathbf{c} \sum_{i \in I} L_i h_i^{\beta_i(I)} \qquad \forall \mathcal{P}' \in \overline{\mathfrak{P}}, \forall I \in \mathcal{P} \circ \mathcal{P}', \forall (h, \eta) \in (0, 1]^d \times [0, 1]^d,$$

where $\mathcal{B}_{h_I,\eta_I}(x_{0,I})$ *is defined in Section 2.4,* $\beta_i(I) := \varkappa(I)\beta_i \varkappa_i^{-1}(I)$, $\varkappa(I) := 1 - \sum_{k \in I}(\beta_k p_k)^{-1}$ *and* $\varkappa_i(I) := 1 - \sum_{k \in I}(p_k^{-1} - p_i^{-1})\beta_k^{-1}$.

The proof of this lemma is given in the Appendix.

5.5.2. *Proof of Theorem* 3

For all $I \in \mathcal{P}$, consider the following system of equations:

$$h_j^{\beta_j(I)} = h_i^{\beta_i(I)} = \sqrt{\frac{1}{nV_{h_I}}}, \qquad i, j \in I,$$

and let $\mathbf{h}_I$ denotes its solution. One can easily check that

$$\mathbf{h}_i = n^{-(\gamma_I(\beta,p)/(2\gamma_I(\beta,p)+1))(1/\beta_i(I))}, \qquad i \in I, I \in \mathcal{P}. \tag{34}$$

Here, we have used that $1/\gamma_I(\beta, p) = \sum_{i \in I} 1/\beta_i(I)$.

We note that $2^{-1}nV(\mathbf{h}, \mathcal{P}) \geq a^{-1}\ln(n)$ for all n large enough. To get the statement of the theorem, we will apply Theorem 1 with $\mathfrak{z} = 1$, $\tau(s) = 1$, $s = 1, \dots, d$, $\mathfrak{h}_I^{(I)} = \mathbf{h}_I$ if $I \in \mathcal{P}$ and $\mathfrak{h}_i^{(I)} = 1$ if $i \in I$, $I \notin \mathcal{P}$, $\overline{\mathfrak{H}} = \{\mathbf{h}\}$, $\overline{\mathfrak{P}} = \{\mathcal{P}\}$. Thus, $\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$ is non-empty for n large enough and we get

$$\mathcal{R}_n^{(q)}\big[\widehat{f}^{(n)}_{(\mathbf{h},\mathcal{P})}, f\big] \leq \alpha_1(\mathbf{c}\overline{L} \vee 1)\left[\sup_{I\in\mathcal{P}} \sum_{i\in I} \mathbf{h}_i^{\beta_i(I)} + \sup_{I\in\mathcal{P}} \sqrt{\frac{1}{nV_{\mathbf{h}_I}}}\right] + \alpha_2 \sup_{I\in\mathcal{P}} \sqrt{\frac{1}{nV_{\mathbf{h}_I}}}, \tag{35}$$

where $\overline{L} := \sup_{i=\overline{1,d}} L_i$. Here, we have used Lemma 3 and the definition of $\mathcal{B}_{(\mathbf{h},\mathcal{P})}(x_0)$.

We deduce from (34) and (35)

$$\mathcal{R}_n^{(q)}\big[\widehat{f}^{(n)}_{(\mathbf{h},\mathcal{P})}, f\big] \leq \big[2\alpha_1(\mathbf{c}\overline{L} \vee 1) + \alpha_2\big] \sup_{I\in\mathcal{P}} n^{-\gamma_I(\beta,p)/(2\gamma_I(\beta,p)+1)} = \big[2\alpha_1(\mathbf{c}\overline{L} \vee 1) + \alpha_2\big] n^{-r/(2r+1)}$$

and the assertion of Theorem 3 follows.

5.5.3. *Proof of Theorem* 4

Set $(\beta, p) \in (0, \beta_{\max}]^d \times [1, \infty]^d$ such that $r(\beta, p, \overline{\varnothing}) > 0$, $\mathcal{P} \in \overline{\mathfrak{P}}$, $L \in (0, \infty)^d$, and $f \in \overline{N}_{p,d}(\beta, L, \mathcal{P})$.

Let us first note the following simple fact. If $\mathcal{P}' \in \overline{\mathfrak{P}}$ and $J = I \cap I'$, $I \in \mathcal{P}$, $I' \in \mathcal{P}'$, we easily prove that $\beta_i(J) \geq \beta_i(I)$ $\forall i \in J$; see, for example, Lepski [29], proof of Theorem 3, for more details. Thus, in view of Lemma 3,

$$\mathcal{B}_{(h,\mathcal{P})}(x_0) \leq \mathbf{c} \sup_{I\in\mathcal{P}} \sum_{i\in I} L_i h_i^{\beta_i(I)} \qquad \forall h \in (0,1]^d. \tag{36}$$

Recall that $\mathfrak{h}_I^{(I)}$, $I \in \mathcal{I}_d$, is the projection on the dyadic grid in $(0,1]^{|I|}$ of $\mathrm{h}_I^{(I)}$ given in (20) and note that $2^{-1}nV_{\mathfrak{h}_I^{(I)}} \geq a^{-1}\ln(n)$ for n large enough. Thus, $\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$ is non-empty and one can apply Theorem 1.

If $r(\beta, p, \mathcal{P}) = r_{\max}$, then it is obvious that $(\beta, p) = (\beta^{(\max)}, p^{(\max)})$ and that $d(\mathcal{P}) = \overline{d}$. Thus, in view of the definition of the multibandwidths $\mathfrak{h}_I^{(I)}$, $I \in \mathcal{P}$, $\inf_{I\in\mathcal{P}} V_{\mathfrak{h}_I^{(I)}} = V_{\max}$. It follows from

Theorem 1 and (36)

$$\mathcal{R}_n^{(q)}[\widehat{f}_n, f] \le \alpha_1(\mathbf{c}\overline{L} \vee 1)\Bigg[\sup_{I\in\mathcal{P}}\sum_{i\in I}\big(\mathfrak{h}_i^{(I)}\big)^{\beta_{\max}} + \sup_{I\in\mathcal{P}}\sqrt{\frac{1}{nV_{\mathfrak{h}_I^{(I)}}}}\Bigg] + \alpha_2[nV_{\max}]^{-1/2},$$

where $\overline{L} := \sup_{i=\overline{1,d}} L_i$. Since $r_{\max} = \beta_{\max}/\overline{d}$, we conclude that there exists a constant $C > 0$ such that

$$\mathcal{R}_n^{(q)}[\widehat{f}_n, f] \le C\big[\alpha_1(\mathbf{c}\overline{L} \vee 1)(d+1) + \alpha_2\big]n^{-r_{\max}/(2r_{\max}+1)}. \tag{37}$$

If $r(\beta, p, \mathcal{P}) < r_{\max}$ we solve, for all $I \in \mathcal{P}$, the system

$$L_j h_j^{\beta_j(I)} = L_i h_i^{\beta_i(I)} = \sqrt{\frac{\ln(n)}{nV_{h_I}}}, \qquad i, j \in I.$$

The solution is

$$\begin{aligned} h_i &= L_i^{-1/\beta_i(I)}\left(\frac{L(I)\ln(n)}{n}\right)^{\gamma_I(\beta,p)/(2\gamma_I(\beta,p)+1)1/\beta_i(I)}, \\ L(I) &= \prod_{i\in I} L_i^{1/\beta_i(I)}, \qquad i \in I, I \in \mathcal{P}. \end{aligned} \tag{38}$$

It is easily seen that $(h, \mathcal{P}) \in \mathfrak{H}[\mathfrak{P}]$ for n large enough. Replacing h by its projection $\bar{h}$ on the dyadic grid $\overline{\mathfrak{H}}$, one has $(\bar{h}, \mathcal{P}) \in \overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$ for n large enough. We deduce from Theorem 1 and (36)

$$\mathcal{R}_n^{(q)}[\widehat{f}_n, f] \le \alpha_1\Bigg[\mathbf{c}\sup_{I\in\mathcal{P}}\sum_{i\in I} L_i \bar{h}_i^{\beta_i(I)} + \sup_{I\in\mathcal{P}}\sqrt{\frac{\ln(n)}{nV_{\bar{h}_I}}}\Bigg] + \alpha_2[nV_{\max}]^{-1/2}. \tag{39}$$

The assertion of Theorem 4 follows from (37), (38) and (39).

5.6. Lower bound for adaptive minimax estimation and optimal rate

5.6.1. *Auxiliary result*

To get the assertion of Theorem 5, we use the following lemma which is due to an oral communication with O. Lepski. This result can be viewed as a generalization of Lemma 2.

Let $(\beta, p) \in (0, \beta_{\max}]^d \times [1, \infty]^d$ such that $r(\beta, p, \overline{\varnothing}) > 0$, $\mathcal{P} \in \overline{\mathfrak{P}}$, $L \in (0, \infty)^d$ and $(\beta', p') \in (0, \beta_{\max}]^d \times [1, \infty]^d$ such that $r(\beta', p', \overline{\varnothing}) > 0$, $\mathcal{P}' \in \overline{\mathfrak{P}}$, $L' \in (0, \infty)^d$ be fixed.

Lemma 4. *Set (a_n) and (b_n) two sequences such that $a_n, b_n, b_n/a_n \to \infty, n \to \infty$. Suppose that exist $f_0 \in N_2 := \overline{N}_{p',d}(\beta', L', \mathcal{P}')$ and $f_1 \in N_1 := \overline{N}_{p,d}(\beta, L, \mathcal{P})$ such that $\mathbb{P}_{f_1}^{(n)}$ is absolutely continuous with respect to $\mathbb{P}_{f_0}^{(n)}$ and*

$$\big|f_1(x_0) - f_0(x_0)\big| = a_n^{-1}; \qquad \mathbb{E}_{f_0}^{(n)}\left[\frac{\mathrm{d}\mathbb{P}_{f_1}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}}\big(X^{(n)}\big)\right]^2 \le \frac{b_n}{a_n}. \tag{40}$$

Then, for any $q \geq 1$,

$$\liminf_{n\to+\infty}\inf_{\widetilde{f}_n}\Big[\sup_{f\in N_1}\mathbb{E}_f^{(n)}\big\{a_n\big|\widetilde{f}_n(x_0)-f(x_0)\big|\big\}^q+\sup_{f\in N_2}\mathbb{E}_f^{(n)}\big\{b_n\big|\widetilde{f}_n(x_0)-f(x_0)\big|\big\}^q\Big]\geq\frac{1}{2},$$

where infimum is taken over all possible estimators.

The proof of this lemma is given in the Appendix.

5.6.2. *Proof of Theorem* 5

(1) Set $N_1 := \overline{N}_{p,d}(\beta, L, \mathcal{P})$, $N_2 := \overline{N}_{p',d}(\beta', L', \mathcal{P}')$, $r_1 := r(\beta, p, \mathcal{P})$ and $r_2 := r(\beta', p', \mathcal{P}')$ such that $0 < r_1 < r_2$. For any τ such that $\frac{r_1}{2r_1+1} < \tau \leq \frac{r_2}{2r_2+1}$, there exists $C(\tau) > 0$ satisfying: $\forall q \geq 1$,

$$\begin{aligned}\liminf_{n\to+\infty}\inf_{\widetilde{f}_n}\Big[\sup_{f\in N_1}\mathbb{E}_f^{(n)}\Big\{\Big(\frac{n}{\ln(n)}\Big)^{r_1/(2r_1+1)}\big|\widetilde{f}_n(x_0)-f(x_0)\big|\Big\}^q\\ +\sup_{f\in N_2}\mathbb{E}_f^{(n)}\big\{n^\tau\big|\widetilde{f}_n(x_0)-f(x_0)\big|\big\}^q\Big]\geq C(\tau).\end{aligned}\tag{41}$$

Let us prove (41). The proof is based on Lemma 4 where we put

$$a_n := \big[2C(\tau)\big]^{-1/q}\Big(\frac{n}{\ln(n)}\Big)^{r_1/(2r_1+1)}, \qquad b_n := \big[2C(\tau)\big]^{-1/q}n^\tau,$$

and the constant $C(\tau) > 0$ will be specified later.

Similarly to the proof of Proposition 1, set $\mathcal{N}(x) := \prod_{i=1}^d \sqrt{2\pi}^{-1}\exp(-x_i^2/2)$ and define $f_0(x) := \sigma^{-1}\mathcal{N}(x/\sigma)$, where σ is chosen in such way that

$$f_0 \in \overline{N}_{p',d}\big(\beta', L', \mathcal{P}'\big) \cap \overline{N}_{p,d}(\beta, \underline{L}/2, \mathcal{P}).$$

Let also f_1 be given in (32). It is obvious that there exists a constant A_0 such that $f_1 \in N_1$ if $A_n \leq A_0$ and

$$A_n\delta_{l,n}^{-\beta_{i_l}}\Big(\prod_{j=1}^m\delta_{j,n}\Big)^{1/p_{i_l}} \leq \frac{\underline{L}_{i_l}}{c_l}, \qquad l=\overline{1,m}, c_l = \|g\|_{p_{i_l}}^{m-1}.\tag{42}$$

Assumptions of Lemma 4 are, respectively, fulfilled if

$$\begin{aligned}c_1^*A_n &\geq \big[2C(\tau)\big]^{1/q}\Big(\frac{\ln(n)}{n}\Big)^{r_1/(2r_1+1)},\\ c_1^* &:= (\sigma\sqrt{2\pi})^{m-d}\big|g(0)\big|^m\prod_{i\notin I}\exp\big(-x_{0,i}^2/2\sigma^2\big);\end{aligned}\tag{43}$$

$$\exp\left[\frac{2\|g\|_2^{2m}}{f_{0,I}(x_{0,I})} n A_n^2 \left(\prod_{l=1}^m \delta_{l,n}\right)\right] \le n^\tau \left(\frac{n}{\ln(n)}\right)^{-r_1/(2r_1+1)}.$$

The latter inequality, in its turn, holds if

$$nA_n^2\left(\prod_{l=1}^m \delta_{l,n}\right) = t^2 \ln(n), \qquad t := \sqrt{\left[c_2^*\right]^{-1}\left(\tau - \frac{r_1}{2r_1+1}\right)}, \qquad c_2^* := \frac{2\|g\|_2^{2m}}{f_{0,I}(x_{0,I})}. \tag{44}$$

Solving the system

$$A_n \delta_{l,n}^{-\beta_{i_l}} \left(\prod_{j=1}^m \delta_{j,n}\right)^{1/p_{i_l}} = \frac{L_{i_l}}{c_l}, \qquad l = \overline{1,m}, \qquad nA_n^2 \left(\prod_{l=1}^m \delta_{l,n}\right) = t^2 \ln(n),$$

we obtain

$$\delta_{l,n} = \left(\frac{c_l}{L_{i_l}}\right)^{1/\beta_{i_l}} \left(\frac{t^2\ln(n)}{n}\right)^{1/(\beta_{i_l} p_{i_l})} A_n^{1/\beta_{i_l} - 2/(\beta_{i_l} p_{i_l})}, \qquad A_n = R\left(\frac{t^2 \ln(n)}{n}\right)^{r_1/(2r_1+1)},$$

$$R = \left[\prod_{l=1}^m \left(\frac{L_{i_l}}{c_l}\right)^{1/(2\beta_{i_l})}\right]^{1/(1-1/s_I - 1/2\beta_I)}.$$

It is easily seen that $A_n, \delta_{l,n} \to 0, l = \overline{1,m}$, if $n \to \infty$. The choice $C(\tau) = \frac{1}{2}[c_1^* R(t^{2r_1/(2r_1+1)})]^q$, completes the proof of the inequality (41). It follows the assertion (i) of Theorem 5.

(2) Let us recall the definition of the set $\mathcal{A} \times \mathfrak{B}$, which is the set of "nuisance" parameters for the considered problem.

$$\mathcal{A} := \left\{(\beta, p) \in (0, \beta_{\max}]^d \times [1,\infty]^d\colon\ r(\beta, p, \overline{\varnothing}) > 0\right\}, \qquad \mathfrak{B} := \overline{\mathfrak{P}}.$$

Let $\widetilde{\psi}_n$ be an admissible family of normalizations and let $\widetilde{f}_n(x_0)$ be $\widetilde{\psi}_n$-adaptive estimator. Define

$$\mathcal{A}^{(0)}[\widetilde{\psi}/\psi] := \left\{(\beta, p) \in \mathcal{A}\colon\ \lim_{n\to\infty} \Upsilon_n(\beta, p) = 0\right\},$$

$$\Upsilon_n(\beta, p) := \inf_{\mathcal{P}\in\overline{\mathfrak{P}}} \Upsilon_n(\beta, p, \mathcal{P}), \qquad \Upsilon_n(\beta, p, \mathcal{P}) := \frac{\widetilde{\psi}_n(\beta, p, \mathcal{P})}{\psi_n(\beta, p, \mathcal{P})},$$

where ψ_n is given in (21). For any $\mathcal{P} \in \overline{\mathfrak{P}}$ put also

$$\mathcal{A}_{\mathcal{P}}^{(\infty)}[\widetilde{\psi}/\psi] := \left\{(\beta, p) \in \mathcal{A}\colon\ \lim_{n\to\infty} \Upsilon_n(\beta_0, p_0)\Upsilon_n(\beta, p, \mathcal{P}) = \infty, \forall(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\Psi}/\Psi]\right\}.$$

In the slight abuse of the notation, we will use later $\psi_n(r)$ instead of $\psi_n(\beta, p, \mathcal{P}), r = r(\beta, p, \mathcal{P})$.

For any $(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ introduce

$$\mathcal{P}_0 := \arg \inf_{\mathcal{P}\in\overline{\mathfrak{P}}} \Upsilon_n(\beta_0, p_0, \mathcal{P}), \qquad r_0 := r(\beta_0, p_0, \mathcal{P}_0). \tag{45}$$

Let us first note that $0 < r_0 < r_{\max}$ for any $(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$. Indeed, if $r_0 = r_{\max}$ then $(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ contradicts to $\psi_n(r_{\max})$ is a minimax rate of convergence. Moreover, for any $r \in (r_0, r_{\max})$, there exists $(\beta, p) \in \mathcal{A}$ and $\mathcal{P} \in \overline{\mathfrak{P}}$ such that $r(\beta, p, \mathcal{P}) = r$. It suffices to choose $\mathcal{P}$ such that $r(\beta^{(\max)}, p^{(\max)}, \mathcal{P}) = r_{\max} = \beta_{\max}/|I|$, $I \in \mathcal{P}$, and $\beta_i = r|I|$, $p_i = \infty$, $i = 1, \dots, d$.

(3) Our goal now is to prove that for any $(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ we have

$$\lim_{n\to\infty} \Upsilon_n(\beta_0, p_0)\Upsilon_n(\beta, p, \mathcal{P}) = \infty \qquad \forall(\beta, p, \mathcal{P})\colon\ r_0 < r(\beta, p, \mathcal{P}) < r_{\max}. \tag{46}$$

Set $N_0 := \overline{N}_{p_0,d}(\beta_0, L_0, \mathcal{P}_0)$ and $N := \overline{N}_{p,d}(\beta, L, \mathcal{P})$ such that $r_0 < r(\beta, p, \mathcal{P}) < r_{\max}$. Applying the inequality (41) with $r_1 = r_0$, $N_1 = N_0$, $r_2 = r$ and $N_2 = N$, we get for any τ satisfying $\frac{r_0}{2r_0+1} < \tau < \frac{r}{2r+1}$

$$\begin{aligned}
&\liminf_{n\to+\infty}\Big[\sup_{f\in N_0} \mathbb{E}_f^{(n)}\big\{\psi_n^{-1}(r_0)\big|\widetilde{f}_n(x_0) - f(x_0)\big|\big\}^q + \sup_{f\in N} \mathbb{E}_f^{(n)}\big\{n^\tau\big|\widetilde{f}_n(x_0) - f(x_0)\big|\big\}^q\Big] \\
&\quad \ge C(\tau).
\end{aligned} \tag{47}$$

Furthermore, by definition of $\widetilde{f}_n(x_0)$ and $\widetilde{\psi}_n$, there exist constants $M_0, M > 0$ such that for all n large enough

$$\sup_{f\in N_0} \mathbb{E}_f^{(n)}\big\{\widetilde{\psi}_n^{-1}(\beta_0, p_0, \mathcal{P}_0)\big|\widetilde{f}_n(x_0) - f(x_0)\big|\big\}^q \le M_0; \tag{48}$$

$$\sup_{f\in N} \mathbb{E}_f^{(n)}\big\{\widetilde{\psi}_n^{-1}(\beta, p, \mathcal{P})\big|\widetilde{f}_n(x_0) - f(x_0)\big|\big\}^q \le M. \tag{49}$$

Note that $\lim_{n\to\infty} \frac{\widetilde{\psi}_n(\beta_0,p_0,\mathcal{P}_0)}{\psi_n(\beta_0,p_0,\mathcal{P}_0)} = 0$ that follows from $(\beta_0, p_0) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ as well as the definition of $\mathcal{P}_0$. Thus, we obtain in view of (48) that

$$\lim_{n\to\infty} \sup_{f\in N_0} \mathbb{E}_f^{(n)}\big\{\psi_n^{-1}(r_0)\big|\widetilde{f}_n(x_0) - f(x_0)\big|\big\}^q = 0.$$

It yields together with (47) and (49) that

$$\liminf_{n\to+\infty} Mn^\tau\widetilde{\psi}_n(\beta, p, \mathcal{P}) \ge C(\tau). \tag{50}$$

Recall that $\psi_n(r) = (\ln(n)/n)^{r/(2r+1)}$. Since $\tau < \frac{r}{2r+1}$ we get for some $a > 0$ satisfying $\tau + a < \frac{r}{2r+1}$ that $n^\tau\psi_n(r) \le n^{-a}$ for n large enough. Hence, we obtain in view of (50)

$$\liminf_{n\to+\infty} n^{-a}\Upsilon_n(\beta, p, \mathcal{P}) := \liminf_{n\to+\infty} n^{-a}\frac{\widetilde{\psi}_n(\beta, p, \mathcal{P})}{\psi_n(\beta, p, \mathcal{P})} \ge \frac{C(\tau)}{M}. \tag{51}$$

Furthermore, since $\varphi_n(\beta_0, p_0, \mathcal{P}_0)$ is a minimax rate of convergence, there exists a constant $M_1 > 0$ such that

$$\Upsilon_n(\beta_0, p_0) := \frac{\widetilde{\psi}_n(\beta_0, p_0, \mathcal{P}_0)}{\psi_n(\beta_0, p_0, \mathcal{P}_0)} \ge M_1\frac{\varphi_n(\beta_0, p_0, \mathcal{P}_0)}{\psi_n(\beta_0, p_0, \mathcal{P}_0)} = M_1\big[\ln(n)\big]^{-r_0/(2r_0+1)} \tag{52}$$

for all n large enough. We deduce from (51) and (52) that $\lim_{n\to\infty} \Upsilon_n(\beta_0, p_0)\Upsilon_n(\beta, p, \mathcal{P}) = \infty$.

(4) Let $(\beta_1, p_1) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ and $(\beta_2, p_2) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$ be arbitrary pairs of parameters. Let also $\mathcal{P}_1$ and $\mathcal{P}_2$ be defined in (45) where (β_0, p_0) is replaced by (β_1, p_1) and (β_2, p_2), respectively. Then necessarily

$$r(\beta_1, p_1, \mathcal{P}_1) = r(\beta_2, p_2, \mathcal{P}_2). \tag{53}$$

Indeed, assume that $r(\beta_1, p_1, \mathcal{P}_1) < r(\beta_2, p_2, \mathcal{P}_2)$. Noting that $\Upsilon_n(\beta_2, p_2) = \Upsilon_n(\beta_2, p_2, \mathcal{P}_2)$, in view of the definition of $\mathcal{P}_2$ we deduce from (46) with $(\beta_1, p_1) = (\beta_0, p_0)$ and $(\beta, p, \mathcal{P}) = (\beta_2, p_2, \mathcal{P}_2)$ that

$$\Upsilon_n(\beta_2, p_2) \to \infty, \qquad n \to \infty. \tag{54}$$

This contradicts to $(\beta_2, p_2) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]$. The case $r(\beta_1, p_1, \mathcal{P}_1) > r(\beta_2, p_2, \mathcal{P}_2)$ is traited similarly.

(5) We are now in position to prove Theorem 5.

First, if $\mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n] \neq \varnothing$, we deduce from (53) that there exists $r_0 \in (0, r_{\max})$ such that

$$r(\beta, p, \mathcal{P}_{(\beta,p)}) = r_0 \qquad \forall (\beta, p) \in \mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]. \tag{55}$$

Here, as previously, $\mathcal{P}_{(\beta,p)} := \arg\inf_{\mathcal{P}\in\overline{\mathfrak{P}}} \Upsilon_n(\beta, p, \mathcal{P})$.

Recall that, for $(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \mathfrak{P}$,

$$r(\beta, p, \mathcal{P}) = \inf_{I\in\mathcal{P}} \gamma_I(\beta, p), \qquad \gamma_I(\beta, p) = \frac{1 - \sum_{i\in I} 1/(\beta_i p_i)}{\sum_{i\in I} 1/\beta_i}, \qquad I \in \mathcal{P}.$$

Thus, obviously

$$\dim\bigl(\mathcal{A}^{(0)}[\widetilde{\psi}_n/\psi_n]\bigr) \le 2d - 1. \tag{56}$$

Next, let $\mathcal{P}^* \in \overline{\mathfrak{P}}$ be a partition satisfying $r(\beta^{(\max)}, p^{(\max)}, \mathcal{P}^*) = r_{\max}$. We deduce from (46) that

$$\mathcal{A}^{(\infty)}_{\mathcal{P}^*}[\widetilde{\psi}/\psi] \supseteq \bigl\{(\beta, p) \in \mathcal{A}\colon\ r_0 < r\bigl(\beta, p, \mathcal{P}^*\bigr) < r_{\max}\bigr\}, \tag{57}$$

where r_0 is defined in (55). Thus, $\mathcal{A}^{(\infty)}_{\mathcal{P}^*}[\widetilde{\psi}/\psi]$ contains an open set of $\mathcal{A}$ since $(\beta, p) \mapsto r(\beta, p, \mathcal{P}^*)$ is continuous. This together with (56) completes the proof of the theorem.

Appendix

A.1. Proof of Proposition 2

Our goal is to establish a uniform bound for the empirical process $\{\xi_h^{(n)}(y_0)\}_h$. Note that the considered family of random fields is a particular case of the generalized empirical processes studied in Lepski [28]. We get the assertions of Proposition 2 from the Theorem 1 in the latter paper since it allows us to assert that, for any $u \geq 1$, $q \geq 1$ and any integer $n \geq 3$

$$\mathbb{E}_g^{(n)}\Big\{\sup_{h\in\mathcal{H}_n^{(s)}}\big[\big|\xi_h^{(n)}(y_0)\big| - \mathcal{U}^{(u,q)}(n,h,y_0)\big]_+\Big\}^q \leq C_s^{(q)}(\mathbf{K},\mathbf{g})[nV_{h^{(\max)}}]^{-q/2}\mathrm{e}^{-u},$$
$$\begin{aligned}\mathcal{U}^{(u,q)}&(n,h,y_0)\\ &:= c(\mathbf{K},s,q)\sqrt{\frac{G_h(y_0)}{nV_h}\Big\{1\vee\ln\Big(\frac{V_{h^{(\max)}}}{V_h}\Big)+2\ln\big(2+\ln G_h(y_0)\big)+u\Big\}}\\ &\quad+\frac{c(\mathbf{K},s,q)}{nV_h}\Big\{1\vee\ln\Big(\frac{V_{h^{(\max)}}}{V_h}\Big)+2\ln\big(2+\ln G_h(y_0)\big)+u\Big\}.\end{aligned} \tag{58}$$

The constants $C_s^{(q)}(\mathbf{K},\mathbf{g})$ and $c(\mathbf{K},s,q)$ are given later.

Thus, we only have to check the Assumptions of Theorem 1 in Lepski [28] and to match the notation used in the present paper and in the latter one. We divide this proof into several steps.

(1) For our case, we first consider that $p=1$, $m=s+1$, $k=s$, $\mathfrak{H}_1^k(n)=\mathcal{H}_n^{(s)}$, $\mathfrak{H}_{k+1}^m(n)=\{y_0\}$, $\mathfrak{h}^{(k)}=h$ and

$$G_\infty(h)=V_h^{-1}\|\mathbf{K}\|_\infty^s,\qquad \underline{G}_n=V_{h^{(\max)}}^{-1}\|\mathbf{K}\|_\infty^s,\qquad \overline{G}_n=V_{h^{(\min)}}^{-1}\|\mathbf{K}\|_\infty^s,$$
$$G_{j,n}(h_j)=\frac{h_j^{(\min)}}{h_j}V_{h^{(\min)}}^{-1}\|\mathbf{K}\|_\infty^s,$$
$$\underline{G}_{j,n}=\frac{h_j^{(\min)}}{h_j^{(\max)}}V_{h^{(\min)}}^{-1}\|\mathbf{K}\|_\infty^s,\qquad j=\overline{1,s},\qquad \varrho_n^{(s)}(\widehat{h},\bar{h})=\max_{j=\overline{1,s}}\big|\ln(\widehat{h}_j)-\ln(\bar{h}_j)\big|.$$

Obviously, Assumption 1(i) in Lepski [28] is fulfilled. Using Assumption (7) (see Section 2.1 of the present paper), we get $\mathrm{supp}(K)\subseteq[-1/2,1/2]^s$ and

$$\big|K(x)-K(y)\big|\leq L_{\mathbf{K}}^{(s)}\max_{j=\overline{1,s}}|x_j-y_j|\qquad\forall x,y\in\mathbb{R}^s,\qquad L_{\mathbf{K}}^{(s)}:=s\|\mathbf{K}\|_\infty^{s-1}L_{\mathbf{K}}>0.$$

Thus, we easily check that, for any $h,h'\in\mathcal{H}_n^{(s)}$ and any $y\in\mathbb{R}^s$,

$$\begin{aligned}&\big|K_h(y-y_0)-K_{h'}(y-y_0)\big|\\ &\quad\leq\Big[\frac{\|\mathbf{K}\|_\infty^s}{V_h}\vee\frac{\|\mathbf{K}\|_\infty^s}{V_{h'}}\Big]\Big\{\exp\big(s\varrho_n^{(s)}(h,h')\big)-1+\frac{L_{\mathbf{K}}^{(s)}}{\|\mathbf{K}\|_\infty^s}\big(\exp\big(\varrho_n^{(s)}(h,h')\big)-1\big)\Big\}.\end{aligned}$$

It implies that Assumption 1(ii) in Lepski [28] holds with

$$D_0(z)=\exp(sz)-1+\frac{L_{\mathbf{K}}^{(s)}}{\|\mathbf{K}\|_\infty^s}\times\big(\exp(z)-1\big),\qquad D_{s+1}\equiv 0,\ L_{s+1}\equiv 0.$$

Furthermore, Assumption 3 in Lepski [28] holds with $N=0$ and $R=1$ since $\mathfrak{H}^m_{k+1}=\mathfrak{H}_{s+1}=\{y_0\}$ and Assumption 2 in Lepski [28] is not needed since $n_1=n_2=n$.

(2) Thus, the application of the Theorem 1 in Lepski [28] is possible. Let us first compute the constants which appear in its proof.

$$\begin{aligned}
C_{N,R,m,k} &= \sup_{\delta>\delta_*}\delta^{-2}s\left[1+\ln\left(\frac{9216(s+1)\delta^2}{[s^*(\delta)]^2}\right)\right]_+ + \sup_{\delta>\delta_*}\delta^{-2}s\left[1+\ln\left(\frac{9216(s+1)\delta}{[s^*(\delta)]}\right)\right]_+ \\
&:= C_s; \\
C_D &= se^s + \frac{seL_{\mathbf{K}}}{\|\mathbf{K}\|_\infty}, \qquad C_{D,b}=\sqrt{2C_D}\vee\left[(2/3)(C_D\vee 8e)\right], \\
\lambda_1 &= 4\sqrt{2eC_D}, \lambda_2=(16/3)(C_D\vee 8e).
\end{aligned}$$

Next, we have to compute the quantities involved in the description of $\mathcal{U}^{(u,q)}_{\mathbf{r}}(n,\mathfrak{h})$.

$$M_q(h)\le C^{(q)}_{s,1}\left[1\vee\ln\left(\frac{V_{h(\max)}}{V_h}\right)\right], \qquad C^{(q)}_{s,1}:=\left[144s\delta_*^{-2}+5q+3+36C_s\right]\vee 1.$$

Since $Y_i, i=\overline{1,n}$, are identically distributed, putting $\mathfrak{h}=(h,y_0)$, $n_1=n_2=n$ and $\mathbf{r}=0$, we have

$$\begin{aligned}
F_{n,\mathbf{r}}(\mathfrak{h}) = 1\vee\left[\int_{\mathbb{R}^s}\left|K_h(y-y_0)\right|g(y)\,\mathrm{d}y\right] := G_h(y_0), \\
F_n = \sup_{h\in\mathcal{H}^{(s)}_n} G_h(y_0)\le 1\vee\mathbf{g}\|\mathbf{K}\|^s_1; \\
\mathcal{U}^{(u,q)}_{\mathbf{r}}(n,\mathfrak{h})\le\mathcal{U}^{(u,q)}(n,h,y_0), \qquad c(\mathbf{K},s,q):=\left[(10C_D)\vee(48e)\right]C^{(q)}_{s,1}\|\mathbf{K}\|^s_\infty.
\end{aligned}$$

Here, we have used that $C^{(q)}_{s,1}\wedge\|\mathbf{K}\|^s_\infty\ge 1$. Thus, we come to the inequality (58) with $C^{(q)}_s(\mathbf{K},\mathbf{g}):=c_q\|\mathbf{K}\|^{sq}_\infty(1\vee\mathbf{g}\|\mathbf{K}\|^s_1)^{q/2}$, $c_q=2^{7q/2+5}3^{q+4}\Gamma(q+1)(C_{D,b})^q$.

(3) If $n\ge 3, nV_h\ge\ln(n), 1\le u\le q\ln(n)$ and $M(h):=1\vee\ln(\frac{V_{h(\max)}}{V_h})$, since $1\le G_h(y_0)\le\|\mathbf{K}\|^s_\infty$, one has

$$\begin{aligned}
(nV_h)^{-1}\left\{M(h)+2\ln\left(2+\ln G_h(y_0)\right)+u\right\} &\le 7(nV_h)^{-1}G_h(y_0)\left\{M(h)+u\right\} \\
&\le 7(1+q)\|\mathbf{K}\|^s_\infty.
\end{aligned} \tag{59}$$

Put finally $\lambda^{(q)}_s[\mathbf{K}]:=c(\mathbf{K},s,q)\sqrt{7}\{\sqrt{7(1+q)\|\mathbf{K}\|^s_\infty}+1\}$. Since $[a^{(q)}_s]^{-1}\ge 1$, the assertion (i) of Proposition 2 follows from (58) and (59). Let us now prove the assertions (ii) and (iii) of Proposition 2.

(4) First, in view of the definition of $\mathfrak{H}^{(q)}_s(n)$, we get the assertion (ii) from the assertion (i) of Proposition 2 since $u\le q\ln(n)$ and $[1\vee\lambda^{(q)}_s]\sqrt{(1+q)a^{(q)}_s}=1/2$. Here, we have used that if $\mathbf{K}$

satisfies the assumption (7), see Section 2.1, $|\mathbf{K}|$ satisfies it as well and, therefore, Proposition 2(i) is applicable to the process $\overline{\xi}_h^{(n)}(y_0)$.

Next, using the trivial inequality $|x\vee a - x\vee b| \le |a-b|$, $x,a,b\in\mathbb{R}$, we easily check that

$$G_h(y_0) \le 2\widetilde{G}_h(y_0) + 2\sup_{h\in\mathfrak{H}_s^{(q)}(n)}\left[\left|\overline{\xi}_h^{(n)}(y_0)\right| - \frac{1}{2}G_h(y_0)\right]_+ \qquad \forall h\in\mathfrak{H}_{a_s}^{(s)}(n). \tag{60}$$

Assertion (iii) of Proposition 2 follows from assertion (ii) and (60).

A.2. Proof of Lemma 1

Note first that, for any $(h,\mathcal{P})\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$, any $(\eta,\mathcal{P}')\in\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$ and any $I\cap I'\in\mathcal{P}\circ\mathcal{P}'$

$$h_{I\cap I'}\vee\eta_{I\cap I'} \in \bigcup_{m=1}^{M_n(I)}\bigcup_{l=1}^{M_n(I')}\mathfrak{H}_{m,l}^{(I\cap I')},$$

$$\mathfrak{H}_{m,l}^{(I\cap I')} := \left\{h_{I\cap I'}\in\prod_{i\in I\cap I'}\left[\frac{1}{n},\mathfrak{h}_i^{(I\cap I',m,l)}\right]:\ nV_{h_I}\ge\left[a_{|I\cap I'|}^{(2q)}\right]^{-1}\ln(n)\right\},$$

where $\mathfrak{h}_i^{(I\cap I',m,l)} := (2^{m\vee l})^{\mathfrak{z}}[\mathfrak{h}_i^{(I)}\vee\mathfrak{h}_i^{(I')}]$, $i\in I\cap I'$.

Set $f\in\mathbb{F}_d[\mathbf{f},\overline{\mathfrak{P}}]$. To get the assertions of Lemma 1, we apply Proposition 2 with $s=|I\cap I'|$, $g=f_{I\cap I'}$, $\mathbf{g}=\mathbf{f}$, $h_i^{(\min)}(n)=\frac{1}{n}$, $h_i^{(\max)}(n)=\mathfrak{h}_i^{(I\cap I',m,l)}$, $\mathfrak{H}_s^{(q)}(n)=\mathfrak{H}_{m,l}^{(I\cap I')}$, $K_h=K_{h_{I\cap I'}}^{(I\cap I')}$, $G_h(y_0)=G_{h_{I\cap I'}}(x_{0,I\cap I'})$, $\widetilde{G}_h(y_0)=\widetilde{G}_{h_{I\cap I'}}(x_{0,I\cap I'})$, $\mathcal{U}_h^{(u)}(y_0)=\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{0,I\cap I'})$, $\xi_h^{(n)}(y_0)=\xi_{h_{I\cap I'}}^{(n)}(x_{0,I\cap I'})$.

Recall that $\overline{\mathfrak{P}}^* := \{\mathcal{P}\circ\mathcal{P}':\ \mathcal{P},\mathcal{P}'\in\overline{\mathfrak{P}}\}$. In view of the definition of $\overline{\mathfrak{H}}[\overline{\mathfrak{P}}]$, we easily check that

$$\xi_n(x_0)\le\sum_{\mathcal{P}\circ\mathcal{P}'\in\overline{\mathfrak{P}}^*}\ \sum_{I\cap I'\in\mathcal{P}\circ\mathcal{P}'}\ \sum_{m=1}^{M_n(I)}\sum_{l=1}^{M_n(I')}\sup_{h_{I\cap I'}\in\mathfrak{H}_{m,l}^{(I\cap I')}}\left[\left|\xi_{h_{I\cap I'}}^{(n)}(x_{0,I})\right| - \lambda_{|I\cap I'|}^{(2q)}\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{0,I\cap I'})\right]_+,$$

with $u=q[1\vee\ln(2^{m\wedge l}V_{\max}/\inf_{I\in\mathcal{P}}V_{\mathfrak{h}_I^{(I)}})]\in[1,2q\ln(n)]$, since $V_{\mathfrak{h}_I^{(I)}}\ge\frac{\ln(n)}{an}$ and $M_n(I)\le\log_2(n)$, $\forall I\in\mathcal{I}_d$.

Therefore, it follows from the assertion (i) of Proposition 2, since $V_{\mathfrak{h}_{I\cap I'}^{(I)}\vee\mathfrak{h}_{I\cap I'}^{(I')}}\ge\inf_{I\in\mathcal{P}}V_{\mathfrak{h}_I^{(I)}}$,

$$\left(\mathbb{E}_f^{(n)}\left\{\sup_{h_{I\cap I'}\in\mathfrak{H}_{m,l}^{(I\cap I')}}\left[\left|\xi_{h_{I\cap I'}}^{(n)}(x_{0,I})\right| - \lambda_{|I\cap I'|}^{(2q)}\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{0,I\cap I'})\right]_+\right\}^{2q}\right)^{1/(2q)}$$

$$\le\left\{C_{|I\cap I'|}^{(2q)}(\mathbf{K},\mathbf{g})\right\}^{1/(2q)}[nV_{\max}]^{-1/2}\left(2^{\mathfrak{z}|I\cap I'|/2}\right)^{-m\vee l}\left(2^{1/2}\right)^{-m\wedge l};$$

$$\left(\mathbb{E}_f^{(n)}\left|\xi_n(x_0)\right|^{2q}\right)^{1/(2q)}\le\mathbf{c}_1[nV_{\max}]^{-1/2},$$

$$\mathbf{c}_1 := \sum_{\mathcal{P}\in\overline{\mathfrak{P}}^*}\sum_{I\in\mathcal{P}}\big\{C^{(2q)}_{|I|}(\mathbf{K},\mathbf{f})\big\}^{1/(2q)}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge 1]/2}}{2^{[(\mathfrak{z}|I|)\wedge 1]/2}-1}\right].$$

Similarly, applying Proposition 2(iii) and using the trivial inequality $[\sup_i x_i - \sup_i y_i]_+ \le \sup_i [x_i - y_i]_+$, we obtain the assertion (ii) of Lemma 1 with $\mathbf{c}_2 := 2\mathbf{c}_1$.

Next, it is easily seen that

$$\overline{G}_n(x_0) \le 2\left(\sum_{\mathcal{P}\circ\mathcal{P}'\in\overline{\mathfrak{P}}^*}\sum_{I\cap I'\in\mathcal{P}\circ\mathcal{P}'}\sum_{m=1}^{M_n(I)}\sum_{l=1}^{M_n(I')}\sup_{h_{I\cap I'}\in\mathfrak{H}^{(I\cap I')}_{m,l}}\left[\left|\overline{\xi}^{(n)}_{h_{I\cap I'}}(x_{0,I})\right| - \frac{1}{2}G_{h_{I\cap I'}}(x_{0,I})\right]_+\right) + 3G(x_0),$$

and that

$$\left(\mathbb{E}^{(n)}_f\left|\bar{\mathbf{f}}_n(x_0)\right|^{2q}\right)^{1/2q} \le 12\lambda d^3 2^{d^2}\left[\left(\mathbb{E}^{(n)}_f\left|\overline{G}_n(x_0)\right|^{2qd^2}\right)^{1/(2qd^2)} + \left(1\vee \mathbf{f}\|\mathbf{K}\|_1^d\right)\right]^{d^2}.$$

Thus, we get assertion (iii) of Lemma 1 from assertion (ii) of Proposition 2 with

$$\mathbf{c}_3 := 12\lambda d^3\left[4\left(\sum_{\mathcal{P}\in\overline{\mathfrak{P}}^*}\sum_{I\in\mathcal{P}\circ\mathcal{P}'}\big\{C^{(2qd^2)}_{|I|}(\mathbf{K},\mathbf{f})\big\}^{1/(2qd^2)}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge 1]/2}}{2^{[(\mathfrak{z}|I|)\wedge 1]/2}-1}\right]\right) + 8\left(1\vee \mathbf{f}\|\mathbf{K}\|_1^d\right)\right]^{d^2}.$$

Similarly, we obtain assertion (iv) of Lemma 1 with

$$\mathbf{c}_4 := 2\left(\sum_{\mathcal{P}\in\overline{\mathfrak{P}}^*}\sum_{I\in\mathcal{P}\circ\mathcal{P}'}\big\{C^{(2q)}_{|I|}(\mathbf{K},\mathbf{f})\big\}^{1/(2q)}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge 1]/2}}{2^{[(\mathfrak{z}|I|)\wedge 1]/2}-1}\right]\right) + 3\left(1\vee \mathbf{f}\|\mathbf{K}\|_1^d\right).$$

This completes the proof of Lemma 1.

A.3. Proof of Lemma 3

The proof of this lemma is based on the embedding theorem for anisotropic Nikolskii classes; see, for example, Theorem 6.9 in Nikolskii [34].

Let $\mathcal{P}' \in \overline{\mathfrak{P}}$ and $I \in \mathcal{P}\circ\mathcal{P}'$ be fixed. Set $\varkappa(I) := 1 - \sum_{k\in I}(\beta_k p_k)^{-1}$ and $\beta_i(I) := \varkappa(I)\beta_i\varkappa_i^{-1}(I)$, where $\varkappa_i(I) := 1 - \sum_{k\in I}(p_k^{-1} - p_i^{-1})\beta_k^{-1}$, $i\in I$. Since $\varkappa(I) > 0$ there exists $c_I := c_I(\mathbf{K}, |I|, p_I, l) > 0$ such that

$$\mathbb{N}_{p_I,|I|}(\beta_I, L_I) \subseteq \mathbb{N}_{\infty,|I|}\big(\beta(I), c_I L_I\big).$$

Introduce the family of $|I|\times|I|$ matrices $E_j := (e_1,\dots,e_j,0,\dots,0)$, $j = \overline{1,|I|}$, and E_0 is zero matrix. For any $(h,\eta)\in(0,1]^d\times[0,1]^d$, using a telescopic sum and the triangle inequality, we get

$$\left|\mathcal{B}_{h_I,\eta_I}(x_{0,I})\right| \le \sum_{j=1}^{|I|}\left|\int K^{(I)}(u)\big[f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_j u\big) - f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_{j-1}u\big)\big]\,\mathrm{d}u\right|.$$

For $j=1,\dots,|I|$ put

$$\mathcal{B}_{h_I,\eta_I,j}(x_{0,I}) := \int_{\mathbb{R}}\mathbf{K}(u_j)\big[f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_j u\big) - f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_{j-1}u\big)\big]\,\mathrm{d}u_j.$$

If $\eta_j \ge h_j$, then $\mathcal{B}_{h_I,\eta_I,j}(x_{0,I})=0$, if not we put $[u]^j := u-u_je_j$, $u\in\mathbb{R}^{|I|}$, and we have

$$\begin{aligned}\mathcal{B}_{h_I,\eta_I,j}(x_{0,I}) = &\int_{\mathbb{R}}\mathbf{K}(u_j)\big[f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_j u\big) - f_I\big(x_{0,I}+[\eta_I u]^j+(h_I\vee\eta_I-\eta_I)E_{j-1}u\big)\big]\,\mathrm{d}u_j\\ &+\int_{\mathbb{R}}\mathbf{K}(u_j)\big[f_I\big(x_{0,I}+[\eta_I u]^j+(h_I\vee\eta_I-\eta_I)E_{j-1}u\big) - f_I\big(x_{0,I}+\eta_I u+(h_I\vee\eta_I-\eta_I)E_{j-1}u\big)\big]\,\mathrm{d}u_j.\end{aligned}$$

Thus, in view of the triangle inequality,

$$\left|\mathcal{B}_{h_I,\eta_I}(x_{0,I})\right| \le 2\sum_{i\in I}c_I L_i h_i^{\beta_i(I)}\int_{\mathbb{R}^{|I|}}\left|K^{(I)}(u)\right||u_i|^{\beta_i(I)}\,\mathrm{d}u \le \mathbf{c}\sum_{i\in I}L_i h_i^{\beta_i(I)},$$

$$\mathbf{c} := \mathbf{c}(\mathbf{K},d,p,l,\mathcal{P}) = 2\|\mathbf{K}\|_1^d \sup_{\mathcal{P}'\in\overline{\mathfrak{P}}}\ \sup_{I\in\mathcal{P}\circ\mathcal{P}'} c_I\big(\mathbf{K},|I|,p_I,l\big).$$

Here, we have used Taylor expansions of $f\in\mathbb{N}_{\infty,|I|}(\beta(I),c_IL_I)$, the product structure of $K^{(I)}$, the Fubini theorem that $\beta(I)\in(0,l]^d$ and (19); see Section 3.2. We have also used that $\mathbf{K}$ is compactly supported on $[-1/2,1/2]$ and that $\|\mathbf{K}\|_1\ge 1$.

A.4. Proof of Lemma 4

Put $T_n := a_n|\widetilde{f}_n(x_0)-f_0(x_0)|$ and

$$\mathcal{R}_n^{(q)}[a_n,b_n,\widetilde{f},f] := \sup_{f\in N_1}\mathbb{E}_f^{(n)}\big\{a_n\big|\widetilde{f}_n(x_0)-f(x_0)\big|\big\}^q + \sup_{f\in N_2}\mathbb{E}_f^{(n)}\big\{b_n\big|\widetilde{f}_n(x_0)-f(x_0)\big|\big\}^q.$$

It is easily seen that $\mathcal{R}_n^{(q)}[a_n, b_n, \widetilde{f}, f] \geq \mathcal{R}_n^{(1)}[a_n, b_n, \widetilde{f}, f]$ and that

$$\mathcal{R}_n^{(1)}[a_n, b_n, \widetilde{f}, f] \geq \mathbb{E}_{f_1}^{(n)}\{|T_n - 1|\} + \frac{b_n}{a_n}\mathbb{E}_{f_0}^{(n)}\{T_n\}.$$

Here, we have used the triangle inequality and the assumption $a_n|f_1(x_0) - f_0(x_0)| = 1$.

Put also $c_n := \frac{b_n}{a_n}$ and $Z_n := \frac{\mathrm{d}\mathbb{P}_{f_1}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}}(X^{(n)})$. We obtain

$$\mathcal{R}_n^{(1)}[a_n, b_n, \widetilde{f}, f] \geq \mathbb{E}_{f_0}^{(n)}\{c_n \wedge Z_n\} \geq \tfrac{1}{2}\Big[c_n + 1 - \sqrt{\mathbb{E}_{f_0}^{(n)}\{c_n - Z_n\}^2}\Big].$$

Here, we have used the trivial equality $a \wedge b = \frac{1}{2}\{a + b - |a - b|\}$, that $\mathbb{E}_{f_0}^{(n)}\{Z_n\} = 1$ and the Cauchy–Schwarz inequality. Using the third assumption, we also have $\mathbb{E}_{f_0}^{(n)}\{c_n - Z_n\}^2 \leq c_n^2 - c_n$. Finally, for n large enough,

$$\inf_{\widetilde{f}} \mathcal{R}_n^{(q)}[a_n, b_n, \widetilde{f}, f] \geq \frac{1}{2}\Big[c_n + 1 - \sqrt{c_n^2 - c_n}\Big] \geq \frac{1}{2}.$$

Acknowledgments

The author is grateful to O. Lepski and the anonymous referees for their very useful remarks and suggestions.

References

[1] Birgé, L. (2008). Model selection for density estimation with $\mathbb{L}_2$-loss. Available at arXiv:0808.1416v2.

[2] Bretagnolle, J. and Huber, C. (1979). Estimation des densités: Risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137. MR0523165

[3] Brown, L.D. and Low, M.G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535. MR1425965

[4] Butucea, C. (2000). Two adaptive rates of convergence in pointwise density estimation. *Math. Methods Statist.* **9** 39–64. MR1772224

[5] Chacón, J.E. and Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *TEST* **19** 375–398. MR2677734

[6] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 569–609. MR3088382

[7] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics*. New York: Wiley. MR0780746

[8] Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimates. *Ann. Statist.* **24** 2499–2512. MR1425963

[9] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25** 2626–2637. MR1604428
[10] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation. Springer Series in Statistics*. New York: Springer. MR1843146
[11] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974
[12] Donoho, D.L. and Low, M.G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970. MR1165601
[13] Efromovich, S. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36** 1127–1155. MR2418652
[14] Efromovich, S.Y. (1985). Non parametric estimation of a density of unknown smoothness. *Theory Probab. Appl.* **30** 557–568.
[15] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344
[16] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673
[17] Goldenshluger, A. and Lepski, O. (2013). On adaptive minimax density estimation on $\mathbb{R}^D$. *Probab. Theory Related Fields*. To appear. Published online 13 July 2013.
[18] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214
[19] Golubev, G.K. (1992). Nonparametric estimation of smooth densities of a distribution in L_2. *Problemy Peredachi Informatsii* **28** 52–62. MR1163140
[20] Hasminskii, R. and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18** 999–1010. MR1062695
[21] Ibragimov, I.A. and Has'minskiĭ, R.Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **98** 61–85, 161–162, 166. MR0591862
[22] Ibragimov, I.A. and Khas'minskiĭ, R.Z. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **108** 72–88, 194, 198. MR0629401
[23] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10** 187–220. MR2046772
[24] Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916
[25] Kerkyacharian, G., Lepski, O. and Picard, D. (2007). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Theory Probab. Appl.* **52** 150–171.
[26] Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli* **2** 229–247. MR1416864
[27] Kluchnikoff, N. (2005). On adaptive estimation of anisotropic functions. Ph.D Thesis, Aix-Marseille 1.
[28] Lepski, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, part 1. *Math. Methods Statist.* **22** 83–99. MR3071956
[29] Lepski, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129
[30] 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. *Ann. Statist.* **25** 929–947. MR1447734
[31] Lepskiĭ, O.V. (1991). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

[32] Mason, D.M. (2009). Risk bounds for kernel density estimators. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **363** 66–104, 183. MR2749118

[33] Massart, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Berlin: Springer. MR2319879

[34] Nikol'skiĭ, S.M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya*, 2nd ed. Moscow: Nauka. MR0506247

[35] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282

[36] Rigollet, P. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12** 351–370. MR2218559

[37] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. MR0079873

[38] Samarov, A. and Tsybakov, A. (2007). Aggregation of density estimators and dimension reduction. In *Advances in Statistical Modeling and Inference. Ser. Biostat.* **3** 233–251. Hackensack, NJ: World Sci. Publ. MR2416118

[39] Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. New York: Wiley. MR1191168

[40] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability*. London: Chapman & Hall. MR0848134

[41] Tsybakov, A.B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239

Received June 2013 and revised March 2014