Electronic Journal of Statistics
Vol. 9 (2015) 106–134
ISSN: 1935-7524
DOI: 10.1214/15-EJS986

$\mathbb{L}_p$ adaptive estimation of an anisotropic 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

Abstract: In this paper, we focus on the problem of a multivariate density estimation under an $\mathbb{L}_p$-loss. We provide a data-driven selection rule from a family of kernel estimators and derive for it $\mathbb{L}_p$-risk oracle inequalities depending on the value of $p \geq 1$. The proposed estimator permits us to take into account approximation properties of the underlying density and its independence structure simultaneously. Specifically, we obtain adaptive upper bounds over a scale of anisotropic Nikolskii classes when the smoothness is also measured with the $\mathbb{L}_p$-norm. It is important to emphasize that the adaptation to unknown independence structure of the estimated density allows us to improve significantly the accuracy of estimation (curse of dimensionality). The main technical tools used in our derivation are uniform bounds on the $\mathbb{L}_p$-norms of empirical processes developed in Goldenshluger and Lepski [13].

Keywords and phrases: Density estimation, oracle inequality, adaptation, independence structure, upper function.

Received January 2014.

Contents

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 density f with respect to the Lebesgue measure. Furthermore, $\mathbb{P}_f$ denotes the probability law of $X^{(n)} = (X_1, \dots, X_n),\ n \in \mathbb{N}^*$, and $\mathbb{E}_f$ is the mathematical expectation with respect to $\mathbb{P}_f$.

Our goal is to estimate the density f using observations $X^{(n)} = (X_1, \dots, X_n)$, $n \in \mathbb{N}^*$. By an estimator, we mean any $X^{(n)}$-measurable mapping $\widehat{f} : \mathbb{R}^n \to \mathbb{L}_p(\mathbb{R}^d)$ and the accuracy of an estimator is measured by its $\mathbb{L}_p$-*risk*:

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] := \left(\mathbb{E}_f \left\|\widehat{f} - f\right\|_p^q\right)^{\frac{1}{q}}, \quad p \in [1, +\infty), \quad q \geq 1.$$

A discussion of traditional methods and a review of the vast literature on the theory and application of density estimation is given in Devroye and Györfi [6], Silverman [37] and Scott [38]. We do not pretend here to provide a detailed overview and mention only the results which are relevant for the problems under consideration. The minimax and adaptive minimax multivariate density estimation under $\mathbb{L}_p$-losses on particular functional classes was studied in Bretagnolle and Huber [3], Ibragimov and Khasminskii ([19, 20]), Devroye and Lugosi ([7, 8, 9]), Efroimovich ([10, 11]), Hasminskii and Ibragimov [17], Golubev [16], Donoho et al. [5], Kerkyacharian, Picard and Tribouley [23], Juditsky and Lambert-Lacroix [22], Rigollet [34], Massart [30] (chapter 7), Samarov and Tsybakov [36], Mason [29], Chacón and Duong [4], Goldenshluger and Lepski [14], and Birgé [2].

Goldenshluger and Lepski [14] developed a data-driven selection rule from a family of kernel estimators. Moreover, the selected estimator is minimax adaptive over a scale of anisotropic Nikolskii classes when the smoothness of the underlying density and the error of estimation are measured with the same $\mathbb{L}_p$-norm.

Lepski [27] proposed an estimator which takes into account the independence between groups of coordinates of the observed vectors, for estimation under the $\mathbb{L}_\infty$-loss. Thus, it was shown that the adaptation to unknown independence structure permits us to reduce the so-called curse of dimensionality. This result was illustrated by application to adaptive minimax estimation over a scale of anisotropic Nikolskii classes.

In Rebelles [33], the same problem was studied in the pointwise setting and some comparisons between the local procedure and the global one in Lepski [27] have been made.

In the present paper, we address the same problem under an $\mathbb{L}_p$-loss, $1 \leq p < \infty$. As in Goldenshluger and Lepski [14], we consider the case where the smoothness of the underlying density is assumed to be also measured in the $\mathbb{L}_p$-norm. Our main goal is to derive optimal minimax adaptive rates in the context of global estimation of a density, by taking advantage of the fact that some coordinates of the observations may be independent from the others. Throughout our article we compare both the results and methods used with those of Goldenshluger and Lepski [14] and Lepski [27].

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:

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, \Sigma\right] := \sup_{f\in\Sigma}\left(\mathbb{E}_f \left\|\widehat{f}-f\right\|_p^q\right)^{\frac{1}{q}}, \quad p \in [1,+\infty), \quad q \geq 1.$$

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

$$\mathcal{R}_p^{(q)}\left[\widehat{f}_*, \Sigma\right] \asymp \inf_{\widehat{f}} \mathcal{R}_p^{(q)}\left[\widehat{f}, \Sigma\right] := \varphi_{n,p}(\Sigma).$$

Here, infimum is taken over all possible estimators.

Smoothness assumption Let Σ be the anisotropic Nikolskii class $N_{p,d}(\beta, L)$ of d-dimensional densities (we recall the definition in Section 3.1). Here, $\beta = (\beta_1, \dots, \beta_d) \in (0,+\infty)^d$ represents the smoothness of the underlying density. Then $\varphi_{n,p}(N_{p,d}(\beta,L)) = n^{-\frac{\gamma_p\overline{\beta}}{\gamma_p+\overline{\beta}}}$, where

$$\overline{\beta} := \left[\sum_{i=1}^d \frac{1}{\beta_i}\right]^{-1}, \qquad \gamma_p := \begin{cases} 1-\frac{1}{p}, & p\in(1,2], \\ \frac{1}{2}, & p>2, \end{cases} \tag{1.1}$$

see, e.g., Ibragimov and Khasminskii ([19], [20]), and Hasminskii and Ibragimov [17].

It is important to emphasize that minimax rates depend heavily on both the dimension d and the index p of the $\mathbb{L}_p$-risk. The dependence on p disappears when we estimate a density belonging to the class $N_{p,d}(\beta,L)$ on a given bounded interval of $\mathbb{R}^d$, see, e.g., Donoho et al. [5] for the case $d=1$.

To reduce the influence of the dimension on the accuracy of estimation (curse of dimensionality), many researchers have studied the possibility of taking into account, not only the smoothness properties of the target function, but also some structural hypothesis on the statistical model. For instance, see the works on the composite function structure in Horowitz and Mamen [18], Iouditski et al. [21] and Baraud and Birgé[1], the works on multi-index structure in Goldenshluger and Lepski [12] and Lepski and Serdyukova [28], and the works on the multiple index model in density estimation in Samarov and Tsybakov [36].

Let us briefly discuss one of the possibilities of facing to this problem in the density model setting. As explained above, the approach which has been recently proposed in Lepski [27] is to take into account the independence structure of the density f, namely its product structure due to the independence structure of the vector X_1.

Structural assumption Denote by $\mathcal{I}_d$ the set of all subsets of $\{1,\ldots,d\}$, except the empty set. Let $\mathfrak{P}$ be a given set of partitions of $\{1,\ldots,d\}$. For all $I \in \mathcal{I}_d$ denote also $\overline{I} = \{1,\ldots,d\}\backslash I$ and $|I|$ =card(I). We will use $\overline{\emptyset}$ for $\{1,\ldots,d\}$. Finally, for all $x \in \mathbb{R}^d$ and $I \in \mathcal{I}_d$ put $x_I := (x_i)_{i\in I}$ and

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

Assume that $f_{\overline{\emptyset}} \equiv f$, that $f_\emptyset \equiv 1$ and note that f_I is the marginal density of $X_{1,I}$. If $\mathcal{P} \in \mathfrak{P}$ is such that the vectors $X_{1,I}$, $I \in \mathcal{P}$, are independent then $f(x) = \prod_{I\in\mathcal{P}} f_I(x_I),\ \forall x \in \mathbb{R}^d$. In the sequel, the possible independence structure of the density f will be represented by a partition belonging to the following set:

$$\mathfrak{P}(f) := \left\{\mathcal{P} \in \mathfrak{P} :\ f(x) = \prod_{I\in\mathcal{P}} f_I(x_I),\ \forall x \in \mathbb{R}^d\right\}. \tag{1.2}$$

Note that $\mathfrak{P}(f)$ is not empty if we consider that $\overline{\emptyset} \in \mathfrak{P}$, or that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of f is known. The possibility of choosing $\mathfrak{P}$, instead of considering all partitions of $\{1,\ldots,d\}$, is introduced for technical purposes. This is explained in more detail in Lepski [27], section 2.1, paragraph "*Extra parameters*".

In this paper, we focus on the problem of minimax estimation with $\mathbb{L}_p$-risk over anisotropic Nikolskii classes $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ (defined by (3.1) in Section 3.1). The definition of these classes is a modification of that of classes $N_{p,d}(\beta, L)$ to take into account the possible independence structure $\mathcal{P}$ of the target density f. Here, we need f and some of its marginals f_I to be uniformly bounded by a real number $\mathbf{f} > 0$. In particular, we will prove in Section 3.2 that, for fixed $\beta \in (0,+\infty)^d$, $L \in (0,+\infty)^d$, $\mathcal{P} \in \mathfrak{P}(f)$ and $\mathbf{f} > 0$,

$$\varphi_{n,p}\left(\ N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right) = n^{-\frac{\gamma_p \overline{r}}{\gamma_p + \overline{r}}}, \qquad \overline{r} := \inf_{I\in\mathcal{P}} \left[\sum_{i\in I} \frac{1}{\beta_i}\right]^{-1}, \tag{1.3}$$

where γ_p is given in (1.1).

If $\mathfrak{P} = \{\overline{\emptyset}\}$, the class $N_{p,d}(\beta, L, \overline{\emptyset}, \mathbf{f})$ coincides with $N_{p,d}(\beta, L) \cap \mathbb{F}[\mathbf{f}]$, where $\mathbb{F}[\mathbf{f}]$ is the set of functions uniformly bounded by $\mathbf{f} > 0$, and we find again the rate given in (1.1). Note however that if $\mathcal{P} \neq \overline{\emptyset}$ then $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}) \subsetneq N_{p,d}(\beta, L) \cap \mathbb{F}[\mathbf{f}]$ and the latter rate can be significantly improved. Indeed, if for instance $\beta = (\boldsymbol{\beta}, \ldots, \boldsymbol{\beta})$ and $\mathcal{P}^* = \{\{1\}, \ldots, \{d\}\}$, then $\overline{r} = \overline{\beta} = \boldsymbol{\beta}$ and

$$n^{-\frac{\gamma_p \boldsymbol{\beta}}{\gamma_p + \boldsymbol{\beta}}} = \varphi_{n,p}\left(\ N_{p,d}(\beta, L, \mathcal{P}^*, \mathbf{f})\right) \ll \varphi_{n,p}\left(\ N_{p,d}(\beta, L)\right) = n^{-\frac{\gamma_p \boldsymbol{\beta}}{d\gamma_p + \boldsymbol{\beta}}}. \tag{1.4}$$

Moreover, $\varphi_{n,p}(\ N_{p,d}(\beta, L, \mathcal{P}^*, \mathbf{f}))$ does not depend on the dimension d.

We remark that minimax rates (accuracy of estimation) depend heavily on the parameter $(\beta, \mathcal{P})$. Knowledge of this parameter cannot be assumed often in particular applications. Hence, it becomes necessary to find an estimator whose construction would be parameter free.

Adaptive minimax estimation In the framework of 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 $\Sigma_\alpha = N_{p,d}(\beta, L)$ then $\alpha = (\beta, L)$ and, if $\Sigma_\alpha = N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ then $\alpha = (\beta, L, \mathcal{P}, \mathbf{f})$ (here, p is fixed).

The first question arising in this framework is the following: does there exist an estimator $\widehat{f}_*$ such that

$$\limsup_{n\to+\infty} \left\{ \varphi_{n,p}^{-1}(\alpha) \mathcal{R}_p^{(q)} \left[\widehat{f}_*, \Sigma_\alpha \right] \right\} < +\infty \quad \forall \alpha \in \mathcal{A}, \tag{1.5}$$

where $\varphi_{n,p}(\alpha)$ is the minimax rate of convergence over Σ_α. If such an estimator exists, it is called an optimal adaptive estimator (O.A.E.).

As mentioned previously, Goldenshluger and Lepski [14] provide an O.A.E. for estimation under $\mathbb{L}_p$-risk, $1 < p < \infty$, over the scale $\{N_{p,d}(\beta, L) \cap \mathbb{F}[\mathbf{f}]\}$.

In this paper, we construct an O.A.E. for estimation under $\mathbb{L}_p$-risk over the scale $\{N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})\}$, $1 < p < \infty$. Therefore, we improve the adaptive rates of convergence found in Goldenshluger and Lepski [14] when the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$.

Furthermore, if $2 \leq p < \infty$, it is easily seen that, by considering the $\mathbb{L}_p$-loss, we also outperform the adaptive rates of convergence obtained in Lepski [27] for estimation under the sup-norm loss over the scale $\{N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})\}$. Indeed,

$$\varphi_{n,p}\left(\, N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right) \ll \varphi_{n,\infty}\left(\, N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right) = \left(\frac{n}{\ln(n)}\right)^{-\frac{r}{2r+1}}, \tag{1.6}$$

$r = \overline{r} - 1/p$, where $\varphi_{n,p}(\, N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}))$ and $\overline{r}$ are given in (1.3). We see that the gain is twofold. We win a factor "$\ln(n)$" and $\overline{r} > r$.

In Rebelles [33], it was shown that there exists no O.A.E. for pointwise estimation over any scale $\{N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})\}$ containing at least two classes. In the pointwise setting, there is a "ln-price" to pay for adaptation both to the smoothness parameter of the target density and to its independence structure.

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 oracle inequalities for the selected estimator. In Section 3, we define anisotropic Nikolskii classes of densities for adaptation with respect to their independence structure and we provide adaptive upper bounds over a scale of those functional classes. It is also established that the quality of estimation we obtain is rate optimal for this problem. Proofs of all main results are given in Section 4. Proofs of technical lemmas are deferred to the Appendix.

2. Estimator's construction and $\mathbb{L}_p$-risk oracle inéqualities

2.1. Kernel estimators related to independence structure

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

$$\exists\, L_{\mathbf{K}}>0: \quad |\mathbf{K}(x)-\mathbf{K}(y)|\le L_{\mathbf{K}}\,|x-y|\,,\ \forall x,y\in\mathbb{R}. \tag{2.1}$$

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

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

$$\widehat{f}_{h_I}(x_I):=n^{-1}\sum_{i=1}^{n}K_{h_I}\left(X_{i,I}-x_I\right).$$

Let h_{max}, h_{min} and V_{min} be fixed numbers satisfying $1/n\le h_{min}\le h_{max}\le 1$ and $h_{max}^d\ge V_{min}>0$. For all $I\in\mathcal{I}_d$, let $\mathcal{H}_I$ be a fixed set of multibandwidths h_I such that

$$\mathcal{H}_I\subseteq\left\{h_I\in\left[h_{min},h_{max}\right]^{|I|}:\ V_{h_I}\ge V_{min}\right\}.$$

Then, define the set of parameters

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

and introduce the family of estimators

$$\mathfrak{F}\,[\,\mathfrak{P}\,]:=\left\{\widehat{f}_{(h,\mathcal{P})}(x)=\prod_{I\in\mathcal{P}}\widehat{f}_{h_I}(x_I),\ (h,\mathcal{P})\in\mathcal{H}\,[\,\mathfrak{P}\,]\right\}. \tag{2.2}$$

Note first that $\widehat{f}_{(h,\overline{\emptyset})}=\widehat{f}_h$ is the Parzen-Rosenblatt estimator (see, e.g., Rosenblatt [35], Parzen [32]) with kernel $K_{\overline{\emptyset}}\equiv K$ and multibandwidth h.

Next, the introduction of the estimator $\widehat{f}_{(h,\mathcal{P})}$ 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. Moreover, by scrutinizing the proof of Theorems 1 and 2 below, we see that

$$\mathcal{R}_p^{(q)}\left[\widehat{f}_{(h,\mathcal{P})},f\right]\le C_1\left(\mathbb{E}_f\sup_{I\in\mathcal{P}}\left\|\widehat{f}_{h_I}-f_I\right\|_{p,I}^q\right)^{\frac{1}{q}}+C_2n^{-1/2},\quad C_1,C_2>0.$$

Here and in the sequel $\|.\|_{s,I}$ denotes the norm $\|.\|_{\mathbb{L}_s(\mathbb{R}^{|I|},\mathrm{d}x_I)}$, $s\in[1,+\infty]$, $I\in\mathcal{I}_d$.

Remark 1. As it is discussed above, if $\mathcal{P} \in \mathfrak{P}(f)$ is known, the initial problem is reduced to the estimation of marginals f_I, $I \in \mathcal{P}$. Therefore, the natural loss that can be used in the definition of the risk for our problem seems to be

$$l\left(\widehat{f}_{(h,\mathcal{P})}, f\right) = \sup_{I \in \mathcal{P}} \left\| \widehat{f}_{h_I} - f_I \right\|_{p,I}.$$

In Section 2.3 we propose a data driven selection from the family $\mathfrak{F}[\,\mathfrak{P}\,]$. The possibility of choosing the sets $\mathcal{H}_I$ is introduced to make our procedure practically feasible. Indeed, $\mathcal{H}_I$ can be chosen as an appropriate grid in $[h_{min}, h_{max}]^{|I|}$. To define our selection rule, we need to introduce some notation and quantities.

2.2. *Auxiliary estimators and quantities*

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

$$\widehat{f}_{h_I,\eta_I}(x_I) := K_{\eta_I} \star \widehat{f}_{h_I}(x_I),$$

where "$\star$" stands for the convolution product on $\mathbb{R}^{|I|}$. Obviously, $\widehat{f}_{h_I,\eta_I} \equiv \widehat{f}_{\eta_I,h_I}$.

We endow the set $\mathfrak{P}$ with the operation "$\diamond$" introduced in Lepski [27]: for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' := \left\{ I \cap I' \neq \emptyset,\ I \in \mathcal{P},\ I' \in \mathcal{P}' \right\},$$

that is, in its turn, a partition of $\{1, \ldots, d\}$.

This allows us to define for $h, \eta \in (0,1]^d$ and $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) := \prod_{I \in \mathcal{P} \diamond \mathcal{P}'} \widehat{f}_{h_I,\eta_I}(x_I). \tag{2.3}$$

The ideas that led to the introduction of the estimators $\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$, based on both the operation "$\star$" and "$\diamond$", are explained in Lepski [27], Section 2.1, paragraph "*Estimation construction*". Note that the arguments given in the latter paper do not depend on the norm used in the definition of the risk and remain valid for estimation under $\mathbb{L}_p$-loss. Here, we give only the following simple explanation. Inspired by the methodology proposed by Goldenshluger and Lepski [14], Section 2.6, we seek auxiliary estimators in the form (2.3) noting that

$$\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} \equiv \widehat{f}_{(\eta,\mathcal{P}'),(h,\mathcal{P})}.$$

Moreover, we remark that $\widehat{f}_{\eta_{I'}}(x_{I'}) - \mathbb{E}_f\{\widehat{f}_{\eta_{I'}}(x_{I'})\}$, $I' \in \mathcal{P}'$, $x_{I'} \in \mathbb{R}^{|I'|}$, is the sum of i.i.d. bounded and centered random variables and, therefore, is "somehow small". Thus, we can expect that

$$\widehat{f}_{(\eta,\mathcal{P}')}(x) = \prod_{I' \in \mathcal{P}'} \widehat{f}_{\eta_{I'}}(x_{I'}) \approx \prod_{I' \in \mathcal{P}'} \mathbb{E}_f \left\{ \widehat{f}_{\eta_{I'}}(x_{I'}) \right\}.$$

For all $\mathcal{P} \in \mathfrak{P}(f)$, where $\mathfrak{P}(f)$ is defined by (1.2), one has

$$\prod_{I' \in \mathcal{P}'} \mathbb{E}_f \left\{ \widehat{f}_{\eta_{I'}}(x_{I'}) \right\} = \prod_{I' \in \mathcal{P}'} \prod_{I \in \mathcal{P}: I \cap I' \neq \emptyset} \mathbb{E}_f \left\{ \widehat{f}_{\eta_{I \cap I'}}(x_{I \cap I'}) \right\} = \prod_{I \in \mathcal{P} \diamond \mathcal{P}'} K_{\eta_I} \star f_I(x_I).$$

Finally, since $\widehat{f}_{h_I}$ is an estimate of f_I, we come to the introduction of $\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$ and we can expect that

$$\widehat{f}_{(\eta,\mathcal{P}')}(x) \approx \widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x).$$

However, we emphasize that the methodology developed by Goldenshluger and Lepski [14] cannot be applied to the selection of a partition $\mathcal{P}$ since it is not based on the selection from a family of linear estimators. Furthermore, the estimation under $\mathbb{L}_p$-loss, $1 \le p < \infty$, instead of sup-norm loss, leads us to modify the method proposed in Lepski [27] by introducing the following quantities and some specifical technical arguments to compute our risk bounds; see the proof of Theorems 1 and 2, Section 4.2.

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

$$\widehat{\mathcal{U}}_p(h_I) := \begin{cases} 128 n^{1/p-1} \left\| K_{h_I} \right\|_{p,I}, & p \in [1,2), \\ \frac{25}{3} n^{-1/2} \left\| K_{h_I} \right\|_{2,I}, & p = 2, \\ 32 \left[\widehat{\rho}_p\left(K_{h_I}\right) \vee n^{-1/2} \left\| K_{h_I} \right\|_{2,I} \right], & p > 2, \end{cases}$$

$$\widehat{\rho}_p\left(K_{h_I}\right) := \frac{15p}{\ln p} \Bigg\{ n^{-\frac{1}{2}} \left[\int_{\mathbb{R}^{|I|}} \left(\frac{1}{n} \sum_{i=1}^{n} \left[K_{h_I}\left(x_I - X_{i,I}\right) \right]^2 \right)^{\frac{p}{2}} \mathrm{d}x_I \right]^{\frac{1}{p}} + 2 n^{\frac{1}{p}-1} \left\| K_{h_I} \right\|_{p,I} \Bigg\}.$$

For $h \in (0,1]^d$ and $\mathcal{P} \in \mathfrak{P}$ put $\widehat{\mathcal{U}}_p(h,\mathcal{P}) := \sup_{I \in \mathcal{P}} \widehat{\mathcal{U}}_p(h_I)$.

We will see in Section 4.1 that the quantities $\widehat{\mathcal{U}}_p(h,\mathcal{P})$ can be viewed as uniform bounds on the $\mathbb{L}_p$-norm of the stochastic errors related to the estimators from the family $\mathfrak{F}[\mathfrak{P}]$. Such "majorants" were developed in Goldenshluger and Lepski [13] and used in Goldenshluger and Lepski [14] for multivariate density estimation under $\mathbb{L}_p$-loss. Let us remark that $\widehat{\mathcal{U}}_p(h,\mathcal{P})$ is a deterministic quantity when $p \in [1,2]$, and a random one when $p > 2$. In both cases, it follows from the results in Lemmas 1 and 2 below, that

$$\left(\mathbb{E}_f \left[\widehat{\mathcal{U}}_p(h,\mathcal{P}) \right]^q \right)^{\frac{1}{q}} \le C_3 \sup_{I \in \mathcal{P}} \left(n V_{h_I} \right)^{-\gamma_p}, \quad \forall (h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,],$$

where $C_3 > 0$ is a constant and γ_p is given in (1.1).

Define finally $\Lambda_p := d[\,\overline{G}_p]^{d(d-1)}$, where

$$\overline{G}_p := 1 \vee \left[\|\mathbf{K}\|_1^d \sup_{(h,\mathcal{P}) \in\, \mathcal{H}[\,\mathfrak{P}\,]} \sup_{\mathcal{P}' \in\, \mathfrak{P}} \left(\sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left\| \widehat{f}_{h_I} \right\|_{p,I} \right) 1_{\{\mathcal{P}' \ne \overline{\emptyset}\} \cup \{\mathcal{P} \ne \overline{\emptyset}\}} \right].$$

We remark that if $\mathfrak{P} = \{\overline{\emptyset}\}$ then $\overline{G}_p = 1$.

2.3. Selection rule and oracle inequalities

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

$$\widehat{\Delta}_p(h,\mathcal{P}) := \sup_{(\eta,\mathcal{P}')\in\, \mathcal{H}[\,\mathfrak{P}\,]} \left[\left\| \widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{(\eta,\mathcal{P}')} \right\|_p - \Lambda_p \widehat{\mathcal{U}}_p(\eta,\mathcal{P}') \right]_+ . \tag{2.4}$$

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

$$\widehat{\Delta}_p(\widehat{h},\widehat{\mathcal{P}}) + \Lambda_p \widehat{\mathcal{U}}_p(\widehat{h},\widehat{\mathcal{P}}) = \inf_{(h,\mathcal{P})\in\, \mathcal{H}[\,\mathfrak{P}\,]} \left[\widehat{\Delta}_p(h,\mathcal{P}) + \Lambda_p \widehat{\mathcal{U}}_p(h,\mathcal{P}) \right]. \tag{2.5}$$

Our selected estimator is $\widehat{f} := \widehat{f}_{(\widehat{h},\widehat{\mathcal{P}})}$.

It is easily checked that $(\widehat{h},\widehat{\mathcal{P}})$ exists, is in $\mathcal{H}[\,\mathfrak{P}\,]$ and is measurable, see, e.g., Lepski [27], section 2.1, paragraph "*Existence and measurability*", for more details.

We also emphasize that the construction of the proposed procedure does not require any condition concerning the density f. However, the following mild assumption will be used for computing its risk:

$$f \in \mathbb{F}\left[\mathbf{f},\mathfrak{P}\,\right] := \left\{ f : \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|f_I\|_\infty \le \mathbf{f},\ \mathfrak{P}(f) \neq \emptyset \right\}, \quad \mathbf{f} > 0, \tag{2.6}$$

where $\mathfrak{P}(f)$ is given in (1.2). Note that the considered class of densities is determined by $\mathfrak{P}$ and in particular

$$\mathbb{F}\left[\mathbf{f},\{\overline{\emptyset}\}\right] = \left\{ f : \|f\|_\infty \le \mathbf{f} \right\},\ \mathbb{F}\left[\mathbf{f},\{\mathcal{P}\}\right] = \left\{ f : \sup_{I\in\mathcal{P}} \|f_I\|_\infty \le \mathbf{f},\ \mathfrak{P}(f) = \{\mathcal{P}\} \right\}.$$

Define, for $(h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$ such that $\mathcal{P} \in \mathfrak{P}(f)$,

$$\mathcal{R}_p^{(q)}\left[(h,\mathcal{P}),f\right] := \left(\mathbb{E}_f \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\| \widehat{f}_{h_I} - f_I \right\|_{p,I}^q \right)^{\frac{1}{q}}, \quad q \ge 1.$$

If the possible independence structure $\mathcal{P}$ of the target density is known, the latter quantity can be viewed as an "$\mathbb{L}_p$-*risk*" of the estimator $\widehat{f}_{(h,\mathcal{P})}$, defined with the loss

$$l\left(\widehat{f}_{(h,\mathcal{P})},f\right) := \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\| \widehat{f}_{h_I} - f_I \right\|_{p,I}.$$

In this case, we see that the effective dimension of estimation is not d, but $d(\mathcal{P}) := \sup_{I\in\mathcal{P}} |I|$. Therefore, the best estimator from the family $\mathfrak{F}[\,\mathfrak{P}\,]$ (the oracle) should be $\widehat{f}_{(h^*,\mathcal{P}^*)}$ such that

$$\mathcal{R}_p^{(q)}\left[(h^*,\mathcal{P}^*),f\right] = \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \mathcal{R}_p^{(q)}\left[(h,\mathcal{P}),f\right].$$

Let us provide the following oracle inequalities for our selected estimator $\widehat{f}$.

Theorem 1. *Assume that $nV_{min} \geq 1$. For all $0 < \mathbf{f} < +\infty$ and all $q \geq 1$:*
(i) if $p \in [1,2)$ and $n \geq 3 \vee 4^{2p(2-p)}$ then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$,

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \left\{\mathcal{R}_p^{(q)}\left[(h,\mathcal{P}), f\right] + \sup_{I\in\mathcal{P}} (nV_{h_I})^{\frac{1}{p}-1}\right\} + \alpha_{p,2} n^{-\frac{1}{2}};$$

(ii) if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$ then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$,

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \left\{\mathcal{R}_p^{(q)}\left[(h,\mathcal{P}), f\right] + \sup_{I\in\mathcal{P}} (nV_{h_I})^{-\frac{1}{2}}\right\} + \alpha_{p,2} n^{-\frac{1}{2}}.$$

The constants $\alpha_{p,i} := \alpha_{p,i}(\boldsymbol{K}, d, q, p, \mathbf{f})$, $p \in [1,2]$, $i = 1,2$, are given in the proof of the theorem.

Theorem 2. *Let $\mathbf{f} > 0$, $q \geq 1$ and $p > 2$. Assume that for some constants $\overline{C}_3$ and $\overline{C}_4$*

$$n \geq \overline{C}_3 \vee 3, \qquad nV_{min} > 1 \vee \overline{C}_4, \qquad n^{-1/(2d)} \leq h_{max} \leq \big[\ln(n)\big]^{-p}.$$

Then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$,

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \left\{\mathcal{R}_p^{(q)}\left[(h,\mathcal{P}), f\right] + \sup_{I\in\mathcal{P}} (nV_{h_I})^{-\frac{1}{2}}\right\} + \alpha_{p,2} n^{-\frac{1}{2}}.$$

The constants $\overline{C}_3$, $\overline{C}_4$ and $\alpha_{p,i}$, $p > 2$, $i = 1,2$, are given in the proof of the theorem and depend on $\boldsymbol{K}, d, q, p$ and $\mathbf{f}$.

Here, we see that the possibility of choosing the set of partitions $\mathfrak{P}$ is interesting for other reasons than the computational one. Indeed, the latter results lead us to consider various problem in the framework of density estimation.

First, it is possible to consider that $\mathfrak{P}$ contains the two elements $\overline{\emptyset}$ and $\{\{1\}, \ldots, \{d\}\}$, if we suppose that the target density has independent components. We may also consider that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of the underlying density is known...

Next, for $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure) we automatically obtain oracle inequalities given in Theorems 1 and 2 in Goldenshluger and Lepski [14], up to numerical constants. The proof of Theorem 3 in the latter paper indicates that, for $p \in [2, \infty)$,

$$\left(\mathbb{E}_f \left\|\widehat{f}_{h_I} - f_I\right\|_{p,I}^q\right)^{\frac{1}{q}} \geq C_4 \left(nV_{h_I}\right)^{-1/2},$$

where $C_4 > 0$ is a constant. This lower bound holds under very weak assumptions on the density f and, together with the result of our Theorem 2, leads to an oracle inequality

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] \leq \overline{\alpha}_{p,1} \inf_{(h,\mathcal{P})\in\mathcal{H}[\mathfrak{P}]:\mathcal{P}\in\mathfrak{P}(f)} \mathcal{R}_p^{(q)}\left[(h,\mathcal{P}), f\right] + \overline{\alpha}_{p,2} n^{-1/2},$$

for some constants $\overline{\alpha}_{p,i} > 0$, $i = 1, 2$.

If $\mathfrak{P} = \{\overline{\emptyset}\}$, then $\mathcal{R}_p^{(q)}[(h,\overline{\emptyset}), f] = \mathcal{R}_p^{(q)}[\widehat{f}_{(h,\overline{\emptyset})}, f]$ and we obtain a so-called $\mathbb{L}_p$-*risk oracle inequality*. Note, however, that for all other cases, $\mathcal{R}_p^{(q)}[(h,\mathcal{P}), f]$ is an upper bound of $\mathcal{R}_p^{(q)}[\widehat{f}_{(h,\mathcal{P})}, f]$, up to a numerical constant. This seems to be a price to pay for taking into account the possible independence structure of the underlying density and, thus, for reducing the influence of the dimension on the quality of estimation.

Furthermore, comparing our results with those in Goldenshluger and Lepski [14], we remark that another price to pay for our problem appears through the constant $\alpha_{p,1}$; see the computations in the proofs of Theorems 1 and 2. Indeed, the prime interest is to obtain oracle inequalities with a constant $\alpha_{p,1}$ close to 1, and this seems to be more difficult whenever we consider that the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$.

However, Theorems 1 and 2 in the present paper lead us to consider various problems arising in the framework of minimax and minimax adaptive estimation. This is the subject of Section 3 below.

2.4. A short simulation study

Consider that we estimate a bivariate density ($d = 2$). Thus, the set of partitions $\mathfrak{P}$ contains the two elements $\mathcal{P}_1 = \{\{1\},\{2\}\}$ and $\mathcal{P}_2 = \{\{1,2\}\}$. Moreover, if we consider that the smoothness parameter $h = (h_1, h_2)$ is fixed, we only have to compare the accuracy of the estimator $\widehat{f}_{(h,\mathcal{P}_1)}$ with that of the classical kernel one $\widehat{f}_{(h,\mathcal{P}_2)} = \widehat{f}_h$. Then, the main question is: does our strategy choose the partition $\mathcal{P}_1$ when the two components of X_1 are independent?

Here, we answer to this question in the following case:

$$f(x_1,x_2) = \frac{1}{2\pi\sigma^2} e^{-(x_1^2+x_2^2)/(2\sigma^2)}, \quad \sigma = 0,1;$$

$$\mathbf{K}_{\mathrm{h}_i}(x_i) = \frac{1}{\mathrm{h}_i\sqrt{2\pi}} e^{-x_i^2/(2\mathrm{h}_i^2)}, \quad \mathrm{h}_i = 0,0313, \ \ i = 1, 2.$$

For simplicity, we estimate f on a grid of 100×100 points in the domain $[-1/2, 1/2]^2$ via Fast Fourier Transform, by using $n = 1000$ simulated random vectors. Because f is an isotropic density, the smoothness parameter $\mathrm{h} = (\mathrm{h}_1, \mathrm{h}_2)$ is an isotropic vector properly chosen in the dyadic grid $\{h = (2^{-k}, 2^{-k}) : \ k \in \mathbb{N}, \log_2(\ln^2(n)) \leq k \leq \log_2(n)\}$, in order to minimize both the $\mathbb{L}_2$-risk (average over 1000 samples) of $\widehat{f}_{(h,\mathcal{P}_1)}$ and the one of $\widehat{f}_{(h,\mathcal{P}_2)}$.

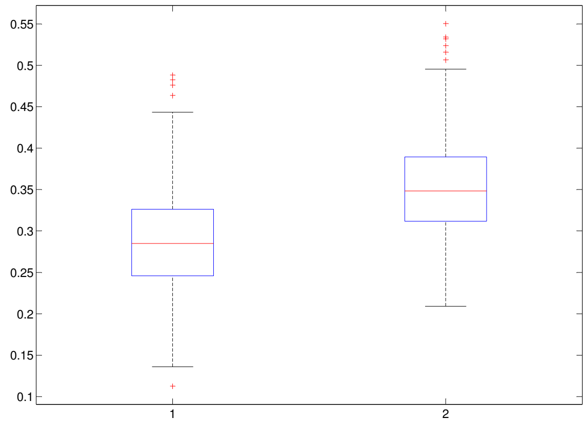

FIG 1. *Comparison of* $\mathbb{L}_2$*-losses.*

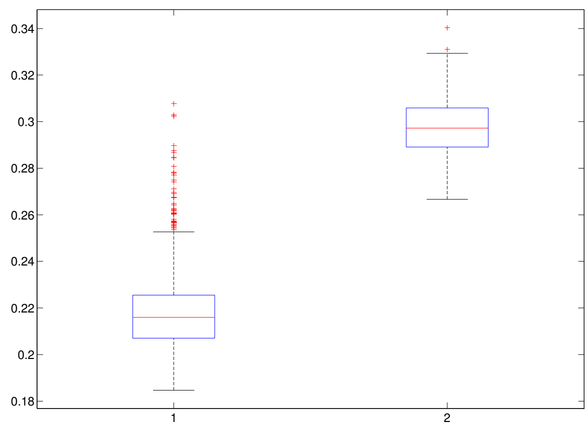

FIG 2. *Comparison of selection criterions.*

Figure 1 shows the boxplots of values of the $\mathbb{L}_2$-loss of both estimators $\widehat{f}_{(\mathrm{h},\mathcal{P}_1)}$ (on the left) and $\widehat{f}_{(\mathrm{h},\mathcal{P}_2)}$ (on the right) over 1000 samples. Note that, in this case, $\widehat{f}_{(\mathrm{h},\mathcal{P}_1)}$ outperforms $\widehat{f}_{(\mathrm{h},\mathcal{P}_2)}$ 991 times and

$$\mathcal{R}_2^{(1)}\left[\widehat{f}_{(\mathrm{h},\mathcal{P}_1)}, f\right] = 0,2884 < 0,3523 = \mathcal{R}_2^{(1)}\left[\widehat{f}_{(\mathrm{h},\mathcal{P}_2)}, f\right],$$

where, here, $\mathcal{R}_2^{(1)}[\widehat{f}_{(\mathrm{h},\mathcal{P}_i)}, f]$ denotes the $\mathbb{L}_2$-risk of $\widehat{f}_{(\mathrm{h},\mathcal{P}_i)}$, $i = 1, 2$, average over 1000 samples.

Figure 2 shows the boxplots of values of both selection criterions $\widehat{\Delta}_2(\mathrm{h},\mathcal{P}_1) + \Lambda_2\widehat{\mathcal{U}}_2(\mathrm{h},\mathcal{P}_1)$ (on the left) and $\widehat{\Delta}_2(\mathrm{h},\mathcal{P}_2) + \Lambda_2\widehat{\mathcal{U}}_2(\mathrm{h},\mathcal{P}_2)$ (on the right) over 1000 samples, with a random quantity Λ_2 multiplied by $c = 0,01$. Here, our strategy chooses the partition $\mathcal{P}_1$ 999 times. We conclude that, for this example, the selected estimator outperforms the classical kernel estimator in almost all cases.

3. $\mathbb{L}_p$ adaptive estimation

In this section, we discuss adaptive minimax estimation over a certain scale of anisotropic Nikolskii classes when the smoothness of the underlying density is assumed to be measured with the same $\mathbb{L}_p$-norm that used to measure the quality of estimation.

3.1. Anisotropic Nikolskii classes of densities related to independence structure

We start with the definition of the *anisotropic Nikolskii class of densities* we use in the sequel. Let $\{e_1, \dots, e_s\}$ denote the canonical basis in $\mathbb{R}^s$, $s \in \mathbb{N}^*$.

Definition 1. Let $p \in [1, \infty)$, $\beta = (\beta_1, \dots, \beta_s)$, $\beta_i > 0$ and $L = (L_1, \dots, L_s)$, $L_i > 0$. A probability density $f : \mathbb{R}^s \to \mathbb{R}$ belongs to the anisotropic Nikolskii class $N_{p,s}(\beta, L)$ if

$$(i)\ \left\| D_i^k f \right\|_p \leq L_i, \quad \forall k = 0, \dots, \lfloor \beta_i \rfloor, \ \forall i = 1, \dots, s;$$

$$(ii)\ \left\| D_i^{\lfloor \beta_i \rfloor} f(\cdot + te_i) - D_i^{\lfloor \beta_i \rfloor} f(\cdot) \right\|_p \leq L_i \left| t \right|^{\beta_i - \lfloor \beta_i \rfloor}, \ \forall t \in \mathbb{R}, \ \forall i = 1, \dots, s.$$

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.

In order to take into account the smoothness of the underlying density and its possible independence structure simultaneously, a collection of anisotropic Nikolskii classes of densities was introduced in Lepski [27], Section 3, Definition 2. However, since the adaptation is not necessarily considered with respect to the set of all partitions of $\{1, \dots, d\}$, the condition imposed therein can be weakened. For instance, if $\mathfrak{P} = \{\bar{\emptyset}\}$ (no independence structure), we want to find again the well known results concerning the adaptive estimation over the scale of anisotropic Nikolskii classes of densities $\{N_{p,d}(\beta, L)\}$, that is not possible with the classes introduced in Lepski [27]. For these reasons, the following collection $\{N_{p,d}(\beta, L, \mathcal{P})\}_{\mathcal{P}}$ was introduced in Rebelles [33], Section 3.1.

Definition 2. Let $(\beta, p, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \mathfrak{P}$ be fixed. A probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $N_{p,d}(\beta, L, \mathcal{P})$ if $f(x) = \prod_{I \in \mathcal{P}} f_I(x_I)$, $\forall x \in \mathbb{R}^d$, and

$$f_I \in N_{p_I, |I|}(\beta_I, L_I), \ \ \forall I \in \mathcal{P}' \diamond \mathcal{P}'', \ \forall \, (\mathcal{P}', \ \mathcal{P}'') \in \mathfrak{P} \times \mathfrak{P}.$$

Finally, recall that the condition $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\ ]$ is required in Theorems 1 and 2, and define

$$N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right) := N_{p,d}\left(\beta, L, \mathcal{P}\right) \cap \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\ \right], \quad 0 < \mathbf{f} < +\infty. \tag{3.1}$$

In the next section, we illustrate the application of Theorems 1 and 2 to adaptive estimation over anisotropic Nikolskii classes of densities $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$.

3.2. Adaptive minimax estimation

For $p \in [1, \infty)$ and $(\beta, \mathcal{P}) \in (0, +\infty)^d \times \mathfrak{P}$ define $\varphi_{n,p}(\beta, \mathcal{P}) := n^{-\frac{\gamma_p \overline{r}}{\gamma_p + \overline{r}}}$, where γ_p is given in (1.1) and

$$\overline{r} := \overline{r}(\beta, \mathcal{P}) = \inf_{I \in \mathcal{P}} \overline{\beta}_I, \qquad \overline{\beta}_I := \left[\sum_{i \in I} \frac{1}{\beta_i}\right]^{-1}, \quad I \in \mathcal{P}. \tag{3.2}$$

We provide the following minimax lower bound.

Theorem 3. *For any $\mathbf{f} > 0$, any $(\beta, L, \mathcal{P}) \in (0, \infty)^d \times (0, \infty)^d \times \mathfrak{P}$ and any $p \in (1, \infty)$*

$$\liminf_{n \to +\infty} \left\{ \varphi_{n,p}^{-1}(\beta, \mathcal{P}) \inf_{\widetilde{f}} \mathcal{R}_p^{(q)} \left[\widetilde{f}, N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right] \right\} > 0,$$

where infimum is taken over all possible estimators.

The proof of Theorem 3 coincides with the one of Theorem 3 in Goldenshluger and Lepski [15], up to minor modifications to take into account the independence structure of the underlying density. Therefore, it is omitted.

Our goal now is to show that $\varphi_{n,p}(\beta, \mathcal{P})$ is the minimax rate of convergence on the anisotropic class $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$, and that a minimax estimator can be selected from the collection $\mathfrak{F}[\,\mathfrak{P}\,]$ given in (2.2).

Assume that $\overline{\emptyset} \in \mathfrak{P}$, that $\mathcal{H}_I$ is the dyadic grid in $\{h_I \in [h_{min}, h_{max}]^{|I|} : V_{h_I} \geq V_{min}\}$, $I \in \mathcal{I}_d$, and consider the estimator $\widehat{f}$ defined by the selection rule (2.4)–(2.5). We show below that the quality of estimation of $\widehat{f}$ is optimal up to a numerical constant on each class $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$, whatever the nuisance parameter $(\beta, L, \mathcal{P}, \mathbf{f})$. We achieve the latter goal with properly chosen kernel $\mathbf{K}$ and numbers h_{max}, h_{min} and V_{min}.

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

$$u_l(z) := \sum_{i=1}^{l} \left(\begin{array}{c} l \\ i \end{array}\right) (-1)^{i+1} \frac{1}{i} u\left(\frac{z}{i}\right), \quad z \in \mathbb{R}. \tag{3.3}$$

Furthermore we use $\mathbf{K} \equiv u_l$ in the definition of the collection of estimators $\mathfrak{F}[\,\mathfrak{P}\,]$. The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [23]. In particular, it was shown that

$$\int_{\mathbb{R}} \mathbf{K}(z)\mathrm{d}z = 1, \qquad \int_{\mathbb{R}} z^k \mathbf{K}(z)\mathrm{d}z = 0, \quad \forall k = 1, \dots, l-1. \tag{3.4}$$

Choose finally $h_{max} := [\ln(n)]^{-(2 \vee p)}$, $h_{min} := n^{-1}$ and $V_{min} := (\overline{C}_4 + 1)n^{-1}$.

Theorem 4. *Let $p \in (1, \infty)$. Then for any $\mathbf{f} > 0$ and any $(\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathfrak{P}$ one has*

$$\limsup_{n \to +\infty} \left\{ \varphi_{n,p}^{-1}(\beta, \mathcal{P}) \mathcal{R}_p^{(q)} \left[\widehat{f}, N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right] \right\} < \infty.$$

It follows that $\varphi_{n,p}(\beta,\mathcal{P})$ is the minimax rate of convergence on each functional class $N_{p,d}(\beta,L,\mathcal{P},\mathbf{f})$ and that our estimator, which is fully data-driven, is an O.A.E. over the scale of functional classes $\{N_{p,d}(\beta,L,\mathcal{P},\mathbf{f})\}_{(\beta,L,\mathcal{P},\mathbf{f})}$. Let us briefly discuss other consequences of Theorem 4.

First, if $\mathfrak{P}=\{\overline{\emptyset}\}$, we obtain automatically the minimax adaptive upper bound given in Goldenshluger and Lepski [14], Theorem 4.

Next, in view of the latter consideration, Theorem 4 allows us to compare the influence of the independence structure on the accuracy of estimation. For example, we see that

$$\varphi_{n,p}\big(\beta,\overline{\emptyset}\big) \gg \varphi_{n,p}\left(\beta,\mathcal{P}\right), \quad \forall \mathcal{P}\neq\overline{\emptyset}.$$

We conclude that the existence of an independence structure improves significantly the accuracy of estimation with $\mathbb{L}_p$-risk. The same conclusion was obtained in Lepski [27] for density estimation under the sup-norm loss and in Rebelles [33] for pointwise density estimation. It is also important to emphasize that there is no price to pay for adaptation to the independence structure in the framework of estimation with an $\mathbb{L}_p$-loss, whereas there is a "ln-price" in the pointwise setting, see Rebelles [33]. Note that, if $\mathfrak{P}=\{\overline{\emptyset}\}$ (no independence structure), there is still a "ln-price" to pay for adaptation to the smoothness parameter when we consider the pointwise criterion. This was shown for the first time in Lepski [26] for the Gaussian white noise model, in the unidimensional case.

Finally, in view of the embedding theorem for anisotropic Nikolskii classes, see, e.g., Theorem 6.9 in Nikolskii [31], if $\sum_{i=1}^{d}1/\beta_i<p$, there exists a number $\mathbf{f}:=\mathbf{f}(\beta,p)>0$ such that $N_{p,d}(\beta,L,\mathcal{P})\subseteq\mathbb{F}[\mathbf{f},\mathfrak{P}\,]$. Therefore, we deduce from Theorems 3 and 4 that our estimator is an O.A.E. over the scale

$$\left\{ N_{p,d}(\beta,L,\mathcal{P}),\ \ (\beta,L,\mathcal{P})\in(0,l]^d\times(0,+\infty)^d\times\mathfrak{P},\ \ \sum_{i=1}^{d}1/\beta_i<p \right\}.$$

4. Proofs of main results

The main technical tools used in our derivations are uniform bounds on the $\mathbb{L}_p$-norm of empirical processes developed in Goldenshluger and Lepski [13]. We start this section by giving corresponding results established in Goldenshluger and Lepski [14] for multivariate-density estimation under $\mathbb{L}_p$-loss.

4.1. Uniform bounds on the $\mathbb{L}_p$-norm of kernel empirical processes

Let $f\in\mathbb{F}[\mathbf{f},\mathfrak{P}\,]$, $\mathbf{f}>0$, and $I\in\mathcal{I}_d$ be fixed. Remind that $\frac{1}{n}\leq h_{min}\leq h_{max}\leq 1$, that

$$\mathcal{H}_I\subseteq\left\{h_I\in\left[h_{min},h_{max}\right]^{|I|}:\ V_{h_I}\geq V_{min}\ \right\},$$

and put

$$A_{\mathcal{H}_I}:=\left[1\vee\ln(h_{max}/h_{min})\right]^{|I|},\quad B_{\mathcal{H}_I}:=1\vee\left[|I|\log_2(h_{max}/h_{min})\right].$$

For $h_I \in (0,1]^{|I|}$ and $x_I \in \mathbb{R}^{|I|}$, define $\xi_{h_I}(x_I) := \widehat{f}_{h_I}(x_I) - \mathbb{E}_f\{\widehat{f}_{h_I}(x_I)\}$, and

$$\rho_p\left(K_{h_I}\right) := \frac{15p}{\ln p}\left\{ n^{-\frac{1}{2}}\left[\int_{\mathbb{R}^{|I|}}\left(\int_{\mathbb{R}^{|I|}}\left[K_{h_I}(x_I - y_I)\right]^2 f_I(y_I)\mathrm{d}y_I\right)^{\frac{p}{2}}\mathrm{d}x_I\right]^{\frac{1}{p}} + 2n^{\frac{1}{p}-1}\left\|K_{h_I}\right\|_{p,I}\right\}.$$

Propositions 1 and 2 below follow immediately from Lemmas 1 and 2 established in Goldenshluger and Lepski [14], Section 4.1. Indeed, assumptions (K1) and (K2) required in the latter paper are satisfied for $L_K = \overline{L}_{\mathbf{K}} := d\|\mathbf{K}\|_\infty^{d-1} L_{\mathbf{K}}$ and $k_\infty = \|\mathbf{K}\|_\infty^d$.

Proposition 1. *(i) If $p \in [1,2)$, then for all integer $n \geq 4^{2p/(2-p)}$*

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I}\left[\left\|\xi_{h_I}\right\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+^q\right\}^{\frac{1}{q}} \leq C_1 A_{\mathcal{H}_I}^{\frac{4}{q}} n^{\frac{1}{p}} \exp\left\{-\frac{2n^{\frac{2}{p}-1}}{37q}\right\}.$$

(ii) Assume that $8[\mathbf{f}^2 h_{max}^{|I|} + 4n^{-1/2}] \leq 1$, then

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I}\left[\left\|\xi_{h_I}\right\|_{2,I} - \widehat{\mathcal{U}}_2(h_I)\right]_+^q\right\}^{\frac{1}{q}} \leq C_2 A_{\mathcal{H}_I}^{\frac{2}{q}} n^{\frac{1}{2}} \exp\left\{-\frac{(16q)^{-1}}{\mathbf{f}^2 h_{max}^{|I|} + 4n^{-\frac{1}{2}}}\right\}.$$

Here $C_i = C_i(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, q)$, $i = 1,2$.

Proposition 2. *Let $p > 2$. Assume that $n \geq C_3$, $nV_{min} > C_4$, and $h_{max}^{|I|} \geq 1/\sqrt{n}$. Then*

$$(i)\quad \left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I}\left[\left\|\xi_{h_I}\right\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+^q\right\}^{\frac{1}{q}} \leq C_5 A_{\mathcal{H}_I}^{\frac{2}{q}} B_{\mathcal{H}_I}^{\frac{1}{q}} n^{\frac{1}{2}} \exp\left\{-\frac{C_6}{\mathbf{f} h_{max}^{\frac{2}{p}}}\right\},$$

$$(ii)\quad \mathbb{E}_f \sup_{h_I \in \overline{\mathcal{H}}_I}\left[\widehat{\mathcal{U}}_p(h_I)\right]^q \leq 32\left(1 + \frac{120p}{\ln p}\right)^q \sup_{h_I \in \overline{\mathcal{H}}_I}\left[\rho_p\left(K_{h_I}\right) \vee \left(n^{-\frac{1}{2}}\left\|K_{h_I}\right\|_{2,I}\right)\right]^q + 32 C_7 A_{\mathcal{H}_I}^2 B_{\mathcal{H}_I} n^{\frac{q(p-2)}{2p}} \exp\left\{-C_8 b_{n,p}\right\}, \qquad \forall \overline{\mathcal{H}}_I \subseteq \mathcal{H}_I,$$

where $b_{n,p} = n^{\frac{4}{p}-1}$ if $p \in (2,4)$ and $b_{n,p} = \{\mathbf{f} h_{max}^{\frac{4}{p}}\}^{-1}$ if $p \in [4,\infty)$. Here $C_i = C_i(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, q, p)$, $i = 4,5,6,7,8$, $C_3 = C_3(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, q, p, \mathbf{f})$.

The following result is obtained straightforwardly by application of Theorem 2 in Goldenshluger and Lepski [13]. All technical arguments are given in the latter paper and its proof is omitted.

Proposition 3. *Let $p>2$. Assume that $n \geq C_3$, $nV_{min} > C_4$, and $h_{max}^{|I|} \geq 1/\sqrt{n}$. Then*

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - 9\rho_p\left(K_{h_I}\right)\right]_+^q\right\}^{\frac{1}{q}} \leq C_9 A_{\mathcal{H}_I}^{\frac{2}{q}} \exp\left\{-\frac{C_{10}}{h_{max}^{\frac{2}{p}}}\right\}.$$

Here $C_i = C_i(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, q, p, \mathbf{f})$, $i = 9, 10$.

All constants appearing in Propositions 1, 2 and 3 can be expressed explicitly, see corresponding results in Goldenshluger and Lepski [14] and in Goldenshluger and Lepski [13].

To compute our risk bounds we need the following technical lemmas. Define

$$\xi_p := \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+,$$

$$\zeta_p := \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - 9\rho_p\left(K_{h_I}\right)\right]_+,$$

$$\overline{\mathbf{f}}_p := d\,\|\mathbf{K}\|_1^d\,\Lambda_p\left(\max\left\{\overline{G}_p, \|\mathbf{K}\|_1^d\,\mathbf{f}^{1-1/p}\right\}\right)^{d-1},$$

and $\widehat{\mathcal{A}}_p(h,\mathcal{P}) := \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \widehat{\mathcal{U}}_p(h_I)$, $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$.

Lemma 1. *Let V_{min} be a fixed number such that $nV_{min} \geq 1$.*

(i) Assume that $p \in [1,2)$ and $n \geq 3 \vee 4^{2p/(2-p)}$. Then,

$$\left(\mathbb{E}_f\,|\xi_p|^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_1(q) n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f\left|\overline{\mathbf{f}}_p\right|^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_2(q), \quad \forall f \in \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\right].$$

(ii) Assume that $p=2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$. Then,

$$\left(\mathbb{E}_f\,|\xi_p|^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_3(q) n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f\left|\overline{\mathbf{f}}_p\right|^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_4(q), \quad \forall f \in \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\right].$$

(iii) If $p \in [1,2]$, then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$,

$$\widehat{\mathcal{A}}_p(h,\mathcal{P}) \leq 128\,\|\boldsymbol{K}\|_\infty^d \sup_{I \in \mathcal{P}} \left(nV_{h_I}\right)^{-\gamma_p}, \quad \forall (h,\mathcal{P}) \in \mathcal{H}\left[\,\mathfrak{P}\,\right].$$

Lemma 2. *Let $p>2$, and assume that for some constants $\overline{C}_3$ and $\overline{C}_4$*

$$n \geq \overline{C}_3, \qquad nV_{min} > 1 \vee \overline{C}_4, \qquad n^{-1/(2d)} \leq h_{max} \leq \left[\ln(n)\right]^{-p}.$$

Then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$,

$$(i)\ \left(\mathbb{E}_f\,|\xi_p|^q\right)^{\frac{1}{q}} \leq \frac{\boldsymbol{c}_5(q)}{n^{\frac{1}{2}}}, \quad \left(\mathbb{E}_f\,|\zeta_p|^q\right)^{\frac{1}{q}} \leq \frac{\boldsymbol{c}_6(q)}{n^{\frac{1}{2}}}, \quad \left(\mathbb{E}_f\left|\overline{\mathbf{f}}_p\right|^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_7(q);$$

$$(ii)\ \left(\mathbb{E}_f\left[\widehat{\mathcal{A}}_p(h,\mathcal{P})\right]^q\right)^{\frac{1}{q}} \leq \boldsymbol{c}_8(q) \sup_{I \in \mathcal{P}} \left(nV_{h_I}\right)^{-\frac{1}{2}} + \frac{\boldsymbol{c}_9(q)}{n^{\frac{1}{2}}}, \ \forall (h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}].$$

All constants involved in the latter lemmas are given in their proofs, those are postponed to the Appendix.

4.2. Oracle inequalities: Proof of Theorems 1 and 2.

Set $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$, $\mathbf{f} > 0$. We divide this proof into six steps.

1) Let $(h, \mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$, $\mathcal{P} \in \mathfrak{P}(f)$, be fixed. Thus, $h_I \in \mathcal{H}_I$, $\forall I \in \mathcal{P}$.
In view of the triangle inequality we have

$$\left\|\widehat{f} - f\right\|_p \leq \left\|\widehat{f}_{(\widehat{h},\widehat{\mathcal{P}})} - \widehat{f}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})}\right\|_p + \left\|\widehat{f}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})} - \widehat{f}_{(h,\mathcal{P})}\right\|_p + \left\|\widehat{f}_{(h,\mathcal{P})} - f\right\|_p,$$

$$\left\|\widehat{f} - f\right\|_p \leq \widehat{\Delta}_p(h, \mathcal{P}) + \Lambda_p \widehat{\mathcal{U}}_p(\widehat{h}, \widehat{\mathcal{P}}) + \widehat{\Delta}_p(\widehat{h}, \widehat{\mathcal{P}}) + \Lambda_p \widehat{\mathcal{U}}_p(h, \mathcal{P}) + \left\|\widehat{f}_{(h,\mathcal{P})} - f\right\|_p.$$

Here we have used that $\widehat{f}_{(h,\mathcal{P}),(\widehat{h},\widehat{\mathcal{P}})} = \widehat{f}_{(\widehat{h},\widehat{\mathcal{P}}),(h,\mathcal{P})}$. By definition of $(\widehat{h}, \widehat{\mathcal{P}})$, we obtain

$$\left\|\widehat{f} - f\right\|_p \leq 2\left[\widehat{\Delta}_p(h, \mathcal{P}) + \Lambda_p \widehat{\mathcal{U}}_p(h, \mathcal{P})\right] + \left\|\widehat{f}_{(h,\mathcal{P})} - f\right\|_p. \tag{4.1}$$

2) Suppose that $\mathcal{P} = \{I_1, \ldots, I_m\}$, $m \in \{1, \ldots, d\}$. Since $\mathcal{P} \in \mathfrak{P}(f)$, for any $x \in \mathbb{R}^d$

$$\left|\widehat{f}_{(h,\mathcal{P})}(x) - f(x)\right| = \left|\prod_{I \in \mathcal{P}} \widehat{f}_{h_I}(x_I) - \prod_{I \in \mathcal{P}} f_I(x_I)\right|$$
$$\leq \sum_{i=1}^{m} \left|\widehat{f}_{h_{I_i}}(x_{I_i}) - f_{I_i}(x_{I_i})\right| \left(\prod_{j=i+1}^{m} \left|\widehat{f}_{h_{I_j}}(x_{I_j})\right|\right) \left(\prod_{k=1}^{i-1} |f_{I_k}(x_{I_k})|\right).$$

Here we have used the trivial equality: for $m \in \mathbb{N}^*$ and $a_i, b_i \in \mathbb{R},\ i = 1, \ldots, m$,

$$\prod_{i=1}^{m} a_i - \prod_{i=1}^{m} b_i = \sum_{i=1}^{m} (a_i - b_i) \left(\prod_{j=i+1}^{m} a_j\right) \left(\prod_{k=1}^{i-1} b_k\right), \tag{4.2}$$

where the product over empty set is assumed to be equal to one.

In view of $\mathcal{P} \in \mathfrak{P}$, the triangle inequality and the Fubini-Tonelli theorem we establish

$$\left\|\widehat{f}_{(h,\mathcal{P})} - f\right\|_p \leq \sum_{i=1}^{m} \left\|\widehat{f}_{h_{I_i}} - f_{I_i}\right\|_{p,I_i} \left(\prod_{j=i+1}^{m} \left\|\widehat{f}_{h_{I_j}}\right\|_{p,I_j}\right) \left(\prod_{k=1}^{i-1} \|f_{I_k}\|_{p,I_k}\right)$$
$$\leq m \left(\overline{G}_p \vee \left\{\mathbf{f}^{1-1/p}\right\}\right)^{m-1} \sup_{I \in \mathcal{P}} \left\|\widehat{f}_{h_I} - f_I\right\|_{p,I}.$$

Here we have used that $\|\mathbf{K}\|_1 \geq \int \mathbf{K} = 1$. Since $\overline{G}_p \geq 1$, it follows

$$\left\|\widehat{f}_{(h,\mathcal{P})} - f\right\|_p \leq d \left(\overline{G}_p \vee \left\{\mathbf{f}^{1-1/p}\right\}\right)^{d-1} \sup_{I \in \mathcal{P}} \left\|\widehat{f}_{h_I} - f_I\right\|_{p,I}. \tag{4.3}$$

3) For any $(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]$ and any $x\in\mathbb{R}^d$

$$\begin{aligned}&\left|\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x)-\widehat{f}_{(\eta,\mathcal{P}')}(x)\right|\\&=\left|\prod_{I'\in\mathcal{P}'}\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}K_{\eta_{I\cap I'}}\star\widehat{f}_{h_{I\cap I'}}(x_{I\cap I'})-\prod_{I'\in\mathcal{P}'}\widehat{f}_{\eta_{I'}}(x_{I'})\right|.\end{aligned}$$

Therefore, by the same method as the one used in step 2, we establish

$$\begin{aligned}&\left\|\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}-\widehat{f}_{(\eta,\mathcal{P}')}\right\|_p\\&\leq d\left[\,\overline{G}_p\,\right]^{d(d-1)}\sup_{I'\in\mathcal{P}'}\left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\widehat{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\widehat{f}_{\eta_{I'}}\right\|_{p,I'}.\end{aligned}\tag{4.4}$$

Here we have used Young's inequality, that $\|\mathbf{K}\|_1\geq\int\mathbf{K}=1$ and that $\overline{G}_p\geq 1$.

4) In view of Young's inequality, for any $I\in\mathcal{I}_d$ and any $\eta\in(0,1]^d$

$$\left\|\mathbb{E}_f\left\{\widehat{f}_{\eta_I}(\cdot)\right\}\right\|_{p,I}=\|K_{\eta_I}\star f_I\|_{p,I}\leq\|K_I\|_{1,I}\,\|f_I\|_{p,I}\leq\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}.\tag{4.5}$$

Then, by the same method as the one used in step 2 and (4.5), for any $(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]$ and any $I'\in\mathcal{P}'$ we get

$$\begin{aligned}&\left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\widehat{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\mathbb{E}_f\left\{\widehat{f}_{\eta_{I\cap I'}}(\cdot)\right\}\right\|_{p,I'}\\&\leq d\left(\overline{G}_p\vee\left\{\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}\right\}\right)^{d-1}\sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|K_{\eta_{I\cap I'}}\star\left(\widehat{f}_{h_{I\cap I'}}-f_{I\cap I'}\right)\right\|_{p,I\cap I'}\\&\leq d\,\|\mathbf{K}\|_1^d\left(\overline{G}_p\vee\left\{\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}\right\}\right)^{d-1}\sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|\widehat{f}_{h_{I\cap I'}}-f_{I\cap I'}\right\|_{p,I\cap I'}.\end{aligned}\tag{4.6}$$

5) For $\eta\in(0,1]^d$ and $I'\in\mathcal{I}_d$, since $\mathcal{P}\in\mathfrak{P}(f)$, we have for any $x\in\mathbb{R}^d$

$$\begin{aligned}\mathbb{E}_f\left\{\widehat{f}_{\eta_{I'}}(x_{I'})\right\}&=\int K_{\eta_{I'}}\left(y_{I'}-x_{I'}\right)\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}f_{I\cap I'}(y_{I\cap I'})\mathrm{d}y_{I'}\\&=\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\mathbb{E}_f\left\{\widehat{f}_{\eta_{I\cap I'}}(x_{I\cap I'})\right\}.\end{aligned}$$

Here we have used the product structure of the kernel K and the Fubini theorem.

Thus, in view of the triangle inequality and (4.4), for any $(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]$, we get

$$
\begin{aligned}
&\|\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{(\eta,\mathcal{P}')}\|_p - \Lambda_p \widehat{\mathcal{U}}_p(\eta,\mathcal{P}') \\
&\quad\leq \quad \Lambda_p \sup_{I'\in\mathcal{P}'} \Bigg\{ \Bigg\| \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widehat{f}_{h_{I\cap I'},\eta_{I\cap I'}} - \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \mathbb{E}_f\left\{\widehat{f}_{\eta_{I\cap I'}}(\cdot)\right\} \Bigg\|_{p,I'} \\
&\qquad\qquad\qquad\qquad\qquad\qquad + \left\|\xi_{\eta_{I'}}\right\|_{p,I'} - \widehat{\mathcal{U}}_p(\eta,\mathcal{P}') \Bigg\}.
\end{aligned}
$$

We deduce, in view of (4.6) and the trivial inequality $[\sup_i x_i - \sup_i y_i]_+ \leq \sup_i [x_i - y_i]_+$,

$$
\widehat{\Delta}_p(h,\mathcal{P}) \leq \overline{\mathbf{f}}_p \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widehat{f}_{h_I} - f_I\right\|_{p,I} + \Lambda_p \xi_p. \tag{4.7}
$$

Finally, since $\|\mathbf{K}\|_1 \geq 1$ and $\overline{\mathbf{f}}_p \geq \Lambda_p \geq 1$, it follows from (4.1), (4.3) and (4.7)

$$
\left\|\widehat{f} - f\right\|_p \leq 3\overline{\mathbf{f}}_p \left\{ \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widehat{f}_{h_I} - f_I\right\|_{p,I} + \widehat{\mathcal{U}}_p(h,\mathcal{P}) + \xi_p \right\}. \tag{4.8}
$$

6) Consider the random event $B_p := \{\overline{G}_p \geq C_p\}$, where C_p is a constant to be specified.

- For $p \in [1,2)$, put $C_p = (1 + 128\|\mathbf{K}\|_\infty^d + \|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p})\|\mathbf{K}\|_1^d + 1$.

Remind that $nV_{h_I} \geq 1$, $\forall h_I \in \mathcal{H}_I$, $I \in \mathcal{I}_d$. In view of Lemma 1 (i)–(iii), Markov's inequality, (4.8), and the Cauchy-Schwarz inequality we get $B_p \subseteq \{\xi_p \geq 1\}$, $[\mathbb{P}_f(B_p)]^{\frac{1}{4q}} \leq \mathbf{c}_1(4q)n^{-1/2}$, and

$$
\begin{aligned}
&\left(\mathbb{E}_f \left\|\widehat{f} - f\right\|_p^q 1_{B_p^c}\right)^{\frac{1}{q}} \\
&\leq 3d^2 \|\mathbf{K}\|_1^d [C_p]^{d^2-1} \left(\mathcal{R}_p^{(q)}[(h,\mathcal{P}),f] + \sup_{I\in\mathcal{P}} \frac{128\|\mathbf{K}\|_\infty^d}{(nV_{h_I})^{\gamma_p}} + \frac{\mathbf{c}_1(q)}{n^{\frac{1}{2}}}\right), \\
&\left(\mathbb{E}_f \left\|\widehat{f} - f\right\|_p^q 1_{B_p}\right)^{\frac{1}{q}} \\
&\leq 3\mathbf{c}_1(4q)\mathbf{c}_2(4q)\left(\mathcal{R}_p^{(2q)}[(h,\mathcal{P}),f] + 128\|\mathbf{K}\|_\infty^d + \mathbf{c}_1(2q)\right) n^{-\frac{1}{2}}, \\
&\mathcal{R}_p^{(2q)}[(h,\mathcal{P}),f] \leq \mathbf{c}_1(2q) + 128\|\mathbf{K}\|_\infty^d + \|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p} + \mathbf{f}^{1-1/p}.
\end{aligned}
$$

Thus, we come to the assertion (i) of Theorem 1 with

$$
\begin{aligned}
\alpha_{p,1} &:= 384d^2 \|\mathbf{K}\|_1^d \|\mathbf{K}\|_\infty^d [C_p]^{d^2-1}, \\
\alpha_{p,2} &:= 3\mathbf{c}_1(4q)\mathbf{c}_2(4q)\left(256\|\mathbf{K}\|_\infty^d + (1 + \|\mathbf{K}\|_1^d)\mathbf{f}^{1-1/p} + 2\mathbf{c}_1(2q)\right) \\
&\qquad\qquad\qquad\qquad\qquad + 3\mathbf{c}_1(q)d^2 \|\mathbf{K}\|_1^d [C_p]^{d^2-1}.
\end{aligned}
$$

• Similarly, for the case $p=2$, we get the assertion (ii) of Theorem 1 with the same $\alpha_{p,1}$ and

$$\alpha_{p,2} := 3\mathbf{c}_3(4q)\mathbf{c}_4(4q)\Big(256\,\|\mathbf{K}\|_\infty^d + (1+\|\mathbf{K}\|_1^d)\mathbf{f}^{1-1/p} + 2\mathbf{c}_3(2q)\Big) + 3\mathbf{c}_3(q)d^2\,\|\mathbf{K}\|_1^d\,[C_p]^{d^2-1}.$$

• For $p>2$, put $C_p = (1+9\overline{c}+\|\mathbf{K}\|_1^d\mathbf{f}^{1-1/p})\|\mathbf{K}\|_1^d + 1$, where $\overline{c}$ is given by (A.5) in the proof of Lemma 2. In view of Lemma 2, Markov's inequality, (4.8), and the Cauchy-Schwarz inequality we establish $B_p \subseteq \{\zeta_p \geq 1\}$, $[\mathbb{P}_f(B_p)]^{\frac{1}{4q}} \leq \mathbf{c}_6(4q)n^{-1/2}$, and

$$\Big(\mathbb{E}_f\left\|\widehat{f}-f\right\|_p^q 1_{B_p^c}\Big)^{\frac{1}{q}} \leq 3d^2\,\|\mathbf{K}\|_1^d\,[C_p]^{d^2-1}\left(\mathcal{R}_p^{(q)}\left[(h,\mathcal{P}),f\right] + \sup_{I\in\mathcal{P}}\frac{\mathbf{c}_8(q)}{(nV_{h_I})^{\frac{1}{2}}} + \frac{\mathbf{c}_9(q)+\mathbf{c}_5(q)}{n^{\frac{1}{2}}}\right),$$

$$\Big(\mathbb{E}_f\left\|\widehat{f}-f\right\|_p^q 1_{B_p}\Big)^{\frac{1}{q}} \leq 3\mathbf{c}_6(4q)\mathbf{c}_7(4q)\Big(\mathcal{R}_p^{(2q)}\left[(h,\mathcal{P}),f\right] + \mathbf{c}_8(2q)+\mathbf{c}_9(2q)+\mathbf{c}_5(2q)\Big)n^{-\frac{1}{2}},$$

$$\mathcal{R}_p^{(2q)}\left[(h,\mathcal{P}),f\right] \leq \mathbf{c}_5(2q)+\mathbf{c}_8(2q)+\mathbf{c}_9(2q)+\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-1/p}+\mathbf{f}^{1-1/p}.$$

Thus, we get the assertion of Theorem 2 with

$$\alpha_{p,1} := 3[1\vee\mathbf{c}_8(q)]d^2\,\|\mathbf{K}\|_1^d\,[C_p]^{d^2-1},$$

$$\alpha_{p,2} := 3\mathbf{c}_6(4q)\mathbf{c}_7(4q)\Big\{2\big[\mathbf{c}_5(2q)+\mathbf{c}_8(2q)+\mathbf{c}_9(2q)\big] + \big(1+\|\mathbf{K}\|_1^d\big)\mathbf{f}^{1-1/p}\Big\} + 3\big[\mathbf{c}_9(q)+\mathbf{c}_5(q)\big]d^2\,\|\mathbf{K}\|_1^d\,[C_p]^{d^2-1}.$$

□

4.3. Upper bounds for adaptive minimax estimation: Proof of Theorem 4

Let $\mathbf{f}>0$, $(\beta,L,\mathcal{P})\in(0,l]^d\times(0,\infty)^d\times\mathfrak{P}$ and $f\in N_{p,d}(\beta,L,\mathcal{P},\mathbf{f})$ be fixed.

In view of the triangle inequality, $\forall h\in(0,1]^d$,

$$\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widehat{f}_{h_I}-f_I\right\|_{p,I} \leq \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\mathbb{E}_f\{\widehat{f}_{h_I}(\cdot)\}-f_I\right\|_{p,I} + \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_I}\|_{p,I}, \quad (4.9)$$

where $\mathbb{E}_f\{\widehat{f}_{h_I}(x_I)\} = K_{h_I}\star f_I(x_I)$ and $\xi_{h_I}(x_I) := \widehat{f}_{h_I}(x_I) - \mathbb{E}_f\{\widehat{f}_{h_I}(x_I)\}$.

Note first that, by applying Proposition 3 in Kerkyacharian, Lepski and Picard [24], it is easily established that, for any $h \in (0,1]^d$, any $\mathcal{P}' \in \mathfrak{P}$ and any $I \in \mathcal{P} \diamond \mathcal{P}'$,

$$\|K_{h_I} \star f_I - f_I\|_{p,I} \leq \sum_{i\in I} c_I(\mathbf{K}, |I|, p, l, L_I) h_i^{\beta_i} \leq \mathbf{c} \sup_{I\in\mathcal{P}} \sum_{i\in I} h_i^{\beta_i}, \quad \mathbf{c} > 0. \quad (4.10)$$

Next, by the choice of h_{max}, we get from Lemma 1 (i)–(iii) and Lemma 2 (i)–(ii)

$$\left(\mathbb{E}_f \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|\xi_{h_I}\|_{p,I}^q\right)^{\frac{1}{q}} \leq O\left(\sup_{I\in\mathcal{P}} (nV_{h_I})^{-\gamma_p}\right). \quad (4.11)$$

Consider now, for all $I \in \mathcal{P}$, the system

$$h_j^{\beta_j} = h_i^{\beta_i} = (nV_{h_I})^{-\gamma_p}, \quad i,j \in I.$$

The solution is given by

$$h_i = n^{-\frac{\gamma_p \overline{\beta_I}}{\gamma_p + \overline{\beta_I}} \frac{1}{\beta_i}}, \quad i \in I, \quad I \in \mathcal{P}, \quad (4.12)$$

where $\overline{\beta_I}$ is given in (3.2).

For all $I \in \mathcal{P}$, $h_I \in [h_{min}, h_{max}]^{|I|}$ and $V_{h_I} \geq V_{min}$ for n large enough and, remember, $\mathcal{H}_I$ is the dyadic grid in $\{h_I \in [h_{min}, h_{max}]^{|I|} : V_{h_I} \geq V_{min}\}$. Then, if $\overline{h}_I$ denotes the projection of h_I on $\mathcal{H}_I$ one has $(\overline{h}, \mathcal{P}) \in \mathcal{H}[\ \mathfrak{P}\ ]$ for n large enough. It follows from Theorems 1 and 2, (4.9), (4.10) and (4.11) that

$$\mathcal{R}_p^{(q)}\left[\widehat{f}, f\right] \leq \mathbf{C}\alpha_{p,1}\left[\sup_{I\in\mathcal{P}} \sum_{i\in I} \overline{h}_i^{\beta_i} + \sup_{I\in\mathcal{P}} \left(nV_{\overline{h}_I}\right)^{-\gamma_p}\right] + \alpha_{p,2} n^{-1/2}, \quad (4.13)$$

$\mathbf{C} > 0$, for n large enough. Indeed, the choice of the numbers h_{max}, h_{min} and V_{min} implies that the conditions required in both theorems are satisfied. Finally, in view of the properties of the dyadic grids, it is easily seen that we get the statement of Theorem 4 from (4.12) and (4.13). □

Appendix: Proof of Lemmas 1 and 2

Set $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\ ]$, $\mathbf{f} > 0$. We obtain Lemmas 1 and 2 by applying Propositions 1 and 2 with $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$.

A.1. Proof of Lemma 1

We divide this proof into several steps.

1) Note that

$$\xi_p \leq \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{h_I\in\mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+.$$

In view of Proposition 1 (i), if $p \in [1,2)$ and $n \geq 3 \vee 4^{2p(2-p)}$,

$$(\mathbb{E}_f |\xi_p|^q)^{\frac{1}{q}} \leq \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_1(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q) A_{\mathcal{H}_I}^{4/q} n^{1/p} \exp\left\{-\frac{2n^{2/p-1}}{37q}\right\}$$

$$\leq \mathbf{c}_1(q) n^{-1/2}, \qquad \mathbf{c}_1(q) := \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_1(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q)\right)$$
$$\times \sup_{n \in \mathbb{N}^*} \left[[\ln(n)]^{\frac{4d}{q}} n^{\frac{1}{2}+\frac{1}{p}} \exp\left\{-\frac{2n^{\frac{2}{p}-1}}{37q}\right\}\right],$$

since $A_{\mathcal{H}_I} \leq [\ln(n)]^{|I|} \leq [\ln(n)]^d$, $\forall I \in \mathcal{I}_d$.

Similarly, in view of Proposition 1 (ii), if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$, $(\mathbb{E}_f |\xi_p|^q)^{\frac{1}{q}} \leq \mathbf{c}_3(q) n^{-1/2}$, with

$$\mathbf{c}_3(q) := \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_2(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q)\right)$$
$$\times \sup_{n \in \mathbb{N}^*} \left[[\ln(n)]^{\frac{2d}{q}} n \exp\left\{-\frac{[\ln(n)]^2 \wedge \sqrt{n}}{16q[\mathbf{f}^2+4]}\right\}\right],$$

since, furthermore, $0 < h_{max}^{|I|} \leq h_{max}$, $\forall I \in \mathcal{I}_d$.

2) For any $p \geq 1$, $\overline{G}_p \leq 1 + \|\mathbf{K}\|_1^d$

$$\times \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left\{\left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+ + \widehat{\mathcal{U}}_p(h_I) + \left\|\mathbb{E}_f\left\{\widehat{f}_{h_I}\right\}\right\|_{p,I}\right\}\right);$$

$$\overline{G}_p \leq 1 + d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_1^{2d} \mathbf{f}^{1-\frac{1}{p}}$$
$$+ \|\mathbf{K}\|_1^d \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left\{\left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+ + \widehat{\mathcal{U}}_p(h_I)\right\}. \tag{A.1}$$

Put $\overline{\mathbf{f}} = 1 \vee \mathbf{f}$. We get from (A.1)

$$\overline{\mathbf{f}}_p \leq d^2 \|\mathbf{K}\|_1^{2d} \Bigg(2d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_1^d \overline{\mathbf{f}}^{1-\frac{1}{p}}$$
$$+ \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left\{\left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I)\right]_+ + \widehat{\mathcal{U}}_p(h_I)\right\}\Bigg)^{d^2}.$$

If $p \in [1,2]$ and $nV_{min} \geq 1$ then

$$\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \widehat{\mathcal{U}}_p(h_I) \leq 128 d \, |\mathfrak{P}|^2 \, \|\mathbf{K}\|_\infty^d , \tag{A.2}$$

since $\operatorname{supp}(\mathbf{K}) \subseteq [-1/2, 1/2]$ and $nV_{h_I} \geq 1$, $\forall h_I \in \mathcal{H}_I$.

Below we use the inequality (A.2) and the following trivial equality:

$$\left(\mathbb{E}_f \left| Y^{d^2} \right|^q\right)^{\frac{1}{q}} = \left[\left(\mathbb{E}_f \, |Y|^{qd^2}\right)^{\frac{1}{qd^2}} \right]^{d^2}, \tag{A.3}$$

for any random variable Y.

In view of Proposition 1 (i), if $p \in [1,2)$ and $n \geq 3 \vee 4^{2p(2-p)}$, $(\mathbb{E}_f |\overline{\mathbf{f}}_p|^q)^{\frac{1}{q}} \leq \mathbf{c}_2(q)$, with

$$\mathbf{c}_2 := d^2 \, \|\mathbf{K}\|_1^{2d} \left[\left(a_p \vee \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_1\Big(\overline{L}_{\mathbf{K}}, k_\infty, |I|, qd^2\Big) \right) b_p \right]^{d^2},$$

$$a_p := 130 d \, |\mathfrak{P}|^2 \, \|\mathbf{K}\|_\infty^d \, \overline{\mathbf{f}}^{1-\frac{1}{p}},$$

$$b_p := 1 + \sup_{n \in \mathbb{N}^*} \left([\ln(n)]^{\frac{4}{qd}} n^{\frac{1}{p}} \exp\left\{ -\frac{2n^{2/p-1}}{37qd^2} \right\} \right).$$

In view of Proposition 1 (ii), if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$, $(\mathbb{E}_f |\overline{\mathbf{f}}_p|^q)^{\frac{1}{q}} \leq \mathbf{c}_4(q)$, with

$$\mathbf{c}_4 := d^2 \, \|\mathbf{K}\|_1^{2d} \left[\left(a_p \vee \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_2\Big(\overline{L}_{\mathbf{K}}, k_\infty, |I|, qd^2\Big) \right) b_p \right]^{d^2},$$

$$a_p := 130 d \, |\mathfrak{P}|^2 \, \|\mathbf{K}\|_\infty^d \, \overline{\mathbf{f}}^{1-\frac{1}{p}},$$

$$b_p := 1 + \sup_{n \in \mathbb{N}^*} \left([\ln(n)]^{\frac{2}{qd}} n^{\frac{1}{2}} \exp\left\{ -\frac{[\ln(n)]^2 \wedge \sqrt{n}}{16qd^2[\mathbf{f}^2+4]} \right\} \right).$$

3) Let $(h, \mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$ be fixed. One has $V_{h_{I \cap I'}} \geq V_{h_I}$, $\forall I \in \mathcal{P}$, $\forall I' \in \mathcal{P}'$, $\forall \mathcal{P}' \in \mathfrak{P}$. Therefore, $\forall p \in [1,2]$,

$$\widehat{\mathcal{A}}_p(h, \mathcal{P}) := \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \widehat{\mathcal{U}}_p(h_I) \leq 128 \, \|\mathbf{K}\|_\infty^d \sup_{I \in \mathcal{P}} \left(\frac{1}{nV_{h_I}} \right)^{1-\frac{1}{p}}.$$

Thus, we finish the proof of Lemma 1. □

A.2. Proof of Lemma 2

Let $p > 2$. Assume that $n \geq \overline{C}_3$, $nV_{min} > 1 \vee \overline{C}_4$ and $n^{-1/(2d)} \leq h_{max} \leq [\ln(n)]^{-p}$, where

$$\overline{C}_3 := \sup_{I \in \mathcal{I}_d} \sup_{i=1,2,3,4} \left[C_3(\overline{L}_{\mathbf{K}}, k_\infty, |I|, iq, p, \mathbf{f}) \vee C_3(\overline{L}_{\mathbf{K}}, k_\infty, |I|, iqd^2, p, \mathbf{f}) \right],$$

$$\overline{C}_4 := \sup_{I \in \mathcal{I}_d} \sup_{i=1,2,3,4} \left[C_4(\overline{L}_{\mathbf{K}}, k_\infty, |I|, iq, p, \mathbf{f}) \vee C_4(\overline{L}_{\mathbf{K}}, k_\infty, |I|, iqd^2, p, \mathbf{f}) \right].$$

First, similarly to the proof of Lemma 1, step 1, it follows from the assertion (i) of Proposition 2 and Proposition 3 that $(\mathbb{E}_f |\xi_p|^q)^{\frac{1}{q}} \leq \mathbf{c}_5(q) n^{-1/2}$ and $(\mathbb{E}_f |\zeta_p|^q)^{\frac{1}{q}} \leq \mathbf{c}_6(q) n^{-1/2}$, with

$$\mathbf{c}_5(q) := \left(\sum_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_5(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q, p) \right) \times \sup_{n \in \mathbb{N}^*} \left([\ln(n)]^{\frac{2d}{q}} [\log_2(n)]^{\frac{2}{2q}} n \exp\left\{ -\overline{\mathbf{c}}_5(q) [\ln(n)]^2 \right\} \right),$$

$$\overline{\mathbf{c}}_5(q) := \mathbf{f}^{-1} \inf_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \inf_{I \in \mathcal{P} \diamond \mathcal{P}'} C_6(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q, p),$$

$$\mathbf{c}_6(q) := \left(\sum_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_9(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q, p, \mathbf{f}) \right) \times \sup_{n \in \mathbb{N}^*} \left([\ln(n)]^{\frac{2d}{q}} n^{\frac{1}{2}} \exp\left\{ -\overline{\mathbf{c}}_6(q) [\ln(n)]^2 \right\} \right),$$

$$\overline{\mathbf{c}}_6(q) := \inf_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \inf_{I \in \mathcal{P} \diamond \mathcal{P}'} C_{10}(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q, p, \mathbf{f}).$$

Next, the assertion (i) of Proposition 2 allows us to assert that

$$\left(\mathbb{E}_f \left| \sum_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widehat{\mathcal{U}}_p(h_I) \right]_+ \right|^{qd^2} \right)^{\frac{1}{qd^2}} \leq \overline{\mathbf{c}}_7(q) [\log_2(n)]^{\frac{3}{qd}} n^{\frac{1}{2}} \exp\left\{ -\overline{\mathbf{c}}_8(q) [\ln(n)]^2 \right\}, \tag{A.4}$$

$$\overline{\mathbf{c}}_7(q) := \sum_{\mathcal{P}, \mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_5(\overline{L}_{\mathbf{K}}, k_\infty, |I|, qd^2, p),$$

$$\overline{\mathbf{c}}_8(q) := \mathbf{f}^{-1} \inf_{\mathcal{P}, \mathcal{P}' \in \overline{\mathfrak{P}}} \inf_{I \in \mathcal{P} \diamond \mathcal{P}'} C_6(\overline{L}_{\mathbf{K}}, k_\infty, |I|, qd^2, p).$$

Note that, for any $\mathcal{P},\mathcal{P}'\in\mathfrak{P}$, any $I\in\mathcal{P}\diamond\mathcal{P}'$ and any $h_I\in\mathcal{H}_I$, in view of Young's inequality,

$$\rho_p\left(K_{h_I}\right)=\frac{15p}{\ln p}\left\{n^{-1/2}\left\|K_{h_I}^2*f_I\right\|_{p/2,I}^{1/2}+2(nV_{h_I})^{1/p-1}\left\|K_I\right\|_{p,I}\right\}$$
$$\leq\frac{15p}{\ln p}\left\{(nV_{h_I})^{-1/2}\left\|\mathbf{K}\right\|_\infty^d\mathbf{f}^{1/2-1/p}+2(nV_{h_I})^{1/p-1}\left\|\mathbf{K}\right\|_\infty^d\right\};$$

$$\rho_p\left(K_{h_I}\right)\leq\overline{\mathbf{c}}(nV_{h_I})^{-1/2},\quad\overline{\mathbf{c}}:=\frac{45p\left\|\mathbf{K}\right\|_\infty^d\overline{\mathbf{f}}^{1/2-1/p}}{\ln p}\geq\left\|\mathbf{K}\right\|_\infty^d\geq1,\tag{A.5}$$

since $\mathrm{supp}(\mathbf{K})\subseteq[-1/2,1/2]$, $nV_{h_I}\geq1$ and $p>2$.

Below we use the trivial inequality $(a+b)^\alpha\leq a^\alpha+b^\alpha$ for any $a,b>0$ and $\alpha\in(0,1)$. Thus, we deduce from the assertion (ii) of Proposition 2 and (A.5) that

$$\left(\mathbb{E}_f\left|\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sum_{I\in\mathcal{P}\diamond\mathcal{P}'}\sup_{h_I\in\mathcal{H}_I}\widehat{\mathcal{U}}_p(h_I)\right|^{qd^2}\right)^{\frac{1}{qd^2}}\tag{A.6}$$
$$\leq\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sum_{I\in\mathcal{P}\diamond\mathcal{P}'}\left(\mathbb{E}_f\left|\sup_{h_I\in\mathcal{H}_I}\widehat{\mathcal{U}}_p(h_I)\right|^{qd^2}\right)^{\frac{1}{qd^2}}$$
$$\leq d\left|\mathfrak{P}\right|^2 32^{\frac{1}{2d^2}}\left[\overline{\mathbf{c}}\left(1+\frac{120p}{\ln p}\right)+\overline{\mathbf{c}}_9(q)[\log_2(n)]^{\frac{3}{qd}}n^{\frac{1}{2}-\frac{1}{p}}\exp\left\{-\frac{\overline{\mathbf{c}}_{10}(q)\overline{b}_{n,p}}{qd^2}\right\}\right],$$

$$\overline{\mathbf{c}}_9(q):=\sup_{\mathcal{P},\mathcal{P}'\in\overline{\mathfrak{P}}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}C_7(\overline{L}_{\mathbf{K}},k_\infty,|I|,qd^2,p),$$
$$\overline{\mathbf{c}}_{10}(q):=\inf_{\mathcal{P},\mathcal{P}'\in\overline{\mathfrak{P}}}\inf_{I\in\mathcal{P}\diamond\mathcal{P}'}C_8(\overline{L}_{\mathbf{K}},k_\infty,|I|,qd^2,p),$$

where $\overline{b}_{n,p}=n^{4/p-1}$ if $p\in(2,4)$ and $\overline{b}_{n,p}=[\ln(n)]^4\mathbf{f}^{-1}$ if $p\in[4,\infty)$.

It follows from (A.1), (A.3), (A.4) and (A.6) that $(\mathbb{E}_f|\overline{\mathbf{f}}_p|^q)^{\frac{1}{q}}\leq\mathbf{c}_7(q)$,

$$\mathbf{c}_7(q):=d^2\|\mathbf{K}\|_1^{2d}$$
$$\times\left[\left(2d\left|\overline{\mathfrak{P}}\right|^2\left\|\mathbf{K}\right\|_1^d\overline{\mathbf{f}}^{1-\frac{1}{p}}+d\left|\overline{\mathfrak{P}}\right|^2 32^{\frac{1}{2qd^2}}\overline{\mathbf{c}}(1+\frac{120p}{\ln p})\right)\vee(\overline{\mathbf{c}}_7(q)\vee\overline{\mathbf{c}}_9(q))\right]^{d^2}$$
$$\times\left[1+2\sup_{n\in\mathbb{N}^*}\left([\log_2(n)]^{\frac{3}{qd}}n^{\frac{1}{2}}\exp\left\{-\left[\overline{\mathbf{c}}_8(q)[\ln(n)]^2\right]\wedge\left[\frac{\overline{\mathbf{c}}_{10}(q)\,\overline{b}_{n,p}}{q\overline{d}}\right]\right\}\right)\right]^{d^2}.$$

Finally, we get the assertion (ii) of Lemma 2 from Proposition 2 (ii) and (A.5), with

$$\mathbf{c}_8(q) := 32^{\frac{1}{q}} d\,|\mathfrak{P}|^2\,(1+120p/\ln p)\overline{\mathbf{c}}$$

$$\mathbf{c}_9(q) := \left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\ \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_7(\overline{L}_{\mathbf{K}}, k_\infty, |I|\,, q, p)\right)$$

$$\times \sup_{n\in\mathbb{N}^*}\left([\log_2(n)]^{\frac{2d+1}{q}} n^{\frac{1}{2}} \exp\left\{-\frac{\overline{c}_{11}(q)\,\overline{b}_{n,p}}{q}\right\}\right),$$

$\overline{c}_{11}(q) := \inf_{\mathcal{P},\mathcal{P}'\in\overline{\mathfrak{P}}} \inf_{I\in\mathcal{P}\diamond\mathcal{P}'} C_8(\overline{L}_{\mathbf{K}}, k_\infty, |I|, q, p)$. □

Acknowledgments

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

References

[1] Baraud, Y. and Birgé, L. (2014). Estimating composite functions by model selection. *Ann. Inst. H. Poincaré Probab. Statist.* **50**, 285–314. MR3161532
[2] Birgé, L. (2013). Model selection for density estimation with $\mathbb{L}_2$-loss. *Probab. Theory and Relat. Fields* **158**, 533–574. MR3176358
[3] Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47**, 119–137. MR0523165
[4] Chacón, J.E. and Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test* **19**, 375–398. MR2677734
[5] Donoho, D.L., Johnston, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539. MR1394974
[6] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The $\mathbb{L}_1$ View*. Wiley Series in Probability and Statistics. Wiley, New York. MR0780746
[7] Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimation. *Ann. Statist.* **24**, 2499–2512. MR1425963
[8] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factor, kernel complexity and Yatracos classes. *Ann. Statist.* **25**, 2626–2637. MR1604428
[9] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York. MR1843146

[10] Efromovich, S.Y. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory. Probab. Appl.* **30**, 557–568. MR0805304

[11] Efromovich, S.Y. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36**, 1127–1155. MR2418652

[12] Goldenshluger, A. and Lepski, O.V. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields*, **143**, 41–71. MR2449122

[13] Goldenshluger, A. and Lepski, O. (2011a). Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* **39**, 2318–2384. MR2932670

[14] Goldenshluger, A. and Lepski, O. (2011b). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608–1632. MR2850214

[15] Goldenshluger, A. and Lepski, O. (2013). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory and Relat. Fields* **159**, 479–543. MR3230001

[16] Golubev, G.K. (1992). Nonparametric estimation of smooth probability densities. *Probl. Inf. Transm.* **1**, 52–62. MR1163140

[17] 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

[18] Horowitz, J.I. and Mamen, I. (1990). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35, 6**, 2589–2619. MR2382659

[19] Ibragimov, I.A. and Khasminski, R.Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **98**, 61–85 (in Russian). MR0591862

[20] Ibragimov, I.A. and Khasminski, R.Z. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **108**, 72–88 (in Russian). MR0629401

[21] Iouditski, A.B., Lepski, O.V. and Tsybakov, A.B. (2009). Nonparametric estimation of composite functions. *Ann. Statist.* **37, 3**, 1360–1440. MR2509077

[22] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10**, 187–220. MR2046772

[23] Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli* **2**, 229–247. MR1416864

[24] Kerkyacharian, G., Lepski, O.V. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory and Relat. Fields* **121**, 137–170. MR1863916

[25] Kerkyacharian, G., Lepski, O.V. and Picard, D. (2007). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Probab. Theory Appl.* **52**, 150–171. MR2354574

[26] Lepski, O.V. (1991). On a problem of adaptive estimation in Gaussian white noise model. *Theory Probab. Appl.* **35**, 454–466. MR1091202

[27] Lepski, O.V. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Ann. Stat.* **40, 2**, 1005–1034. MR3099129

[28] Lepski, O.V. and Serdyukova, N. (2014). Adaptive estimation under single-index constraint in a regression model. *Ann. Stat.* **40, 1**, 1–28. MR3161459

[29] Mason, D.M. (2009). Risk bounds for density kernel estimators. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **363**, 66–104. Available at `http://www.pdmi.ras.ru/znsl/`. MR2749118

[30] Massart, P. (2007). *Concentration Inequalities and Model Selection*. Lecture from the 33rd Summer School on Probability Theory held in Saint-Floor, July 6–23, 2003. Lecture Notes in Mathematics, 1886. Springer, Berlin. MR2319879

[31] Nikol'skii, S.M. (1977). Priblizheni Funktsii Mnogikh Peremennykh i Teoremy Vlosheniya (in Russian). [Approximation of functions of several variables and imbedding theorems.] Naukka, Moscow, 1977. MR0506247

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

[33] Rebelles, G. (2014). Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, forthcoming.

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

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

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

[37] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. MR0848134

[38] Scott, D. (1992). *Multivariate Density Estimation*. Wiley, New York. MR1191168

[39] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York. MR2724359