Structural adaptive deconvolution under $L_p$-losses

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 address the problem of estimating a multidimensional density $f$ by using indirect observations from the statistical model $Y = X + \varepsilon$. Here, $\varepsilon$ is a measurement error independent of the random vector $X$ of interest, and having a known density with respect to the Lebesgue measure. Our aim is to obtain optimal accuracy of estimation under $L_p$-losses when the error $\varepsilon$ has a characteristic function with a polynomial decay. To achieve this goal, we first construct a kernel estimator of $f$ which is fully data driven. Then, we derive for it an oracle inequality under very mild assumptions on the characteristic function of the error $\varepsilon$. As a consequence, we get minimax adaptive upper bounds over a large scale of anisotropic Nikolskii classes and we prove that our estimator is asymptotically rate optimal when $p \in [2, +\infty]$. Furthermore, our estimation procedure adapts automatically to the possible independence structure of $f$ and this allows us to improve significantly the accuracy of estimation.

Keywords and phrases: density estimation, deconvolution, kernel estimator, oracle inequality, adaptation, independence structure, concentration inequality.

1. Introduction

Let $X_k = (X_{k,1}, \ldots, X_{k,d})$, $k \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^d$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having an unknown density $f$ with respect to the Lebesgue measure. Assume that we have at our disposal indirect observations given by

$$Y_k = X_k + \varepsilon_k, \quad k = 1, \ldots, n,$$

where the errors $\varepsilon_k$ are also i.i.d. $d$-dimensional random vectors, independent of the $X_k$’s, with a known density $g$.

The goal is to estimate the density $f$ by using observations $Y^{(n)} = (Y_1, \ldots, Y_n)$. By an estimator we mean any $Y^{(n)}$-measurable mapping $\hat{f} : \mathbb{R}^n \rightarrow L_p(\mathbb{R}^d)$. The accuracy of an estimator is measured by its $L_p$-risk

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

Here and in the sequel $\mathbb{E}_f$ denotes the expectation with respect to the probability measure $\mathbb{P}_f$ of the observations $Y^{(n)} = (Y_1, \ldots, Y_n)$ and $\left\| g \right\|_r$ is the $L_r$-norm of $g \in L_r(\mathbb{R}^s)$, $s \in \mathbb{N}^*$, $r \in [1, +\infty]$. We will also denote by $\hat{g}$ the Fourier transform of $g \in L_1(\mathbb{R}^s)$, defined by $\hat{g}(x) = \int e^{i<s,x>} g(x) dx$, where $<\cdot, \cdot>$ is the euclidean scalar product on $\mathbb{R}^s$.

The aforementioned deconvolution model, which is more realistic than the density model (with direct observations), exists in many different fields and is the subject of many theoretical studies. In most of them, the main interest is to provide estimators which achieve optimal rates of convergence on particular functional classes in a minimax sense. For instance, the problem of minimax estimation in the deconvolution model with pointwise and $L_2$ risks was investigated by Fan ([8],[9]), Pensky and Vidakovic [25], Butucea and Tsybakov ([3],[4]), Butucea and Comte [2]. The global estimation in the deconvolution model was also considered under a weighted...
the same Lepski [9]. Minimax estimation from the others, but in a unified way.

In the framework of the minimax estimation it is assumed that \( f \) belongs to a certain set of functions \( \Sigma \), and then the accuracy of an estimator \( \hat{f} \) is measured by its \textit{maximal risk} over \( \Sigma \):

\[
\mathcal{R}_p \left[ \hat{f}, \Sigma \right] := \sup_{\hat{f} \in \Sigma} \mathcal{R}_p \left[ \hat{f}, \tilde{f} \right].
\]

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

\[
\mathcal{R}_p \left[ \tilde{f}_n, \Sigma \right] \asymp \inf_{\tilde{f}_n} \mathcal{R}_p \left[ \tilde{f}_n, \Sigma \right] := \varphi_{n,p}(\Sigma), \quad n \to +\infty,
\]

where infimum is taken over all possible estimators. Such an estimator is called minimax on \( \Sigma \).

In this paper, we focus on the problem of minimax estimation over anisotropic Nikolskii classes of densities \( N_{r,d}(\beta, L) \), see the definition in Section 2.3. Whereas the vector \( \beta = (\beta_1, \ldots, \beta_d) \) represents the smoothness of the target density, \( r = (r_1, \ldots, r_d) \) represents the index of homogeneity. For the case where \( p \) is finite we will assume that the smoothness of \( f \) is measured in the same \( \mathbb{L}_p \)-norm that the accuracy of estimation, that means \( r_j = p \) for \( j = 1, \ldots, d \). In the latter case, the vector \( r \) will be replaced by \( p \) in the notation of the functional class.

In Comte and Lacour [5] it was shown that

\[
\varphi_{n,2}(N_{2,d}(\beta, L)) \asymp n^{-\frac{\tau}{d+1}}, \quad \tau := \left[ \sum_{j=1}^{d} \frac{2\lambda_j + 1}{\beta_j} \right]^{-1}, \tag{2}
\]

when the common density \( q \) of the errors (which is assumed to be known) satisfies

\[
A_1 \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{\tau}} \leq |\hat{q}(t)| \leq A_2 \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{\tau}}, \quad \forall t \in \mathbb{R}^d,
\]

for some constants \( A_1, A_2, \lambda_j > 0, j = 1, \ldots, d \). Such a density is usually called ordinary smooth of order \( \lambda = (\lambda_1, \ldots, \lambda_d) \).

Note that the latter result was proved in the one dimensional setting in Fan [9]. However, whereas Fan [9] provided an estimator whose construction depends on the smoothness parameter \( \beta \) of the functional class \( N_{2,1}(\beta, L) \) (which is not known in practice), Comte and Lacour [5] proposed an adaptive strategy. Indeed, they have constructed a single estimator which is fully data driven and minimax on each class \( N_{2,d}(\beta, L) \), whatever the nuisance parameter \( (\beta, L) \) in a large range. Such an estimator is called optimal adaptive over the scale \( \{N_{2,d}(\beta, L)\}_{(\beta, L)} \).

Lounici and Nickl [22] considered the problem of adaptive deconvolution of a univariate density under sup-norm loss and proved that

\[
\varphi_{n,\infty}(N_{\infty,1}(\beta, L)) \asymp \left( \frac{n}{\ln(n)} \right)^{-\frac{\tau}{d+1}}, \quad \tau := \left[ \frac{2\lambda + 1}{\beta} \right]^{-1}, \tag{3}
\]
when the common density \( q \) of the errors is ordinary smooth of order \( \lambda > 0 \). Moreover, they provided an optimal adaptive estimator over the scale of Hölder classes \( \{N_{\infty,1}(\beta,L)\}_{(\beta,L)} \).

It is worth mentioning that Fan [9], Lounici and Nickl [22] and Comte and Lacour [5], as in many other papers, considered also the case of errors having a common density whose Fourier transform has exponential decay, usually called super smooth. In the multidimensional setting, Comte and Lacour [5] showed that, in presence of super smooth noise, the rates of convergence on anisotropic Nikoliskii classes (considered as classes of ordinary smooth densities) are logarithmic and achieved by a kernel estimator whose bandwidth depends only on the smoothness parameters of the noise. Thus, in the latter case, no bandwidth selection procedure is required to get adaptive properties. In this paper, we only study the case of ordinary smooth noise, namely a "moderately ill-posed" case in inverse problems.

However, the rates of convergence given in (2)-(3) are recovered from the results we obtain in the present paper. Indeed, we provide adaptive kernel estimators which achieve the following minimax rates of convergence respectively:

\[
\varphi_{n,p}(N_{p,d}(\beta,L)) \asymp n^{-\frac{\tau}{2\tau+1}}, \quad \forall p \in [2, +\infty); \tag{4}
\]

\[
\varphi_{n,\infty}(N_{r,d}(\beta,L)) \asymp \left(\frac{n}{\ln(n)}\right)^{-\frac{\tau}{2\tau+1}}, \quad \Upsilon^{-1} := \tau^{-1} + [\omega \tau]^{-1}, \tag{5}
\]

where \( \tau \) is given in (2), \( \omega := \left[\sum_{j=1}^d \frac{2\lambda_j+1}{\beta_j \rho_j}\right]^{-1} \) and \( \kappa := \left(1 - \sum_{j=1}^d \frac{1}{\beta_j \rho_j}\right) \left[\sum_{j=1}^d \frac{1}{\beta_j} \right]^{-1} > 0 \).

Here, the optimality is a direct consequence of minimax lower bounds recently obtained by Lepski and Willer [21]. As usually, these lower bounds hold under additional assumptions on the common density of the errors, see Section 2.4. Moreover, they proved that there is no uniformly consistent estimator on \( N_{r,d}(\beta,L) \) under sup-norm loss if \( \kappa \leq 0 \). Note also that, for \( p \in (1,2) \), our estimator does not achieve the minimax lower bound on \( N_{p,d}(\beta,L) \) they have found. Finally, we will not consider the case \( p = 1 \) since the results in Lepski and Willer [21] show that there is no uniformly consistent estimator on \( N_{1,d}(\beta,L) \) under \( L_1 \)-loss.

It is important to emphasize that minimax rates depend heavily on the dimension \( d \). 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 Mammen [15], Iouditski et al. [16] and Baraud and Birgé[1], the works on multi-index structure in Golden-shluger and Lepski [10] and Lepski and Serdyukova [20], and the works on the multiple index model in density estimation in Samarov and Tsybakov [28].

Below, we discuss one of the possibilities of facing to this problem in the framework of density estimation. The approach which has been recently proposed in Lepski [19] is to take into account the independence structure of the target density \( f \), namely its product structure due to the independence structure of the vector \( X_1 \).

**Organization of the paper** In Section 2, we describe assumptions on the densities involved in the statistical model (1) and we recall the minimax lower bounds obtained in Lepski and Willer [21] useful in this paper. In Section 3, we introduce the family of kernel estimators we use for our procedure and then we describe the selection rule that leads to the construction of our final estimator. In Section 4, we provide some oracle inequalities and, as consequences, minimax adaptive upper bounds under \( L_p \)-losses over scales of anisotropic Nikoliskii classes. Further, we discuss the optimality of our estimator and the influence of the independence structure of the target density on the accuracy of estimation. Proofs of all main results are given in Section 5. Proofs of technical useful results are deferred to the Appendix.
2. Assumptions on densities \( f \) and \( q \)

2.1. Structural assumption on the target density

Denote by \( \mathcal{I}_d \) the set of all subsets of \( \{1, \ldots, d\} \), except the empty set. Let \( \Psi \) be a given set of partitions of \( \{1, \ldots, d\} \). For all \( I \in \mathcal{I}_d \), denote also \( \mathcal{T} = \{1, \ldots, d\} \setminus I \) and \( |I| = \text{card}(I) \). We will use \( \emptyset \) 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, for any probability density \( g : \mathbb{R}^d \to \mathbb{R}_+ \),

\[
g_I(x_I) := \int_{\mathcal{T}} g(x) dx_T.
\]

Assume that \( g_{\emptyset} \equiv g \) and that \( g_{\emptyset} \equiv 1 \). Note also that \( f_I \) and \( q_I \) are the marginal densities of \( X_{1,I} \) and \( \varepsilon_{1,I} \) respectively.

If \( \mathcal{P} \in \Psi \) 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:

\[
\Psi(f) := \left\{ \mathcal{P} \in \Psi : f(x) = \prod_{I \in \mathcal{P}} f_I(x_I), \ \forall x \in \mathbb{R}^d \right\}.
\]

Remark that \( \Psi(f) \) is not empty if we consider that \( \emptyset \in \Psi \), or that \( \Psi = \{ \mathcal{P} \} \) if the independence structure of \( f \) is known. The possibility of choosing \( \Psi \), instead of considering all partitions of \( \{1, \ldots, d\} \), is introduced for technical purposes. This is explained in more detail in Lepski [19], section 2.1, paragraph "Extra parameters".

Finally, we endow the set \( \Psi \) with the operation "\( \circ \)" introduced in Lepski [19]: for any \( \mathcal{P}, \mathcal{P}' \in \Psi \)

\[
\mathcal{P} \circ \mathcal{P}' := \{ I \cap I' \neq \emptyset, I \in \mathcal{P}, I' \in \mathcal{P}' \}.
\]

The use of this operation for the estimation procedure allows us to construct an estimator which adapts automatically to the independence structure of the underlying density.

2.2. Noise assumptions for upper bounds

Both the definition of our estimation procedure and the computation of the \( L_p \)-risk, \( p \in (1, +\infty] \), lead us to consider that the density \( q \) of the noise random vector \( \varepsilon_1 \) satisfies following assumptions.

Assumption (N1). Assume that, for any \( I \in \mathcal{P} \circ \mathcal{P}' \), \( (\mathcal{P}, \mathcal{P}') \in \Psi \times \Psi \):

(i) if \( p = 2 \), then \( \| q_I \|_1 < +\infty \);

(ii) if \( p \in (2, +\infty] \), then \( \| q_I \|_\infty < +\infty \).

Assumption (N2). Assume that, for some constants \( A > 0, \lambda_j > 0, j = 1, \ldots, d \), one has for any \( I \in \mathcal{P} \circ \mathcal{P}' \), \( (\mathcal{P}, \mathcal{P}') \in \Psi \times \Psi \):

(i) if \( p = 2 \),

\[
|q_I(t)| \geq A^{-1} \prod_{j \in I} (1 + t_j^2)^{-\frac{\lambda_j}{2}}, \ \forall t \in \mathbb{R}^d;
\]

(ii) if \( p \in (1, +\infty) \setminus \{2\} \), \( \hat{q}_I(t_I) \neq 0, \ \forall t \in \mathbb{R}^d, \ \hat{q}_I^{-1} \in \mathcal{C}^{|I|}(\mathbb{R}^{|I|}) \) and

\[
\left| \left[D^\alpha \hat{q}_I^{-1}\right](t_I) \prod_{j \in I} t_j^{\alpha_j} \right| \leq A \prod_{j \in I} (1 + t_j^2)^{\frac{\lambda_j}{2}}, \ \forall t \in \mathbb{R}^d, \ \forall \alpha_I = (\alpha_j)_{j \in I} \in \mathbb{N}^{|I|}, \sum_{j \in I} \alpha_j \leq |I|;
\]

4
(iii) if \( p = +\infty, \hat{q}_I(t_I) \neq 0, \forall t_I \in \mathbb{R}^d, \hat{q}_I^{-1} \in C^1(\mathbb{R}^d) \) and
\[
\left| D_k^{\alpha_k} \hat{q}_I^{-1}(t_I) \right| \leq A \prod_{j \in I} (1 + t_j^2)^{\alpha_j}, \forall t_I \in \mathbb{R}^d, \forall k \in I, \forall \alpha_k \in \{0, 1\}.
\]

Here and in the sequel, \( D_k^{\alpha_k} g \) denotes the \( \alpha_k \)th order partial derivative of \( g \) with respect to the \( k \)th variable, \( D_0^{\alpha_k} g \equiv g \) and, for any multi-index \( \alpha = (\alpha_1, \ldots, \alpha_s) \in \mathbb{N}^s, D^{\alpha} g \) denotes the derivative \( D_{1}^{\alpha_1} \ldots D_{s}^{\alpha_s} g \) of \( g : \mathbb{R}^s \to \mathbb{R} \).

Assumption (N1) is satisfied for many distributions like centered Gaussian, Cauchy, Laplace or Gamma type multivariate ones. Assumption (N2) is quite restrictive since it does not hold for the classical Cauchy and Gaussian densities, whose characteristic functions have exponential decay. However, it is verified by the centered Laplace and Gamma type distributions, whose characteristic functions have polynomial decay. As mentioned in Comte and Lacour [5], the latter case keep a great interest in particular physical contexts; see, for instance, the study of the pile-up model in Comte and Rebafka [6].

In what follows, we assume that \( q \) satisfies Assumptions (N1)-(N2).

### 2.3. Smoothness assumption on the target density

In the literature there are several definitions of the **anisotropic Nikolskii class of densities** which are equivalent. Let us recall the definition we use in the present paper. Set \( \{e_1, \ldots, e_s\} \), the canonical basis in \( \mathbb{R}^s \), \( s \in \mathbb{N}^* \).

**Definition 1.** Assume that \( r = (r_1, \ldots, r_s) \in [1, +\infty]^s \), \( \beta = (\beta_1, \ldots, \beta_s) \in (0, +\infty)^s \) and \( L = (L_1, \ldots, L_s) \in (0, +\infty)^s \). A probability density \( g : \mathbb{R}^s \to \mathbb{R}_+ \) belongs to the **anisotropic Nikolskii class** \( N_{r,s}(\beta, L) \) if
\[
(i) \left\| D^k_{r_1} g \right\|_{r_1} \leq L_j, \forall k = 0, \ldots, [\beta_j], \forall j = 1, \ldots, s;
\]
\[
(ii) \left\| D^{[\beta_j]}_{r_1} g(\cdot + ze_j) - D^{[\beta_j]}_{r_1} g(\cdot) \right\|_{r_1} \leq L_j \left| z \right|^{\beta_j - [\beta_j]}, \forall z \in \mathbb{R}, \forall j = 1, \ldots, s.
\]

Here and in the sequel, \( [a] \) is the largest integer strictly less than the real number \( a \). Furthermore, we use the notation \( N_{r,s}(\beta, L) \) for \( N_{r,s}(\beta, L) \) when \( r = (r, \ldots, r) \).

In order to take into account the smoothness of the underlying density and its possible independence structure simultaneously, a certain collection of anisotropic Nikolskii classes of densities was introduced in Lepski [19], Section 3, Definition 2. However, since the adaptation is not necessarily considered with respect to the set of all partitions of \( \{1, \ldots, d\} \), the condition imposed therein can be weakened. For instance, if \( \mathcal{P} = \{\emptyset\} \) (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 \( \{N_{r,d}(\beta, L)\} \), that is not possible with the classes introduced in Lepski [19]. For these reasons, the following collection \( \{N_{r,d}(\beta, L, \mathcal{P})\}_P \) was introduced in Rebels [26], Section 3.1.

**Definition 2.** Let \( r \in [1, +\infty]^d \) and \( (\beta, L, \mathcal{P}) \in (0, +\infty)^d \times (0, +\infty)^d \times \mathcal{P} \) be fixed. A probability density \( g : \mathbb{R}^d \to \mathbb{R}_+ \) belongs to the class \( N_{r,d}(\beta, L, \mathcal{P}) \) if
\[
g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \forall x \in \mathbb{R}^d; \quad g_I \in N_{r_I,|I|}(\beta_I, L_I), \forall I \in \mathcal{P}' \circ \mathcal{P}'', \forall (\mathcal{P}', \mathcal{P}'') \in \mathcal{P} \times \mathcal{P}. \tag{8}
\]

Note that, if \( \mathcal{P} = \{\emptyset\} \), the class \( N_{r,d}(\beta, L, \emptyset) \) coincides with the classical anisotropic Nikolskii class of densities \( N_{r,d}(\beta, L) \).
2.4. Noise assumptions for minimax lower bounds

Recently, Lepski and Willer [21] have obtained minimax lower bounds for \( \varphi_{n,p}(N_{r,d}(\beta,L)) \), \( p \in [1, +\infty] \), when the density \( q \) of the noise random vector \( \varepsilon_1 \) satisfies the following assumption.

**Assumption (N3).** For any multi-index \( \alpha = (\alpha_1, \ldots, \alpha_d) \in \{0, 1\}^d \) satisfying \( \alpha_1 + \ldots + \alpha_d \geq 1 \), \( D^\alpha \bar{q} \) exists. Furthermore, there exist constants \( B > 0 \) and \( \lambda_j > 0, j = 1, \ldots, d \), such that:

\[
\begin{align*}
(i) & \quad |\bar{q}(t)| \leq B \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d; \\
(ii) & \quad \|\tilde{q}^{-1} D^\alpha \bar{q}\|_\infty \leq B, \quad \forall \alpha = (\alpha_1, \ldots, \alpha_d) \in \{0, 1\}^d, \alpha_1 + \ldots + \alpha_d \geq 1.
\end{align*}
\]

Note first that Assumption (N3) is also verified for centered Laplace or Gamma-type distributions. Next, if \( \mathcal{P} = \{\emptyset\} \) (no independence structure), any density \( q \) that satisfies both the condition (i) of Assumption (N3) and Assumptions (N2) verifies

\[
A^{-1} \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{2}} \leq |\bar{q}(t)| \leq B \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d,
\]

and hence is ordinary smooth of order \( \lambda = (\lambda_1, \ldots, \lambda_d) \). Furthermore, the condition imposed in the left hand side of the latter inequalities, together with the condition (ii) of Assumption (N3) (or Condition 1 in Lounici and Nickl [22] for the one dimensional setting), implies that condition (iii) of Assumption (N2) is satisfied.

The results below follow from Theorems 2 and 3 in Lepski and Willer [21] and allow us to assert the optimality of our estimators when \( \mathcal{P} = \{\emptyset\} \) (no independence structure).

**Theorem 1.** Let \( L_0 > 0 \) and \( p \in [2, +\infty) \) be fixed. Suppose that Assumptions (N3) is satisfied. Then, for any \( (\beta, L) \in (0, \infty)^d \times [L_0, \infty)^d \)

\[
\liminf_{n \to +\infty} \inf_{f_n} \left\{ \varphi_{n,p}^{-1}(N_{p,d}(\beta,L)) R_p \left[ \hat{f}_n, N_{p,d}(\beta,L) \right] \right\} > 0,
\]

where infimum is taken over all possible estimators and \( \varphi_{n,p}(N_{p,d}(\beta,L)) \) is given in (4).

**Theorem 2.** Let \( L_0 > 0 \) and \( (\beta, L, r) \in (0, \infty)^d \times [L_0, \infty)^d \times [1, \infty)^d \) be fixed. Suppose that Assumptions (N3) is satisfied. Then,

(i) there is no uniformly consistent estimator if \( 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} \leq 0 \);

(ii) if \( 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0 \)

\[
\liminf_{n \to +\infty} \inf_{f_n} \left\{ \varphi_{n,\infty}^{-1}(N_{r,d}(\beta,L)) R_\infty \left[ \hat{f}_n, N_{r,d}(\beta,L) \right] \right\} > 0,
\]

where infimum is taken over all possible estimators and \( \varphi_{n,\infty}(N_{r,d}(\beta,L)) \) is given in (5).

3. Estimation procedure

In this section, we construct an estimator following a scheme of selection rule introduced in Lepski [19] to take into account the possible independence structure of the underlying density. If \( \mathcal{P} = \{\emptyset\} \) this scheme coincides with a version of the methodology proposed by Goldenshluger and Lepski [12]. This methodology, employed in many areas of nonparametric statistics, has been recently used by Comte and Lacour [5] in the framework of the deconvolution model.
3.1. Kernel-type estimators

Let $K : \mathbb{R} \to \mathbb{R}$ be a fixed symmetric kernel ($\int K = 1$) belonging to the well known Schwartz class $\mathbb{S}(\mathbb{R})$. For instance, $K$ may be a Gaussian kernel. For all $I \in I_d$, $h \in (0,1]^d$ and $x \in \mathbb{R}^d$ put

$$K_I(x_I) := \prod_{j \in I} K(x_j), \quad K_{h_I}(x_I) := V_{h_I}^{-1} \prod_{j \in I} K(x_j/h_j), \quad V_{h_I} := \prod_{j \in I} h_j.$$  

Therefore, in view of the definition of both the kernel $K$ and Assumption (N2) on the errors, one can define the kernel-type estimator

$$\tilde{f}_{h_I}(x_I) := n^{-\frac{1}{2}} \sum_{k=1}^{n} L_{(h_I)}(x_I - Y_{k,I}), \quad L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i(t_I \cdot x_I)} \frac{K_{h_I}(t_I)}{\hat{q}_I(t_I)} dt_I.$$  

(9)

The ideas that led to the introduction of the estimators $\tilde{f}_{h_I}$ are explained in Fan [9] in the one-dimensional setting and, in Comte and Lacour [5] in the multivariate context.

Family of estimators  Below we propose a data driven selection from the family of estimators

$$\mathfrak{F}[\mathfrak{P}] := \left\{ \tilde{f}_{(h,P)}(x) = \prod_{I \in \mathfrak{P}} \tilde{f}_{h_I}(x_I), \ x \in \mathbb{R}^d, (h,P) \in \mathcal{H}_p[\mathfrak{P}] \right\},$$

where the set $\mathcal{H}_p[\mathfrak{P}]$ of parameters $(h,P)$ is constructed as follows.

For $I \in I_d$, consider first the set of multibandwidths

$$\mathcal{H}_{p,I} := \left\{ h_I \in \left[h_{\min}^{(p)}, h_{\max}^{(p)}\right]^{[|I|]} : h_j = 2^{-k_j}, k_j \in \mathbb{N}^*, j \in I \right\},$$

$$h_{\min}^{(p)} := \begin{cases} n^{-\left(1 + \frac{1}{p}\right)}, & p \in (1, +\infty), \\ n^{-1}, & p = +\infty, \end{cases} \quad h_{\max}^{(p)} := \begin{cases} [\ln(n)]^{-\frac{1}{p}}, & p \in (1, +\infty), \\ 1, & p = +\infty. \end{cases}$$

Then define

$$\mathcal{H}_{p,I} := \left\{ h_I \in \mathcal{H}_{p,I} : (nV_{h_I})^{\frac{1}{2}} \prod_{j \in I} h_j^{\lambda_j} \geq c_p 1_{\{p < +\infty\}} + \sqrt{\ln(n)} 1_{\{p = +\infty\}} \right\},$$  

(11)

$$b_p := \frac{1}{2} \wedge (1 - \frac{1}{p}), \quad c_p := 1 \wedge \left\{ \frac{p}{\pi} \left[ 1 + \lambda_{\max} \left( 2 \vee \frac{p}{p - 1} \right) \right]^{-p[b_p + \lambda_{\max}]} \right\}, \quad \lambda_{\max} := \max_{j=1\ldots,d} \lambda_j.$$  

The constant $c_p$ is chosen in order to have $\mathcal{H}_{p,I} \neq \emptyset$, $\forall n \geq 3$.

Put finally

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

The introduction of the estimator $\tilde{f}_{(h,P)}$ is based on the following simple observation. If there exists $P \in \mathfrak{P}(f)$, the idea is to estimate separately each marginal density corresponding to $I \in \mathfrak{P}$. Since the estimated density possesses the product structure we seek its estimator in the same form.
**Auxiliary estimators** We mimic the procedure of Lepski [19] by introducing the following auxiliary estimators. Consider first the classical kernel auxiliary estimators

\[ \tilde{f}_{h, \eta}(x_I) := K_{\eta} \ast f_{h, \eta}(x_I), \quad h, \eta \in (0, 1]^d, \quad I \in I_d, \]

where, here and in the sequel, "\( \ast \)" stands for the standard convolution product on \( \mathbb{R}^s, \ s \in \mathbb{N}^* \).

Then put, for \( h, \eta \in (0, 1]^d \) and \( P, P' \in \mathfrak{P} \),

\[ \tilde{f}(h, P, (\eta, P')) := \prod_{I \in P \circ P'} \tilde{f}_{h, \eta}(x_I), \]

where the operation "\( \circ \)" is defined by (7).

The ideas that led to the introduction of the estimators \( \tilde{f}(h, P, (\eta, P')) \), based on both the operation "\( \ast \)" and "\( \circ \)" are explained in Lepski [19], 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 \( \|_{\mathfrak{P}} \)–loss.

### 3.2. Selection rule

For \( I \in I_d \) and \( h \in (0, 1]^d \), define

\[ \mathcal{U}_p(h_I) := \begin{cases} \frac{1}{n^{p-1}} \| L(h_I) \|_p, & p \in (1, 2), \\ n^{-\frac{p}{2}} \prod_{j \in I} h_j^{-\frac{\lambda_j}{2} - \frac{1}{2}}, & p = 2, \\ n^{-\frac{p}{2}} \left[ \prod_{j \in I} h_j^{-\frac{\lambda_j}{2} + \sqrt{\ln(n)} \| L(h_I) \|_{\frac{2p}{p+2}}} \right], & p \in (2, +\infty), \\ n^{-\frac{p}{2}} \sqrt{\ln(n)} \prod_{j \in I} h_j^{-\frac{\lambda_j}{2}}, & p = +\infty. \end{cases} \]

Put also \( \Lambda_p := \varnothing \gamma_p \left[ \overline{\mathcal{G}}_p \right]^{(p-1)} \), where \( \varnothing := \sup_{\mathfrak{P}} |\mathfrak{P}| \),

\[ \overline{\mathcal{G}}_p := 1 \lor \left[ \| K \|_1^d \sup_{(h, P) \in \mathcal{H}_p[\mathfrak{P}]} \sup_{P' \in \mathfrak{P}} \sup_{I \in P \circ P'} \| \tilde{f}_{h, I} \|_p \right] \]

and \( \gamma_p > 0 \) is a numerical constant whose expression is given in Section 5.1 below.

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

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

Define finally \( \tilde{\rho}(h, P) \) satisfying

\[ \tilde{\Delta}_p(\tilde{\rho}(h, P)) + \Lambda_p \mathcal{U}_p(h, P) = \inf_{(h, P) \in \mathcal{H}_p[\mathfrak{P}]} \left[ \Delta_p(h, P) + \Lambda_p \mathcal{U}_p(h, P) \right]. \]

Our selected estimator is \( \tilde{f} := \tilde{f}(\tilde{\rho}(h, P)) \).

Note first that the existence of the quantities involved in the selection procedure is ensured by both the finiteness of the set \( \mathcal{H}_p[\mathfrak{P}] \) and the following result. The first statement given in Proposition 1 is a simple consequence of Marcinkiewicz Multiplier Theorem; see Theorem 5.2.4. and Corollary 5.2.5. in Grafakos [14].
Assume that Assumptions (N1)-(N2) are verified. Let

\[ P \text{ is determined by the choice of } f, \]

and in particular

\[ f \in \mathbb{F}_p [\mathcal{P}^2] := \left\{ g \in \mathbb{F} : \sup_{p \in \mathcal{P} \cap \mathcal{P}^2} \sup_{I \in \mathcal{P} \cap \mathcal{P}^2} \| g \|_p < \infty \right\}, \quad (15) \]

where \( \mathbb{F} \) denotes the set of all probability densities \( g : \mathbb{R}^d \to \mathbb{R}_+ \). The considered class of densities is determined by the choice of \( \mathcal{P} \) and in particular

\[ \mathbb{F}_p [\emptyset] = \left\{ g \in \mathbb{F} : \| g \|_p < \infty \right\}, \quad \mathbb{F}_p [\{ \mathcal{P} \}] = \left\{ g \in \mathbb{F} : \sup_{I \in \mathcal{P}} \| g \|_p < \infty \right\}. \]

\section{Main results}

In this section, we first provide oracle inequalities for our estimator \( \tilde{f}_h \). Then, we discuss adaptive minimax estimation over scales of anisotropic Nikolskii classes.

\subsection{Oracle inequalities}

Note 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:

\[ f \in \mathbb{F}_p [\mathcal{P}] := \left\{ g \in \mathbb{F} : \sup_{p, p' \in \mathcal{P}} \sup_{I \in \mathcal{P} \cap \mathcal{P}^2} \| g \|_p < \infty \right\}, \quad (16) \]

\[ \mathbb{E} \sup_{h \in \mathcal{H}_{p, l}} \left[ \| \xi_{h_l}(x) - g \|_p - \gamma_{p, l}(r) \mathcal{U}_p(h_l) \right]^r \leq c_p(r)n^{-\frac{r}{2}}, \quad c_p(r) > 0. \quad (14) \]

The proof of this proposition is postponed to Appendix. It is important to emphasize that the first bound was not used for the definition of \( \mathcal{U}_p(h_l) \) since a dimensional constant is not explicitly done in Theorem 5.2.4. of Grafakos [14].

Next, we also emphasize that the quantity \( \mathcal{U}_p(h_l) \) can be viewed, up to a numerical constant, as a uniform bound on the \( \mathbb{L}_p \)-norm of the stochastic error provided by the kernel-type estimator \( \tilde{f}_h \). This is explained by the following result. For \( I \in \mathcal{I}_d, h \in (0, 1]^d \) and \( x \in \mathbb{R}^d \), define

\[ \xi_{h_l}(x) := \tilde{f}_h(x) - \mathbb{E}\{\tilde{f}_h(x)\}. \]

\[ (14) \]

The constants \( \gamma_{p, l}(r) \) and \( c_p(r) \) do not depend on the sample size \( n \). Their explicit expressions can be found in the proof of the latter result, which is also postponed to Appendix.

Finally, in view of the assumptions on the kernel \( \mathbb{K} \), since \( \mathcal{H}_p [\mathcal{P}] \) is a finite set, \( (\tilde{h}, \tilde{\mathcal{P}}) \) exists, is in \( \mathcal{H}_p [\mathcal{P}] \) and is \( Y^{(n)} \)-measurable. It follows that \( \tilde{f} : \mathbb{R}^n \to \mathbb{L}_p (\mathbb{R}^d) \) is an \( Y^{(n)} \)-measurable mapping.

\section{Main results}

In this section, we first provide oracle inequalities for our estimator \( \tilde{f}_h \). Then, we discuss adaptive minimax estimation over scales of anisotropic Nikolskii classes.

\subsection{Oracle inequalities}

Note 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:

\[ f \in \mathbb{F}_p [\mathcal{P}] := \left\{ g \in \mathbb{F} : \sup_{p, p' \in \mathcal{P}} \sup_{I \in \mathcal{P} \cap \mathcal{P}^2} \| g \|_p < \infty \right\}, \quad (16) \]

where \( \mathbb{F} \) denotes the set of all probability densities \( g : \mathbb{R}^d \to \mathbb{R}_+ \). The considered class of densities is determined by the choice of \( \mathcal{P} \) and in particular

\[ \mathbb{F}_p [\emptyset] = \left\{ g \in \mathbb{F} : \| g \|_p < \infty \right\}, \quad \mathbb{F}_p [\{ \mathcal{P} \}] = \left\{ g \in \mathbb{F} : \sup_{I \in \mathcal{P}} \| g \|_p < \infty \right\}. \]
Define, for \((h, \mathcal{P}) \in \mathcal{H}_p \cup \mathcal{P}\) such that \(\mathcal{P} \in \Psi(f)\),

\[
\mathcal{R}_p \left( (h, \mathcal{P}), f \right) := \left( E f \sup_{\mathcal{P}' \in \Psi} \sup_{I \in \mathcal{P} \cap \mathcal{P}'} \left\| \hat{f}_{h_I} - f_I \right\|_p^p \right)^{\frac{1}{p}}, \quad p \in (1, +\infty),
\]

\[
\mathcal{R}_p \left( (h, \mathcal{P}), f \right) := E f \sup_{\mathcal{P}' \in \Psi} \sup_{I \in \mathcal{P} \cap \mathcal{P}'} \left\| \hat{f}_{h_I} - f_I \right\|_\infty.
\]

If the possible independence structure \(\mathcal{P}\) of the target density is known, the latter quantity can be viewed as an "\(L_p\)-risk" of the estimator \(\hat{f}(h, \mathcal{P})\), defined with the loss

\[
l \left( \hat{f}(h, \mathcal{P}), f \right) := \sup_{\mathcal{P}' \in \Psi} \sup_{I \in \mathcal{P} \cap \mathcal{P}'} \left\| \hat{f}_{h_I} - f_I \right\|_p.
\]

In this case, we see that the effective dimension of estimation is not \(d\), but \(d(\mathcal{P}) := \sup f \in \mathcal{P} \left\| f \right\|_\infty\).

Therefore, the best estimator from the family \(\mathcal{F}[\Psi]\) (the oracle) should be \(\hat{f}(\mathcal{P}^*_{\text{opt}})\) such that

\[
\mathcal{R}_p \left( (h, \mathcal{P}^*_{\text{opt}}), f \right) = \inf_{(h, \mathcal{P}) \in \mathcal{H}_p \cup \mathcal{P}} \mathcal{R}_p \left( (h, \mathcal{P}), f \right).
\]

Let us provide the following oracle inequalities for our selected estimator \(\hat{f}\).

**Theorem 3.** Assume that Assumptions (N1)-(N2) are satisfied.

If \(n \geq 3\) and \(p \in (1, +\infty)\) then: \(\forall f \in \mathcal{F}[\Psi]\),

\[
\mathcal{R}_p \left( \hat{f}, f \right) \leq C_{p,1}(f_p) \inf_{(h, \mathcal{P}) \in \mathcal{H}_p \cup \mathcal{P}} \left\{ \mathcal{R}_p \left( (h, \mathcal{P}), f \right) + \gamma_p \mathcal{H}_p(h, \mathcal{P}) \right\} + C_{p,2}(f_p)n^{-\frac{1}{2}},
\]

where \(f_p := 1 + \sup_{\mathcal{P}' \in \Psi} \sup_{I \in \mathcal{P} \cap \mathcal{P}'} \left\| f_I \right\|_p\).

The explicit expression of \(C_{p,1}(f_p) = C_{p,1}(d, \Psi, K, q, f_p)\) and \(C_{p,2}(f_p) = C_{p,2}(d, \Psi, K, q, f_p)\) is given in the proof of the theorem. It is worth to note that the maps \(f_p \mapsto C_{p,1}(f_p)\) and \(f_p \mapsto C_{p,2}(f_p)\) are bounded on any bounded interval of \(\mathbb{R}_+\).

If \(\Psi = \{\emptyset\}\) we obtain automatically some oracle inequalities for estimation on \(\mathbb{R}^d\) under \(L_p\)-loss, without considering any independence structure. In this case, the result above can be improved. Indeed, by scrutinizing its proof, one can easily see that the following theorem is true.

**Theorem 4.** Assume that \(\Psi = \{\emptyset\}\) and that Assumptions (N1)-(N2) are satisfied.

If \(n \geq 3\) and \(p \in (1, +\infty)\) then: \(\forall f \in \mathcal{F}\),

\[
\mathcal{R}_p \left( \hat{f}, f \right) \leq \inf_{h \in \mathcal{H}_p \cup \Psi} \left\{ \left( 1 + 2 \|K\|_1^q \right) \mathcal{R}_p \left( \hat{f}_h, f \right) + 2\gamma_p \mathcal{H}_p(h) \right\} + 2C_p n^{-\frac{1}{2}}.
\]

The explicit expression of the absolute constant \(C_p = C_p(d, \Psi, K, q) > 0\) is given in the proof of the theorem.

Note first that the statement of Theorem 4 holds for all probability densities \(f \in \mathcal{F}\), that is not true for Theorem 3. Next, the constant \(1 + 2 \|K\|_1^q\) is more suitable than \(C_{p,1}(f_p)\). Indeed, the prime interest in the oracle approach is to obtain a constant that does not depend on the target density and close to one. However, Theorem 3 allows us to consider both the smoothness properties and the independence structure of the target density and then to reduce the influence of the dimension on the accuracy of estimation. Indeed, if \(f\) has an independence structure \(\mathcal{P} \neq \emptyset\) and the smoothness parameter \(h\) is fixed and properly chosen then our procedure should choose the true partition \(\mathcal{P}\) and the estimator \(\hat{f}(h, \mathcal{P})\) should provide a better accuracy of estimation than the classical kernel-type estimator \(\hat{f}_h\). This was illustrated by a short simulation study in Rebelles [27] for the density model (with direct observations), under the \(L_2\)-loss.
4.2. L\textsubscript{p}-adaptive minimax estimation

In what follows, we illustrate the application of Theorems 3 and 4 to adaptive estimation over anisotropic Nikolskii classes of densities \( N_{r,d}(\beta, L) \) and \( N_{r,d}(\beta, L) \) respectively. To compute an L\textsubscript{p}-risk of a kernel-type estimator, we first compute its bias. Thus, we need to enforce the assumptions imposed on the kernel \( K \). One of the possibilities is the following, proposed in Kerkyacharian, Lepski and Picard [17].

For a given integer \( l \geq 2 \) and a given symmetric function \( u : \mathbb{R} \rightarrow \mathbb{R} \) belonging to the Schwartz class \( \mathbb{S}(\mathbb{R}) \) and satisfying \( \int_{\mathbb{R}} u(z)dz = 1 \) set

\[
 u_l(z) := \sum_{j=1}^{l} \binom{l}{j} (-1)^{j+1} \frac{1}{j} u \left( \frac{z}{j} \right), \quad z \in \mathbb{R}.
\]

(18)

Furthermore we use \( K \equiv u_l \) in the definition of the collection of estimators \( \mathcal{K}(\mathbb{P}) \). The relation of kernel \( u_l \) to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [17]. In particular, it has been shown that

\[
 \int_{\mathbb{R}} K(z)dz = 1, \quad \int_{\mathbb{R}} z^k K(z)dz = 0, \quad \forall k = 1, \ldots, l - 1.
\]

(19)

4.2.1. Minimax adaptive estimation under an \( L_p \)-loss

For \( (\beta, \mathcal{P}) \in (0, +\infty)^d \times \mathbb{P} \) define \( \phi_{n,p}(\beta, \mathcal{P}) := n^{-\tau/\tau+1} \), where

\[
 \tau := \tau(\beta, \mathcal{P}) = \inf_{I \subseteq \mathcal{P}} \tau_I, \quad \tau_I := \left[ \sum_{j \in I} \frac{b_p^{-1} \lambda_j + 1}{\beta_j} \right]^{-1},
\]

(20)

where \( b_p \) is given in (11).

Assume that \( \overline{\theta} \in \mathbb{P} \) and consider the estimator \( \tilde{f} \) defined by the selection rule (12)-(13) with \( p \in (1, +\infty) \).

**Theorem 5.** Let \( p \in (1, +\infty) \) be arbitrary fixed. Suppose that Assumptions (N1)-(N2) are satisfied. Then for any \( (\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathbb{P} \) one has

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

To get the statement of the latter theorem we apply Theorem 3. If \( \mathbb{P} = \{ \overline{\theta} \} \) (no independence structure), we obtain the following theorem by applying Theorem 4.

**Theorem 6.** Let \( p \in (1, +\infty) \) be arbitrary fixed. Suppose that \( \mathbb{P} = \{ \overline{\theta} \} \) and that Assumptions (N1)-(N2) are satisfied. Then for any \( (\beta, L) \in (0, l]^d \times (0, \infty)^d \) one has

\[
 \limsup_{n \to +\infty} \left\{ \phi_{n,p}^{-1}(\beta, \overline{\theta}) \mathcal{R}_p \left[ \tilde{f}, N_{p,d}(\beta, L) \right] \right\} < \infty.
\]

To the best of our knowledge, the latter results are new. In view of the assertion of Theorem 1, if \( p \in [2, +\infty) \) and Assumptions (N1)-(N3) on the errors are satisfied, we deduce from Theorem 6 that \( \phi_{n,p}(\beta, \overline{\theta}) \) is the minimax rate of convergence on the anisotropic Nikolskii class \( N_{p,d}(\beta, L) \) and that a minimax estimator can be selected from the collection of kernel-type estimators introduced in Section 3.1. Moreover, if \( \mathbb{P} = \{ \overline{\theta} \} \) (no independence structure), the quality of estimation of our estimator \( \tilde{f} \) is optimal, up to a numerical constant, on each class \( N_{p,d}(\beta, L) \),
whatever the nuisance parameter \((\beta, L)\). Thus, in the aforementioned case, \(\tilde{f}\) is an optimal adaptive estimator over the scale \(\{N_{p,d}(\beta, L)\}_{(\beta, L)}\).

Remark that \(L_p\)-estimation of an anisotropic density in the deconvolution model does not require that this density is uniformly bounded, whereas it is imposed in all the works concerning the density model (with direct observations); see, e.g. in Goldenshluger and Lepski [12].

Unfortunately, if \(p \in (1, 2)\), our estimator does not achieve the minimax lower bound on \(N_{p,d}(\beta, L)\) obtained in Lepski and Willer [21] under the \(L_p\)-loss. We conclude that our estimator is not minimax on \(N_{p,d}(\beta, L)\) or the lower bound in Lepski and Willer [21] is not the minimax rate of convergence on the latter functional class.

It is important to emphasize that both Theorems 5-6 allow us to analyze the influence of the independence structure on the accuracy of estimation under an \(L_p\)-loss in the deconvolution model. Indeed, we see that

\[
\phi_{n,p}(\beta, \emptyset) \gg \phi_{n,p}(\beta, \mathcal{P}), \quad \mathcal{P} \neq \emptyset,
\]

whatever the independence structure of the common density of the errors. Thus, our estimation procedure allows us to improve significantly the accuracy of estimation if the target density has an independence structure \(\mathcal{P} \neq \emptyset\).

Having said that, the question is: is \(\phi_{n,p}(\beta, \mathcal{P})\) the minimax rate of convergence on the functional class \(N_{p,d}(\beta, L, \mathcal{P})\)? For the density model (that corresponds to \(\lambda_j = 0, j = 1, \ldots, d\)), it is asserted in Rebelles [27] that the answer is positive and that the proof of the corresponding minimax lower bound coincides with the one of Theorem 3 in Goldenshluger and Lepski [13], up to minor modifications to take into account the independence structure. For the deconvolution model, we conjecture that the answer is also positive if \(p \in [2, +\infty)\) and that a minimax lower bound on \(N_{p,d}(\beta, L, \mathcal{P})\) can be obtained, up to straightforward modifications, as in Lepski and Willer [21].

### 4.2.2. Minimax adaptive estimation under sup-norm loss

For \((\beta, r, \mathcal{P}) \in (0, +\infty)^d \times [1, +\infty]^d \times \Phi\) define \(\phi_{n,\infty}(\beta, r, \mathcal{P}) := \left(\frac{n}{\ln(n)}\right)^{-\tau_1^{-1}}\), where

\[
\tau_1 := \left[\sum_{j \in I} \frac{2\lambda_j + 1}{\beta_j}\right]^{-1}, \quad \omega_1 := \left[\sum_{j \in I} \frac{2\lambda_j + 1}{\beta_j \lambda_j}\right]^{-1}, \quad \kappa_1 := \frac{1 - \sum_{j \in I} \frac{1}{\lambda_j}}{\sum_{j \in I} \frac{1}{\lambda_j}}.
\]  

(21)

Assume that \(\emptyset \in \Phi\) and consider the estimator \(\tilde{f}\) defined by the selection rule (12)-(13) with \(p = +\infty\). As previously, we obtain the following two theorems:

**Theorem 7.** Suppose that Assumptions (N1)-(N2) are satisfied. Then for any \((\beta, L, r, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d \times \Phi\) satisfying \(1 - \sum_{j=1}^{d} \frac{1}{\beta_j \lambda_j} > 0\) one has

\[
\limsup_{n \to +\infty} \left\{ \phi_{n,\infty}^{-1}(\beta, r, \mathcal{P}) \mathcal{R}_{p} \left[ \tilde{f}, N_{r,d}(\beta, L, \mathcal{P}) \right] \right\} < \infty.
\]

**Theorem 8.** Suppose that \(\emptyset \in \{\emptyset\}\) and that Assumptions (N1)-(N2) are satisfied. Then for any \((\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d\) satisfying \(1 - \sum_{j=1}^{d} \frac{1}{\beta_j \lambda_j} > 0\) one has

\[
\limsup_{n \to +\infty} \left\{ \phi_{n,\infty}^{-1}(\beta, r, \emptyset) \mathcal{R}_{p} \left[ \tilde{f}, N_{r,d}(\beta, L) \right] \right\} < \infty.
\]
To the best of our knowledge, the latter results are also new. Note first that in the case of direct observations we find again the results obtained in Lepski [19]. Next, if $\mathcal{Q} = \{\emptyset\}$ and $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$, it follows from Theorems 2 and 8 that, in presence of the noise satisfying Assumptions (N1)-(N3), $\phi_{n,\infty}(\beta, r, \emptyset)$ is the minimax rate of convergence on the anisotropic class $N_{r,d}(\beta, L)$. In this case, our estimator is an optimal adaptive one over the scale

$$\left\{ N_{r,d}(\beta, L), (\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d, 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0 \right\}.$$

It is worth to note that our estimator can be used for pointwise estimation. Moreover, it follows from Theorem 2 that our estimator achieves the adaptive rates of convergence found in Comte and Lacour [5.1. Quantities and technical lemma]

The rate of convergence $\phi_{n,\infty}(\beta, r, \emptyset)$ in presence of the noise satisfying Assumptions (N1)-(N3), $\phi_{n,\infty}(\beta, r, \mathcal{Q})$ is the minimax rate of convergence on the anisotropic class $N_{r,d}(\beta, L)$. In this case, our estimator is an optimal adaptive one over the scale

$$\left\{ N_{r,d}(\beta, L), (\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d, 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0 \right\}.$$

As previously, Theorems 7-8 allow us to conclude that our procedure leads to a better accuracy of estimation under sup-norm loss whenever the target density has an independence structure $\mathcal{P} \neq \emptyset$. In this case, we improve significantly the results obtained in Comte and Lacour [5] under a pointwise loss. Furthermore, we emphasize that if the target density has a known independence structure $\mathcal{P}$, $\mathcal{Q} = \{\mathcal{P}\}$, $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} \leq 0$ and $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0, \forall \mathcal{I} \in \mathcal{P}$, our estimator achieves the rate of convergence $\phi_{n,\infty}(\beta, r, \mathcal{P})$ on $N_{r,d}(\beta, L, \mathcal{P})$ whereas there is no uniformly consistent estimator on $N_{r,d}(\beta, L)$.

Finally, we conjecture that $\phi_{n,\infty}(\beta, r, \mathcal{P})$ is the minimax rate of convergence on $N_{r,d}(\beta, L, \mathcal{P})$ when $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$ and that a proof of the corresponding lower bound can be obtained by a minor modification of that in Lepski and Willer [21] to take into account the possible independence structure of the underlying density.

5. Proofs of main results

5.1. Quantities and technical lemma

For brevity, introduce first

$$\mathcal{I}^c := \{ I \in \mathcal{P} \circ \mathcal{P}', (\mathcal{P}, \mathcal{P}') \in \mathcal{Q} \times \mathcal{Q}\}, \quad \overline{U}_{r} := \sup_{n \in \mathbb{N}^*} \sup_{I \in \mathcal{I}^c} \sup_{h_I \in \mathcal{H}_{p,l}} U_{r}(h_I) < \infty,$$

Note that the finiteness of $\overline{U}_{r}$ is due both to the definition of the sets of multibandwidths $\mathcal{H}_{p,l}$ and to the bounds given in Proposition 1.

Next, define the constant $\gamma_p$ involved in the selection rule. For $I \in \mathcal{I}^c$ and $r \geq 1$, put

$$\gamma_{p,I}(r) := \begin{cases} 4 + \sqrt{\frac{37c-1}{2-p}}, & p \in (1, 2), \\ \left(7C_I + 3A (2\pi)^{\frac{1}{r}} \left\| \bar{K}_I g_l \right\|_\infty \| g_l \|_{\frac{1}{2}} \right) r, & p = 2, \\ \left(\frac{46c(p)[p+\epsilon]}{3e} \right)^{\frac{1}{2}} \left[ 1 \vee C_I \left( 1 \vee \left\| g_l \right\|_\infty \right)^{\frac{3}{2}} \right] r, & p \in (2, +\infty), \\ 6C_I(K, q) \left[ 1 \vee \left\| g_l \right\|_\infty \right]^{\frac{7}{2}} \left[ 93|I| \ln(|I|) + 69r \right], & p = +\infty, 
\end{cases}$$

where $g_l(t_l) := \prod_{l \in I} \left( 1 + t_l^2 \right)^{\frac{1}{2}}, C_I := A \left\{ (2\pi)^{\frac{1}{2r}} \left( \left\| \bar{K}_I g_l \right\|_2 \vee \left\| \bar{K}_I g_l \right\|_1 \right) \right\}, c_p$ is given in the definition of $\mathcal{H}_{p,l}, c(p) := 15p / \ln(p)$ and

$$C_I(K, q) := \frac{A}{(2\pi)^{\frac{1}{2r}}} \left\{ \left\| \bar{K}_I g_l \right\|_2 \vee \left\| \bar{K}_I g_l \right\|_1 \vee \left( \max_{j \in I} \left\| D_j^1 \bar{K}_I g_l \right\|_1 \right) \vee \left\| \bar{K}_I \varphi_I \right\|_2 \vee \left\| \bar{K}_I \varphi_I \right\|_1 \right\}.$$
with \( \varphi_f(t_j) := \sup_{j\in I} |t_j| g_j(t_I) \). Then, put \( r_k := k p \mathbf{1}_{\{p<\infty\}} + k \mathbf{1}_{\{p=+\infty\}}, k \geq 1 \), and

\[
\gamma_p := \begin{cases} 
\sup_{P,P'\in \mathcal{P}} \sup_{I\in \mathcal{I}^{P,P'}} \{ \gamma_{p,I}(r_4) \}, & \mathcal{P} \neq \{\emptyset\}, \\
\gamma_{p,\emptyset}(r_1), & \mathcal{P} = \{\emptyset\}.
\end{cases}
\]

Finally, we need the following technical lemma in order to compute our risk bounds. Define

\[
\xi_p := \sup_{I \in \mathcal{I}^{P,\emptyset}, h_I \in \mathcal{H}_{P,I}} \left( \sup_{I \in \mathcal{I}^{P,\emptyset}} \left( \|h_I\|_p - \gamma_{p,I}(h_I) \right) \right) + ,
\]

\[
\Phi_p := \hat{\alpha}^2 \|\mathbf{K}\|_1 \left( \sup_{I \in \mathcal{I}^{P,\emptyset}} \left( \|\mathbf{K}\|_1^d \mathbf{P}_I \right) \right)^{\frac{3}{2}}, \quad \mathbf{f}_p := 1 \vee \left( \sup_{I \in \mathcal{I}^{P,\emptyset}} \|f_I\|_p \right).
\]

**Lemma 1.** Assume that \( \mathcal{P} \neq \{\emptyset\} \). Set \( \mathbf{r} \in \{r_1, r_2, r_4\} \). Under Assumptions (N1)-(N2), if \( p \in (1, +\infty) \) then, for all integer \( n \geq 3 \),

\[
\left( \mathbb{E} f \right| \xi_p) \leq c_{p,1}(\mathbf{r}) n^{-\frac{1}{2}}, \quad \left( \mathbb{E} f \right| \Phi_p)^{\frac{1}{2}} \leq c_{p,2}(\mathbf{r}, \mathbf{f}_p), \quad \forall f \in \mathcal{F}_p[\mathcal{P}].
\]

The absolute constants \( c_{p,1}(\mathbf{r}) > 0 \) and \( c_{p,2}(\mathbf{r}, \mathbf{f}_p) > 0 \) can be explicitly expressed and the maps \( \mathbf{f}_p \mapsto c_{p,2}(\mathbf{r}, \mathbf{f}_p) \) are bounded on any bounded interval of \( \mathbb{R}_+ \); see the proof of the latter result, which is postponed to Appendix.

### 5.2. Oracle inequalities : proof of Theorems 3 and 4.

1) Set \( p \in [1, +\infty] \) and \( f \in \mathcal{F}_p[\mathcal{P}] \). Let \( (h, \mathcal{P}) \in \mathcal{H}_p[\mathcal{P}], \mathcal{P} \in \mathcal{P}(f) \), be fixed.

In view of the triangle inequality we have

\[
\left\| \tilde{f} - f \right\|_p \leq \left\| \tilde{f}(h, \mathcal{P}) - \tilde{f}(h, \mathcal{P}, (h, \mathcal{P})) \right\|_p + \left\| \tilde{f}(h, \mathcal{P}, (h, \mathcal{P})) - \tilde{f}(h, \mathcal{P}) \right\|_p + \left\| \tilde{f}(h, \mathcal{P}) - f \right\|_p
\]

\[
\leq \Delta_p(h, \mathcal{P}) + \Lambda_p \mathcal{M}_p(h, \mathcal{P}) + \Delta_p(h, \mathcal{P}) + \Lambda_p \mathcal{M}_p(h, \mathcal{P}) + \left\| \tilde{f}(h, \mathcal{P}) - f \right\|_p.
\]

Here we have used the equality \( \tilde{f}(h, \mathcal{P}, (h, \mathcal{P})) = \tilde{f}(h, \mathcal{P}, (h, \mathcal{P})) \). By definition of \( (h, \mathcal{P}) \), we obtain

\[
\left\| \tilde{f} - f \right\|_p \leq 2 \left[ \Delta_p(h, \mathcal{P}) + \Lambda_p \mathcal{M}_p(h, \mathcal{P}) \right] + \left\| \tilde{f}(h, \mathcal{P}) - f \right\|_p.
\]

(22)

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

\[
\left\| \tilde{f}(h, \mathcal{P})(x) - f(x) \right\| = \left| \prod_{I \in \mathcal{P}} \tilde{f}_{h_I}(x_I) - \prod_{I \in \mathcal{P}} f_I(x_I) \right|
\]

\[
\leq \sum_{j=1}^m \left| \tilde{f}_{h_{I_j}}(x_{I_j}) - f_{I_j}(x_{I_j}) \right| \left( \prod_{k=j+1,m} \left| \tilde{f}_{h_{I_k}}(x_{I_k}) \right| \right) \left( \prod_{l=1,j-1} \left| f_{I_l}(x_{I_l}) \right| \right).
\]

Here we have used the trivial equality: for \( m \in \mathbb{N}^* \) and \( a_j, b_j \in \mathbb{R}, j = 1, m, \)

\[
\prod_{j=1}^m a_j - \prod_{j=1}^m b_j = \sum_{j=1}^m (a_j - b_j) \left( \prod_{k=j+1,m} a_k \right) \left( \prod_{l=1,j-1} b_l \right),
\]

(23)
where the product over empty set is assumed to be equal to one.

In view of \( P \in \mathfrak{P} \), the triangle inequality and the Fubini-Tonelli theorem (used for the case \( p < \infty \)) we establish

\[
\|\tilde{f}_{(h,P)} - f\|_p \leq \sum_{j=1}^{m} \|\tilde{f}_{h_{I_j}} - f_{I_j}\|_p \left( \prod_{k=j+1,m} \|\tilde{f}_{h_{I_k}}\|_p \right) \left( \prod_{l=1,j-1} \|f_{I_l}\|_p \right)
\]

\[
\leq m \left( \max \{\overline{c}_p, f_p\} \right)^{m-1} \sup_{I \in \mathcal{P}} \|\tilde{f}_{h_I} - f_I\|_p,
\]

since \( \|K\|_1 \geq \int K = 1 \). Remind that \( \delta = \sup_{P \in \mathfrak{P}} |P| \) and \( \overline{c}_p \geq 1 \). It follows

\[
\|\tilde{f}_{(h,P)} - f\|_p \leq \delta \left( \max \{\overline{c}_p, f_p\} \right)^{\delta-1} \sup_{I \in \mathcal{P}} \|\tilde{f}_{h_I} - f_I\|_p.
\]

(24)

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

\[
|\tilde{f}_{(h,P),(\eta,P')}(x) - \tilde{f}_{(\eta,P')}(x)| = \prod_{I' \in P'} \prod_{I \in \mathcal{P} : I \cap I' \neq \emptyset} K_{\eta_I \cap I'} \ast \sum_{I \in \mathcal{P} : I \cap I' = \emptyset} \tilde{f}_{h_{I'}}(x_{I' \cap I}).
\]

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

\[
\|\tilde{f}_{(h,P),(\eta,P')} - \tilde{f}_{(\eta,P')}\|_p \leq \delta \left( \overline{c}_p \right)^{\delta^2-1} \sup_{I' \in \mathcal{P}'} \|\tilde{f}_{h_{I'}} - f_{I'}\|_p.
\]

(25)

Here we have used Young’s inequality and the inequalities \( \|K\|_1 \geq \int K = 1 \) and \( \overline{c}_p \geq 1 \).

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

\[
\|E_f \{\tilde{f}_{\eta_I}(\cdot)\}\|_p = \|K_{\eta_I} \ast f_I\|_p \leq \|K_I\|_1 \|f_I\|_p \leq \|K\|_1 \|f_p\|.
\]

(26)

Then, by the same method as the one used in step 2 and (26), for any \((\eta, P') \in \mathcal{H}_p[\mathfrak{P}]\) and any \(I' \in \mathcal{P}'\) we get

\[
\left\| \prod_{I' \in \mathcal{P} : I \cap I' \neq \emptyset} \tilde{f}_{h_{I'} \cap I'} - \prod_{I \in \mathcal{P} : I \cap I' \neq \emptyset} E_f \{\tilde{f}_{\eta_{I'}}(\cdot)\} \right\|_p
\]

\[
\leq \delta \left( \max \{\overline{c}_p, \|K\|_1^d f_p\} \right)^{\delta'-1} \sup_{I' \in \mathcal{P} \cap I' \neq \emptyset} \|K_{\eta_{I'} \cap I'} \ast \left( \tilde{f}_{h_{I'} \cap I'} - f_{I'}\right)\|_p
\]

\[
\leq \delta \|K\|_1 \left( \max \{\overline{c}_p, \|K\|_1^d f_p\} \right)^{\delta'-1} \sup_{I' \in \mathcal{P} \cap I' \neq \emptyset} \|\tilde{f}_{h_{I'} \cap I'} - f_{I'}\|_p.
\]

(27)

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

\[
E_f \{\tilde{f}_{\eta_I}(x_{I'})\} = \int K_{\eta_{I'}} (y_{I'} - x_{I'}) \prod_{I' \in \mathcal{P} : I' \neq I} f_{I'}(y_{I' \cap I})dy_{I'} = \prod_{I' \in \mathcal{P} : I' \neq I} E_f \{\tilde{f}_{\eta_I}(x_{I'})\}.
\]

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

Thus, in view of the triangle inequality, (25), (27) and the trivial inequality \( \sup_i x_i - \sup_i y_i \leq \sup_i |x_i - y_i| \), for any \((\eta, P') \in \mathcal{H}_p[\mathfrak{P}]\), we get
\[ \left\| \tilde{f}(\eta, P) - \tilde{f}(\eta, P') \right\|_p - \Lambda_p M_p(\eta, P') \right\|_p + \left\| \xi_{\eta'} \right\|_p - \gamma_p M_p(\eta') \right\|_p \right] \]

\[ \leq \delta \left( G_{p} \right)^{2(0-1)} \sup_{I' \in P'} \left\| \prod_{I' \in P'} \tilde{f}_{\eta'-I'} - \prod_{I' \in P'} \mathbb{E}_f \left\{ \tilde{f}_{\eta'-I'}(\cdot) \right\} \right\|_p + \left\| \xi_{\eta'} \right\|_p - \gamma_p M_p(\eta') \right\|_p \]

\[ \left\| \tilde{f}(\eta, P) - \tilde{f}(\eta, P') \right\|_p - \Lambda_p M_p(\eta, P') \right\|_p + \sup_{P' \in P} \sup_{I' \in P'} \left\| \tilde{f}_{\eta'} - f_{I'} \right\|_p + \bar{f}_p \xi_p, \]

since \( \bar{f}_p \geq \delta \left( G_{p} \right)^{2(0-1)} \geq 1 \). We deduce

\[ \bar{\Delta}_p(h, P) \leq \bar{f}_p \sup_{P' \in P} \sup_{I' \in P'} \left\| \tilde{f}_{\eta'} - f_{I'} \right\|_p + \bar{f}_p \xi_p. \] 

Finally, it follows from (22), (24) and (28)

\[ \left\| \tilde{f} - f \right\|_p \leq 3\bar{f}_p \left\{ \sup_{P' \in P} \sup_{I' \in P'} \left\| \tilde{f}_{\eta'} - f_{I'} \right\|_p + \gamma_p M_p(h, P) + \xi_p \right\}. \] (29)

6) Consider the random event \( B_p := \left\{ \bar{G}_p \geq C_p \right\}, C_p(f_p) = (1 + \gamma_p M_p + \|K\|_1^d f_p^d) \|K\|_1^d + 1 \).

Put also

\[ \mathcal{R}_{p}^{(r)} [(h, P), f] := \left( \mathbb{E}_f \sup_{P' \in P} \sup_{I' \in P'} \left\| \tilde{f}_{\eta'} - f_{I'} \right\|_p \right)^{1/r}, \quad r \geq 1. \]

In view of (26), Lemma 1, Markov’s inequality, (29), and the Cauchy-Schwarz inequality we get \( B_p \subseteq \{ \xi_p \geq 1 \} \), \( \mathbb{E}_f \left( B_p \right) \leq c_{p,1}(r) n^{-1/2} \) and

\[ \left( \mathbb{E}_f \left\| \tilde{f} - f \right\|_p^{r_1} \right)^{1/r_1} \leq 3\delta^2 \|K\|_1^d \|C_p(f_p)\|^{2^2-1} \left( \mathcal{R}_{p}^{(r_1)} [(h, P), f] + \gamma_p M_p(h, P) + c_{p,1}(r_1) \sqrt{n} \right), \]

\[ \left( \mathbb{E}_f \left\| \tilde{f} - f \right\|_p^{r_1} \right)^{1/r_1} \leq 3c_{p,1}(r_4)c_{p,2}(r_4, f) \left( \mathcal{R}_{p}^{(r_2)} [(h, P), f] + \gamma_p M_p + c_{p,1}(r_2) \right) n^{-1/2}, \]

\[ \mathcal{R}_{p}^{(r_2)} [(h, P), f] \leq c_{p,1}(r_2) + \gamma_p M_p + \|K\|_1^d f_p + f_p. \]

Thus, we come to the assertion of Theorem 3 with \( C_{p,1}(f_p) := 3\delta^2 \|K\|_1^d \|C_p(f_p)\|^{2^2-1} \) and

\[ C_{p,2}(f_p) := 3c_{p,1}(r_4)c_{p,2}(r_4, f) \left( 2\gamma_p M_p + (1 + \|K\|_1^d)^d f_p + 2c_{p,1}(r_2) \right) + 3c_{p,1}(r_1) \delta^2 \|K\|_1^d \|C_p(f_p)\|^{2^2-1}, \]

since \( \mathcal{R}_{p}^{(r_1)} [(h, P), f] = \mathcal{R}_{p} [(h, P), f] \). The constants \( c_{p,1}(r_k) \) and \( c_{p,2}(r_k, f_p), k = 1, 2, 4 \), are given in the proof of Lemma 1.

7) **Particular case:** \( \mathcal{P} = \{ \emptyset \} \) (no independence structure)

Set \( f \in \mathcal{F} \) and let \( h \in \mathcal{H}_{p, \emptyset} \) be arbitrary fixed. By scrutinizing the steps 1)-5) we easily see that

\[ \left\| \tilde{f} - f \right\|_p \leq (1 + 2\|K\|_1^d) \left\| \tilde{h} - f \right\|_p + 2\gamma_p, \emptyset (r_1) M_p(h) + 2 \left\| \xi h_p - \gamma_{p, \emptyset} (r_1) M_p(h) \right\|_p. \]
Thus, we get from Proposition 2
\[
\left(\mathbb{E}_f \left\| \tilde{f} - f \right\|_{p}^{\frac{1}{p}} \right) \leq (1 + 2 \left\| \mathbf{K} \right\|_{1}^{d}) \left( \mathbb{E}_f \left\| \tilde{f}_h - f \right\|_{p}^{\frac{1}{p}} \right) + 2 \gamma_{p,\mathbf{g}}(\mathbf{r}_1) \mathcal{U}_p(h) + 2c_p(\mathbf{r}_1)n^{-1/2},
\]
where the constants \( \gamma_{p,\mathbf{g}}(\mathbf{r}_1) \) and \( c_p(\mathbf{r}_1) \) are given in the proof of Proposition 2.

5.3. Adaptive minimax upper bounds: Proof of Theorems 5-8

1) Case \( p \in (1, +\infty) \): let \( (\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathfrak{P} \) and \( f \in \mathcal{N}_{p,d}(\beta, L, \mathcal{P}) \subset \mathbb{F}_p[\mathfrak{P}] \) be arbitrary fixed.

In view of the triangle inequality, \( \forall h \in (0, 1]^d \),
\[
\sup_{\mathcal{P} \in \mathfrak{P}} \sup_{J \in \mathcal{P} \circ \mathcal{P}'} \left\| \tilde{f}_{h,j} - f_{J} \right\|_{p} \leq \sup_{\mathcal{P} \in \mathfrak{P}} \sup_{J \in \mathcal{P} \circ \mathcal{P}'} \left\| \mathbb{E}_f \{ \tilde{f}_{h,j}(\cdot) \} - f_{J} \right\|_{p} + \sup_{\mathcal{P} \in \mathfrak{P}} \sup_{J \in \mathcal{P} \circ \mathcal{P}'} \left\| \xi_{h,j} \right\|_{p}
\]
where \( \mathbb{E}_f \{ \tilde{f}_{h,j}(x,J) \} = K_{h,j} * f_J(x,J) \) and, remind, \( \xi_{h,j}(x,J) := \tilde{f}_{h,j}(x,J) - \mathbb{E}_f \{ \tilde{f}_{h,j}(x,J) \} \).

Note first that, by applying Proposition 3 in Kerkyacharian, Lepski and Picard [17], it is easily established that, for any \( h \in (0, 1]^d \), any \( \mathcal{P}' \in \mathfrak{P} \) and any \( J \in \mathcal{P} \circ \mathcal{P}' \),
\[
\left\| K_{h,j} * f_J - f_J \right\|_{p} \leq \sum_{j \in J} c_f(K, |J|, p, l, L) h_j^{\beta_j} \leq c \sum_{j \in J} h_j^{\beta_j} = c \sum_{i \in \mathcal{P}'} \sum_{j \in I} h_j^{\beta_j}, \quad c > 0.
\]

Next, if \( (h, \mathcal{P}) \in \mathcal{H}_p[\mathfrak{P}] \), we easily get from Propositions 1-2
\[
\left( \mathbb{E}_f \sup_{\mathcal{P} \in \mathfrak{P}} \sup_{J \in \mathcal{P} \circ \mathcal{P}'} \left\| \xi_{h,j} \right\|_{p}^{\frac{1}{p}} \right) \leq O \left( \sup_{i \in \mathcal{P}'} \frac{1}{\sqrt{n \prod_{j \in I} h_j^{\beta_j}}} \right).
\]

Consider now, for all \( I \in \mathcal{P} \), the system
\[
h_j^{\beta_j} = h_k^{\beta_k} = \frac{1}{\sqrt{n \prod_{j \in I} h_j^{\beta_j}}} h_j^{b_j + \gamma_j + 1}, \quad j, k \in I.
\]
The solution is given by
\[
h_j = n^{-\frac{d}{2\gamma_j + 1} - 1}, \quad j \in I, \quad I \in \mathcal{P},
\]
where \( \tau_I \) is given in (20).

Note that, for all \( I \in \mathcal{P} \), \( h_I \in [h_{\min}^{(p)}, h_{\max}^{(p)}]|I| \) and \( n \prod_{j \in I} h_j^{b_j + \gamma_j + 1} \geq 1 \) for \( n \) large enough. Denote by \( \mathbf{h}_I \) the projection of \( h_I \) on the dyadic grid \( \mathcal{H}_{p,I} \). It is easily checked that \( (h, \mathcal{P}) \in \mathcal{H}_p[\mathfrak{P}] \) for \( n \) large enough. Thus, it follows from Theorem 3, (30), (31) and (32) that
\[
\mathcal{R}_p \left[ \tilde{f}, f \right] \leq C \left[ \sup_{I \in \mathcal{P}} \sum_{j \in I} h_j^{\beta_j} + \sup_{I \in \mathcal{P}} \frac{1}{\sqrt{n \prod_{j \in I} h_j^{\beta_j}}} \right] + \alpha_{p,2n^{-1/2}}, \quad C > 0,
\]
for \( n \) large enough. Finally, it is easily seen that we get the statement of Theorem 5 from (33) and (34). Similarly, the statement of Theorem 6 is obtained by applying Theorem 4.
2) **Case** $p = +\infty$: let $(\beta, L, r, \mathcal{P}) \in (0, 1]^d \times (0, \infty)^d \times [1, +\infty)^d \times \mathcal{P}$ such that $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$ and $f \in N_{r,d}(\beta, L, \mathcal{P})$ be arbitrary fixed. It follows from the definition of the latter functional class and the embedding theorem for anisotropic Nikolskii classes, see, e.g., Theorem 6.9 in Nikolskii [24], that $N_{r,d}(\beta, L, \mathcal{P}) \subset \mathcal{F}_{\infty}[\mathcal{P}]$, since $1 - \sum_{j \in I} \frac{1}{\beta_j r_j} > 0, \forall I \in \mathcal{I}_d$.

Note first that, in view of the arguments given in the proof of Theorem 3 in Lepski [19], it follows from Lemme 4 in the latter paper that, for any $h \in (0,1]^d$, any $\mathcal{P} \in \mathcal{P}$ and any $J \in \mathcal{P} \circ \mathcal{P}'$,

$$
\|K_{h_J} \ast f_J - f_J\|_\infty \leq c \sup_{I \in \mathcal{P}} \sum_{j \in I} h_j^{\beta(I)} c := c(K, d, l, L) > 0,
$$

$$
\beta_j(I) := \sigma(I) \beta_j \sigma^{-1}(I), \quad \sigma(I) := 1 - \sum_{k \in I} (\beta_k p_k)^{-1}, \quad \sigma_j(I) := 1 - \sum_{k \in I} (p_k^{-1} - p_j^{-1}) \beta_k^{-1}.
$$

Next, if $(h, \mathcal{P}) \in H_{\infty}[\mathcal{P}]$, we easily get from Propositions 2

$$
\mathbb{E}_f \sup_{\mathcal{P}' \in \mathcal{P}} \sup_{h \in \mathcal{P}'(I)} \|\xi_{h_I}\|_\infty \leq O\left(\sup_{I \in \mathcal{P}} \frac{\ln(n)}{n \prod_{j \in I} h_j^{2\lambda_j + 1}}\right).
$$

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

$$
h_j^{\beta(I)} = h_k^{\beta(I)} = \sqrt[2\lambda_j + 1]{\frac{\ln(n)}{n \prod_{j \in I} h_j^{2\lambda_j + 1}}}, \quad j, k \in I.
$$

The solution is given by

$$
h_j = \left(\frac{n}{\ln(n)}\right)^{\frac{1}{2\lambda_j + 1}} \tau_j, \quad j \in I, \quad I \in \mathcal{P},
$$

where $\tau_j$ is given in (21).

Note that, for all $I \in \mathcal{P}$, $n \prod_{j \in I} h_j^{2\lambda_j + 1} \geq \ln(n)$ for $n$ large enough. Thus, as previously, we get the statement of Theorem 7 from Theorem 3, (35), (36) and (37). Similarly, the statement of Theorem 8 is obtained by applying Theorem 4.

6. **Appendix**

6.1. **Proof of Proposition 1**

Let $h \in (0,1]^d$ and $I \in \mathcal{I}_d^+$ be arbitrary fixed. Note that

$$
L(h_I)(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i(t_I \cdot x_I)} \frac{\overline{K_I(h_I t_I)g_I(h_I t_I)}}{g_I(h_I t_I) \overline{q_I(t_I)}} dt_I, \quad g_I(t_I) := \prod_{j \in I} (1 + t_j^2)^{-\frac{1}{2}},
$$

where $h_I t_I$ denotes the coordinate-wise product of the vectors $h_I$ and $t_I$. 

18
1) **Proof of assertion (i)** Set $p \in (1, 2)$. Here, we apply the Marcinkiewicz Multiplier Theorem on $\mathbb{R}^{|I|}$, given in Grafakos [14] p. 363, with

$$ m(t_I) = g_I^{-1}(h_I t_I) \hat{g}_I^{-1}(t_I). $$

In view of Assumption (N2) on $q$, $m$ is a bounded function defined away from the coordinates axes on $\mathbb{R}^{|I|}$ and is $C^{|I|}$ on this region. Moreover,

$$ \sup_{t_I \in \mathbb{R}^{|I|}} |m(t_I)| \leq A \sup_{u_I \in \mathbb{R}^{|I|}} \left[ \prod_{j \in I} \left( 1 + u_j \right)^{-\lambda_j} \prod_{j \in I} \left( 1 + [u_j/h_j]^2 \right)^{\frac{\gamma_j}{2}} \right] \leq A \prod_{j \in I} h_j^{-\lambda_j}. \quad (39) $$

Set $\alpha_I = (\alpha_j)_{j \in I} \in \mathbb{N}^{|I|}$ satisfying $|\alpha_I| := \sum_{j \in I} \alpha_j \leq |I|$. In view of Leibniz’s rule, one has

$$ [D^{\alpha_I} m] (t_I) = \sum_{\gamma_I \leq \alpha_I} \left( \alpha_I \right) \left\{ \prod_{j \in I} h_j^{\gamma_j} \right\} \left[ D^{\gamma_I} g_I^{-1} \right] (h_I t_I) \left[ D^{\alpha_I - \gamma_I} \hat{g}_I^{-1} \right] (t_I), \quad \forall t \in \mathbb{R}^d. $$

Here, $\gamma_I \leq \alpha_I$ means $\gamma_j \leq \alpha_j, \forall j \in I$.

Let $t_I$ be chosen such that $t_I \neq 0$ if $\alpha_j \neq 0$. In this case, for any multi-index $\gamma_I \leq \alpha_I$,

$$ \left\{ \prod_{j \in I} h_j^{\gamma_j} \right\} \left[ D^{\gamma_I} g_I^{-1} \right] (h_I t_I) \left[ D^{\alpha_I - \gamma_I} \hat{g}_I^{-1} \right] (t_I) = \left\{ \prod_{j \in I} (t_I h_I)^{\gamma_j} \right\} \left[ D^{\gamma_I} g_I^{-1} \right] (h_I t_I) \left[ D^{\alpha_I - \gamma_I} \hat{g}_I^{-1} \right] (t_I) \left\{ \prod_{j \in I} t_I^{-\alpha_j} \right\} \left( \prod_{j \in I} t_I^{\alpha_j} \right). $$

Here, we assume that $0^0$ is equal to one.

Since $q$ satisfies Assumption (N2), we obtain similarly as in (39)

$$ |[D^{\alpha_I} m] (t_I)| \leq C(|I|, q_I) A \left\{ \prod_{j \in I} h_j^{-\lambda_j} \right\} \left( \prod_{j \in I} |t_I|^{-\alpha_j} \right), \quad (40) $$

$$ C(|I|, q_I) := \max_{|\alpha_I| \leq |I|} \left\{ \sum_{\gamma_I \leq \alpha_I} \left( \alpha_I \right) \sup_{u_I \in \mathbb{R}^{|I|}} \left\{ \prod_{j \in I} u_j^{\gamma_j} \right\} \left[ D^{\gamma_I} g_I^{-1} \right] (u_I) g_I(u_I) \right\} < \infty. \quad (41) $$

Put $\widehat{S_I}(t_I) := \widehat{K_I}(t_I) g_I(t_I)$, $t \in \mathbb{R}^d$. Since $K \in \mathcal{S}(\mathbb{R})$, $\widehat{S_I} \in \mathcal{S}(\mathbb{R}^{|I|})$ is the Fourier transform of a function $S_I \in \mathcal{S}(\mathbb{R}^{|I|}) \subset L_p(\mathbb{R}^{|I|})$. As

$$ L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i(t_I, x_I)} m(t_I) \widehat{S_I}(h_I t_I) dt_I, $$

it follows from Corollary 5.2.5. in Grafakos [14], (39) and (40)

$$ \|L_{(h_I)}\|_p \leq 2AC_{|I|} C(|I|, q_I) \max (p, (p - 1)^{-1}) \|S_I\|_p (V_{h_I})^{-(1-1/p)} \prod_{j \in I} h_j^{-\lambda_j}, $$

where $C_{|I|} < \infty$ is a dimensional constant which is not explicitly done in the aforementioned result. Thus, assertion (i) of Proposition 1 is proved with

$$ C_{p,I} := 2A \left\{ C_{|I|} C(|I|, q_I) \max (p, (p - 1)^{-1}) \|S_I\|_p \right\}. $$
2) Proof of assertion (ii) Note first that

\[ \left\| L(h_1) \right\|_2 = (2\pi)^{-\frac{|I|}{2}} \left\| \tilde{K}_{h_1} / \hat{q}_I \right\|_2, \quad \| L(h_1) \|_\infty \leq (2\pi)^{-|I|} \left\| L(h_1) \right\|_1. \]

In view of Assumption (N2) on the errors,

\[ \left\| \tilde{K}_{h_1} / \hat{q}_I \right\|_2^2 \leq A^2 \int_{|I|} \left| \tilde{K}_I(h_j t_I) \right|^2 \prod_{j \in I} (1 + t_j^2)^{\lambda_j} dt_I \]

\[ \leq A^2 \left( \int_{|I|} \left| \tilde{K}_I(u_I) \right|^2 \prod_{j \in I} (1 + u_j^2)^{\lambda_j} du_I \right) V_{h_1}^{-1} \prod_{j \in I} h_j^{2\lambda_j}; \]

\[ \left\| \tilde{K}_{h_1} / \hat{q}_I \right\|_1 \leq A \left( \int_{|I|} \left| \tilde{K}_I(u_I) \right| \prod_{j \in I} (1 + u_j^2)^{\lambda_j / 2} du_I \right) V_{h_1}^{-1} \prod_{j \in I} h_j^{-\lambda_j}. \]

Thus, assertion (ii) of Proposition 1 is proved with

\[ C_I := A \left\{ (2\pi)^{-\frac{|I|}{2}} \left( \left\| \tilde{K}_I g_I \right\|_2 \vee \left\| \tilde{K}_I \hat{q}_I \right\|_1 \right) \right\}, \]

where \( g_I \) is given in (38).

\[ \square \]

6.2. Proof of Proposition 2: case \( p < \infty. \)

Let \( I \in \mathcal{I}_d \) be arbitrary fixed. We get the statement of Proposition 2 by applying Theorem 1 and Corollaries 2 and 3 in Goldenshluger and Lepski [11] with \( s = p, \mathcal{X} = \mathcal{T} = \mathbb{R}^{|I|}, \nu = \tau \) is the Lebesgue measure on \( \mathbb{R}^{|I|}, \) \( w(\cdot, \cdot) = n^{-1} L(h_1)(\cdot - \cdot) \) and \( M_s(w) \leq n^{-1} L(h_1) \|_p < \infty. \) Here, the i.i.d. random vectors are the \( Y_{k,1} \)'s and their common density is \( f_I * q_I. \) By using the continuity property of \( L(h_1)(\cdot) \), it is easily proved that Assumption (A1) in the aforementioned paper is fulfilled.

1) Case \( p \in (1, 2). \) Let \( r \geq 1 \) and \( h_1 \in \mathcal{H}_{p, I} \) be arbitrary fixed.

By application of Corollary 2 in Goldenshluger and Lepski [11], one has

\[ \mathbb{P} \left\{ \left\| \xi_{h_1} \right\|_p \geq U_p(h_1) + z \right\} \leq \exp \left\{ -\frac{z^2}{A^2_p(h_1)} \right\}, \quad \forall z > 0, \forall n \geq 1, \tag{42} \]

where \( U_p(h_1) = 4n^{-\frac{1}{p}} \| L(h_1) \|_p \) and \( A^2_p(h_1) = 37n^{-1} \| L(h_1) \|_p^2. \)

By integration of (42) we easily get, for all integer \( n \geq 3, \)

\[ \mathbb{E} \left[ \left\| \xi_{h_1} \right\|_p - U_p(h_1) - A_p(h_1) \sqrt{\ln(n)} \right]^r \geq \Gamma(r + 1) [A_p(h_1)]^r e^{-r \ln(n)} \]

\[ \leq \Gamma(r + 1)^r \sup_{h_1 \in \mathcal{H}_{p, I}} \left[ n^{-\frac{r}{p}} \| L(h_1) \|_p \right]^r n^{-r}, \]

where \( \Gamma(\cdot) \) is the well known Gamma function.

Note that, for all integer \( n \geq 3, \)

\[ U_p(h_1) + A_p(h_1) \sqrt{\ln(n)} \leq \left\{ 4 + \frac{37e^{-1} p r}{2 - p} \right\} n^{-\frac{1}{p}} \| L(h_1) \|_p : =: \gamma_{p, I} U_p(h_1). \]
Since $\text{card}(\mathcal{H}_{p, I}) \leq \left[ \left( 1 + \frac{p}{|I|} \right) \log_2(n) \right]^{\frac{1}{|I|}}$, we obtain, for all integer $n \geq 3$,

$$
\left\{ \mathbb{E} \sup_{h_I \in \mathcal{H}_{p, I}} \left[ \| h_{I, I} \|_{p, I} - \gamma_{p, I}(r) \mathcal{U}_p(h_I) \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}} \leq c_p(r)n^{-\frac{1}{2}},
$$

c_p(r) := 7 \left[ \Gamma(r + 1) \right]^{\frac{1}{2}} \sup_{n \in \mathbb{N}} \sup_{I \in \mathbb{F}_n^*} \sup_{h_I \in \mathcal{H}_{p, I}} \left\{ \frac{n^{-1}[2\log_2(n)]^{|I|}}{\| L(h_I) \|_p} \right\},
$$

which is finite in view of Proposition 1 and the definition of the set $\mathcal{H}_{p, I}$.

2) Case $p = 2$. Let $r \geq 1$ and $h_I \in \mathcal{H}_{p, I}$ be arbitrary fixed. Here, we apply Theorem 1 in Goldenshluger and Lepski [11] but we compute differently the upper bound on the "dual" variance $\sigma^2$ by using the arguments given in the proof of Proposition 7 in Comte and Lacour [5]. Indeed, we obtain

$$
\sigma^2 \leq n^{-2} (2\pi)^{-|I|} \frac{\tilde{K}_{h_I}}{\tilde{q}_I}^2 \left( \int_{\mathbb{R}^{|I|}} \left| \widehat{f}_1(t_I) \right| dt_I \right) \leq n^{-2} (2\pi)^{-|I|} \| \tilde{q}_I \|_1 \frac{\tilde{K}_{h_I}}{\tilde{q}_I}^2 \left( \int_{\mathbb{R}^{|I|}} \left| \widehat{f}_1(t_I) \right| dt_I \right),
$$

since $\| \widehat{f}_I \|_{\infty} \leq \| f_I \|_1 = 1$.

Taking into account the latter inequality, the result of Theorem 1 in Goldenshluger and Lepski [11] should be

$$
P \left\{ \| h_{I, I} \|_{p, I} \geq U_p(h_I) + z \right\} \leq \exp \left\{ -\frac{z^2}{A_p^2(h_I) + B_p(h_I)z} \right\}, \quad \forall z > 0, \forall n \geq 1, \quad (43)
$$

$$
U_p(h_I) = n^{-\frac{1}{2}} \| L(h_I) \|_2 ,
$$

$$
A_p^2(h_I) = \frac{6}{(2\pi)^{|I|}} \| \tilde{q}_I \|_1 n^{-1} \| \tilde{K}_{h_I}/\tilde{q}_I \|_{\infty}^2 + 24n^{-\frac{3}{2}} \| L(h_I) \|_2^2 , \quad B_p(h_I) = \frac{4}{3}n^{-1} \| L(h_I) \|_2 .
$$

By integration of (43) we get, for all integer $n \geq 3$,

$$
\mathbb{E} \left[ \| h_{I, I} \|_{p, I} - U_p(h_I) - A_p(h_I)\sqrt{r \ln(n)} - B_p(h_I)r \ln(n) \right]^{\frac{1}{r}} \leq \Gamma(r + 1) \left\{ A_p(h_I) + B_p(h_I) \right\} e^{-r \ln(n)}
$$

$$
\leq \Gamma(r + 1) \left( 6 \| \tilde{q}_I \|_1 \right)^{\frac{1}{r}} \sup_{h_I \in \mathcal{H}_{p, I}} \left\{ \| \tilde{K}_{h_I}/\tilde{q}_I \|_{\infty} + \| L(h_I) \|_2 \right\}^{\frac{1}{r}} n^{-\frac{3}{2} - \frac{r}{2}}.
$$

Note that, in view of Assumption (N2) on the errors,

$$
\| \tilde{K}_{h_I}/\tilde{q}_I \|_{\infty} \leq A \| \tilde{K}_{l, I} \|_{\infty} \prod_{j \in I} h_j^{-\lambda_i}, \quad (44)
$$

where $g_I$ is given in (38). Thus, in view of Proposition 1, (44) and the definition of $\mathcal{H}_{p, I}$, for all integer $n \geq 3$,

$$
U_p(h_I) + A_p(h_I)\sqrt{r \ln(n)} + B_p(h_I)r \ln(n)
$$

$$
\leq \left\{ C_I \left( 1 + \frac{8r}{3e} + \sqrt{\frac{48r}{e}} \right) + A \| \tilde{K}_{l, I} \|_{\infty} \sqrt{\frac{6r}{(2\pi)^{|I|}}} \right\} n^{-\frac{1}{2}} \prod_{j \in I} h_j^{-\lambda_i - \frac{1}{2}}
$$

$$
= : \gamma_{p, I}(r) \mathcal{U}_p(h_I).
$$
Finally, we obtain for all integer $n \geq 3$

$$
\left\{ \mathbb{E} \sup_{h_I \in \mathcal{H}_{p,I}} \left[ \| \xi_{h_I} \|_p - \gamma_{p,I}(r) \mathcal{U}_p(h_I) \right]^r \right\}^+ \leq c_p(r)n^{-\frac{1}{r}}, \quad c_p(r) := \Gamma(r + 1)^{-\frac{1}{r}}
$$

$$
\times \sup_n \sup_{I \in \mathcal{I}^3} \sup_{h_I \in \mathcal{H}_{p,I}} \left[ 6 \vee \| \hat{q_I} \|^\frac{2}{r} i \right] \left\{ \left\| \mathcal{K}_{h_I} \hat{q_I} \right\|_\infty + \left\| L(h_I) \right\|_2 \right\} \left[ 2 \log_2(n) \right] n^{-1},
$$

which is finite in view of Proposition 1, (44) and the definition of the set $\mathcal{H}_{p,I}$.

**3) Case $p > 2$.** Let $r \geq 1$ and $h_I \in \mathcal{H}_{p,I}$ be arbitrary fixed.

By application of Corollary 3 in Goldenshlag and Lepski [11], one has

$$
\mathbb{P} \left\{ \| \xi_{h_I} \|_p \geq U_p(h_I) + z \right\} \leq \exp\left\{ -\frac{z^2}{A^2_p(h_I) + B_p(h_I)\gamma} \right\}, \quad \forall z > 0, \forall n \geq 1, \quad (45)
$$

$$
U_p(h_I) = 3c(p) \| q_I \|^\frac{2}{2p} \left\{ n^{-\frac{4}{2p}} \left\| L(h_I) \right\|_2 + n^{-\frac{2}{p}} \left\| L(h_I) \right\|_p \right\},
$$

$$
A^2_p(h_I) = 16c(p) \| q_I \|^\frac{2}{2p} \left\{ n^{-1} \left\| L(h_I) \right\|_2^2 + n^{-\frac{2}{p}} \left\| L(h_I) \right\|_p \right\} + B_p(h_I)
$$

$$
B_p(h_I) = \frac{4}{3} c(p)n^{-1} \left\| L(h_I) \right\|_p, \quad c(p) = \frac{15p}{\ln(p)}.
$$

Here, we have used the following inequalities, which are consequences of Young’s inequality.

$$
\| f \ast q_I \|_\infty \leq \| f \|_1 \| q_I \|_\infty \leq \| q_I \|_\infty, \quad \left\| \sqrt{f \ast q_I} \right\|_p \leq \| f \ast q_I \|^\frac{2}{\frac{4}{p}}_\infty \left\| f \ast q_I \right\|_1 \leq \| q_I \|^\frac{4}{\frac{4}{p} - 1}_\infty.
$$

By integration of (45) we get, for all integer $n \geq 3$,

$$
\mathbb{E} \left[ \| \xi_{h_I} \|_p - U_p(h_I) - A_p(h_I)\sqrt{r \ln(n)} - B_p(h_I)\ln(n) \right]^r
$$

$$
\leq \Gamma(r + 1) \left\{ A_p(h_I) + B_p(h_I) \right\}^r e^{-r \ln(n)}
$$

$$
\leq \Gamma(r + 1) \left\{ 6c(p) \left( 1 \vee \| q_I \|_\infty^\frac{2}{p} \right) \right\}^r
$$

$$
\times \sup_{h_I \in \mathcal{H}_{p,I}} \left\{ \left\| L(h_I) \right\|_2^2 + \sqrt{\left\| L(h_I) \right\|_2 \left\| L(h_I) \right\|_p} + \left\| L(h_I) \right\|_p \right\} n^{-\frac{4}{2p}}.
$$

In view of Proposition 1, we get

$$
\left\| L(h_I) \right\|_p \leq \left\| L(h_I) \right\|_\infty^\frac{2}{2} \left\| L(h_I) \right\|_2^\frac{2}{2} \leq C_I V_{h_I}^\frac{1}{2} \prod_{j \in I} h_{\lambda,j}^\frac{1}{2}. \quad (46)
$$
Thus, in view of Proposition 1, (46) and the definition of $\mathcal{H}_{p,l}$, for all integer $n \geq 3$,
\[
U_p(h_I) + A_p(h_I)\sqrt{r\ln(n)} + B_p(h_I)r\ln(n)
\leq c_p \frac{1}{\sqrt{n}} \left( 1 + C_I \right) \left\{ 6c(p)\|q_I\|_\infty \frac{1}{\sqrt{n}} + 8\sqrt{\frac{rv_p(p\lor q)}{e}}\|q_I\|_\infty + \frac{4rc(p)(p\lor q)}{3c} \right\}
\times n^{-\frac{1}{2}} \left[ \prod_{j \in I} h_j^{-\lambda_j - \frac{1}{2}} + \sqrt{\ln(n)} \left\| L(h_I) \right\|_{p, \infty} \right]
:= \gamma_{p,l}(r)U_p(h_I).
\]

Finally, we obtain for all integer $n \geq 3$
\[
\left\{ \mathbb{E} \sup_{h_I \in \mathcal{H}_{p,l}} \left\| \xi_{h_I} \right\|_p - \gamma_{p,l}(r)U_p(h_I) \right\}^{\frac{1}{r}} \leq c_p(r)n^{-\frac{1}{2}}, \quad c_p(r) := 6c(p)\Gamma(r + 1)^{\frac{1}{r}}
\times \sup_{n \in \mathbb{N}} \sup_{I \in \mathcal{I}_n} \sup_{h_I \in \mathcal{H}_{p,l}} \left[ \left( 1 + \|q_I\|_\infty \right) \left\{ \|L(h_I)\|_{p, \infty} + \sqrt{\ln(n)} \left\| L(h_I) \right\|_p \right\} \right]^{n-1} \left[ \left( 1 + \frac{D}{\|I\|} \right) \log_2(n) \right]^{\frac{1}{r}},
\]
which is finite in view of Proposition 1, (46) and the definition of the set $\mathcal{H}_{p,l}$.

6.3. Proof of Proposition 2: case $p = +\infty$.

Let $n \geq 3$, $I \in \mathcal{I}_n$ and $h_I \in [1/n, 1]^{\|I\|}$ be arbitrary fixed. Assume that $n \prod_{j \in I} h_j^{2\lambda_j + 1} \geq \ln(n)$. We divide this proof into several steps.

1) Preliminaries: First, since $q$ satisfies Assumption (N2) and the $Y_{k,I}$’s are i.i.d. random vectors with density $f_I * q_I$, we get from Proposition 1
\[
\sup_{x_I \in \mathbb{R}^{\|I\|}} \sup_{y_I \in \mathbb{R}^{\|I\|}} |L(h_I)(x_I - y_I)| \leq \left\| L(h_I) \right\|_\infty \leq C_l(K, q) \prod_{j \in I} h_j^{-\lambda_j - 1} < \infty,
\]
\[
C_l(K, q) := \frac{A}{(2\pi)^{\frac{\|I\|}{2}}} \left\{ \left\| \widehat{K}_l g_I \right\|_2 \lor \left\| \widehat{K}_l g_I \right\|_1 \lor \left( \max_{j \in I} \left\| D_j^1 \widehat{K}_l g_I \right\|_1 \right) \lor \left\| \widehat{K}_l \varphi_l \right\|_2 \lor \left\| \widehat{K}_l \varphi_l \right\|_1 \right\},
\]
where $\varphi_l(t_I) := \sup_{j \in I} |t_j| g_I(t_I)$ and $g_I$ is given in (38);
\[
\sup_{x_I \in \mathbb{R}^{\|I\|}} \left( \mathbb{E} \left| L(h_I)(x_I - Y_{1,I}) \right|^2 \right)^{\frac{1}{2}} \leq \sqrt{\|f_I * q_I\|_\infty} \left\| L(h_I) \right\|_2 \leq \sqrt{\|q_I\|_\infty} C_l(K, q) \prod_{j \in I} h_j^{-\lambda_j - \frac{1}{2}}.
\]

Next, set $x_I$ and $\overline{x}_I$ be arbitrary fixed in $\mathbb{R}^{\|I\|}$. For any $t_I \in \mathbb{R}^{\|I\|}$
\[
|e^{-i\langle t_I, x_I \rangle} - e^{-i\langle t_I, \overline{x}_I \rangle}| = \left| \prod_{j \in I} e^{-it_j x_j} - \prod_{j \in I} e^{-it_j \overline{x}_j} \right|
\leq |I| \sup_{j \in I} |e^{-it_j x_j} - e^{-it_j \overline{x}_j}|
\leq |I| \sup_{j \in I} |t_j| \sup_{j \in I} |x_j - \overline{x}_j|.
\]

23
Therefore, for any \( y_I \in \mathbb{R}^{[I]} \)

\[
| L_{(h_I)}(x_I - y_I) - L_{(h_I)}(\bar{x}_I - y_I) |
\leq \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} \left| \hat{K}_{h_I}(t_I) \right| | e^{-i < t_I, x_I>} - e^{-i < t_I, \bar{x}_I>} | \, dt_I
\leq n|I|C_I(K, q) \prod_{j \in I} h_j^{-\lambda_j-1} \sup_{j \in I} |x_j - \bar{x}_j|;
\] (49)

\[
\left( \mathbb{E} \left| L_{(h_I)}(x_I - Y_{1,I}) - L_{(h_I)}(\bar{x}_I - Y_{1,I}) \right|^2 \right)^{\frac{1}{2}}
\leq \left( \frac{\| f_I * q_I \|_\infty}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} \left| \hat{K}_{h_I}(t_I) \right|^2 | e^{-i < t_I, x_I>} - e^{-i < t_I, \bar{x}_I>}^2 | \, dt_I \right)^{\frac{1}{2}}
\leq n|I| \sqrt{1 \vee q_I} \| C_I(K, q) \|_\infty \prod_{j \in I} h_j^{-\lambda_j-\frac{1}{2}} \sup_{j \in I} |x_j - \bar{x}_j|;
\] (50)

Consider now the normalized empirical process

\[
\bar{\xi}_{h_I}(x_I) := \left( C_I(K, q) \left[ \frac{2(1 \vee q_I)\|_\infty}{n \prod_{j \in I} h_j^{2\lambda_j+1}} \right] \right)^{-1} \xi_{h_I}(x_I).
\]

In view of Bernstein inequality, (49), (48), (49) and (50), \( \forall \, z > 0 \),

\[
P \{ | \bar{\xi}_{h_I}(x_I) | > z \} \leq 2 \exp \left\{ -\frac{z^2}{A(x_I) + zB(x_I)} \right\}; \] (51)

\[
P \{ | \bar{\xi}_{h_I}(x_I) - \bar{\xi}_{h_I}(\bar{x}_I) | > z \} \leq 2 \exp \left\{ -\frac{z^2}{a^2(x_I, \bar{x}_I) + zb(x_I, \bar{x}_I)} \right\}, \] (52)

where \( A(x_I) := 1, B(x_I) := \left( \prod_{j \in I} h_j^{2\lambda_j+1} \right)^{-\frac{1}{2}} \leq 1 \) and

\[
a(x_I, \bar{x}_I) = b(x_I, \bar{x}_I) := 2 \wedge \left\{ n|I| \sup_{j \in I} |x_j - \bar{x}_j| \right\}. \] (53)

It is easily seen that \( a(\cdot, \cdot) \) is a semi-metric on \( \mathbb{R}^{[I]} \).

2) Supremum-norm over totally bounded sets: In this step we obtain bounds of the supremum-norm of the normalized empirical process \( \bar{\xi}_{h_I}(\cdot) \) over totally bounded sets by applying Proposition 1 in Lepski [18] with \( \mathcal{T} = \mathbb{R}^{[I]}, \mathcal{C} = \mathbb{R}, \mathcal{X} = \bar{\xi}_{h_I} \) and \( \Psi(\cdot) = | \cdot | \). Then we have to check Assumptions 1, 2 and 3 required in the latter Proposition and to match the notations used in the present paper and in Lepski [18].

Note first that, in view of (51), (52) and (53), Assumption 1 is fulfilled with \( c = 2 \). Next, consider the family of closed balls

\[
\mathbb{B}_R(t_I) := \left\{ x_I \in \mathbb{R}^{[I]} : \sup_{j \in I} |x_j - t_j| \leq R/2 \right\}, \quad R \geq 1, \ t_I \in \mathbb{R}^{[I]}.
\]
In view of the continuity property of the Fourier transforms and the definition of the semi-metrics \(a\) and \(b\), it is obvious that Assumption 2 is also satisfied with \(\Theta = B_{2 J}(I)\), \(\overline{A}_\Theta = 1\) and 
\[
\overline{B}_\Theta = \left( \prod_{j \in I} h_j^{2 \lambda_j + 1} \right)^{-\frac{1}{2}}.
\]

Let \(s : \mathbb{R} \to \mathbb{R}_+ \setminus \{0\}\) defined by \(s(z) := (0, 01 + z^8)^{-1}\). Obviously \(\sum_{k \geq 0} s(2^k/2) \leq 1\) and, for any \(z > 0\),
\[
\mathcal{E}_{\Theta,a}(z(4\delta)^{-1}s(\delta)) \leq |I| \left\lfloor \ln \left( \frac{Rn|I|}{z(4\delta)^{-1}s(\delta)} \right) \right\rfloor_+, \quad \forall \delta > 0,
\]
where \(\mathcal{E}_{\Theta,a}(\delta), \delta > 0\), denotes the entropy of \(\Theta\) measured in \(a\). Then, for any \(z > 0\), there exists \(\delta_\varepsilon > 0\) small enough such that
\[
e_{s}^{(a)}(z, \Theta) := \sup_{\delta > 0} \delta^{-2} \mathcal{E}_{\Theta,a}(z(4\delta)^{-1}s(\delta)) = \sup_{\delta > \delta_\varepsilon} \delta^{-2} \mathcal{E}_{\Theta,a}(z(4\delta)^{-1}s(\delta)) < \infty;
\]
\[
e_{s}^{(b)}(z, \Theta) := \sup_{\delta > 0} \delta^{-1} \mathcal{E}_{\Theta,b}(z(4\delta)^{-1}s(\delta)) = \sup_{\delta > \delta_\varepsilon} \delta^{-1} \mathcal{E}_{\Theta,b}(z(4\delta)^{-1}s(\delta)) < \infty.
\]
Thus, Assumption 3 in Lepski [18] is fulfilled and Proposition 1 in the latter paper can be applied. Let us compute the quantities which appear in this result.

Choose \(\tilde{s} = (s, s)\), \(\kappa = (2\overline{A}_\Theta, 2\overline{B}_\Theta)\) and \(\varepsilon = \sqrt{2} - 1\). Since \(\overline{A}_\Theta \cap \overline{B}_\Theta \leq 1\) and \(a(x_1, \overline{x}_I) = b(x_1, \overline{x}_I) \leq 2, \forall x_1, \overline{x}_I \in \mathbb{R}^{|I|}\), we straightforwardly get
\[
e_{s}(\kappa, \Theta) := e_{s}^{(a)}(2\overline{A}_\Theta, \Theta) + e_{s}^{(b)}(2\overline{B}_\Theta, \Theta) \leq \sup_{\delta > 0} \delta^{-2} \mathcal{E}_{\Theta,a}(2(4\delta)^{-1}s(\delta)) + \sup_{\delta > 0} \delta^{-1} \mathcal{E}_{\Theta,b}(2(4\delta)^{-1}s(\delta)) \leq 4,5|I| \left\lfloor \ln \left( Rn|I| \right) \right\rfloor_+, + \frac{8}{5};
\]
\[
U_{s}^{(c)}(y, \kappa, \Theta) := \kappa_1 \sqrt{2[1 + \varepsilon^{-1}]^2 e_{s}(\kappa, \Theta) + y} + \kappa_2 \left( 2[1 + \varepsilon^{-1}]^2 e_{s}(\kappa, \Theta) + y \right) \leq 2\sqrt{31/|I| \ln (Rn|I|)} + 59 + \frac{2(31/|I| \ln (Rn|I|) + 59 + y)}{\sqrt{n \prod_{j \in I} h_j^{2 \lambda_j + 1}}}.
\]

Thus, it follows from Proposition 1 in Lepski [18] that, for any \(y \geq 1\) and any \(r \geq 1\),
\[
\mathbb{E} \left\{ \sup_{x_1 \in \mathbb{B}_{\mathbb{Z}^{|I|}}(I)} \left[ \xi_{h_1}(x_1) \right] - U_{s}^{(c)}(y, \kappa, \Theta) \right\}^r \leq 4 \Gamma(r + 1) \left( 2 y^{-1} U_{s}^{(c)}(y, \kappa, \Theta) \right)^r e^{-\frac{r}{4}}.
\]

3) **Supremum-norm over the whole space:** Let \(x_I \in \mathbb{R}^{|I|}\) be arbitrary fixed and \(y_I \in \mathbb{R}^{|I|}\) be such that \(\sup_{j \in I} |x_j - y_j| \geq n\). By integration by parts, we easily get
\[
|L(h_I)(x_I - y_I)| \leq \max_{j \in I} \left\| D_j^1 \left( \frac{K_{h_I}/q_I}{q_I} \right) \right\|_1 \leq \frac{C_I(K, q)}{n \prod_{j \in I} h_j^{2 \lambda_j + 1}} \leq \frac{C_I(K, q)}{n \prod_{j \in I} h_j^{2 \lambda_j + 1}},
\]
in view of Assumption (N2) on the errors.

Consider the collection of closed balls \(\mathcal{B}_n : \mathbb{B}_{2 J}(n j) \subset \mathbb{Z}^{|I|}\). Obviously this collection is a countable cover of \(\mathbb{R}^{|I|}\). Put, for any \(j \in \mathbb{Z}^{|I|}\),
\[
\xi_j := \int_{\mathcal{B}_n} f_I * q_I(x_I) dx_I, \quad \mathbb{B}(j) := \bigcup_{k \in \mathbb{Z}^{|I|}: \mathbb{B}_{2 J}(n k) \cap \mathbb{B}_n(k) \neq \emptyset} \mathcal{B}_{2 J}(n k).
\]
It is easily checked that
\[
\sum_{j \in \mathbb{Z}^{|I|}} f_j = \int_{\mathbb{R}^{|I|}} f_I \ast q_I(x_I) \left[ \sum_{j \in \mathbb{Z}^{|I|}} 1_{B(j)}(x_I) \right] \, dx_I \leq 4^{|I|}. \tag{57}
\]

Set \( j \in \mathbb{Z}^{|I|} \) such that \( f_j \geq n^{-v} \), where \( v \geq 1 \) is specified later. If \( y = 2 \ln(1/f_j) + (r + 1) \ln(n) \), we get from (55)
\[
\mathbb{E} \left\{ \sup_{x_I \in \mathbb{B}_2(nj)} |\xi_{h_I}(x_I)| - \gamma_{\infty,I}^{(v)}(r) \right\} \leq 2^{r+2} \Gamma(r+1) \left[ \gamma_{\infty,I}^{(v)}(r) \right] f_j n^{-\frac{r+1}{r}},
\]
where \( \gamma_{\infty,I}^{(v)}(r) := 4C_I(K,q)\sqrt{2(1 \vee q)}(93|I| \ln(|I|) + 60 + 2v + r) \), since \( n \prod_{j \in I} h_j^{2\lambda_j+1} \geq \ln(n) \).

Thus, in view of (57), we obtain
\[
\mathbb{E} \left\{ \sup_{x_I \in \Theta_1} |\xi_{h_I}(x_I)| - \gamma_{\infty,I}^{(v)}(r) \right\} \leq 2^{r+2}|I| \Gamma(r+1) \left[ \gamma_{\infty,I}^{(v)}(r) \right] n^{-\frac{r+1}{r}}, \tag{58}
\]
where \( \Theta_1 := \bigcup_{j \in \mathbb{Z}^{|I|} : f_j \geq n^{-v}} \mathbb{B}_2(nj) \).

Set \( j \in \mathbb{Z}^{|I|} \) such that \( f_j < n^{-v} \) and \( x_I \in \mathbb{B}_2(nj) \). In view of (42) and (56) we get, for any \( k = 1, \ldots, n, \)
\[
\mathbb{E} \left| L_{(h_I)}(x_I - Y_{k,I}) \right| = \mathbb{E} \left\{ \left| L_{(h_I)}(x_I - Y_{k,I}) \right| 1_{B(j)}(Y_{k,I}) \right\} + \mathbb{E} \left\{ \left| L_{(h_I)}(x_I - Y_{k,I}) \right| 1_{\mathbb{R}^{|I|}\setminus B(j)}(Y_{k,I})\right\}
\leq \mathbb{P} \{ Y_{k,I} \in \mathbb{B}(j) \} \frac{C_I(K,q)}{\prod_{j \in I} h_j^{2\lambda_j+1}} + \frac{C_I(K,q)}{n \prod_{j \in I} h_j^{2\lambda_j+1}}
\leq \frac{2C_I(K,q)}{n \prod_{j \in I} h_j^{2\lambda_j+1}}, \tag{59}
\]
since \( f_j := \mathbb{P} \{ Y_{k,I} \in \mathbb{B}(j) \} \leq n^{-v}, v \geq 1 \) and \( \sup_{j \in I} |x_j - Y_{k,j}| \geq n \) when \( Y_{k,I} \in \mathbb{R}^{|I|}\setminus \mathbb{B}(j) \).

Introduce random events
\[
D_j := \left\{ \sum_{k=1}^{n} 1_{B(j)}(Y_{k,I}) \geq 2 \right\}, \quad j \in \mathbb{Z}^{|I|}, \quad D := \bigcup_{j \in \mathbb{Z}^{|I|} : f_j < n^{-v}} D_j.
\]

Let \( \overline{D} \) be the complementary to \( D \). If \( \overline{D} \) holds then, in view of (42) and (56),
\[
n^{-1} \sum_{k=1}^{n} \left| L_{(h_I)}(x_I - Y_{k,I}) \right| \leq \frac{2C_I(K,q)}{n \prod_{j \in I} h_j^{2\lambda_j+1}}, \quad \forall x_I \in \Theta_2 := \mathbb{R}^{|I|}\setminus \Theta_1. \tag{60}
\]

Since \( n \prod_{j \in I} h_j^{2\lambda_j+1} \geq \ln(n) \), we get from (59) and (60)
\[
\sup_{x_I \in \Theta_2} |\xi_{h_I}(x_I)| 1_{\overline{D}} \leq \gamma_{\infty,I}^{(v)}(r) \left[ \frac{\ln(n)}{n \prod_{j \in I} h_j^{2\lambda_j+1}} \right].
\]
and, taking into account that \( \sup_{x_I \in \Theta_2} |\xi_{h_t}(x_I)| \leq 2C_I(K, q)n \),

\[
E \left\{ \sup_{x_I \in \Theta_2} |\xi_{h_t}(x_I)| - \gamma_{\infty,I}^{(r)}(r) \left( \frac{\ln(n)}{n \prod_{j \in I} h_j^{2 \lambda_j + 1}} \right)^r \right\} \leq [2C_I(K, q)]^r n^r P(D). \tag{61}
\]

Let \( j \in \mathbb{Z}^{|I|} \) satisfying \( f_j < n^{-v} \) be arbitrary fixed. In view of Markov inequality one has for any \( z > 0 \)

\[
P(D_j) \leq e^{-2z} \left[ E \left\{ e^{z \mathbb{1}_{D_j}(Y_1, I)} \right\} \right]^n \leq \exp \{ -2z + n(e^z - 1)f_j \},
\]

since the \( Y_{k,I}'s \) are i.i.d. random vectors. Minimizing the right hand side in \( z > 0 \) we obtain

\[
P(D_j) \leq (e/2)^2(nf_j)^2 \leq 2f_j n^{-2v}.
\]  \( \tag{62} \)

Thus, choosing \( v = 1, 5r + 2, 5 \), it follows from (57), (58), (61) and (62)

\[
E \left\{ \|\xi_{h_t}\|_{\infty} - \gamma_{\infty,I}^{(r)}(r) \left( \frac{\ln(n)}{n \prod_{j \in I} h_j^{2 \lambda_j + 1}} \right)^r \right\} \leq 2^{r+3+2|I|} \Gamma(r + 1) \left[ \gamma_{\infty,I}^{(r)}(r) \right]^r n^{r - \frac{r + 1}{2}}, \tag{63}
\]

where \( \gamma_{\infty,I}^{(r)}(r) := \gamma_{\infty,I}^{(1.5r+2.5)}(r) \).

Finally, in view of the definition of \( \mathcal{H}_{\infty,I} \),

\[
\left\{ \frac{E \sup_{h_I \in \mathcal{H}_{\infty,I}} \left( \left[ \|\xi_{h_I}\|_{\infty} - \gamma_{\infty,I}^{(r)}(r) \mathcal{U}_{\infty}(h_I) \right]_+^r \right)^{\frac{1}{r}}}{c_\infty(r) n^{-\frac{1}{2}}} \right\} \leq c_\infty(r) n^{-\frac{1}{2}}, \tag{64}
\]

\[
c_\infty(r) := \left[ \Gamma(r + 1) \right]^{\frac{1}{2}} \sup_{n \in \mathbb{N}} \sup_{I \in \mathbb{Z}^{|I|}} \left\{ \gamma_{\infty,I}^{(r)}(r) \left[ 2^{r+3+2|I|} \right]^{\frac{1}{r}} \left[ \log_2(n) \right]^{\frac{1}{r}} n^{-\frac{1}{2}} \right\} < \infty.
\]

\[
6.4. \text{Proof of Lemma 1}
\]

Assume that \( \mathfrak{P} \neq \{\emptyset\} \). Set \( f \in \mathbb{F}_p[\mathfrak{P}] \) and let \( r \in \{r_1, r_2, r_4\} \) be arbitrary fixed. We obtain Lemma 1 by applying Proposition 2. We divide this proof into two steps.

1) Note that

\[
\xi_p \leq \sum_{I \in \mathbb{Z}^{|I|}} \sup_{h_I \in \mathcal{H}_{p,I}} \left( \left[ \|\xi_{h_I}\|_p - \gamma_{p,I}^{(r)}(r) \mathcal{U}_p(h_I) \right]_+ \right),
\]

since \( \gamma_{p,I}^{(r)}(r) \) increase with \( r \). In view of Proposition 2, if \( p \in (1, +\infty] \) and \( n \geq 3 \),

\[
(E_f |\xi_p|^r)^{\frac{1}{r}} \leq c_{p,1}(r) n^{-\frac{1}{2}}, \quad c_{p,1}(r) := d|\mathfrak{P}|^2 c_p(r).
\]

27
Thus, we finish the proof of Lemma 1.

Below we use the following trivial equality: for any random variable $Y$

$$\left( E_f \left| Y^{r^2} \right| \right)^{\frac{1}{r}} = \left[ \left( E_f \left| Y^{r^2} \right| \right)^{\frac{1}{r^2}} \right]^{\frac{1}{r^2}} \right], \quad (65)$$

In view of Proposition 2, if $p \in (1, +\infty]$ and $n \geq 3$, \( \left( E_f \left| \tilde{F}_p \right|^r \right)^{\frac{1}{r}} \leq c_{p,2}(r, f_p) \) with

$$c_{p,2}(r, f_p) := d^2 \left\{ K \right\}^r \left[ 1 + \left\{ K \right\}^r \left( C_1 + \left| K \right| \right) + \gamma_p \right\}^{\frac{1}{r^2}} \] \right].$$

Thus, we finish the proof of Lemma 1.

References

[1] Baraud, Y. and Birgé, L. (2014). Estimating composite functions by model selection. *Ann. Inst. H. Poincaré Probab. Statist.* **50**, 285–314.
[2] Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15**, 1, 69–88.
[3] Butucea, C. and Tsybakov, A.B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.* **52**, 1, 111–128.
[4] Butucea, C. and Tsybakov, A.B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52**, 2, 237–249.
[5] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 2, 569–609.
[6] Comte, F. and Rebafka, T. (2012). Adaptive density estimation in the pile-up model involving measurement errors. *Electronic Journal of Statist.* **6**, 2002–2037.
[7] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factor, kernel complexity and Yatracos classes. *Ann. Statist.* **25**, 2626–2637.
[8] Fan, J. (1991). On the optimal rates of convergence for non parametric deconvolution problems. *Ann. Statist.* **19**(3), 1257–1272.
[9] Fan, J. (1991). Global behavior of deconvolution kernel estimates. *Statis. Sinica* **1**, 541–551.
[10] Goldenshluger, A. and Lepski, O.V. (2009). Structural adaptation via $L_p$ – norm oracle inequalities. *Probab. Theory Related Fields.* **143**, 41–71.
[11] Goldenshluger, A. and Lepski, O. (2011a). Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* **39**, 2318–2384.
[12] Goldenshluger, A. and Lepski, O. (2011b). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* 39, 1608–1639.

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

[14] Grafakos, L. (2008). *Classical Fourier Analysis*. Graduate Texts in Mathematics, 249. Springer

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

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

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

[18] Lepski, O. (2013). Upper functions for positive random functionals. I. General setting and Gaussian Random functions. *Math. Methods Statist.* 22 (1), 1–27.

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

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

[21] Lepski, O. and Willer, T. (2014). Lower bounds in the convolution structure density model. *manuscript*.

[22] Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* 39, 2, 201–231.

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

[24] Nikol’skii, S.M. (1977). *Priblizheni Funktsii Mnogikh Peremennykh i Teoremy Vlosheniya* (in Russian). [Approximation of functions of several variables and embedding theorems.] Nauka, Moscow, 1977.

[25] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27, 6, 2033–2053.

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

[27] Rebelles, G. (2015). $L_p$ adaptive estimation of an anisotropic density under independence hypothesis. *Electronic Journal of Statist.* 9, 106–134.

[28] 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.

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