ISSN 1066-5307, Mathematical Methods of Statistics, 2016, Vol. 25, No. 1, pp. 26–53. © Allerton Press, Inc., 2016.

Structural Adaptive Deconvolution under $\mathbb{L}_p$-Losses

G. Rebelles[1*]

[1] *Inst. Math. de Marseille, Aix-Marseille Univ., Marseille, France*

Received July 23, 2015; in final form, February 20, 2016

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, ε is a measurement error independent of the random vector X of interest and having a known density with respect to Lebesgue measure. Our aim is to obtain optimal accuracy of estimation under $\mathbb{L}_p$-losses when the error ε 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 ε. 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: density estimation, deconvolution, kernel estimator, oracle inequality, adaptation, independence structure, concentration inequality.

2000 Mathematics Subject Classification: 62G05, 62G20.

DOI: 10.3103/S1066530716010026

1. INTRODUCTION

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

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

where the errors ε_k are also i.i.d. d-dimensional random vectors, independent of the X_k's, with a known density q.

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 $\widetilde{f}\colon (\mathbb{R}^d)^n \to \mathbb{L}_p(\mathbb{R}^d)$. The accuracy of an estimator is measured by its $\mathbb{L}_p$-*risk*

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

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 $\|g\|_{\mathbf{r}} = (\int |g(x)|^{\mathbf{r}} \, dx)^{1/\mathbf{r}}$ is the $\mathbb{L}_{\mathbf{r}}$-norm of $g \in \mathbb{L}_{\mathbf{r}}(\mathbb{R}^s)$, $s \in \mathbb{N}^*$, $\mathbf{r} \in [1, +\infty)$, with the usual modification for $\mathbf{r} = \infty$. We will also denote by $\widehat{g}$ the Fourier transform of $g \in \mathbb{L}_1(\mathbb{R}^s)$, defined by $\widehat{g}(x) = \int e^{i\langle t, x\rangle} g(x)\, dx$, where $\langle \cdot, \cdot \rangle$ 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 various 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 nonparametric estimation in the deconvolution model with pointwise and $\mathbb{L}_2$ risks was investigated by Carroll and Hall [8], Stefanski [38], Stefanski and

*E-mail: rebelles.gilles@neuf.fr

Carroll [39], Fan ([12],[13],[14]), Pensky and Vidakovic [34], Butucea [4], Hall and Meister [20], Meister [32], Butucea and Tsybakov ([6],[7]), Butucea and Comte [5]. Global density deconvolution was also considered under a weighted $\mathbb{L}_p$-norm (defined with an integrable weight function) by Fan [13]-[14] and under the sup-norm loss by Stefanski [38], Bissantz, Dümbgen, Holzmann and Munk [3] and Lounici and Nickl [28]. Whereas all the works cited above are in the one-dimensional setting, the problem of deconvolving a multidimensional density under pointwise or $\mathbb{L}_2$ loss has been addressed by Masry ([29], [30]), Youndjé and Wells [41], and Comte and Lacour [9].

Our aim here is twofold. First, we deal with optimal deconvolution of a multivariate density under $\mathbb{L}_p$ and sup-norm losses. Next, as in Lepski [25] (under sup-norm loss) and in Rebelles [36] (under $\mathbb{L}_p$-losses) for the density model, we also take advantage of the fact that some coordinates of the X_k's may be independent from the others, but in a unified way. To analyze the accuracy of our estimation procedures, we use the minimax criterion.

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

$$\mathcal{R}_p[\widetilde{f},\Sigma] := \sup_{f\in\Sigma}\mathcal{R}_p[\widetilde{f},f].$$

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

$$\mathcal{R}_p[\widetilde{f}_*,\Sigma] \asymp \inf_{\widetilde{f}_n}\mathcal{R}_p[\widetilde{f},\Sigma] =: \varphi_{n,p}(\Sigma), \quad n\to+\infty,$$

where the infimum is taken over all possible estimators. Such an estimator is called minimax on Σ.

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 2.2. Whereas the vector $\beta=(\beta_1,\dots,\beta_d)$ represents the smoothness of the target density, $r=(r_1,\dots,r_d)$ represents the index of homogeneity. When p is finite we assume that both the smoothness of f and the accuracy of estimation are measured in the same $\mathbb{L}_p$-norm, which means $r_j=p$ for $j=1,\dots,d$. This restriction permits us to use a global selection procedure from a family of linear estimators that leads to optimal accuracy in the minimax sense. Otherwise, this is not possible. In the latter case, the vector r is replaced by p in the notation of the functional class. If $\beta_j=\beta_0$, $L_j=L_0$ and $r_j=r_0$ for all $j=1,\dots,d$, any function belonging to $N_{r,d}(\beta,L)$ is called isotropic function. Let us briefly present some interesting results and the novelties that we propose in the context described above.

In Comte and Lacour [9] it was shown that

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

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

$$\mathbf{A}_1\prod_{j=1}^{d}(1+t_j^2)^{-\frac{\lambda_j}{2}} \le |\widehat{q}(t)| \le \mathbf{A}_2\prod_{j=1}^{d}(1+t_j^2)^{-\frac{\lambda_j}{2}}, \quad \forall t\in\mathbb{R}^d,$$

for some constants $\mathbf{A}_1,\mathbf{A}_2,\lambda_j>0$, $j=1,\dots,d$. Such a density is usually called ordinary smooth of order $\lambda=(\lambda_1,\dots,\lambda_d)$.

Note that this result was proved in the one-dimensional setting by Fan [13]. However, whereas Fan [13] provided an estimator whose construction depends on the smoothness parameter β of the functional class $N_{2,1}(\beta,L)$ (which is not known in practice), Comte and Lacour [9] proposed an adaptive strategy. Indeed, they provided a single estimator which is fully data driven and minimax on each class $N_{2,d}(\beta,L)$, whatever the nuisance parameter (β,L) in a large range. Such an estimator is called optimal adaptive over the scale of functional classes $\{N_{2,d}(\beta,L)\}_{(\beta,L)}$.

Lounici and Nickl [28] considered the problem of adaptive deconvolution of a univariate density under sup-norm loss and proved that

$$\varphi_{n,\infty}(N_{\infty,1}(\beta,L)) \asymp \left(\frac{n}{\log(n)}\right)^{-\frac{\tau}{2\tau+1}}, \quad \tau := \left[\frac{2\lambda+1}{\beta}\right]^{-1}, \tag{3}$$

when the common density q of the errors is ordinary smooth of order $\lambda > 0$. Moreover, they constructed an optimal adaptive estimator over the scale of Hölder classes $\{N_{\infty,1}(\beta,L)\}_{(\beta,L)}$.

It is worth mentioning that Fan [13], Lounici and Nickl [28] and Comte and Lacour [9], as in most of the aforementioned 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 [9] showed that, in the presence of super smooth noise, the rates of convergence on anisotropic Nikolskii classes (considered as classes of ordinary smooth densities) are logarithmic and achieved by a single kernel estimator which is optimal adaptive and whose construction does not require any bandwidth selection procedure. Note that Youndjé and Wells [41] considered the problem of adaptive deconvolution of an isotropic density in the ordinary smooth case, namely the "moderately ill-posed" case in inverse problems. The results obtained in Comte and Lacour [9] under $\mathbb{L}_2$-loss generalize considerably those of Youndjé and Wells [41].

In the present paper, we deal with the problem of minimax adaptive deconvolution of an anisotropic density in the ordinary smooth case with $\mathbb{L}_p$-risks, $p \in [1,\infty]$. The rates of convergence given in (2)-(3) are recovered from the results we obtain. Indeed, we provide adaptive kernel estimators which achieve the following minimax rates of convergence respectively:

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

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

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

Here, the optimality is a direct consequence of minimax lower bounds recently obtained by Lepski and Willer [27]. As usual, 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 the sup-norm loss if $\kappa \leq 0$ and under the $\mathbb{L}_1$-loss. Therefore, we do not consider the case $p = 1$. When $p \in (1,2)$, our estimation procedure is adaptive, but does not achieve the minimax lower bound on $N_{p,d}(\beta,L)$ found by Lepski and Willer [27].

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 hypotheses on the statistical model. For instance, see the works on the composite function structure in Horowitz and Mammen [21], Juditsky et al. [22] and Baraud and Birgé [1], the works on multi-index structure in Goldenshluger and Lepski [15] and Lepski and Serdyukova [26], the works on the multiple index model in density estimation in Samarov and Tsybakov [37] and the works on anisotropic denoising in functional deconvolution model in Benhaddou, Pensky and Picard [2].

Below, we discuss one of the possibilities of dealing with this problem in the framework of density estimation. The approach which has been recently proposed in Lepski [25] 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 of interest.

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 [27] and used in this paper. In Section 3, we introduce the family of kernel estimators we consider for our procedure and then we describe the selection rule that leads to the construction of our final estimators. In Section 4, we provide some oracle inequalities and, as consequences, minimax adaptive upper bounds under $\mathbb{L}_p$-losses over scales of anisotropic Nikolskii classes. Further, we discuss the optimality of our estimators 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 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 $\{1,\dots,d\}$, except the empty set. Let $\mathfrak{P}$ be a given set of partitions of $\{1,\dots,d\}$. For all $I\in\mathcal{I}_d$ denote also $\overline{I}=\{1,\dots,d\}\setminus I$ and $|I|=\mathrm{card}(I)$. We will use $\overline{\emptyset}$ for $\{1,\dots,d\}$. Finally, for all $x\in\mathbb{R}^d$ and $I\in\mathcal{I}_d$ put $x_I:=(x_j)_{j\in I}$ and, for any probability density $g\colon\mathbb{R}^d\to\mathbb{R}_+$,

$$g_I(x_I):=\int_{\mathbb{R}^{|\overline{I}|}}g(x)\,dx_{\overline{I}}.$$

Assume that $g_{\overline{\emptyset}}\equiv g$ and that $g_{\emptyset}\equiv 1$. Note also that f_I and q_I are the marginal densities of X_I and ε_I respectively.

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

$$\mathfrak{P}(f):=\Big\{\mathcal{P}\in\mathfrak{P}\colon f(x)=\prod_{I\in\mathcal{P}}f_I(x_I),\ \forall x\in\mathbb{R}^d\Big\}. \tag{6}$$

Remark that $\mathfrak{P}(f)$ is not empty if we consider that $\overline{\emptyset}\in\mathfrak{P}$, or that $\mathfrak{P}=\{\mathcal{P}\}$ if the independence structure of f is known. The possibility of choosing $\mathfrak{P}$, instead of considering all partitions of $\{1,\dots,d\}$, is introduced for technical purposes. This is explained in more detail in Lepski [25], Section 2.1, paragraph "*Extra parameters*".

Finally, we endow the set $\mathfrak{P}$ with the operation "$\diamond$" introduced in Lepski [25]: for any $\mathcal{P},\mathcal{P}'\in\mathfrak{P}$

$$\mathcal{P}\diamond\mathcal{P}':=\big\{I\cap I'\neq\emptyset,\ I\in\mathcal{P},\ I'\in\mathcal{P}'\big\}. \tag{7}$$

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. Smoothness Assumption on the Target Density

In the literature we can find several definitions of the *anisotropic Nikolskii class of densities*. Let us recall the one we use in the present paper. Set $\{e_1,\dots,e_s\}$, the canonical basis in $\mathbb{R}^s$, $s\in\mathbb{N}^*$.

Definition 1. Set $r=(r_1,\dots,r_s)\in[1,+\infty]^s$, $\beta=(\beta_1,\dots,\beta_s)\in(0,+\infty)^s$ and $L=(L_1,\dots,L_s)\in(0,+\infty)^s$. A probability density $g\colon\mathbb{R}^s\to\mathbb{R}_+$ belongs to the anisotropic Nikolskii class $N_{r,s}(\beta,L)$ if

(i) $\|D_j^k g\|_{r_j}\le L_j,\qquad\forall k=0,\dots,\lfloor\beta_j\rfloor,\quad\forall j=1,\dots,s;$

(ii) $\|D_j^{\lfloor\beta_j\rfloor}g(\cdot+ze_j)-D_j^{\lfloor\beta_j\rfloor}g(\cdot)\|_{r_j}\le L_j|z|^{\beta_j-\lfloor\beta_j\rfloor},\qquad\forall z\in\mathbb{R},\quad\forall j=1,\dots,s.$

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

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 [25], Section 3, Definition 2. However, since the adaptation is not necessarily considered with respect to the set of all partitions of $\{1,\dots,d\}$, the condition imposed therein can be weakened. For instance, if $\mathfrak{P}=\{\overline{\emptyset}\}$ (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 $\big\{N_{r,d}(\beta,L)\big\}$, which is not possible with the classes introduced in Lepski [25]. For these reasons, the following collection $\big\{N_{r,d}(\beta,L,\mathcal{P})\big\}_{\mathcal{P}}$ was introduced in Rebelles [35], Section 3.1.

Definition 2. Let $r \in [1,+\infty]^d$ and $(\beta, L, \mathcal{P}) \in (0,+\infty)^d \times (0,+\infty)^d \times \mathfrak{P}$ be fixed. A probability density $g\colon \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $N_{r,d}(\beta, L, \mathcal{P})$ if

$$g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \quad \forall x \in \mathbb{R}^d; \qquad g_I \in N_{r_I,|I|}(\beta_I, L_I), \quad \forall I \in \mathcal{P}' \diamond \mathcal{P}'', \quad \forall (\mathcal{P}',\ \mathcal{P}'') \in \mathfrak{P} \times \mathfrak{P}. \tag{8}$$

Note that, if $\mathfrak{P} = \{\overline{\emptyset}\}$, the class $N_{r,d}(\beta, L, \overline{\emptyset})$ coincides with the classical anisotropic Nikolskii class of densities $N_{r,d}(\beta, L)$.

2.3. Noise Assumptions for Upper Bounds

Both the definition of our estimation procedure and the computation of the $\mathbb{L}_p$-risk, $p \in (1, +\infty]$, lead us to consider that the density q of the noise random vector ε_1 satisfies the following assumptions.

Assumption (N1). *Assume that, for any $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$:*

(i) *if $p = 2$, then $\|\widehat{q_I}\|_1 < +\infty$;*

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

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

(i) *if $p = 2$, then*

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

(ii) *if $p \in (1, +\infty) \setminus \{2\}$, then $\widehat{q_I}(t_I) \neq 0$, $\forall t \in \mathbb{R}^d$, $\widehat{q_I}^{\,-1} \in \mathcal{C}^{|I|}(\mathbb{R}^{|I|})$ and*

$$\Big| [D^{\alpha_I} \widehat{q_I}^{\,-1}](t_I) \prod_{j \in I} t_j^{\alpha_j} \Big| \leq \mathbf{A} \prod_{j \in I} (1 + t_j^2)^{\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d,\ \forall \alpha_I = (\alpha_j)_{j \in I} \in \mathbb{N}^{|I|},\ \sum_{j \in I} \alpha_j \leq |I|;$$

(iii) *if $p = +\infty$, then $\widehat{q_I}(t_I) \neq 0$, $\forall t \in \mathbb{R}^d$, $\widehat{q_I}^{\,-1} \in \mathcal{C}^1(\mathbb{R}^{|I|})$ and*

$$\big| [D_k^{\alpha_k} \widehat{q_I}^{\,-1}](t_I) \big| \leq \mathbf{A} \prod_{j \in I} (1 + t_j^2)^{\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d,\ \forall k \in I,\ \forall \alpha_k \in \{0, 1\}.$$

Here and in the sequel, $D_k^{\alpha_k} g$ denotes the α_kth order partial derivative of g with respect to the kth variable, $D_k^0 g \equiv g$ 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 $g\colon \mathbb{R}^s \to \mathbb{R}$.

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 holds for the centered Laplace and Gamma type distributions, whose characteristic functions have polynomial decay. As mentioned in Comte and Lacour [9], the latter case is of great interest in particular physical contexts; see, for instance, the study of the pile-up model in Comte and Rebafka [10].

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

2.4. Noise Assumptions for Minimax Lower Bounds

Recently, Lepski and Willer [27] 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 ε satisfies the following assumption.

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

$$\text{(i)} \quad |\widehat{q}(t)| \leq \mathbf{B} \prod_{j=1}^{d} (1 + t_j^2)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d;$$

$$\text{(ii)} \quad \|\widehat{q}^{\,-1} D^\alpha \widehat{q}\|_\infty \leq \mathbf{B}, \quad \forall \alpha = (\alpha_1, \ldots, \alpha_d) \in \{0,1\}^d, \quad \alpha_1 + \ldots + \alpha_d \geq 1.$$

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

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

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 condition (ii) of Assumption (N3) (or Condition 1 in Lounici and Nickl [28] for the one-dimensional setting), implies that condition (iii) of Assumption (N2) is satisfied.

In the setting of the deconvolution density model, Lepski and Willer [27] provided minimax lower bounds on $N_{r,d}(\beta, L, M) = N_{r,d}(\beta, L) \cap \{f \colon \|f\|_\infty \leq M\}$ in four different zones described in terms of parameters p, r, β and λ, namely the *tail zone*, the *dense zone* and the *sparse zone* which is divided in two zones.

Since $N_{r,d}(\beta, L, M) \subset N_{r,d}(\beta, L)$, the results below follow from Theorems 2 and 3 in Lepski and Willer [27] and allow us to assert the optimality of our estimators when $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure) in some particular cases.

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

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

where the 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 Assumption* (N3) *is satisfied. Then,*

(i) *there is no uniformly consistent estimator over $N_{r,d}(\beta, L)$ if $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} \leq 0$;*

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

$$\liminf_{n \to +\infty} \inf_{\widetilde{f}_n} \left\{ \varphi_{n,\infty}^{-1}(N_{r,d}(\beta, L)) \mathcal{R}_\infty[\widetilde{f}_n, N_{r,d}(\beta, L)] \right\} > 0,$$

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

The settings of Theorems 1 and 2 correspond to particular cases of the dense zone and the sparse zone respectively. Further, when the problem of minimax estimation under $\mathbb{L}_p$-loss on the class $N_{p,d}(\beta, L)$ is considered and $p \in (1, 2)$, this corresponds to a particular case of the tail zone.

3. ESTIMATION PROCEDURE

In this section, we construct an estimator following a scheme of selection rule introduced in Lepski [25] to take into account the possible independence structure of the underlying density. If $\mathfrak{P} = \{\overline{\emptyset}\}$, this scheme coincides with a version of the methodology proposed by Goldenshluger and Lepski [17]. This methodology, employed in many areas of nonparametric statistics, has been recently used by Comte and Lacour [9] in the framework of the deconvolution model.

3.1. Kernel-Type Estimators

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

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

Therefore, in view of the definition of the kernel $\mathbf{K}$ and Assumption (N2) on the errors, one can define the *kernel-type estimator*

$$\widetilde{f}_{h_I}(x_I) := n^{-1} \sum_{k=1}^{n} L_{(h_I)}(x_I - Y_{k,I}), \quad L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i\langle t_I, x_I\rangle} \frac{\widehat{K_{h_I}}(t_I)}{\widehat{q_I}(t_I)}\, dt_I. \tag{9}$$

The ideas that led to the introduction of the estimators $\widetilde{f}_{h_I}$ are explained in Fan [13] in the one-dimensional setting and in Comte and Lacour [9] in the multivariate context.

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

$$\mathfrak{F}[\,\mathfrak{P}\,] := \left\{ \widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I),\ x \in \mathbb{R}^d,\ (h,\mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,] \right\}, \tag{10}$$

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

For $I \in \mathcal{I}_d$, consider first the set of multibandwidths

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

$$h_{\min}^{(p)} := \begin{cases} n^{-(1\vee \frac{p}{|I|})}, & p \in (1,+\infty), \\ n^{-1}, & p = +\infty, \end{cases} \qquad h_{\max}^{(p)} := \begin{cases} [\log(n)]^{-\frac{p}{|I|}}, & p \in (1,+\infty), \\ 1, & p = +\infty. \end{cases}$$

Then define

$$\mathcal{H}_{p,I} := \left\{ h_I \in \mathfrak{H}_{p,I}\colon (nV_{h_I})^{\frac{1}{2}\wedge(1-\frac{1}{p})} \prod_{j\in I} h_j^{\lambda_j} \geq c_p \mathbf{1}_{\{p<\infty\}} + \sqrt{\log(n)}\,\mathbf{1}_{\{p=+\infty\}} \right\}, \tag{11}$$

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

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

Put finally

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

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

Auxiliary estimators. We mimic the procedure of Lepski [25] by introducing the following auxiliary estimators. Consider first the classical kernel auxiliary estimators

$$\widetilde{f}_{h_I,\eta_I}(x_I) := K_{\eta_I} \star \widetilde{f}_{h_I}(x_I), \qquad h,\eta \in (0,1]^d, \quad I \in \mathcal{I}_d,$$

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

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

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

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

The ideas that led to the introduction of the estimators $\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$, based on both the operations "$\star$" and "$\diamond$", are explained in Lepski [25], Section 2.1, paragraph "*Estimation construction*". Note that the arguments given in that paper do not depend on the norm used in the definition of the risk and remain valid for estimation under $\mathbb{L}_p$-loss.

3.2. Selection Rule

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

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

Put also $\Lambda_p := \mathfrak{d}\gamma_p[\,\overline{G}_p]^{\mathfrak{d}(\mathfrak{d}-1)}$, where $\mathfrak{d} := \sup_{\mathcal{P}\in\mathfrak{P}}|\mathcal{P}|$,

$$\overline{G}_p := 1 \vee \Big[\ \|\mathbf{K}\|_1^d \sup_{(h,\mathcal{P})\in\,\mathcal{H}_p[\,\mathfrak{P}\,]}\ \sup_{\mathcal{P}'\in\,\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I}\|_p\Big]$$

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

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

$$\widetilde{\Delta}_p(h,\mathcal{P}) := \sup_{(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]} \Big[\ \|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\|_p - \Lambda_p\mathcal{U}_p(\eta,\mathcal{P}')\Big]_+. \tag{12}$$

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

$$\widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \Lambda_p\mathcal{U}_p(\widetilde{h},\widetilde{\mathcal{P}}) = \inf_{(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]}\big[\widetilde{\Delta}_p(h,\mathcal{P}) + \Lambda_p\mathcal{U}_p(h,\mathcal{P})\big]. \tag{13}$$

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

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 [19].

Proposition 1. *Assume that Assumption* (N2) *is satisfied.*

(i) *If* $p \in (1,\infty) \setminus \{2\}$, *for any* $\mathbf{r} \in (1,2)$ *and any* $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$, *there exists a constant* $C_{\mathbf{r},I} := C_{\mathbf{r},I}(|I|,\mathbf{K},q) > 0$ *such that*

$$\|L_{(h_I)}\|_{\mathbf{r}} \le C_{\mathbf{r},I}(V_{h_I})^{-(1-1/\mathbf{r})} \prod_{j\in I} h_j^{-\lambda_j}, \quad \forall h \in (0,1]^d.$$

(ii) *For any* $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$, *there exists a constant* $C_I := C_I(|I|,\mathbf{K},q) > 0$ *such that*

$$\|L_{(h_I)}\|_2 \le C_I \prod_{j\in I} h_j^{-\lambda_j - \frac{1}{2}}, \qquad \|L_{(h_I)}\|_\infty \le C_I \prod_{j\in I} h_j^{-\lambda_j - 1}, \quad \forall h \in (0,1]^d.$$

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 $\mathcal{U}_p(h_I)$ since a dimensional constant is not explicitly done in Theorem 5.2.4 of Grafakos [19].

Next, we also emphasize that the quantity $\mathcal{U}_p(h_I)$ 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 $\widetilde{f}_{h_I}$. 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

$$\xi_{h_I}(x_I) := \widetilde{f}_{h_I}(x_I) - \mathbb{E}\{\widetilde{f}_{h_I}(x_I)\}.$$

Proposition 2. *Assume that Assumptions* (N1)–(N2) *hold. Let* $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$, *be arbitrary fixed. If* $p \in (1,+\infty]$, $\mathbf{r} \ge 1$ *and* $n \ge 3$ *then*

$$\Big\{\mathbb{E} \sup_{h_I \in \mathcal{H}_{p,I}} \big[\, \|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\, \big]_+^{\mathbf{r}} \Big\}^{\frac{1}{\mathbf{r}}} \le c_p(\mathbf{r}) n^{-\frac{1}{2}}, \quad c_p(\mathbf{r}) > 0. \tag{14}$$

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

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

4. MAIN RESULTS

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

4.1. 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:

$$f \in \mathbf{F}_p[\,\mathfrak{P}\,] := \Big\{ g \in \mathbf{F} : \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \|g_I\|_p < \infty \Big\}, \tag{15}$$

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

$$\mathbf{F}_p[\,\{\overline{\emptyset}\}\,] = \big\{ g \in \mathbf{F} \colon \|g\|_p < \infty \big\}, \qquad \mathbf{F}_p[\,\{\mathcal{P}\}\,] = \Big\{ g \in \mathbf{F} \colon \sup_{I \in \mathcal{P}} \|g_I\|_p < \infty \Big\}.$$

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

$$\mathcal{R}_p[(h,\mathcal{P}),f] := \Big(\mathbb{E}_f \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I} - f_I\|_p^p \Big)^{\frac{1}{p}}, \quad p \in (1,+\infty),$$

$$\mathcal{R}_\infty[(h,\mathcal{P}),f] := \mathbb{E}_f \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I} - f_I\|_\infty.$$

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

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

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

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

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

Theorem 3. *Suppose that Assumptions* (N1)–(N2) *are satisfied.*

If $n \geq 3$ *and* $p \in (1, +\infty]$ *then:* $\forall f \in \mathbf{F}_p\,[\mathfrak{P}]$,

$$\mathcal{R}_p[\widetilde{f}, f] \leq \mathbf{C}_{p,1}(\mathbf{f}_p) \inf_{(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]\inf\mathcal{P}\in\mathfrak{P}(f)} \big\{\mathcal{R}_p[(h,\mathcal{P}), f] + \gamma_p\mathcal{U}_p(h,\mathcal{P})\big\} + \mathbf{C}_{p,2}(\mathbf{f}_p)n^{-\frac{1}{2}}, \qquad (16)$$

where $\mathbf{f}_p := 1 \vee [\sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|f_I\|_p]$.

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

If $\mathfrak{P} = \{\overline{\emptyset}\}$, we obtain automatically some oracle inequalities for estimation on $\mathbb{R}^d$ under $\mathbb{L}_p$-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. *Suppose that* $\mathfrak{P} = \{\overline{\emptyset}\}$ *and that Assumptions* (N1)–(N2) *are satisfied.*

If $n \geq 3$ *and* $p \in (1, +\infty]$ *then:* $\forall f \in \mathbf{F}$,

$$\mathcal{R}_p[\widetilde{f}, f] \leq \inf_{h\in\mathcal{H}_{p,\overline{\emptyset}}} \big\{(1 + 2\|\boldsymbol{K}\|_1^d)\mathcal{R}_p[\widetilde{f}_h, f] + 2\gamma_p\mathcal{U}_p(h)\big\} + 2\mathbf{C}_p n^{-\frac{1}{2}}. \qquad (17)$$

The explicit expression of the absolute constant $\mathbf{C}_p = \mathbf{C}_p(d, \mathfrak{P}, \mathbf{K}, q) > 0$ is given in the proof of the theorem.

Note first that the statement of Theorem 4 holds for all probability densities $f \in \mathbf{F}$, which is not true for Theorem 3. Next, the constant $1 + 2\|\mathbf{K}\|_1^d$ is more suitable than $\mathbf{C}_{p,1}(\mathbf{f}_p)$. 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 $\mathcal{P} \neq \overline{\emptyset}$ and the smoothness parameter h is fixed and properly chosen, then our procedure should select the true partition $\mathcal{P}$ and the estimator $\widetilde{f}_{(h,\mathcal{P})}$ should provide a better accuracy of estimation than the classical kernel-type estimator $\widetilde{f}_h$. This was illustrated by a short simulation study in Rebelles [36] for the density model (with direct observations), under the $\mathbb{L}_2$-loss.

4.2. $\mathbb{L}_p$-Adaptive Minimax Estimation

In what follows, we illustrate the application of Theorems 3 and 4 to adaptive estimation over anisotropic Nikolskii classes of densities $N_{r,d}(\beta, L, \mathcal{P})$ and $N_{r,d}(\beta, L)$, respectively. To compute the $\mathbb{L}_p$-risk of a kernel-type estimator, we first compute its bias. Thus we need to enforce the assumptions imposed on the kernel $\mathbf{K}$. One of the possibilities is the following, proposed in Kerkyacharian, Lepski and Picard [23].

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

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

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

The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [23]. In particular, it has been shown that

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

4.2.1. Minimax adaptive estimation under an $\mathbb{L}_p$-loss

For $(\beta, \mathcal{P}) \in (0, +\infty)^d \times \mathfrak{P}$ define $\phi_{n,p}(\beta, \mathcal{P}) := n^{-\frac{[\frac{1}{2} \wedge (1-\frac{1}{p})]\tau}{\tau + [\frac{1}{2} \wedge (1-\frac{1}{p})]}}$, where

$$\tau := \tau(\beta, \mathcal{P}) = \inf_{I \in \mathcal{P}} \tau_I, \qquad \tau_I := \left[\sum_{j \in I} \frac{[\frac{1}{2} \wedge (1-\frac{1}{p})]^{-1} \lambda_j + 1}{\beta_j}\right]^{-1}. \tag{20}$$

Assume that $\overline{\emptyset} \in \mathfrak{P}$ and consider the estimator $\widetilde{f}$ 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 $(\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathfrak{P}$ one has*

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

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

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

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

To the best of our knowledge, these results are new. Below, we briefly discuss several consequences of Theorems 5 and 6.

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 $\phi_{n,p}(\beta, \overline{\emptyset})$ 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 $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure), the quality of estimation of our estimator $\widetilde{f}$ is optimal, up to a numerical constant, on each class $N_{p,d}(\beta, L)$, whatever the nuisance parameter (β, L). Thus, in the aforementioned case, $\widetilde{f}$ is an optimal adaptive estimator over the scale $\{N_{p,d}(\beta, L)\}_{(\beta,L)}$.

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 [27] under the $\mathbb{L}_p$-loss. We conclude that either our estimator is not minimax on $N_{p,d}(\beta, L)$ or the lower bound in Lepski and Willer [27] is not the minimax rate of convergence on the latter functional class.

Further, our results show that $\mathbb{L}_p$-estimation of an anisotropic density in the deconvolution model does not necessarily require that the target function is uniformly bounded, as is assumed in all the works concerning the density model (with direct observations); see, e.g., Goldenshluger and Lepski [17]. See also the discussion in Lepski and Willer [27] concerning the deconvolution model, Section 3, paragraph "*Deconvolution density model. Bounded case.*".

It is also important to emphasize that both Theorems 5 and 6 allow us to analyze the influence of the independence structure on the accuracy of estimation under an $\mathbb{L}_p$-loss in the deconvolution model. Indeed, we see that

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

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 $\mathcal{P} \neq \overline{\emptyset}$. For instance, if $p \in [2, \infty)$, $\beta = (\boldsymbol{\beta}, \dots, \boldsymbol{\beta})$ and $\boldsymbol{\mathcal{P}} = \{\{1\}, \dots, \{d\}\}$, then

$$n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+2(\sum_{j=1}^d \lambda_j)+d}} = \phi_{n,p}(\beta, \overline{\emptyset}) \gg \phi_{n,p}(\beta, \boldsymbol{\mathcal{P}}) = n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+2\lambda_{\max}+1}}, \quad \lambda_{\max} := \max_{j=1,\dots,d} \lambda_j, \tag{21}$$

and $\phi_{n,p}(\beta, \boldsymbol{\mathcal{P}})$ is the minimax rate of convergence in the one-dimensional setting.

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 $\lambda_j = 0$, $j = 1, \dots, d$), it is asserted in Rebelles [36] 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 [18], up to minor modifications to take into account the independence structure. Specifically, in the proof of the lower bound for any given partition $\mathcal{P}$ one can perturb only the density of the group of variables corresponding to the index set $I \in \mathcal{P}$ with minimal value of $\overline{\beta}_I = [\sum_{j \in I} 1/\beta_j]^{-1}$. 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 [27].

4.2.2. Minimax adaptive estimation under sup-norm loss

For $(\beta, r, \mathcal{P}) \in (0, +\infty)^d \times [1, +\infty]^d \times \mathfrak{P}$ define $\phi_{n,\infty}(\beta, r, \mathcal{P}) := (n/\log(n))^{-\frac{\Upsilon}{2\Upsilon+1}}$, where

$$\Upsilon := \Upsilon(\beta, r, \mathcal{P}) = \inf_{I \in \mathcal{P}} \Upsilon_I, \qquad \Upsilon_I := \left(\tau_I^{-1} + [\omega_I \kappa_I]^{-1}\right)^{-1},$$

$$\tau_I := \left[\sum_{j \in I} \frac{2\lambda_j + 1}{\beta_j}\right]^{-1}, \quad \omega_I := \left[\sum_{j \in I} \frac{2\lambda_j + 1}{\beta_j r_j}\right]^{-1}, \quad \kappa_I := \frac{1 - \sum_{j \in I} \frac{1}{\beta_j r_j}}{\sum_{j \in I} \frac{1}{\beta_j}}. \tag{22}$$

Assume that $\overline{\emptyset} \in \mathfrak{P}$ and consider the estimator $\widetilde{f}$ defined by the selection rule (12)–(13) with $p = +\infty$. As before, we obtain the following two theorems:

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

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

Theorem 8. *Suppose that* $\mathfrak{P} = \{\overline{\emptyset}\}$ *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* $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$ *one has*

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

To the best of our knowledge, these results are also new. As before, we briefly discuss several consequences of Theorems 7 and 8.

If $\mathfrak{P} = \{\overline{\emptyset}\}$ and $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$, it follows from Theorems 2 and 8 that, in the presence of the noise satisfying Assumptions (N1)–(N3), $\phi_{n,\infty}(\beta, r, \overline{\emptyset})$ 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

$$\Big\{ N_{r,d}(\beta, L),\ (\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d,\ 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0 \Big\}. \tag{23}$$

Thus our results generalize considerably those of Lounici and Nickl [28] when the target density has Hölder-type regularity and the noise is ordinary-smooth.

It is worth to note that our method of estimation can be used for pointwise estimation. Moreover, it follows from Theorem 8 that our estimator achieves the adaptive rates of convergence found in Comte and Lacour [9] with a pointwise criterion over the scale of anisotropic Hölder classes $\{N_{\infty,d}(\beta, L)\}_{(\beta,L)}$. Thus, in the case of ordinary smooth density and ordinary smooth noise, we extend their results to the scale of anisotropic Nikolskii classes given in (23). Note that the logarithmic term in the rates $\phi_{n,\infty}(\beta, r, \mathcal{P})$ is known to be an "optimal payment" for adaptation to the regularity of the target density in the pointwise setting; see, e.g., Butucea and Comte [5].

As before (under $\mathbb{L}_p$-loss), Theorems 7 and 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 $\mathcal{P} \neq \overline{\emptyset}$. In particular, our method of estimation outperforms that of Comte and Lacour [9] in the pointwise setting when both the estimated density and the noise are ordinary smooth. Another interesting fact related to the consideration of the eventual independence structure of the target density f is the following. Suppose that f belongs to the functional class $N_{r,d}(\beta, L, \mathcal{P})$ satisfying $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} \leq 0$ and $1 - \sum_{j \in I}^{d} \frac{1}{\beta_j r_j} > 0$, $\forall I \in \mathcal{P}$, and that $\mathfrak{P} = \{\mathcal{P}\}$. Scrutinizing the proof of Theorem 7, one can see that it is possible to construct a kernel estimator that achieves the rate $\phi_{n,\infty}(\beta, r, \mathcal{P})$, 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 [27] to take into account the possible independence structure of the underlying density.

5. PROOFS OF THE MAIN RESULTS

5.1. Quantities and Technical Lemma

For brevity, introduce first

$$\mathcal{I}_d^{\diamond} := \{I \in \mathcal{P} \diamond \mathcal{P}',\ (\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}\}, \qquad \overline{\mathcal{U}_p} := \sup_{n \in \mathbb{N}^*} \sup_{I \in \mathcal{I}_d^{\diamond}} \sup_{h_I \in \mathcal{H}_{p,I}} \mathcal{U}_p(h_I) < \infty.$$

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

Next, define the constant γ_p involved in the selection rule. For $I \in \mathcal{I}_d^\diamond$ and $\mathbf{r} \geq 1$, put

$$\gamma_{p,I}(\mathbf{r}) := \begin{cases} 4 + \sqrt{\frac{37e^{-1}p\mathbf{r}}{2-p}}, & p \in (1,2), \\ \big(7C_I + 3\mathbf{A}(2\pi)^{-\frac{|I|}{2}}\|\widehat{K_I}g_I\|_\infty \|\widehat{q_I}\|_{\frac{1}{2}}\big)\mathbf{r}, & p = 2, \\ \big(\frac{46c(p)[p\vee e]}{3e}\big)c_p^{\frac{1}{p}-\frac{1}{2}}[1 \vee C_I](1 \vee \|q_I\|_\infty)^{\frac{3}{4}}\mathbf{r}, & p \in (2,+\infty), \\ 6C_I(\mathbf{K},q)(1 \vee \|q_I\|_\infty)^{\frac{1}{2}}\big[93|I|\log(|I|) + 69\mathbf{r}\big], & p = +\infty, \end{cases}$$

where $g_I(t_I) := \prod_{j\in I}(1+t_j^2)^{\frac{\lambda_j}{2}}$, $C_I := \mathbf{A}\big\{(2\pi)^{-\frac{|I|}{2}}(\|\widehat{K_I}g_I\|_2 \vee \|\widehat{K_I}g_I\|_1)\big\}$, c_p is given in the definition of $\mathcal{H}_{p,I}$, $c(p) := 15p/\log(p)$ and

$$C_I(\mathbf{K},q) := \frac{\mathbf{A}}{(2\pi)^{\frac{|I|}{2}}}\Big\{\|\widehat{K_I}g_I\|_2 \vee \|\widehat{K_I}g_I\|_1 \vee (\max_{j\in I}\|D_j^1\widehat{K_I}g_I\|_1) \vee \|\widehat{K_I}\varphi_I\|_2 \vee \|\widehat{K_I}\varphi_I\|_1\Big\},$$

with $\varphi_I(t_I) := \sup_{j\in I}|t_j|g_I(t_I)$. Then, put $\mathbf{r}_k := kp\mathbf{1}_{\{p<\infty\}} + k\mathbf{1}_{\{p=+\infty\}}$, $k \geq 1$, and

$$\gamma_p := \begin{cases} \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\{\gamma_{p,I}(\mathbf{r}_4)\}, & \mathfrak{P} \neq \{\overline{\emptyset}\}, \\ \gamma_{p,\overline{\emptyset}}(\mathbf{r}_1), & \mathfrak{P} = \{\overline{\emptyset}\}. \end{cases}$$

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

$$\xi_p := \sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}[\,\|\xi_{h_I}\|_p - \gamma_p\mathcal{U}_p(h_I)]_+,$$

$$\overline{\mathbf{f}}_p := \mathfrak{d}^2\|\mathsf{K}\|_1^d[\overline{G}_p]^{\mathfrak{d}(\mathfrak{d}-1)}\big(\max\{\overline{G}_p, \|\mathsf{K}\|_1^d\mathbf{f}_p\}\big)^{\mathfrak{d}-1}, \quad \mathbf{f}_p := 1 \vee \Big[\sup_{I\in\mathcal{I}_d^\diamond}\|f_I\|_p\Big].$$

Lemma 1. *Assume that* $\mathfrak{P} \neq \{\overline{\emptyset}\}$. *Set* $\mathbf{r} \in \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4\}$. *Under Assumptions* (N1)–(N2)*, if* $p \in (1,+\infty]$ *then, for all integer* $n \geq 3$*,*

$$(\mathbb{E}_f|\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_{p,1}(\mathbf{r})n^{-\frac{1}{2}}, \quad \big(\mathbb{E}_f|\overline{\mathbf{f}}_p|^{\mathbf{r}}\big)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_{p,2}(\mathbf{r},\mathbf{f}_p), \quad \forall f \in \mathbf{F}_p[\,\mathfrak{P}\,].$$

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

5.2. Oracle Inequalities: Proof of Theorems 3 and 4

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

In view of the triangle inequality we have

$$\begin{aligned} \|\widetilde{f} - f\|_p &\leq \|\widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}\|_p + |\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P})}\|_p + \|\widetilde{f}_{(h,\mathcal{P})} - f\|_p \\ &\leq \widetilde{\Delta}_p(h,\mathcal{P}) + \Lambda_p\mathcal{U}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \Lambda_p\mathcal{U}_p(h,\mathcal{P}) + \|\widetilde{f}_{(h,\mathcal{P})} - f\|_p. \end{aligned}$$

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

$$\|\widetilde{f} - f\|_p \leq 2[\widetilde{\Delta}_p(h,\mathcal{P}) + \Lambda_p\mathcal{U}_p(h,\mathcal{P})] + \|\widetilde{f}_{(h,\mathcal{P})} - f\|_p. \tag{24}$$

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

$$
\begin{aligned}
|\widetilde{f}_{(h,\mathcal{P})}(x)-f(x)| &= \Big|\prod_{I\in\mathcal{P}}\widetilde{f}_{h_I}(x_I)-\prod_{I\in\mathcal{P}}f_I(x_I)\Big| \\
&\le \sum_{j=1}^{m}|\widetilde{f}_{h_{I_j}}(x_{I_j})-f_{I_j}(x_{I_j})|\Big(\prod_{k=j+1,\dots,m}|\widetilde{f}_{h_{I_k}}(x_{I_k})|\Big)\Big(\prod_{l=1,\dots,j-1}|f_{I_l}(x_{I_l})|\Big).
\end{aligned}
$$

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

$$
\prod_{j=1}^{m}a_j-\prod_{j=1}^{m}b_j=\sum_{j=1}^{m}(a_j-b_j)\Big(\prod_{k=j+1,\dots,m}a_k\Big)\Big(\prod_{l=1,\dots,j-1}b_l\Big), \tag{25}
$$

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

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

$$
\begin{aligned}
\|\widetilde{f}_{(h,\mathcal{P})}-f\|_p &\le \sum_{j=1}^{m}\|\widetilde{f}_{h_{I_j}}-f_{I_j}\|_p\Big(\prod_{k=j+1,\dots,m}\|\widetilde{f}_{h_{I_k}}\|_p\Big)\Big(\prod_{l=1,\dots,j-1}\|f_{I_l}\|_p\Big) \\
&\le m\big(\max\{\overline{G}_p,\mathbf{f}_p\}\big)^{m-1}\sup_{I\in\mathcal{P}}\|\widetilde{f}_{h_I}-f_I\|_p,
\end{aligned}
$$

since $\|\mathbf{K}\|_1\ge\int\mathbf{K}=1$. Recall that $\mathfrak{d}=\sup_{\mathcal{P}\in\mathfrak{P}}|\mathcal{P}|$ and $\overline{G}_p\ge 1$. It follows

$$
\|\widetilde{f}_{(h,\mathcal{P})}-f\|_p\le\mathfrak{d}\big(\max\{\overline{G}_p,\mathbf{f}_p\}\big)^{\mathfrak{d}-1}\sup_{I\in\mathcal{P}}\|\widetilde{f}_{h_I}-f_I\|_p. \tag{26}
$$

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

$$
|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x)-\widetilde{f}_{(\eta,\mathcal{P}')}(x)|=\Big|\prod_{I'\in\mathcal{P}'}\prod_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}K_{\eta_{I\cap I'_j}}\star\widetilde{f}_{h_{I\cap I'_j}}(x_{I\cap I'})-\prod_{I'\in\mathcal{P}'}\widetilde{f}_{\eta_{I'}}(x_{I'})\Big|.
$$

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

$$
\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}-\widetilde{f}_{(\eta,\mathcal{P}')}\|_p\le\mathfrak{d}[\,\overline{G}_p\,]^{\mathfrak{d}(\mathfrak{d}-1)}\sup_{I'\in\mathcal{P}'}\Big\|\prod_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}\widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\widetilde{f}_{\eta_{I'}}\Big\|_p. \tag{27}
$$

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

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

$$
\|\mathbb{E}_f\{\widetilde{f}_{\eta_I}(\cdot)\}\|_p=\|K_{\eta_I}\star f_I\|_p\le\|K_I\|_1\|f_I\|_p\le\|\mathbf{K}\|_1^d\mathbf{f}_p. \tag{28}
$$

Then, by the same method as the one used in step 2 and (28), for any $(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]$ and any $I'\in\mathcal{P}'$ we get

$$
\begin{aligned}
&\Big\|\prod_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}\widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\prod_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}\mathbb{E}_f\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\}\Big\|_p \\
&\le\mathfrak{d}\big(\max\{\overline{G}_p,\|\mathbf{K}\|_1^d\mathbf{f}_p\}\big)^{\mathfrak{d}-1}\sup_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}\big\|K_{\eta_{I\cap I'}}\star(\widetilde{f}_{h_{I\cap I'}}-f_{I\cap I'})\big\|_p \\
&\le\mathfrak{d}\|\mathbf{K}\|_1^d\big(\max\{\overline{G}_p,\|\mathbf{K}\|_1^d\mathbf{f}_p\}\big)^{\mathfrak{d}-1}\sup_{I\in\mathcal{P}:\,I\cap I'\neq\emptyset}\|\widetilde{f}_{h_{I\cap I'}}-f_{I\cap I'}\|_p.
\end{aligned} \tag{29}
$$

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

$$\mathbb{E}_f\{\widetilde{f}_{\eta_{I'}}(x_{I'})\} = \int K_{\eta_{I'}}(y_{I'} - x_{I'}) \prod_{I \in \mathcal{P}:\, I\cap I' \neq \emptyset} f_{I\cap I'}(y_{I\cap I'})\, dy_{I'} = \prod_{I \in \mathcal{P}:\, I\cap I' \neq \emptyset} \mathbb{E}_f\{\widetilde{f}_{\eta_{I\cap I'}}(x_{I\cap I'})\}.$$

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

Thus, in view of the triangle inequality, (27), (29) and the trivial inequality $[\sup_i x_i - \sup_i y_i]_+ \le \sup_i [x_i - y_i]_+$, for any $(\eta, \mathcal{P}') \in \mathcal{H}_p[\,\mathfrak{P}\,]$, we get

$$\begin{aligned}
&\big[\, \|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\|_p - \Lambda_p \mathcal{U}_p(\eta, \mathcal{P}')\big]_+ \\
&\qquad \le \mathfrak{d}[\overline{G}_p]^{\mathfrak{d}(\mathfrak{d}-1)} \sup_{I' \in \mathcal{P}'} \Bigg[\Big\| \prod_{I\in\mathcal{P}:\, I\cap I' \neq\emptyset} \widetilde{f}_{h_{I\cap I'}, \eta_{I\cap I'}} - \prod_{I\in\mathcal{P}:\, I\cap I'\neq\emptyset} \mathbb{E}_f\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\}\Big\|_p \\
&\qquad\qquad + \|\xi_{\eta_{I'}}\|_p - \gamma_p \mathcal{U}_p(\eta_{I'})\Bigg]_+ ;
\end{aligned}$$

$$\big[\, \|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\|_p - \Lambda_p \mathcal{U}_p(\eta, \mathcal{P}')\big]_+ \le \overline{\mathbf{f}}_p \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in \mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I} - f_I\|_p + \overline{\mathbf{f}}_p \xi_p,$$

since $\overline{\mathbf{f}}_p \ge \mathfrak{d}[\overline{G}_p]^{\mathfrak{d}(\mathfrak{d}-1)} \ge 1$. We deduce

$$\widetilde{\Delta}_p(h,\mathcal{P}) \le \overline{\mathbf{f}}_p \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in \mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I} - f_I\|_p + \overline{\mathbf{f}}_p \xi_p. \tag{30}$$

Finally, it follows from (24), (26) and (30)

$$\|\widetilde{f} - f\|_p \le 3\overline{\mathbf{f}}_p \big\{ \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in \mathcal{P}\diamond\mathcal{P}'} \|\widetilde{f}_{h_I} - f_I\|_p + \gamma_p \mathcal{U}_p(h,\mathcal{P}) + \xi_p \big\}. \tag{31}$$

(6) Consider the random event $B_p := \big\{\overline{G}_p \ge \mathcal{C}_p(\mathbf{f}_p)\big\}$, $\mathcal{C}_p(\mathbf{f}_p) = \big(1 + \gamma_p\overline{\mathcal{U}_p} + \|\mathbf{K}\|_1^d \mathbf{f}_p\big)\|\mathbf{K}\|_1^d + 1$. Put also

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

In view of (28), Lemma 1, Markov's inequality, (31), and the Cauchy–Schwarz inequality we get $B_p \subseteq \big\{\xi_p \ge 1\big\}$, $\big[\mathbb{P}_f(B_p)\big]^{\frac{1}{\mathbf{r}_4}} \le \mathbf{c}_{p,1}(\mathbf{r}_4) n^{-1/2}$ and

$$\begin{aligned}
(\mathbb{E}_f \|\widetilde{f} - f\|_p^{\mathbf{r}_1} 1_{B_p^c})^{\frac{1}{\mathbf{r}_1}} &\le 3\mathfrak{d}^2 \|\mathbf{K}\|_1^d [\mathcal{C}_p(\mathbf{f}_p)]^{\mathfrak{d}^2-1} \left(\mathcal{R}_p^{(\mathbf{r}_1)}[(h,\mathcal{P}), f] + \gamma_p \mathcal{U}_p(h,\mathcal{P}) + \frac{\mathbf{c}_{p,1}(\mathbf{r}_1)}{\sqrt{n}} \right), \\
(\mathbb{E}_f \|\widetilde{f} - f\|_p^{\mathbf{r}_1} 1_{B_p})^{\frac{1}{\mathbf{r}_1}} &\le 3\mathbf{c}_{p,1}(\mathbf{r}_4)\mathbf{c}_{p,2}(\mathbf{r}_4, \mathbf{f}_p)\big(\mathcal{R}_p^{(\mathbf{r}_2)}[(h,\mathcal{P}), f] + \gamma_p\overline{\mathcal{U}_p} + \mathbf{c}_{p,1}(\mathbf{r}_2)\big) n^{-1/2}, \\
\mathcal{R}_p^{(\mathbf{r}_2)}[(h,\mathcal{P}), f] &\le \mathbf{c}_{p,1}(\mathbf{r}_2) + \gamma_p\overline{\mathcal{U}_p} + \|\mathbf{K}\|_1^d \mathbf{f}_p + \mathbf{f}_p.
\end{aligned}$$

Thus we come to the assertion of Theorem 3 with $\mathbf{C}_{p,1}(\mathbf{f}_p) := 3\mathfrak{d}^2 \|\mathbf{K}\|_1^d [\mathcal{C}_p(\mathbf{f}_p)]^{\mathfrak{d}^2-1}$ and

$$\begin{aligned}
\mathbf{C}_{p,2}(\mathbf{f}_p) &:= 3\mathbf{c}_{p,1}(\mathbf{r}_4)\mathbf{c}_{p,2}(\mathbf{r}_4, \mathbf{f}_p)\big(2\gamma_p\overline{\mathcal{U}_p} + (1 + \|\mathbf{K}\|_1^d)\mathbf{f}_p + 2\mathbf{c}_{p,1}(\mathbf{r}_2)\big) \\
&\qquad + 3\mathbf{c}_{p,1}(\mathbf{r}_1)\mathfrak{d}^2 \|\mathbf{K}\|_1^d [\mathcal{C}_p(\mathbf{f}_p)]^{\mathfrak{d}^2-1},
\end{aligned}$$

since $\mathcal{R}_p^{(\mathbf{r}_1)}[(h,\mathcal{P}), f] = \mathcal{R}_p[(h,\mathcal{P}), f]$. The constants $\mathbf{c}_{p,1}(\mathbf{r}_k)$ and $\mathbf{c}_{p,2}(\mathbf{r}_k, \mathbf{f}_p)$, $k = 1, 2, 4$, are given in the proof of Lemma 1.

(7) Particular case: $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure).

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

$$\|\widetilde{f} - f\|_p \le (1 + 2\|\mathbf{K}\|_1^d)\|\widetilde{f}_h - f\|_p + 2\gamma_{p,\overline{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h) + 2\big[\, \|\xi_h\|_p - \gamma_{p,\overline{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h)\big]_+.$$

Thus we get from Proposition 2

$$\left(\mathbb{E}_f\|\widetilde{f}-f\|_p^{\mathbf{r}_1}\right)^{\frac{1}{\mathbf{r}_1}} \le (1+2\|\mathbf{K}\|_1^d)(\mathbb{E}_f\|\widetilde{f}_h-f\|_p^{\mathbf{r}_1})^{\frac{1}{\mathbf{r}_1}} + 2\gamma_{p,\overline{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h) + 2c_p(\mathbf{r}_1)n^{-1/2},$$

where the constants $\gamma_{p,\overline{\emptyset}}(\mathbf{r}_1)$ 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 $f \in N_{p,d}(\beta, L, \mathcal{P}) \subset \mathbf{F}_p[\,\mathfrak{P}\,]$ be arbitrary fixed.

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

$$\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{J\in\mathcal{P}\diamond\mathcal{P}'}\|\widetilde{f}_{h_J}-f_J\|_p \le \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{J\in\mathcal{P}\diamond\mathcal{P}'}\|\mathbb{E}_f\{\widetilde{f}_{h_J}(\cdot)\}-f_J\|_p + \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{J\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_J}\|_p, \tag{32}$$

where $\mathbb{E}_f\{\widetilde{f}_{h_J}(x_J)\} = K_{h_J}\star f_J(x_J)$ and, recall, $\xi_{h_J}(x_J) := \widetilde{f}_{h_J}(x_J) - \mathbb{E}_f\{\widetilde{f}_{h_J}(x_J)\}$.

Note first that, by applying Proposition 3 in Kerkyacharian, Lepski and Picard [23], it is easily established that, for any $h \in (0,1]^d$, any $\mathcal{P}' \in \mathfrak{P}$ and any $J \in \mathcal{P}\diamond\mathcal{P}'$,

$$\|K_{h_J}\star f_J - f_J\|_p \le \sum_{j\in J} c_J\big(\mathbf{K},|J|,p,l,L_J\big)h_j^{\beta_j} \le \mathbf{c}\sum_{j\in J}h_j^{\beta_j} \le \mathbf{c}\sup_{I\in\mathcal{P}}\sum_{j\in I}h_j^{\beta_j}, \quad \mathbf{c}>0. \tag{33}$$

Next, if $(h,\mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,]$, we easily get from Propositions 1–2

$$\left(\mathbb{E}_f\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{J\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_J}\|_p^p\right)^{\frac{1}{p}} = \mathcal{O}\left(\sup_{I\in\mathcal{P}}\left[n\prod_{j\in I}h_j^{[\frac{1}{2}\wedge(1-\frac{1}{p})]^{-1}\lambda_j+1}\right]^{-[\frac{1}{2}\wedge(1-\frac{1}{p})]}\right). \tag{34}$$

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

$$h_j^{\beta_j} = h_k^{\beta_k} = \left[n\prod_{j\in I}h_j^{[\frac{1}{2}\wedge(1-\frac{1}{p})]^{-1}\lambda_j+1}\right]^{-[\frac{1}{2}\wedge(1-\frac{1}{p})]}, \qquad j,k\in I.$$

The solution is given by

$$h_j = n^{-\frac{[\frac{1}{2}\wedge(1-\frac{1}{p})]\tau_I}{\tau_I+[\frac{1}{2}\wedge(1-\frac{1}{p})]}\frac{1}{\beta_j}}, \qquad j\in I,\ I\in\mathcal{P}, \tag{35}$$

where τ_I is given in (20).

Note that $h_I \in [h_{\min}^{(p)}, h_{\max}^{(p)}]^{|I|}$ and $n\prod_{j\in I}h_j^{[\frac{1}{2}\wedge(1-\frac{1}{p})]^{-1}\lambda_j+1} \ge 1$ for all $I \in \mathcal{P}\diamond\mathcal{P}'$, $\mathcal{P}'\in\mathfrak{P}$, and n large enough. Denote by $\overline{h}_I$ the projection of h_I on the dyadic grid $\mathcal{H}_{p,I}$. It is easily checked that $(\overline{h},\mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,]$ for n large enough. Thus it follows from Theorem 3, (32), (33) and (34) that

$$\mathcal{R}_p[\widehat{f},f] \le \mathbf{C}\left[\sup_{I\in\mathcal{P}}\sum_{j\in I}\overline{h}_j^{\beta_j} + \sup_{I\in\mathcal{P}}\left[n\prod_{j\in I}h_j^{[\frac{1}{2}\wedge(1-\frac{1}{p})]^{-1}\lambda_j+1}\right]^{-[\frac{1}{2}\wedge(1-\frac{1}{p})]}\right] + \mathbf{C}'n^{-1/2}, \tag{36}$$

for n large enough. Finally, it is easily seen that we get the statement of Theorem 5 from (35) and (36). Similarly, the statement of Theorem 6 is obtained by applying Theorem 4.

(2) Case $p=+\infty$**.** Let $(\beta, L, r, \mathcal{P}) \in (0,l]^d \times (0,\infty)^d \times [1,+\infty]^d \times \mathfrak{P}$ 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 [33], that $N_{r,d}(\beta,L,\mathcal{P}) \subset \mathbf{F}_\infty[\,\mathfrak{P}\,]$, 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 [25], it follows from Lemma 4 in [25] that, for any $h \in (0,1]^d$, any $\mathcal{P}' \in \mathfrak{P}$ and any $J \in \mathcal{P} \diamond \mathcal{P}'$,

$$\|K_{h_J} \star f_J - f_J\|_\infty \le \mathbf{c} \sup_{I \in \mathcal{P}} \sum_{j \in I} h_j^{\beta_j(I)}, \quad \mathbf{c} := c(\mathbf{K}, d, l, L) > 0, \tag{37}$$

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

Next, if $(h, \mathcal{P}) \in \mathcal{H}_\infty[\,\mathfrak{P}\,]$, we easily get from Proposition 2

$$\mathbb{E}_f \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{J \in \mathcal{P} \diamond \mathcal{P}'} \|\xi_{h_I}\|_\infty = \mathcal{O}\left(\sup_{I \in \mathcal{P}} \sqrt{\frac{\log(n)}{n \prod_{j \in I} h_j^{2\lambda_j + 1}}} \right). \tag{38}$$

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

$$h_j^{\beta_j(I)} = h_k^{\beta_k(I)} = \sqrt{\frac{\log(n)}{n \prod_{j \in I} h_j^{2\lambda_j + 1}}}, \qquad j, k \in I.$$

The solution is given by

$$h_j = \left(\frac{n}{\log(n)} \right)^{-\frac{\Upsilon_I}{2\Upsilon_I + 1} \frac{1}{\beta_j(I)}}, \qquad j \in I,\ I \in \mathcal{P}, \tag{39}$$

where Υ_I is given in (22).

Note that $n \prod_{j \in I} h_j^{2\lambda_j + 1} \ge \log(n)$ for all $I \in \mathcal{P} \diamond \mathcal{P}'$, $\mathcal{P}' \in \mathfrak{P}$, and n large enough. Thus, as before, we get the statement of Theorem 7 from Theorem 3, (37), (38) and (39). Similarly, the statement of Theorem 8 is obtained by applying Theorem 4. $\square$

6. APPENDIX

6.1. Proof of Proposition 1

Assume that Assumption (N2) is satisfied. Let $h \in (0,1]^d$ and $I \in \mathcal{I}_d^\diamond$ be arbitrary fixed. Note that

$$L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i\langle t_I, x_I \rangle} \frac{\widehat{K_I}(h_I t_I) g_I(h_I t_I)}{g_I(h_I t_I) \widehat{q_I}(t_I)}\, dt_I, \qquad g_I(t_I) := \prod_{j \in I} (1 + t_j^2)^{\frac{\lambda_j}{2}}, \tag{40}$$

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

(1) Proof of assertion (i). Suppose that $p \in (1, \infty) \setminus \{2\}$. Let $\mathbf{r} \in (1, 2)$ be arbitrary fixed. Here we apply the Marcinkiewicz Multiplier Theorem on $\mathbb{R}^{|I|}$, given in Grafakos [19], p. 363, with

$$m(t_I) = g_I^{-1}(h_I t_I) \widehat{q_I}^{-1}(t_I).$$

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

$$\sup_{t_I \in \mathbb{R}^{|I|}} |m(t_I)| \le \mathbf{A} \sup_{u_I \in \mathbb{R}^{|I|}} \left[\prod_{j \in I} (1 + u_j^2)^{-\frac{\lambda_j}{2}} \prod_{j \in I} (1 + [u_j / h_j]^2)^{\frac{\lambda_j}{2}} \right] \le \mathbf{A} \prod_{j \in I} h_j^{-\lambda_j}. \tag{41}$$

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

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

Here, $\gamma_I \le \alpha_I$ means $\gamma_j \le \alpha_j$, $\forall j \in I$, and $\binom{\alpha_I}{\gamma_I} = \prod_{j \in I} \binom{\alpha_j}{\gamma_j}$.

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

$$\Big\{\prod_{j\in I} h_j^{\gamma_j}\Big\}[D^{\gamma_I} g_I^{-1}](h_I t_I)[D^{\alpha_I-\gamma_I}\widehat{q_I}^{-1}](t_I)$$
$$= \Big\{\prod_{j\in I}(t_j h_j)^{\gamma_j}\Big\}[D^{\gamma_I} g_I^{-1}](h_I t_I)[D^{\alpha_I-\gamma_I}\widehat{q_I}^{-1}](t_I)\Big\{\prod_{j\in I} t_j^{\alpha_j-\gamma_j}\Big\}\Big(\prod_{j\in I} t_j^{-\alpha_j}\Big).$$

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

Since q satisfies Assumption (N2), we obtain similarly to (41)

$$|\,[D^{\alpha_I} m](t_I)| \leq C(|I|, q_I)\mathbf{A}\Big\{\prod_{j\in I} h_j^{-\lambda_j}\Big\}\Big(\prod_{j\in I}|t_j|^{-\alpha_j}\Big), \tag{42}$$

$$C(|I|, q_I) := \max_{|\alpha_I|\leq|I|}\Big\{\sum_{\gamma_I\leq\alpha_I}\binom{\alpha_I}{\gamma_I}\sup_{u_I\in\mathbb{R}^{|I|}}\Big|\Big\{\prod_{j\in I} u_j^{\gamma_j}\Big\}[D^{\gamma_I} g_I^{-1}](u_I) g_I(u_I)\Big|\Big\} < \infty. \tag{43}$$

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

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

it follows from Corollary 5.2.5 in Grafakos [19], (41) and (42)

$$\|L_{(h_I)}\|_{\mathbf{r}} \leq 2\mathbf{A}C_{|I|}C(|I|, q_I)\max(\mathbf{r}, (\mathbf{r}-1)^{-1})^{6|I|}\|S_I\|_{\mathbf{r}}(V_{h_I})^{-(1-1/\mathbf{r})}\prod_{j\in I} h_j^{-\lambda_j},$$

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

$$C_{\mathbf{r},I} := 2\mathbf{A}\big\{C_{|I|}C(|I|, q_I)\max(\mathbf{r}, (\mathbf{r}-1)^{-1})^{6|I|}\|S_I\|_{\mathbf{r}}\big\}.$$

(2) Proof of assertion (ii). Note first that

$$\|L_{(h_I)}\|_2 = (2\pi)^{-\frac{|I|}{2}}\|\widehat{K_{h_I}}/\widehat{q_I}\|_2, \qquad \|L_{(h_I)}\|_\infty \leq (2\pi)^{-|I|}\|L_{(h_I)}\|_1.$$

In view of Assumption (N2) on the errors,

$$\|\widehat{K_{h_I}}/\widehat{q_I}\|_2^2 \ \leq\ \mathbf{A}^2\int_{\mathbb{R}^{|I|}}|\widehat{K_I}(h_I t_I)|^2\prod_{j\in I}(1+t_j^2)^{\lambda_j}\, dt_I$$
$$\leq\ \mathbf{A}^2\Big(\int_{\mathbb{R}^{|I|}}|\widehat{K_I}(u_I)|^2\prod_{j\in I}(1+u_j^2)^{\lambda_j}\, du_I\Big)V_{h_I}^{-1}\prod_{j\in I} h_j^{-2\lambda_j};$$
$$\|\widehat{K_{h_I}}/\widehat{q_I}\|_1 \ \leq\ \mathbf{A}\Big(\int_{\mathbb{R}^{|I|}}|\widehat{K_I}(u_I)|\prod_{j\in I}(1+u_j^2)^{\lambda_j/2}\, du_I\Big)V_{h_I}^{-1}\prod_{j\in I} h_j^{-\lambda_j}.$$

Thus assertion (ii) of Proposition 1 is proved with

$$C_I := \mathbf{A}\big\{(2\pi)^{-\frac{|I|}{2}}\big(\|\widehat{K_I}g_I\|_2 \vee \|\widehat{K_I}g_I\|_1\big)\big\},$$

where g_I is given in (40). □

6.2. Proof of Proposition 2: Case $p < \infty$

Let $I \in \mathcal{I}_d^\diamond$ be arbitrary fixed. We get the statement of Proposition 2 by applying Theorem 1 and Corollaries 2 and 3 in Goldenshluger and Lepski [16] with $s = p$, $\mathcal{X} = \mathcal{T} = \mathbb{R}^{|I|}$, $\nu = \tau$ is the Lebesgue measure on $\mathbb{R}^{|I|}$, $w(\cdot,\cdot) = n^{-1}L_{(h_I)}(\cdot - \cdot)$ and $M_s(w) = \|n^{-1}L_{(h_I)}\|_p < \infty$. Here, the i.i.d. random vectors are the $Y_{k,I}$'s and their common density is $f_I \star q_I$. By using the continuity property of $L_{(h_I)}(\cdot)$, it is easily proved that Assumption (A1) in the aforementioned paper is fulfilled.

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

By application of Corollary 2 in Goldenshluger and Lepski [16], one has

$$\mathbb{P}_f\big\{\|\xi_{h_I}\|_p \geq U_p(h_I) + z\big\} \leq \exp\left\{-\frac{z^2}{A_p^2(h_I)}\right\}, \qquad \forall z > 0, \quad \forall n \geq 1, \tag{44}$$

where $U_p(h_I) = 4n^{\frac{1}{p}-1}\|L_{(h_I)}\|_p$ and $A_p^2(h_I) = 37n^{-1}\|L_{(h_I)}\|_p^2$.

By integration of (44) we easily get, for all integer $n \geq 3$,

$$\begin{aligned}\mathbb{E}_f\Big[\,\|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{r\log(n)}\Big]_+^{\mathbf{r}} &\leq \Gamma(\mathbf{r}+1)\big[U_p(h_I) + A_p(h_I)\big]^{\mathbf{r}} e^{-\mathbf{r}\log(n)}\\ &\leq \Gamma(\mathbf{r}+1)11^{\mathbf{r}} \sup_{h_I \in \mathcal{H}_{p,I}} \big[n^{-\frac{1}{2}}\|L_{(h_I)}\|_p\big]^{\mathbf{r}} n^{-\mathbf{r}},\end{aligned}$$

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

Note that, for all integer $n \geq 3$,

$$U_p(h_I) + A_p(h_I)\sqrt{\mathbf{r}\log(n)} \leq \left\{4 + \sqrt{\frac{37e^{-1}p\mathbf{r}}{2-p}}\right\} n^{\frac{1}{p}-1}\,\|L_{(h_I)}\|_p =: \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I).$$

Since $\mathrm{card}(\mathcal{H}_{p,I}) \leq \Big[\Big(1 \vee \frac{p}{|I|}\Big)\log_2(n)\Big]^{|I|}$, we obtain, for all integer $n \geq 3$,

$$\Big\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_{p,I}} \big[\,\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\big]_+^{\mathbf{r}}\Big\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}},$$

$$c_p(\mathbf{r}) := 11\big[\Gamma(\mathbf{r}+1)\big]^{\frac{1}{\mathbf{r}}} \sup_{n \in \mathbb{N}^*} \sup_{I \in \mathcal{I}_d^\diamond} \sup_{h_I \in \mathcal{H}_{p,I}} \big\{n^{\frac{1}{p}-\frac{3}{2}}[2\log_2(n)]^{\frac{|I|}{\mathbf{r}}}\|L_{(h_I)}\|_p\big\},$$

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

(2) Case $p = 2$. Let $\mathbf{r} \geq 1$ and $h_I \in \mathcal{H}_{p,I}$ be arbitrary fixed. Here we apply Theorem 1 in Goldenshluger and Lepski [16], but we compute differently the upper bound on the "*dual*" variance σ^2 by using the arguments given in the proof of Proposition 7 in Comte and Lacour [9]. Indeed, we obtain

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

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

Taking into account the latter inequality, the result of Theorem 1 in Goldenshluger and Lepski [16] should be: $\forall z > 0$, $\forall n \geq 1$,

$$\mathbb{P}_f\big\{\|\xi_{h_I}\|_p \geq U_p(h_I) + z\big\} \leq \exp\left\{-\frac{z^2}{A_p^2(h_I) + B_p(h_I)z}\right\}, \tag{45}$$

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

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

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

$$\mathbb{E}_f\Big[\ \|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{\mathbf{r}\log(n)} - B_p(h_I)\mathbf{r}\log(n)\Big]_+^{\mathbf{r}}$$

$$\leq \Gamma(\mathbf{r}+1)\big[U_p(h_I) + A_p(h_I) + B_p(h_I)\big]^{\mathbf{r}} e^{-\mathbf{r}\log(n)}$$

$$\leq \Gamma(\mathbf{r}+1)\big(8 \vee \|\widehat{q}_I\|_1^{\frac{1}{2}}\big)^{\mathbf{r}} \sup_{h_I\in\mathcal{H}_{p,I}} \big\{\|\widehat{K_{h_I}}/\widehat{q}_I\|_\infty + \|L_{(h_I)}\|_2\big\}^{\mathbf{r}} n^{-\frac{\mathbf{r}}{2}-\mathbf{r}}.$$

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

$$\|\widehat{K_{h_I}}/\widehat{q}_I\|_\infty \leq \mathbf{A}\|\widehat{K}_I g_I\|_\infty \prod_{j\in I} h_j^{-\lambda_j}, \tag{46}$$

where g_I is given in (40). Thus, in view of Proposition 1, (46) and the definition of $\mathcal{H}_{p,I}$, for any integer $n \geq 3$,

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

$$\leq \left\{C_I\left(1 + \frac{8\mathbf{r}}{3e} + \sqrt{\frac{48\mathbf{r}}{e}}\right) + \mathbf{A}\|\widehat{K}_I g_I\|_\infty\sqrt{\frac{6\mathbf{r}\|\widehat{q}_I\|_1}{(2\pi)^{|I|}}}\right\} n^{-\frac{1}{2}} \prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}$$

$$=: \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I).$$

Finally, we obtain for any integer $n \geq 3$

$$\Big\{\mathbb{E}_f \sup_{h_I\in\mathcal{H}_{p,I}} \big[\ \|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\big]_+^{\mathbf{r}}\Big\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}},$$

$$c_p(\mathbf{r}) := [\Gamma(\mathbf{r}+1)]^{\frac{1}{\mathbf{r}}}$$

$$\times \sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\Big[(8 \vee \|\widehat{q}_I\|_1^{\frac{1}{2}})\{\|\widehat{K_{h_I}}/\widehat{q}_I\|_\infty + \|L_{(h_I)}\|_2\}\big[2\log_2(n)\big]^{\frac{|I|}{\mathbf{r}}} n^{-1}\Big],$$

which is finite in view of Proposition 1, (46) and the definition of the set $\mathcal{H}_{p,I}$.

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

By application of Corollary 3 in Goldenshluger and Lepski [16], one has: $\forall z > 0$, $\forall n \geq 1$,

$$\mathbb{P}_f\big\{\|\xi_{h_I}\|_p \geq U_p(h_I) + z\big\} \leq \exp\left\{-\frac{z^2}{A_p^2(h_I) + B_p(h_I)z}\right\}, \tag{47}$$

$$U_p(h_I) = 3c(p)\|q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}\Big\{n^{-\frac{1}{2}}\|L_{(h_I)}\|_2 + n^{\frac{1}{p}-1}\|L_{(h_I)}\|_p\Big\},$$

$$A_p^2(h_I) = 16c(p)\|q_I\|_\infty^{\frac{3}{2}}\Big\{n^{-1}\|L_{(h_I)}\|_{\frac{2p}{p+2}}^2 + n^{-\frac{3}{2}}\|L_{(h_I)}\|_2\|L_{(h_I)}\|_p + n^{\frac{1}{p}-2}\|L_{(h_I)}\|_p^2\Big\},$$

$$B_p(h_I) = \frac{4}{3}c(p)n^{-1}\|L_{(h_I)}\|_p, \qquad c(p) = \frac{15p}{\log(p)}.$$

Here we have used the following inequalities, which are consequences of Young's inequality:

$$\|f_I \star q_I\|_\infty \leq \|f_I\|_1\|q_I\|_\infty \leq \|q_I\|_\infty,$$

$$\big\|\sqrt{f_I \star q_I}\big\|_p \leq \|f_I \star q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}\|f_I \star q_I\|_1 \leq \|q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}.$$

By integration of (47) we get, for any integer $n \geq 3$,

$$\mathbb{E}_f\Big[\ \|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{\mathbf{r}\log(n)} - B_p(h_I)\mathbf{r}\log(n)\Big]_+^{\mathbf{r}}$$

$$\leq \Gamma(\mathbf{r}+1)\big[U_p(h_I) + A_p(h_I) + B_p(h_I)\big]^{\mathbf{r}} e^{-\mathbf{r}\log(n)}$$

$$\leq \Gamma(\mathbf{r}+1)\big\{7c(p)(1\vee\|q_I\|_\infty)^{\frac{3}{4}}\big\}^{\mathbf{r}}$$

$$\times \sup_{h_I\in\mathcal{H}_{p,I}}\Big\{\|L_{(h_I)}\|_2 + \|L_{(h_I)}\|_{\frac{2p}{p+2}} + \sqrt{\|L_{(h_I)}\|_2\|L_{(h_I)}\|_p} + \|L_{(h_I)}\|_p\Big\}^{\mathbf{r}} n^{-\frac{\mathbf{r}}{2}-\mathbf{r}}.$$

In view of Proposition 1, we get

$$\|L_{(h_I)}\|_p \leq \|L_{(h_I)}\|_\infty^{1-\frac{2}{p}}\|L_{(h_I)}\|_2^{\frac{2}{p}} \leq C_I V_{h_I}^{\frac{1}{p}-\frac{1}{2}}\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}. \tag{48}$$

Thus, in view of Proposition 1, (48) and the definition of $\mathcal{H}_{p,I}$, for all integer $n \geq 3$,

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

$$\leq c_p^{\frac{1}{p}-\frac{1}{2}}(1\vee C_I)\bigg\{6c(p)\|q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}} + 8\sqrt{\frac{\mathbf{r}c(p)[p\vee e]}{e}}\|q_I\|_\infty^{\frac{3}{4}} + \frac{4\mathbf{r}c(p)[p\vee e]}{3e}\bigg\}$$

$$\times n^{-\frac{1}{2}}\bigg[\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}} + \sqrt{\log(n)}\|L_{(h_I)}\|_{\frac{2p}{p+2}}\bigg] =: \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I).$$

Finally, we obtain for any integer $n \geq 3$

$$\Big\{\mathbb{E}_f \sup_{h_I\in\mathcal{H}_{p,I}}\big[\ \|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\big]_+^{\mathbf{r}}\Big\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}},$$

$$c_p(\mathbf{r}) := 7c(p)[\Gamma(\mathbf{r}+1)]^{\frac{1}{\mathbf{r}}}\sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\bigg[\ \big(1\vee\|q_I\|_\infty\big)^{\frac{3}{4}}\Big\{\|L_{(h_I)}\|_2 + \|L_{(h_I)}\|_{\frac{2p}{p+2}}$$

$$+\sqrt{\|L_{(h_I)}\|_2\|L_{(h_I)}\|_p} + \|L_{(h_I)}\|_p\Big\}n^{-1}\Big[\Big(1\vee\frac{p}{|I|}\Big)\log_2(n)\Big]^{\frac{|I|}{\mathbf{r}}}\ \bigg],$$

which is finite in view of Proposition 1, (48) and the definition of the set $\mathcal{H}_{p,I}$. □

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

Let $n \geq 3$, $I \in \mathcal{I}_d^\diamond$ and $h_I \in [1/n, 1]^{|I|}$ be arbitrary fixed. Assume that $n\prod_{j\in I} h_j^{2\lambda_j+1} \geq \log(n)$. 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 $f_I \star q_I$, we get from Proposition 1

$$\sup_{x_I\in\mathbb{R}^{|I|}}\sup_{y_I\in\mathbb{R}^{|I|}}|L_{(h_I)}(x_I - y_I)| \leq \|L_{(h_I)}\|_\infty \leq C_I(\mathbf{K}, q)\prod_{j\in I} h_j^{-\lambda_j-1} < \infty, \tag{49}$$

$$C_I(\mathbf{K}, q) := \frac{\mathbf{A}}{(2\pi)^{\frac{|I|}{2}}}\Big\{\|\widehat{K_I}g_I\|_2 \vee \|\widehat{K_I}g_I\|_1 \vee (\max_{j\in I}\|D_j^1\widehat{K_I}g_I\|_1) \vee \|\widehat{K_I}\varphi_I\|_2 \vee \|\widehat{K_I}\varphi_I\|_1\Big\},$$

where $\varphi_I(t_I) := \sup_{j\in I}|t_j|g_I(t_I)$ and g_I is given in (40);

$$\sup_{x_I\in\mathbb{R}^{|I|}}\big(\mathbb{E}_f|L_{(h_I)}(x_I-Y_{1,I})|^2\big)^{\frac{1}{2}} \le \sqrt{\|f_I\star q_I\|_\infty}\|L_{(h_I)}\|_2 \le \sqrt{\|q_I\|_\infty}C_I(\mathbf{K},q)\prod_{j\in I}h_j^{-\lambda_j-\frac{1}{2}}. \quad (50)$$

Next, set x_I and $\overline{x}_I$ be arbitrary fixed in $\mathbb{R}^{|I|}$. For any $t_I\in\mathbb{R}^{|I|}$

$$\begin{aligned}|e^{-i\langle t_I,x_I\rangle}-e^{-i\langle t_I,\overline{x}_I\rangle}| &= \Big|\prod_{j\in I}e^{-it_jx_j}-\prod_{j\in I}e^{-it_j\overline{x}_j}\Big|\\ &\le |I|\sup_{j\in I}|e^{-it_jx_j}-e^{-it_j\overline{x}_j}| \le |I|\sup_{j\in I}|t_j|\sup_{j\in I}|x_j-\overline{x}_j|.\end{aligned}$$

Therefore, for any $y_I\in\mathbb{R}^{|I|}$

$$\begin{aligned}|L_{(h_I)}(x_I-y_I)-L_{(h_I)}(\overline{x}_I-y_I)| &\le \frac{1}{(2\pi)^{|I|}}\int_{\mathbb{R}^{|I|}}\left|\frac{\widehat{K_{h_I}}(t_I)}{\widehat{q_I}(t_I)}\right||e^{-i\langle t_I,x_I\rangle}-e^{-i\langle t_I,\overline{x}_I\rangle}|\,dt_I\\ &\le n|I|C_I(\mathbf{K},q)\prod_{j\in I}h_j^{-\lambda_j-1}\sup_{j\in I}|x_j-\overline{x}_j|; \quad (51)\end{aligned}$$

$$\begin{aligned}&\big(\mathbb{E}_f|L_{(h_I)}(x_I-Y_{1,I})-L_{(h_I)}(\overline{x}_I-Y_{1,I})|^2\big)^{\frac{1}{2}}\\ &\le\left(\frac{\|f_I\star q_I\|_\infty}{(2\pi)^{|I|}}\int_{\mathbb{R}^{|I|}}\left|\frac{\widehat{K_{h_I}}(t_I)}{\widehat{q_I}(t_I)}\right|^2\big|e^{-i\langle t_I,x_I\rangle}-e^{-i\langle t_I,\overline{x}_I\rangle}\big|^2\,dt_I\right)^{\frac{1}{2}}\\ &\le n|I|\sqrt{1\vee\|q_I\|_\infty}C_I(\mathbf{K},q)\prod_{j\in I}h_j^{-\lambda_j-\frac{1}{2}}\sup_{j\in I}|x_j-\overline{x}_j|; \quad (52)\end{aligned}$$

Consider now the normalized empirical process

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

In view of Bernstein's inequality, (49), (50), (51) and (52), $\forall z>0$,

$$\mathbb{P}_f\{|\overline{\xi}_{h_I}(x_I)|>z\} \le 2\exp\left\{-\frac{z^2}{A^2(x_I)+zB(x_I)}\right\}, \quad (53)$$

$$\mathbb{P}_f\big\{|\overline{\xi}_{h_I}(x_I)-\overline{\xi}_{h_I}(\overline{x}_I)|>z\big\} \le 2\exp\left\{-\frac{z^2}{\mathrm{a}^2(x_I,\overline{x}_I)+z\mathrm{b}(x_I,\overline{x}_I)}\right\}, \quad (54)$$

where $A(x_I):=1$, $B(x_I):=\big(n\prod_{j\in I}h_j^{2\lambda_j+1}\big)^{-\frac{1}{2}}\le 1$ and

$$\mathrm{a}(x_I,\overline{x}_I)=\mathrm{b}(x_I,\overline{x}_I):=2\wedge\big\{n|I|\sup_{j\in I}|x_j-\overline{x}_j|\big\}. \quad (55)$$

It is easily seen that $\mathrm{a}(\cdot,\cdot)$ 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 $\overline{\xi}_{h_I}(\cdot)$ over totally bounded sets by applying Proposition 1 in Lepski [24] with $\mathfrak{T}=\mathbb{R}^{|I|}$, $\mathfrak{S}=\mathbb{R}$, $\chi=\overline{\xi}_{h_I}$ and $\Psi(\cdot)=|\cdot|$. Then we have to check Assumptions 1, 2 and 3 required in the latter Proposition and to match the notation used in the present paper and in Lepski [24].

Note first that, in view of (53), (54) and (55), Assumption 1 is fulfilled with $c = 2$. Next, consider the family of closed balls

$$\mathbb{B}_{\frac{R}{2}}(t_I) := \Big\{ x_I \in \mathbb{R}^{|I|}: \sup_{j\in I} |x_j - t_j| \le R/2 \Big\}, \qquad R \ge 1, \quad t_I \in \mathbb{R}^{|I|}.$$

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 $\Theta = \mathbb{B}_{\frac{R}{2}}(t_I)$, $\overline{A}_\Theta = 1$ and $\overline{B}_\Theta = \big(n \prod_{j\in I} h_j^{2\lambda_j+1}\big)^{-\frac{1}{2}}$.

Let $s\colon \mathbb{R} \to \mathbb{R}_+ \setminus \{0\}$ be defined by $s(z) := (0.01 + z^8)^{-1}$. Obviously $\sum_{k\ge 0} s\big(2^{k/2}\big) \le 1$ and, for any $z > 0$,

$$\mathfrak{E}_{\Theta,\mathrm{a}}\big(z(48\delta)^{-1}s(\delta)\big) \le |I| \left[\log\left(\frac{Rn|I|}{z(48\delta)^{-1}s(\delta)} \right) \right]_+, \qquad \forall \delta > 0, \tag{56}$$

where $\mathfrak{E}_{\Theta,\mathrm{a}}(\delta)$, $\delta > 0$, denotes the entropy of Θ measured in a. Then, for any $z > 0$, there exists $\delta_* > 0$ small enough such that

$$e_s^{(\mathrm{a})}(z,\Theta) := \sup_{\delta>0} \delta^{-2} \mathfrak{E}_{\Theta,\mathrm{a}}\big(z(48\delta)^{-1}s(\delta)\big) = \sup_{\delta>\delta_*} \delta^{-2} \mathfrak{E}_{\Theta,\mathrm{a}}\big(z(48\delta)^{-1}s(\delta)\big) < \infty,$$

$$e_s^{(\mathrm{b})}(z,\Theta) := \sup_{\delta>0} \delta^{-1} \mathfrak{E}_{\Theta,\mathrm{b}}\big(z(48\delta)^{-1}s(\delta)\big) = \sup_{\delta>\delta_*} \delta^{-1} \mathfrak{E}_{\Theta,\mathrm{b}}\big(z(48\delta)^{-1}s(\delta)\big) < \infty.$$

Thus Assumption 3 in Lepski [24] is fulfilled and Proposition 1 in that paper can be applied. Let us compute the quantities which appear in this result.

Choose $\vec{s} = (s,s)$, $\varkappa = (2\overline{A}_\Theta, 2\overline{B}_\Theta)$ and $\varepsilon = \sqrt{2} - 1$. Since $\overline{A}_\Theta \vee \overline{B}_\Theta \le 1$ and $\mathrm{a}(x_I, \overline{x}_I) = \mathrm{b}(x_I, \overline{x}_I) \le 2$, $\forall x_I, \overline{x}_I \in \mathbb{R}^{|I|}$, we straightforwardly get

$$\begin{aligned} e_{\vec{s}}(\varkappa,\Theta) &:= e_s^{(\mathrm{a})}(2\overline{A}_\Theta,\Theta) + e_s^{(\mathrm{b})}(2\overline{B}_\Theta,\Theta) \\ &\le \sup_{\delta>0.61} \delta^{-2} \mathfrak{E}_{\Theta,\mathrm{a}}\big(2(48\delta)^{-1}s(\delta)\big) + \sup_{\delta>0.61} \delta^{-1} \mathfrak{E}_{\Theta,\mathrm{b}}\big(2(48\delta)^{-1}s(\delta)\big) \\ &\le 4.5|I|\big[\log(Rn|I|)\big]_+ + 8.5; \end{aligned}$$

$$\begin{aligned} U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta) &:= \varkappa_1 \sqrt{2[1+\varepsilon^{-1}]^2 e_{\vec{s}}(\varkappa,\Theta) + y} + \varkappa_2\big(2[1+\varepsilon^{-1}]^2 e_{\vec{s}}(\varkappa,\Theta) + y\big) \\ &\le 2\sqrt{31|I|\log(Rn|I|) + 59 + y} + \frac{2(31|I|\log(Rn|I|) + 59 + y)}{\sqrt{n\prod_{j\in I} h_j^{2\lambda_j+1}}}. \end{aligned}$$

Thus it follows from Proposition 1 in Lepski [24] that, for any $y \ge 1$ and any $\mathbf{r} \ge 1$,

$$\mathbb{E}_f \Big\{ \sup_{x_I \in \mathbb{B}_{\frac{R}{2}}(t_I)} |\overline{\xi}_{h_I}(x_I)| - U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta) \Big\}_+^r \le 4\Gamma(\mathbf{r}+1)\big[2y^{-1}U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta)\big]^{\mathbf{r}} e^{-\frac{y}{2}}. \tag{57}$$

(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 $\sup_{j\in I} |x_j - y_j| \ge n$. By integration by parts, we easily get

$$|L_{(h_I)}(x_I - y_I)| \le \frac{\max_{j\in I} \|D_j^1(\widehat{K_{h_I}}/\widehat{q_I})\|_1}{(2\pi)^{|I|} \sup_{j\in I} |x_j - y_j|} \le \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I} h_j^{\lambda_j+1}} \le \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}, \tag{58}$$

in view of Assumption (N2) on the errors.

Consider the collection of closed balls $\{\mathbb{B}_{\frac{n}{2}}(n\mathbf{j}),\ \mathbf{j}\in\mathbb{Z}^{|I|}\}$. Obviously this collection is a countable cover of $\mathbb{R}^{|I|}$. Put, for any $\mathbf{j}\in\mathbb{Z}^{|I|}$,

$$\mathrm{f}_{\mathbf{j}} := \int_{\mathbb{B}(\mathbf{j})} f_I\star q_I(x_I)\,dx_I, \qquad \mathbb{B}(\mathbf{j}) := \bigcup_{\mathbf{k}\in\mathbb{Z}^{|I|}:\ \mathbb{B}_{\frac{n}{2}}(n\mathbf{j})\cap\mathbb{B}_{\frac{n}{2}}(n\mathbf{k})\neq\emptyset} \mathbb{B}_{\frac{n}{2}}(n\mathbf{k}).$$

It is easily checked that

$$\sum_{\mathbf{j}\in\mathbb{Z}^{|I|}} \mathrm{f}_{\mathbf{j}} = \int_{\mathbb{R}^{|I|}} f_I\star q_I(x_I)\Big[\sum_{\mathbf{j}\in\mathbb{Z}^{|I|}} \mathbf{1}_{\mathbb{B}(\mathbf{j})}(x_I)\Big]\,dx_I \leq 4^{|I|}. \tag{59}$$

Set $\mathbf{j}\in\mathbb{Z}^{|I|}$ such that $\mathrm{f}_{\mathbf{j}}\geq n^{-v}$, where $v\geq 1$ is specified later. If $y=2\log(1/\mathrm{f}_{\mathbf{j}})+(\mathbf{r}+1)\log(n)$, we get from (57)

$$\mathbb{E}_f\Bigg\{\sup_{x_I\in\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})}|\xi_{h_I}(x_I)| - \gamma^{(v)}_{\infty,I}(\mathbf{r})\sqrt{\frac{\log(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\ \Bigg\}_+^{\mathbf{r}} \leq 2^{\mathbf{r}+2}\Gamma(\mathbf{r}+1)\big[\gamma^{(v)}_{\infty,I}(\mathbf{r})\big]^{\mathbf{r}}\mathrm{f}_{\mathbf{j}} n^{-\frac{\mathbf{r}+1}{2}},$$

where

$$\gamma^{(v)}_{\infty,I}(\mathbf{r}) := 4C_I(\mathbf{K},q)\sqrt{2(1\vee\|q_I\|_\infty)}\,(93|I|\log(|I|)+60+2v+\mathbf{r}),$$

since $n\prod_{j\in I}h_j^{2\lambda_j+1}\geq\log(n)$.

Thus, in view of (59), we obtain

$$\mathbb{E}_f\Bigg\{\sup_{x_I\in\Theta_1}|\xi_{h_I}(x_I)| - \gamma^{(v)}_{\infty,I}(\mathbf{r})\sqrt{\frac{\log(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\ \Bigg\}_+^{\mathbf{r}} \leq 2^{\mathbf{r}+2+2|I|}\Gamma(\mathbf{r}+1)\big[\gamma^{(v)}_{\infty,I}(\mathbf{r})\big]^{\mathbf{r}} n^{-\frac{\mathbf{r}+1}{2}}, \tag{60}$$

where $\Theta_1 := \cup_{\mathbf{j}\in\mathbb{Z}^{|I|}:\ \mathrm{f}_{\mathbf{j}}\geq n^{-v}}\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})$.

Set $\mathbf{j}\in\mathbb{Z}^{|I|}$ such that $\mathrm{f}_{\mathbf{j}}<n^{-v}$ and $x_I\in\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})$. In view of (49) and (58) we get, for any $k=1,\ldots,n$,

$$\begin{aligned}
\mathbb{E}_f|L_{(h_I)}(x_I-Y_{k,I})| &= \mathbb{E}_f\big\{|L_{(h_I)}(x_I-Y_{k,I})|\mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{k,I})\big\} \\
&\quad + \mathbb{E}_f\big\{|L_{(h_I)}(x_I-Y_{k,I})|\mathbf{1}_{\mathbb{R}^{|I|}\setminus\mathbb{B}(\mathbf{j})}(Y_{k,I})\big\} \\
&\leq \mathbb{P}_f\{Y_{k,I}\in\mathbb{B}(\mathbf{j})\}\frac{C_I(\mathbf{K},q)}{\prod_{j\in I}h_j^{2\lambda_j+1}} + \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}} \\
&\leq \frac{2C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}},
\end{aligned} \tag{61}$$

since $\mathrm{f}_{\mathbf{j}} := \mathbb{P}_f\{Y_{k,I}\in\mathbb{B}(\mathbf{j})\}\leq n^{-v}$, $v\geq 1$ and $\sup_{j\in I}|x_j-Y_{k,j}|\geq n$ when $Y_{k,I}\in\mathbb{R}^{|I|}\setminus\mathbb{B}(\mathbf{j})$.

Introduce random events

$$D_{\mathbf{j}} := \Bigg\{\sum_{k=1}^n \mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{k,I})\geq 2\Bigg\}, \quad \mathbf{j}\in\mathbb{Z}^{|I|}, \qquad D := \bigcup_{\mathbf{j}\in\mathbb{Z}^{|I|}:\ \mathrm{f}_{\mathbf{j}}<n^{-v}} D_{\mathbf{j}}.$$

Let $\overline{D}$ be the complementary to D. If $\overline{D}$ holds then, in view of (49) and (58),

$$n^{-1}\sum_{k=1}^n |L_{(h_I)}(x_I-Y_{k,I})| \leq \frac{2C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}, \quad \forall x_I\in\Theta_2 := \mathbb{R}^{|I|}\setminus\Theta_1. \tag{62}$$

Since $n\prod_{j\in I}h_j^{2\lambda_j+1}\geq\log(n)$, we get from (61) and (62)

$$\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)|\mathbf{1}_{\overline{D}}\leq\gamma^{(v)}_{\infty,I}(\mathbf{r})\sqrt{\frac{\log(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}$$

and, taking into account that $\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)|\leq 2C_I(\mathbf{K},\mathbf{q})n$,

$$\mathbb{E}_f\biggl\{\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)|-\gamma^{(v)}_{\infty,I}(\mathbf{r})\sqrt{\frac{\log(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\ \biggr\}_+^{\mathbf{r}}\leq[2C_I(\mathbf{K},\mathbf{q})]^{\mathbf{r}}n^{\mathbf{r}}\mathbb{P}_f(D). \tag{63}$$

Let $\mathbf{j}\in\mathbb{Z}^{|I|}$ satisfying $\mathrm{f}_{\mathbf{j}}<n^{-v}$ be arbitrary fixed. In view of Markov's inequality one has for any $z>0$

$$\mathbb{P}_f(D_{\mathbf{j}})\leq e^{-2z}\bigl[\mathbb{E}_f\{e^{z\mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{1,I})}\}\bigr]^n\leq\exp\{-2z+n(e^z-1)\mathrm{f}_{\mathbf{j}}\},$$

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

$$\mathbb{P}_f(D_{\mathbf{j}})\leq(e/2)^2(n\mathrm{f}_{\mathbf{j}})^2\leq 2\mathrm{f}_{\mathbf{j}}n^{2-v}. \tag{64}$$

Thus, choosing $v=1.5\mathbf{r}+2.5$, it follows from (59), (60), (63) and (64)

$$\mathbb{E}_f\biggl\{\|\xi_{h_I}\|_\infty-\gamma_{\infty,I}(\mathbf{r})\sqrt{\frac{\log(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\ \biggr\}_+^{\mathbf{r}}\leq 2^{\mathbf{r}+3+2|I|}\Gamma(\mathbf{r}+1)[\gamma_{\infty,I}(\mathbf{r})]^{\mathbf{r}}n^{-\frac{\mathbf{r}+1}{2}}, \tag{65}$$

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

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

$$\Bigl\{\mathbb{E}_f\sup_{h_I\in\mathcal{H}_{\infty,I}}\bigl[\ \|\xi_{h_I}\|_\infty-\gamma_{\infty,I}(\mathbf{r})\mathcal{U}_\infty(h_I)\bigr]_+^{\mathbf{r}}\Bigr\}^{\frac{1}{\mathbf{r}}}\leq c_\infty(\mathbf{r})n^{-\frac12}, \tag{66}$$

$$c_\infty(\mathbf{r}):=[\Gamma(\mathbf{r}+1)]^{\frac{1}{\mathbf{r}}}\sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^\diamond}\Bigl\{\gamma_{\infty,I}(\mathbf{r})[2^{\mathbf{r}+3+2|I|}]^{\frac{1}{\mathbf{r}}}[\log_2(n)]^{\frac{|I|}{\mathbf{r}}}n^{-\frac{1}{2\mathbf{r}}}\Bigr\}<\infty.$$

□

6.4. Proof of Lemma 1

Assume that $\mathfrak{P}\neq\{\overline{\emptyset}\}$. Set $f\in\mathbf{F}_p[\,\mathfrak{P}\,]$ and let $\mathbf{r}\in\{\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_4\}$ be arbitrary fixed. We obtain Lemma 1 by applying Proposition 2. We divide this proof into two steps.

(1) Note that

$$\xi_p\leq\sum_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\bigl[\ \|\xi_{h_I}\|_p-\gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\ \bigr]_+,$$

since $\gamma_{p,I}(\mathbf{r})$ increase with $\mathbf{r}$. In view of Proposition 2, if $p\in(1,+\infty]$ and $n\geq3$,

$$\bigl(\mathbb{E}_f|\xi_p|^{\mathbf{r}}\bigr)^{\frac{1}{\mathbf{r}}}\leq\mathbf{c}_{p,1}(\mathbf{r})n^{-\frac12},\qquad\mathbf{c}_{p,1}(\mathbf{r}):=d|\mathfrak{P}|^2c_p(\mathbf{r}).$$

(2) For any $p\geq1$

$$\begin{aligned}\overline{G}_p\ &\leq\ 1+\|\mathbf{K}\|_1^d\sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\Bigl\{\bigl[\ \|\xi_{h_I}\|_p-\overline{\gamma}_p\mathcal{U}_p(h_I)\ \bigr]_++\overline{\gamma}_p\mathcal{U}_p(h_I)+\|\mathbb{E}_f\{\widetilde{f}_{h_I}\}\|_p\Bigr\}\\ &\leq\ 1+\|\mathbf{K}\|_1^d\bigl(\overline{\xi}_p+\overline{\gamma}_p\overline{\mathcal{U}}_p+\|\mathbf{K}\|_1^d\mathbf{f}_p\bigr),\\ \overline{\gamma}_p\ &:=\ \sup_{I\in\mathcal{I}_d^\diamond}\gamma_{p,I}(\mathbf{r}_4\mathfrak{d}^2),\qquad\overline{\xi}_p:=\sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\bigl[\ \|\xi_{h_I}\|_p-\gamma_{p,I}(\mathbf{r}_4\mathfrak{d}^2)\mathcal{U}_p(h_I)\ \bigr]_+;\\ \overline{\mathbf{f}}_p\ &\leq\ \mathfrak{d}^2\|\mathbf{K}\|_1^d\Bigl[\overline{G}_p+\|\mathbf{K}\|_1^d\mathbf{f}_p\Bigr]^{\mathfrak{d}^2-1}\leq\mathfrak{d}^2\|\mathbf{K}\|_1^d\Bigl[1+\|\mathbf{K}\|_1^d\bigl(\overline{\xi}_p+\overline{\gamma}_p\overline{\mathcal{U}}_p+\|\mathbf{K}\|_1^d\mathbf{f}_p+\mathbf{f}_p\bigr)\ \Bigr]^{\mathfrak{d}^2}.\end{aligned}$$

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

$$\big(\mathbb{E}_f|Y^{\mathfrak{d}^2}|^{\mathbf{r}}\big)^{\frac{1}{\mathbf{r}}} = \big[\big(\mathbb{E}_f|Y|^{\mathbf{r}\mathfrak{d}^2}\big)^{\frac{1}{\mathbf{r}\mathfrak{d}^2}}\big]^{\mathfrak{d}^2}. \tag{67}$$

In view of Proposition 2, if $p \in (1, +\infty]$ and $n \geq 3$, $(\mathbb{E}_f|\overline{\mathbf{f}}_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_{p,2}(\mathbf{r}, \mathbf{f}_p)$ with

$$\mathbf{c}_{p,2}(\mathbf{r}, \mathbf{f}_p) := \mathfrak{d}^2\|\mathbf{K}\|_1^d\Big[1 + \|\mathbf{K}\|_1^d\big(d|\mathfrak{P}|^2 c_p(\mathbf{r}\mathfrak{d}^2) + \overline{\gamma}_p\overline{\mathcal{U}}_p + \|\mathbf{K}\|_1^d\mathbf{f}_p + \mathbf{f}_p\big)\Big]^{\mathfrak{d}^2}.$$

Thus, we finish the proof of Lemma 1. □

REFERENCES

1. Y. Baraud, and L. Birgé, "Estimating Composite Functions by Model Selection", Ann. Inst. H. Poincaré Probab. Statist. **50**, 285–314 (2014).
2. R. Benhaddou, M. Pensky, and D. Picard, "Anisotropic Denoising in Functional Deconvolution Model with Dimension-Free Convergence Rates", Electronic J. Statist. **7**, 1686–1715 (2013).
3. N. Bissantz, L. Dümbgen, H. Holzmann, and A. Munk, "Nonparametric Confidence Bands in Deconvolution Density Estimation", J. Roy. Statist. Soc., Ser. B **69**, 483–506 (2007).
4. C. Butucea, "Deconvolution of Supersmooth Densities with Smooth Noise", Canad. J. Statist. **32** (2), 181–192 (2004).
5. C. Butucea and F. Comte, "Adaptive Estimation of Linear Functionals in the Convolution Model and Applications", Bernoulli **15** (1), 69–68 (2009).
6. C. Butucea and A. B. Tsybakov, "Sharp Optimality in Density Deconvolution with Dominating Bias. I", Teor. Veroyatn. Primen. **52** (1), 111–128 (2007).
7. C. Butucea and A. B. Tsybakov, "Sharp Optimality in Density Deconvolution with Dominating Bias. II", Theory Probab. Appl. **52** (2), 237–249 (2008).
8. R. J. Carroll and P. Hall, "Optimal Rates of Convergence for Deconvolving a Density", J. Amer. Statist. Assoc. **83**, 1184–1186 (1988).
9. F. Comte and C. Lacour, "Anisotropic Adaptive Kernel Deconvolution", Ann. Inst. H. Poincaré Probab. Statist. **49** (2), 569–609 (2013).
10. F. Comte and T. Rebafka, "Adaptive Density Estimation in the Pile-Up Model Involving Measurement Errors", Electronic J. Statist. **6**, 2002–2037 (2012).
11. L. Devroye and G. Lugosi, "Nonasymptotic Universal Smoothing Factor, Kernel Complexity and Yatracos Classes", Ann. Statist. **25**, 2626–2637 (1997).
12. J. Fan, "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems", Ann. Statist. **19** (3), 1257–1272 (1991).
13. J. Fan, "Global Behavior of Deconvolution Kernel Estimates", Statist. Sinica **1**, 541–551 (1991).
14. J. Fan, "Adaptively Local One-Dimensional Subproblems with Application to a Deconvolution Problem", Ann. Statist. **21** (2), 600–610 (1993).
15. A. Goldenshluger and O. V. Lepski, "Structural Adaptation via $\mathbb{L}_p$-Norm Oracle Inequalities", Probab. Theory Rel. Fields **143**, 41–71 (2009).
16. A. Goldenshluger and O. Lepski, "Uniform Bounds for Norms of Sums of Independent Random Functions", Ann. Probab. **39**, 2318–2384 (2011a).
17. A. Goldenshluger and O. Lepski, "Bandwidth Selection in Kernel Density Estimation: Oracle Inequalities and Adaptive Minimax Optimality", Ann. Statist. **39**, 1608–1639 (2011b).
18. A. Goldenshluger and O. Lepski, "On Adaptive Minimax Density Estimation on $\mathbb{R}^d$", Probab. Theory and Rel. Fields **159**, 479–543 (2013).
19. L. Grafakos, *Classical Fourier Analysis*, in *Graduate Texts in Mathematics* (Springer, 2008), Vol. 249.
20. P. Hall and A. Meister, "A Ridge-Parameter Approach to Deconvolution", Ann. Statist. **35** (4), 1535–1558 (2007).
21. J. I. Horowitz and E. Mammen, "Rate-Optimal Estimation for a General Class of Nonparametric Regression Models with Unknown Link Functions", Ann. Statist. **35** (6), 2589–2619 (1990).
22. A. B. Juditsky, O. V. Lepski, and A. B. Tsybakov, "Nonparametric Estimation of Composite Functions", Ann. Statist. **37** (3), 1360–1440 (2009).
23. G. Kerkyacharian, O. V. Lepski, and D. Picard, "Nonlinear Estimation in Anisotropic Multi-Index Denoising", Probab. Theory and Rel. Fields **121**, 137–170 (2001).
24. O. Lepski, "Upper Functions for Positive Random Functionals. I. General Setting and Gaussian Random Functions", Math. Methods Statist. **22** (1), 1–27 (2013).
25. O. Lepski, "Multivariate Density Estimation Under Sup-Norm Loss: Oracle Approach, Adaptation and Independence Structure", Ann. Statist. **40** (2), 1005–1034 (2013).

26. O. V. Lepski and N. Serdyukova, "Adaptive Estimation under Single-Index Constraint in a Regression Model", Ann. Stat. **40** (1), 1–28 (2014).
27. O. Lepski and T. Willer, "Lower Bounds in the Convolution Structure Density Model", Bernoulli (2014) (forthcoming paper).
28. K. Lounici and R. Nickl, "Global Uniform Risk Bounds for Wavelet Deconvolution Estimators", Ann. Statist. **39** (2), 201–231 (2011).
29. E. Masry, "Multivariate Probability Density Deconvolution for Stationary Random Processes", IEEE Trans. Inform. Theory **37** (4), 1105–1115 (1991).
30. E. Masry, "Strong Consistency and Rates for Deconvolution of Multivariate Densities of Stationary Processes", Stochastic Proc. and Their Appl. **47**, 53–74 (1993).
31. P. Massart, *Concentration Inequalities and Model Selection*, in *Lecture Notes in Mathematics*, Vol. 1886: *Lecture from the 33rd Summer School on Probability Theory Held in Saint-Floor, July 6-23, 2003* (Springer, Berlin, 2007).
32. A. Meister, "Deconvolution from Fourier-Oscillating Error Densities under Decay and Smoothness Restrictions", Inverse Problems **24** (1), (2008).
33. S. M. Nikol'skii, *Approximation of Functions of Several Variables and Embedding Theorems* (Nauka, Moscow, 1977) [in Russian].
34. M. Pensky and B. Vidakovic, "Adaptive Wavelet Estimator for Nonparametric Density Deconvolution", Ann. Statist. **27** (6), 2033–2053 (1999).
35. G. Rebelles, "Pointwise Adaptive Estimation of a Multivariate Density under Independence Hypothesis", Bernoulli **21** (4), 1984–2023 (2014).
36. G. Rebelles, "$\mathbb{L}_p$ Adaptive Estimation of an Anisotropic Density under Independence Hypothesis", Electronic J. Statist. **9**, 106–134 (2015).
37. A. Samarov and A. B. Tsybakov, "Aggregation of Density Estimators and Dimension Reduction", in *Advances in Statistical Modeling and Inference*, *Ser. Biostat.*, 3, (World Sci. Publ., Hackensack, NJ, 2007), pp. 233–251.
38. L. A. Stefanski, "Rates of Convergence of Some Estimators in a Class of Deconvolution Problems", Statist. Probab. Lett. **9**, 229–235 (1990).
39. L. A. Stefanski and R. J. Carroll, "Deconvoluting Kernel Density Estimators", Statistics **21**, 169–184 (1990).
40. A. Tsybakov, *Introduction to Nonparametric Estimation*, in *Springer Series in Statistics* (Springer, New York, 2009).
41. E. Youndjé and M. T. Wells, "Optimal Bandwidth Selection for Multivariate Kernel Deconvolution Density Estimation", TEST **17** (1), 138–162 (2008).