

A new approach to estimator selection

O. V. LEPSKI

Aix Marseille Univ, CNRS, Centrale Marseille, I2M, 39, rue F. Joliot-Curie, 13453 Marseille, France.
E-mail: oleg.lepski@univ-amu.fr

In the framework of an abstract statistical model, we discuss how to use the solution of one estimation problem (*Problem A*) in order to construct an estimator in another, completely different, *Problem B*. As a solution of *Problem A* we understand a data-driven selection from a given family of estimators $\mathbf{A}(\mathfrak{H}) = \{\widehat{A}_{\mathfrak{h}}, \mathfrak{h} \in \mathfrak{H}\}$ and establishing for the selected estimator so-called oracle inequality. If $\hat{\mathfrak{h}} \in \mathfrak{H}$ is the selected parameter and $\mathbf{B}(\mathfrak{H}) = \{\widehat{B}_{\mathfrak{h}}, \mathfrak{h} \in \mathfrak{H}\}$ is an estimator's collection built in *Problem B*, we suggest to use the estimator $\widehat{B}_{\hat{\mathfrak{h}}}$. We present very general selection rule led to selector $\hat{\mathfrak{h}}$ and find conditions under which the estimator $\widehat{B}_{\hat{\mathfrak{h}}}$ is reasonable. Our approach is illustrated by several examples related to adaptive estimation.

Keywords: adaptive estimation; density model; generalized deconvolution model; oracle approach; upper function

1. Introduction

Let $(\mathcal{X}^{(n)}, \mathfrak{T}^{(n)}, \mathbb{P}_f^{(n)}, f \in \mathbb{F})$ be the statistical experiment generated by the observation $X^{(n)}$.

Let $A : \mathbb{F} \rightarrow \mathfrak{S}_1$ and $B : \mathbb{F} \rightarrow \mathfrak{S}_2$ be two mappings to be estimated and $\mathfrak{S}_1, \mathfrak{S}_2$ be sets endowed with semi-metrics ℓ and ρ , respectively.

For any $X^{(n)}$ -measurable \mathfrak{S}_j -valued map $\tilde{Q}_j, j = 1, 2$, and any $q \geq 1$ introduce

$$\mathcal{R}_A^q[\tilde{Q}_1, f] = \mathbb{E}_f^{(n)}[\ell(\tilde{Q}_1, A(f))]^q; \quad \mathcal{R}_B^q[\tilde{Q}_2, f] = \mathbb{E}_f^{(n)}[\rho(\tilde{Q}_2, B(f))]^q, \quad f \in \mathbb{F}.$$

Here and later $\mathbb{E}_f^{(n)}$ denotes the mathematical expectation w.r.t to the $\mathbb{P}_f^{(n)}$ and the number q is supposed to be fixed.

The main objective of the present paper can be described as follows. Assume that the problem of estimation of $A(\cdot)$, called furthermore *Problem A*, is much easier than the estimation of $B(\cdot)$ (*Problem B*). We will not precise here the exact meaning of “easier”, which may be the theoretical difficulty or computational complexity or something else. One can imagine, for instance, that *Problem A* has been already solved while *Problem B* is still not. It is also important to realize that *Problems A* and *B* may have completely different natures. For example, $A(f)$ is f , where $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is the multivariate probability density and $\ell = \|\cdot\|_p$ is \mathbb{L}_p -norm on \mathbb{R}^d , while $B(f)$ is functional, i.e. $\mathfrak{S}_2 = \mathbb{R}$. It can be the value of f or its derivatives at a given point, some norm of this function, the entropy functional or Fisher information and so on. Even if both problems have the same nature, for instance, $A(f) = B(f) = f$, the loss functions (semi-metrics ℓ and ρ) can be different. In particular, one can consider $\ell(\cdot) = \|\cdot\|_p$ and $\rho(\cdot) = \|\cdot\|_s, p \neq s$.

The problem which we address consists in finding hypotheses under which some elements of the solution of *Problem A* could be used in the construction of an estimation procedure for solving *Problem B*. Let us discuss this approach more in detail.

The variety of statistical procedures developed last quarter of century deal with the data-driven selection from the particular family of estimators, Barron *et al.* [3], Baraud *et al.* [2], Birgé and Massart [4], Bunea *et al.* [5], Cai [6], Cavalier and Tsybakov [8], Cavalier and Golubev [7], Dalalyan and Tsybakov [11], Devroye and Lugosi [12], Goldenshluger [16], Goldenshluger and Lepski [18,20,22], Kerkyacharian *et al.* [25], Lepski and Levit [34], Nemirovski [37], Rigollet and Tsybakov [41], Tsybakov [43], Wegkamp [47] among many others. A very detailed overview on this topic can be found in the recent paper Lepski [32].

Suppose that we are given by the collection of estimators $\mathbf{A}(\mathfrak{H}) = \{\widehat{A}_h, h \in \mathfrak{H}\}$, used for the estimation of the map $A(\cdot)$, parameterized by some parameter set \mathfrak{H} (\widehat{A}_h depend usually on n but we will omit this dependence in the notations). The quality of estimation is measured by the family of risks $\{\mathcal{R}_A[\widehat{A}_h, f], h \in \mathfrak{H}, f \in \mathbb{F}\}$. Let us say that *Problem A* is solved if one can find $X^{(n)}$ -measurable element $\hat{h} \in \mathfrak{H}$ (data-driven selector) such that the selected estimator $\widehat{A}_{\hat{h}}$ satisfies so-called *oracle-type* inequality:

$$\mathcal{R}_A[\widehat{A}_{\hat{h}}, f] \leq \inf_{h \in \mathfrak{H}} \mathcal{A}_\ell^{(n)}(f, h) + cr_n, \quad \forall f \in \mathbb{F}, \forall n \geq 1. \tag{1.1}$$

Here $c > 0$ is a numerical constant independent on n and f and $r_n \rightarrow 0, n \rightarrow \infty$ is given sequence. As to the quantity $\mathcal{A}_\ell^{(n)}(\cdot, \cdot)$ it is explicitly expressed and for some particular problems one can prove the inequality (1.1) with

$$\mathcal{A}_\ell^{(n)}(f, h) = C\mathcal{R}_A[\widehat{A}_h, f],$$

where C is as previously a universal constant.

Let $\mathbf{B}(\mathfrak{H}) = \{\widehat{B}_h, h \in \mathfrak{H}\}$ be a given collection of statistical procedures related to the estimation of $B(\cdot)$. It is extremely important for us that both collections $\mathbf{A}(\mathfrak{H})$ and $\mathbf{B}(\mathfrak{H})$ are parameterized by one and the same set \mathfrak{H} . As it was mentioned above *Problem A* and *Problem B* may have different nature and often only the set \mathfrak{H} relates both problems.

The questions which we want to answer are: *under which conditions the selector \hat{h} provides a reasonable choice from the collection $\mathbf{B}(\mathfrak{H})$? Is it possible to establish an oracle inequality similar to (1.1) for the selected estimator $\widehat{B}_{\hat{h}}$?*

We do not think that the answers on aforementioned questions can be obtained when an arbitrary selection rule led to \hat{h} is considered. So, in the next section in the framework of an abstract statistical model we propose a quite general selection scheme and establish for it the oracle inequality (1.1). This part of the paper has an independent interest since all results will be obtained under few very general assumptions which can be verified for many statistical models and problems. The proposed approach can be viewed as a generalization of several estimation procedures developed by the author and his collaborators during last twenty years, see Lepski and Levit [34], Kerkyacharian *et al.* [25], Juditsky *et al.* [24], Goldenshluger and Lepski [17,18,20–22], Lepski [28,32].

Coming back to the study of the behavior of the “plug-in” estimator $\widehat{B}_{\hat{h}}$ we would like to emphasize that it will be done under the following assumption imposed on the statistical experiment. We will assume that $X^{(n)} = (X_1^{(n)}, X_2^{(n)})$, where $X_1^{(n)}, X_2^{(n)}$ are independent random elements. Our selection rule (led in particular to the solution of *Problem A*) is based on the observation $X_1^{(n)}$ while the estimator’s collection $\mathbf{B}(\mathfrak{H})$ is built from the observation $X_2^{(n)}$. This

fundamental assumption may correspond to splitting data on two independent samples or to the situation when a statistician disposes an independent copy of the considered statistical model.

The following notations will be used in the sequel: $\mathbb{P}_{1,f}^{(n)}$ and $\mathbb{P}_{2,f}^{(n)}$ denote marginal laws of $X_1^{(n)}$ and $X_2^{(n)}$, respectively and $\mathbb{E}_{i,f}^{(n)}, i = 1, 2$, will be used for the mathematical expectation w.r.t. $\mathbb{P}_{i,f}^{(n)}$.

We finish this introduction presenting several examples in which the discussed above strategy can be applied. These considerations will be continued in Sections 4 and 5.

Estimation in the deconvolution density model under \mathbb{L}_p -loss. Consider the following observation scheme:

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \tag{1.2}$$

where $X_i, i = 1, \dots, n$, are *i.i.d.* d -dimensional random vectors with common density f to be estimated. The noise variables $Y_i, i = 1, \dots, n$, are *i.i.d.* d -dimensional random vectors with known common density g . The sequences $\{X_i, i = 1, \dots, n\}$ and $\{Y_i, i = 1, \dots, n\}$ are supposed to be mutually independent.

Let $X^{(n)} = (Z_1, \dots, Z_n)$ and let $B(f) = f$ and $\rho(\cdot) = \ell(\cdot) = \|\cdot\|_p$ that means that we are interested in estimation of the underlying density f under \mathbb{L}_p -loss.

Let $K : \mathbb{R}^d \rightarrow \mathbb{R}$ be the function belonging to $\mathbb{L}_1(\mathbb{R}^d)$ and $\int_{\mathbb{R}} K = 1$. Let \mathcal{H}^d be the diadic grid of $(0, 1]^d$ and define for any $\vec{h} = (h_1, \dots, h_d) \in \mathcal{H}^d$

$$K_{\vec{h}}(t) = V_{\vec{h}}^{-1} K(t_1/h_1, \dots, t_d/h_d), \quad t \in \mathbb{R}^d, V_{\vec{h}} = \prod_{j=1}^d h_j. \tag{1.3}$$

For any $\vec{h} \in \mathcal{H}^d$ let $M(\cdot, \vec{h}) : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfy the operator equation

$$K_{\vec{h}}(y) = \int_{\mathbb{R}^d} g(t - y)M(t, \vec{h}) dt, \quad y \in \mathbb{R}^d. \tag{1.4}$$

Introduce the following estimator's collection

$$\mathbf{B}(\mathcal{H}^d) = \left\{ \widehat{B}_{\vec{h}}(\cdot) = n^{-1} \sum_{i=1}^n M(Z_i - \cdot, \vec{h}), \vec{h} \in \mathcal{H}^d \right\}. \tag{1.5}$$

In Comte and Lacour [10] and in Rebelles [39], the data-driven selection rules from $\mathbf{B}(\mathcal{H}^d)$, based on the methodology developed in Goldenshluger and Lepski [22], were proposed. The authors established oracle inequalities of type (1.1) and deduced from them several results related to adaptive estimation.

Our idea is quite different and consists in the following. Consider first the estimation of $A(f) = g \star f$, where here and later “ \star ” denotes the convolution operator. Note that $g \star f$ is the density of Z_1 and, therefore, can be easily estimated from the observation $X^{(n)}$ using standard kernel estimator. Consider the collection

$$\mathbf{A}(\mathcal{H}^d) = \left\{ \widehat{A}_{\vec{h}}(\cdot) = \sum_{i=1}^n K_{\vec{h}}(Z_i - \cdot), \vec{h} \in \mathcal{H}^d \right\}. \tag{1.6}$$

The problem of bandwidth selection from $\mathbf{A}(\mathcal{H}^d)$ was studied in Goldenshluger and Lepski [20, 21] where several oracle inequalities were proved. Let $\vec{\mathbf{h}}_n$ be a selected bandwidth. We propose then to use $\widehat{B}_{\vec{\mathbf{h}}_n}(\cdot)$ as the final estimator.

We remark that similar strategy in the linear regression model (selection from the family of spectral cut-off estimators, $d = 1$) was adopted in Chernousova and Golubev [9]. In the sequence space Gaussian white noise model the approach discussed above was applied in Knapik and Solomond [27] in order to find the posterior contraction rate in inverse problems in the context of the Bayesian nonparametrics.

Estimation of derivatives in the density model under \mathbb{L}_p -loss. Let $X^{(n)} = (X_1, \dots, X_n)$ be *i.i.d.* random variables with unknown common density f (the multidimensional version of the problem will be presented in Section 5). We are interested in estimating of $f^{(m)}$, $m \in \mathbb{N}^*$, where $f^{(m)}$ denotes m th derivative of f .

Thus, $B_m(f) = f^{(m)}$ and let $\rho(\cdot) = \ell(\cdot) = \|\cdot\|_p$. Set $A(f) = f$ and consider the family of kernel estimators

$$\mathbf{A}(\mathcal{H}^1) = \left\{ \widehat{A}_h(\cdot) = \sum_{i=1}^n K_h(X_i - \cdot), h \in \mathcal{H}^1 \right\}.$$

It is worth noting that for any $m \in \mathbb{N}^*$ the estimator $\widehat{B}_{h,m}(\cdot) = \widehat{A}_h^{(m)}(\cdot)$ is usually used for estimating $B_m(f)$. Hence, one of the possibilities consists in the selection from the collection

$$\mathbf{B}_m(\mathcal{H}^1) = \{ \widehat{B}_{h,m}(\cdot), h \in \mathcal{H}^1 \}$$

in order to construct an estimator for $f^{(m)}$. Note, however that in this case the corresponding selector as well as the selected estimator will depend on m . In particular, the selection schemes are different for different values of m .

Our approach to considered problem consists in selecting from the family $\mathbf{A}(\mathcal{H}^1)$ that provides us with the selector \mathbf{h}_n . Then for any $m \in \mathbb{N}^*$, we suggest to use $\widehat{B}_{\mathbf{h}_n,m}(\cdot)$ as an estimator for $B_m(f)$. In other words the problem we address can be formulated as follows. Is it possible to differentiate the kernel estimator with data-driven bandwidth in order to get the estimator for any derivative of the underlying density simultaneously?

In the framework of Gaussian white noise model, this problem was studied in Efromovich [13]. It was shown that the answer is positive when $p = 2$ and the adaptation is considered over the collection of Sobolev classes. More precisely, it was shown that the differentiation of the Efromovich–Pinsker estimator leads to the efficient adaptive estimator of any derivative. In Abramovich *et al.* [1], the same answer was obtained when the collection of Besov classes is considered. It was shown ($p = 2, d = 1$) that the differentiation of special wavelet based adaptive estimator leads to the adaptive estimation of the corresponding derivative.

In Section 5, we prove that the answer on aforementioned question is positive when the adaptive estimation of partial derivatives of a function belonging to an anisotropic Nikol’skii class in $\mathbb{L}_p(\mathbb{R}^d)$ is considered under an arbitrary \mathbb{L}_p -loss and in an arbitrary dimension $d \geq 1$.

2. Selection scheme for solving of *Problem A*

Let $\mathfrak{H}_n, n \in \mathbb{N}^*$, be a sequence of countable subsets of \mathfrak{H} . Let $\{\widehat{A}_h, h \in \mathfrak{H}\}$ and $\{\widehat{A}_{h,\eta}, h, \eta \in \mathfrak{H}\}$ be the families of $X_1^{(n)}$ -measurable \mathfrak{S}_1 -valued mappings possessing the properties formulated below. Both \widehat{A}_h and $\widehat{A}_{h,\eta}$ depend usually on n but we will omit this dependence for the sake of simplicity of notations. Let $\varepsilon_n \rightarrow 0, n \rightarrow \infty$ be a given sequence.

Suppose there exist collections of \mathfrak{S}_1 -valued functionals $\{\Lambda_h(f), h \in \mathfrak{H}\}, \{\Lambda_{h,\eta}(f), h, \eta \in \mathfrak{H}\}$ and a collection of *positive* $X_1^{(n)}$ -measurable random variables $\Psi_n = \{\Psi_n(h), h \in \mathfrak{H}\}$ for which the following conditions hold. (The functionals Λ_h and $\Lambda_{h,\eta}$ may depend on n (not necessarily) but we will omit this dependence in the notations.)

$\mathbf{A}^{\text{permute}}$. $\widehat{A}_{h,\eta} \equiv \widehat{A}_{\eta,h}$, for any $h, \eta \in \mathfrak{H}_n$.

$\mathbf{A}^{\text{upper}}$. For any $n \geq 1$

$$(i) \sup_{f \in \mathbb{F}} \mathbb{E}_{1,f}^{(n)} \left(\sup_{h \in \mathfrak{H}_n} [\ell(\widehat{A}_h, \Lambda_h(f)) - \Psi_n(h)]_+^q \right) \leq \varepsilon_n^q;$$

$$(ii) \sup_{f \in \mathbb{F}} \mathbb{E}_{1,f}^{(n)} \left(\sup_{h,\eta \in \mathfrak{H}_n} [\ell(\widehat{A}_{h,\eta}, \Lambda_{h,\eta}(f)) - \{\Psi_n(h) \wedge \Psi_n(\eta)\}]_+^q \right) \leq \varepsilon_n^q.$$

Remark 1. Often the collection $\{\Psi_n(h), h \in \mathfrak{H}\}$ satisfying the hypothesis $\mathbf{A}^{\text{upper}}$ is not random. This is typically the case when a statistical problem is studied in the framework of white gaussian noise or regression model. See also Lemma 1 below, case $p \leq 2$.

2.1. Discussion

For many statistical models and problems $\Lambda_h(f) = \mathbb{E}_{1,f}^{(n)}(\widehat{A}_h)$ and $\Lambda_{h,\eta}(f) = \mathbb{E}_{1,f}^{(n)}(\widehat{A}_{h,\eta})$. In this case $\ell(\widehat{A}_h, \Lambda_h(f))$ and $\ell(\widehat{A}_{h,\eta}, \Lambda_{h,\eta}(f))$ can be viewed as stochastic errors related to the estimators \widehat{A}_h and $\widehat{A}_{h,\eta}$ respectively. Hence, following the terminology used in Lepski [33] we can say that $\{\Psi_n(h), h \in \mathfrak{H}\}$ and $\{\Psi_n(h) \wedge \Psi_n(\eta), h, \eta \in \mathfrak{H}\}$ are upper functions of level ε_n for the collection of corresponding stochastic errors.

The development of the probabilistic tools allowing to control the behavior of stochastic errors related to statistical procedures is the subject of vast literature, see, for instance, the books van der Vaart and Wellner [45], van de Geer [44], Massart [36]. The inequalities similar to those appeared in the hypothesis $\mathbf{A}^{\text{upper}}$ can be found in Egishyants and Ostrovskii [14] and Lepski [29–31]. The upper functions for \mathbb{L}_p -norm of “kernel-type” empirical and Gaussian processes were obtained in Goldenshluger and Lepski [19] and Lepski [33].

We conclude that from the one hand the verification of the hypothesis $\mathbf{A}^{\text{upper}}$ is in some sense necessary task when the oracle approach or adaptive estimation is considered. On the other hand the very developed probabilistic machinery is in our disposal.

Let us also remark that if the analogues of hypotheses $\mathbf{A}^{\text{upper}}$ and $\mathbf{A}^{\text{permute}}$ can be checked in *Problem B*, then the selection rule from the estimator’s collection $\{\widehat{B}_h, h \in \mathfrak{H}\}$ presented below will provide the solution of *Problem B*. However in some cases the verification of these hypotheses is much more difficult in *Problem B* than in *Problem A* and it is one of the reasons why we propose to proceed differently.

Let us now discuss some examples of estimator’s collections for which the hypothesis $\mathbf{A}^{\text{permute}}$ is fulfilled.

Example 1. Let $\mathbf{A}(\mathcal{H}^d)$ and $\mathbf{B}(\mathcal{H}^d)$ be the families defined in (1.5) and (1.6). For any $\vec{h}, \vec{h}' \in \mathcal{H}^d$ set $\vec{h} \vee \vec{h}' = (h_1 \vee h'_1, \dots, h_d \vee h'_d)$ and introduce

$$\widehat{A}_{\vec{h}, \vec{h}'}(\cdot) = \widehat{A}_{\vec{h} \vee \vec{h}'}(\cdot), \quad \widehat{B}_{\vec{h}, \vec{h}'}(\cdot) = \widehat{B}_{\vec{h} \vee \vec{h}'}(\cdot).$$

It is obvious that the hypothesis $\mathbf{A}^{\text{permute}}$ is verified for these estimators. The selection rules based on this construction of families of auxiliary estimators can be found in Kerkyacharian *et al.* [25, 26], Rebelles [39].

Example 2. Consider either the density model generated by the observation $X^{(n)} = (X_1, \dots, X_n)$, where $X_i \in \mathbb{R}^d, i = 1, \dots, n$, are i.i.d. random vectors or the observation model (1.2).

Let $K_{\vec{h}}(\cdot)$ and $M(\cdot, \vec{h}), \vec{h} \in \mathcal{H}^d$, be defined in (1.3) and (1.4), respectively. For any $\vec{h}, \vec{h}' \in \mathcal{H}^d$ set

$$K_{\vec{h}, \vec{h}'}(\cdot) = \int_{\mathbb{R}^d} K_{\vec{h}}(\cdot - t) K_{\vec{h}'}(t) dt, \quad M(\cdot, \vec{h}, \vec{h}') = \int_{\mathbb{R}^d} M(\cdot - t, \vec{h}) M(t, \vec{h}') dt$$

and define

$$\widehat{A}_{\vec{h}, \vec{h}'}(\cdot) = \sum_{i=1}^n K_{\vec{h}, \vec{h}'}(X_i - \cdot) \quad \text{or} \quad \widehat{A}_{\vec{h}, \vec{h}'}(\cdot) = \sum_{i=1}^n K_{\vec{h}, \vec{h}'}(Z_i - \cdot),$$

$$\widehat{B}_{\vec{h}, \vec{h}'}(\cdot) = \sum_{i=1}^n M(Z_i - \cdot, \vec{h}, \vec{h}').$$

Since, $K_{\vec{h}, \vec{h}'} \equiv K_{\vec{h}', \vec{h}}$ and $M(\vec{h}, \vec{h}') \equiv M(\vec{h}', \vec{h})$ the hypothesis $\mathbf{A}^{\text{permute}}$ holds. The selection rules based on this construction of auxiliary estimators were proposed in Lepski and Levit [34], Goldenshluger and Lepski [17,18,20,21], Comte and Lacour [10], Reynaud-Bouret *et al.* [40].

Example 3. Let \mathcal{D} be a set endowed with the Borel measure μ and let $\{\psi_k, k \in \mathbb{N}^d\}$ be an orthogonal basis in $\mathbb{L}_2(\mathcal{D}, \mu)$ possessing the following properties.

$$\psi_{\mathbf{0}} \equiv \mathbf{c} \neq 0, \quad \int_{\mathcal{D}} \psi_k(t) \mu(dt) = 0, \quad \forall k \neq \mathbf{0} = (0, \dots, 0).$$

Let $\mathbb{T} = \{\tau = (\tau_k, k \in \mathbb{N}^d)\}$ be a given subset of l_2 such that $\tau_{\mathbf{0}} c^2 \mu(\mathcal{D}) = 1$ for all $\tau \in \mathbb{T}$. Introduce

$$K_\tau(t, x) = \sum_{k \in \mathbb{N}^d} \tau_k \psi_k(t) \psi_k(x), \quad \tau \in \mathbb{T},$$

and define for any $\tau, \tau' \in \mathbb{T}$

$$K_{\tau, \tau'}(t, x) = \int_{\mathcal{D}} K_{\tau}(t, y) K_{\tau'}(y, x) \mu(dy).$$

Consider the statistical experiment generated by the observation $X^{(n)} = (X_1, \dots, X_n)$, where $X_i, i = 1, \dots, n$, are i.i.d. \mathcal{D} -valued random variables having unknown density f with respect to the measure μ . Introduce the following collection of estimators

$$\widehat{A}_{\tau, \tau'}(\cdot) = \sum_{i=1}^n K_{\tau, \tau'}(X_i, \cdot), \quad \tau, \tau' \in \mathbb{T}.$$

As it was shown in Goldenshluger and Lepski [22], $K_{\tau, \tau'} \equiv K_{\tau', \tau}$ for any $\tau, \tau' \in \mathbb{T}$ and, therefore, the hypothesis $\mathbf{A}^{\text{permute}}$ is fulfilled.

We remark that all estimator’s construction discussed above can be applied in other statistical models where kernel-type estimators are used, for example, white Gaussian noise model and the regression model.

2.2. (Ψ_n, ℓ) -selection rule and corresponding oracle inequality

Our objective is to propose the selection rule from an arbitrary collection $\mathbf{A}(\mathfrak{H}_n) = \{\widehat{A}_h, h \in \mathfrak{H}_n\}$ satisfying hypotheses $\mathbf{A}^{\text{permute}}$ and $\mathbf{A}^{\text{upper}}$, and establish for it the oracle-type inequality (1.1).

Define for any $h \in \mathfrak{H}_n$

$$\widehat{R}_n(h) = \sup_{\eta \in \mathfrak{H}_n} [\ell(\widehat{A}_{h, \eta}, \widehat{A}_{\eta}) - 2\Psi_n(\eta)]_+.$$

Let $\widehat{h}^{(n)} \in \mathfrak{H}_n$ be an arbitrary $X_1^{(n)}$ -measurable random element satisfying

$$\widehat{R}_n(\widehat{h}^{(n)}) + 2\Psi_n(\widehat{h}^{(n)}) \leq \inf_{h \in \mathfrak{H}_n} \{\widehat{R}_n(h) + 2\Psi_n(h)\} + \varepsilon'_n,$$

where ε'_n is an arbitrary sequence and we will assume that $\varepsilon'_n \leq \varepsilon_n$.

Introduce the following notation: for any $f \in \mathbb{F}, h \in \mathfrak{H}_n$ and $n \geq 1$

$$\mathcal{B}_A^{(n)}(f, h) = \ell(\Lambda_h(f), A(f)) + 2 \sup_{\eta \in \mathfrak{H}_n} \ell(\Lambda_{h, \eta}(f), \Lambda_{\eta}(f)), \quad \psi_n(f, h) = [\mathbb{E}_{1, f}^{(n)} \{\Psi_n^q(h)\}]^{\frac{1}{q}}.$$

Theorem 1. *Let $\mathbf{A}^{\text{permute}}$ and $\mathbf{A}^{\text{upper}}$ be fulfilled. Then, for any $f \in \mathbb{F}$ and $n \geq 1$*

$$\mathcal{R}_A[\widehat{A}_{\widehat{h}^{(n)}}, f] \leq \inf_{h \in \mathfrak{H}_n} \{\mathcal{B}_A^{(n)}(f, h) + 5\psi_n(f, h)\} + 6\varepsilon_n.$$

Note that the proposed (Ψ_n, ℓ) selection rule is built using only observation part $X_1^{(n)}$. This means that the application of the result presented in Theorem 1 does not require any splitting of data since one can formally suppose that $X_1^{(n)}$ is the original data set.

2.3. Proof of Theorem 1

For the simplicity of notations throughout the proof, we will write \hat{h} instead of $\hat{h}^{(n)}$.

Fix $h \in \mathfrak{H}_n$. We have obviously in view of the definition of $\widehat{R}_n(\cdot)$

$$\begin{aligned} \ell(\widehat{A}_{\hat{h}}, \widehat{A}_{h, \hat{h}}) &\leq 2\Psi_n(\hat{h}) + [\ell(\widehat{A}_{\hat{h}}, \widehat{A}_{h, \hat{h}}) - 2\Psi_n(\hat{h})]_+ \\ &\leq 2\Psi_n(\hat{h}) + \widehat{R}_n(h). \end{aligned} \tag{2.1}$$

Here we have also used that $\hat{h} \in \mathfrak{H}_n$ by its definition. Taking into account $\mathbf{A}^{\text{permute}}$ we get

$$\begin{aligned} \ell(\widehat{A}_h, \widehat{A}_{h, \hat{h}}) &= \ell(\widehat{A}_h, \widehat{A}_{\hat{h}, h}) \leq 2\Psi_n(h) + [\ell(\widehat{A}_h, \widehat{A}_{\hat{h}, h}) - 2\Psi_n(h)]_+ \\ &\leq 2\Psi_n(h) + \widehat{R}_n(\hat{h}). \end{aligned} \tag{2.2}$$

We get from (2.1), (2.2) and the definition of \hat{h}

$$\begin{aligned} \ell(\widehat{A}_{\hat{h}}, \widehat{A}_{h, \hat{h}}) + \ell(\widehat{A}_h, \widehat{A}_{h, \hat{h}}) &\leq \widehat{R}_n(\hat{h}) + 2\Psi_n(\hat{h}) + \widehat{R}_n(h) + 2\Psi_n(h) \\ &\leq 2\{\widehat{R}_n(h) + 2\Psi_n(h)\} + \varepsilon'_n. \end{aligned} \tag{2.3}$$

We have in view of the triangle inequality for any $h \in \mathfrak{H}_n$

$$\widehat{R}_n(h) \leq \sup_{\eta \in \mathfrak{H}_n} \ell(\Lambda_{h, \eta}(f), \Lambda_{\eta}(f)) + \xi_1 + \xi_2, \tag{2.4}$$

where we have put

$$\begin{aligned} \xi_1 &= \sup_{\eta \in \mathfrak{H}_n} [\ell(\widehat{A}_{\eta}, \Lambda_{\eta}) - \Psi_n(\eta)]_+, \\ \xi_2 &= \sup_{h, \eta \in \mathfrak{H}_n} [\ell(\widehat{A}_{h, \eta}, \Lambda_{h, \eta}) - \{\Psi_n(h) \wedge \Psi_n(\eta)\}]_+. \end{aligned} \tag{2.5}$$

Thus, we obtain from (2.3) and (2.4) for any $h \in \mathfrak{H}_n$

$$\ell(\widehat{A}_{\hat{h}}, \widehat{A}_{h, \hat{h}}) + \ell(\widehat{A}_h, \widehat{A}_{h, \hat{h}}) \leq 2 \sup_{\eta \in \mathfrak{H}_n} \ell(\Lambda_{h, \eta}(f), \Lambda_{\eta}(f)) + 4\Psi_n(h) + 2\xi_1 + 2\xi_2 + \varepsilon'_n. \tag{2.6}$$

Obviously for any $h \in \mathfrak{H}_n$

$$\ell(\widehat{A}_h, A(f)) \leq \ell(\Lambda_h(f), A(f)) + \Psi_n(h) + \xi_1.$$

It yields together with (2.6) by the triangle inequality

$$\ell(\widehat{A}_{\hat{h}}, A(f)) \leq \mathcal{B}_A^{(n)}(f, h) + 5\Psi_n(h) + 3\xi_1 + 2\xi_2 + \varepsilon'_n, \quad \forall h \in \mathfrak{H}_n, \forall f \in \mathbb{F}. \tag{2.7}$$

Taking into account the hypothesis $\mathbf{A}^{\text{upper}}$ we get for any $h \in \mathfrak{H}_n$ and any $f \in \mathbb{F}$

$$\{\mathbb{E}_{1, f}^{(n)}[\ell(\widehat{A}_{\hat{h}}, A(f))]^q\}^{\frac{1}{q}} \leq \mathcal{B}_A^{(n)}(f, h) + 5\psi_n(f, h) + 6\varepsilon_n.$$

Here we have used that $\varepsilon'_n \leq \varepsilon_n$. Noting that the left-hand side of the obtained inequality is independent of \mathfrak{h} we come to the assertion of the theorem.

3. “Plug-in” estimator for solving of *Problem B* and its properties

Let $\mathbf{B}(\mathfrak{H}) = \{\widehat{B}_{\mathfrak{h}}, \mathfrak{h} \in \mathfrak{H}\}$ be the family of $X_2^{(n)}$ -measurable \mathfrak{S}_2 -valued mappings. The goal of this section is to bound from above the risk of “plug-in” estimator $\widehat{B}_{\widehat{\mathfrak{h}}^{(n)}}$, where $\widehat{\mathfrak{h}}^{(n)}$ is obtained by (Ψ_n, ℓ) -selection rule.

We add the following assumption to the hypothesis $\mathbf{A}^{\text{upper}}$.

$\mathcal{A}^{\text{upper}}$. *There exist the constant $C_\Psi \geq 1$ such that for any $n \geq 1$*

$$\begin{aligned} \sup_{f \in \mathbb{F}} \mathbb{E}_{1,f}^{(n)} \left(\sup_{\mathfrak{h} \in \mathfrak{H}_n} [\Psi_n(\mathfrak{h}) - C_\Psi \psi_n(f, \mathfrak{h})]_+^q \right) &\leq \varepsilon_n^q, \\ \sup_{f \in \mathbb{F}} \mathbb{E}_{1,f}^{(n)} \left(\sup_{\mathfrak{h} \in \mathfrak{H}_n} [\psi_n(f, \mathfrak{h}) - C_\Psi \Psi_n(\mathfrak{h})]_+^q \right) &\leq \varepsilon_n^q. \end{aligned}$$

We note that the hypothesis $\mathcal{A}^{\text{upper}}$ is fulfilled with $C_\Psi = 1$ if $\Psi_n(\mathfrak{h})$ is non-random and in this case $\psi_n(\cdot, \cdot)$ is independent of f . Actually the hypothesis $\mathcal{A}^{\text{upper}}$ guarantees that the random function $\Psi_n(\cdot)$ is well-concentrated around some non-random mapping. We would like to stress that for all known to the author problems in order to check $\mathbf{A}^{\text{upper}}$ one first verifies that the required inequalities hold for the “non-random” mapping $\psi_n(f, \mathfrak{h})$. However, it may depend on unknown f and, therefore, cannot be used in the estimation construction. In this case this quantity is replaced by its empirical counterpart $\Psi_n(\mathfrak{h})$ satisfying the hypothesis $\mathcal{A}^{\text{upper}}$.

Let $\{\Upsilon_{\mathfrak{h}}, \mathfrak{h} \in \mathfrak{H}\}$ be a collection of \mathfrak{S}_2 -valued functionals. Set for any $\mathfrak{h} \in \mathfrak{H}$, $f \in \mathbb{F}$ and $n \geq 1$

$$\mathcal{E}_n(f, \mathfrak{h}) = \left(\mathbb{E}_{2,f}^{(n)} \left\{ \rho^q \left(\widehat{B}_{\mathfrak{h}}, \Upsilon_{\mathfrak{h}}(f) \right) \right\} \right)^{\frac{1}{q}}. \tag{3.1}$$

Introduce the following sets of parameters.

$$\begin{aligned} \mathfrak{W}_n(f) &= \left\{ \mathfrak{h} \in \mathfrak{H}_n : \psi_n(f, \mathfrak{h}) < 2C_\Psi^2 \inf_{\eta \in \mathfrak{H}_n} [\mathcal{B}_A^{(n)}(f, \eta) + 2\psi_n(f, \eta)] \right\}; \\ \mathfrak{L}_n(f) &= \left\{ \mathfrak{h} \in \mathfrak{H}_n : \ell(\Lambda_{\mathfrak{h}}, A(f)) < 4C_\Psi \inf_{\eta \in \mathfrak{H}_n} [\mathcal{B}_A^{(n)}(f, \eta) + 2\psi_n(f, \eta)] \right\}. \end{aligned}$$

Set $\delta_n = \inf_{f \in \mathbb{F}} \inf_{\mathfrak{h} \in \mathfrak{H}_n} \psi(f, \mathfrak{h})$ and define for any $f \in \mathbb{F}$ and $n \geq 1$

$$\begin{aligned} \tau_n(f) &= \sup_{\mathfrak{h} \in \mathfrak{W}_n(f)} \mathcal{E}_n(f, \mathfrak{h}) + (5\varepsilon_n/\delta_n) \sup_{\mathfrak{h} \in \mathfrak{H}_n} \mathcal{E}_n(f, \mathfrak{h}); \\ \nu_n(f) &= \sup_{\mathfrak{h} \in \mathfrak{L}_n(f)} \rho(\Upsilon_{\mathfrak{h}}, B(f)) + (30\varepsilon_n/\delta_n) \sup_{\mathfrak{h} \in \mathfrak{H}_n} \rho(\Upsilon_{\mathfrak{h}}, B(f)). \end{aligned}$$

Let $\widehat{\mathfrak{h}}^{(n)}$ is obtained by (Ψ_n, ℓ) -selection rule with $\varepsilon'_n \leq \delta_n/4$.

Theorem 2. Let $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$ hold. Then, for any $f \in \mathbb{F}$ and $n \geq 1$

$$\mathcal{R}_B[\widehat{B}_{\mathfrak{h}^{(n)}}(f)] \leq \nu_n(f) + \tau_n(f).$$

We will see that in particular examples the sequence ε_n decreases to zero very rapidly. It allows often to assert that

$$\varkappa_n := (5\varepsilon_n/\delta_n) \sup_{f \in \mathbb{F}} \left(\sup_{\mathfrak{h} \in \mathfrak{H}_n} \mathcal{E}_n(f, \mathfrak{h}) + 6 \sup_{\mathfrak{h} \in \mathfrak{H}_n} \rho(\Upsilon_{\mathfrak{h}}, B(f)) \right) \rightarrow 0, \quad n \rightarrow \infty$$

very fast and the statement of the theorem is reduced to

$$\mathcal{R}_B[\widehat{B}_{\mathfrak{h}^{(n)}}(f)] \leq \sup_{\mathfrak{h} \in \mathfrak{L}_n(f)} \rho(\Upsilon_{\mathfrak{h}}, B(f)) + \sup_{\mathfrak{h} \in \mathfrak{V}_n(f)} \mathcal{E}_n(f, \mathfrak{h}) + \varkappa_n. \quad (3.2)$$

Remark 2. Note that the sets $\mathfrak{V}_n(f)$ and $\mathfrak{L}_n(f)$ are completely determined by the quantities appeared in the solution of *Problem A* while $\mathcal{E}_n(f, \mathfrak{h})$ and $\rho(\Upsilon_{\mathfrak{h}}, B(f))$ represent respectively the standard deviation of stochastic error and approximation error of the estimator $\widehat{B}_{\mathfrak{h}}$ used in *Problem B*. The functionals $\tau_n(f)$ and $\nu_n(f)$ explain how all these quantities are related.

Although, (3.2) is simpler to analyze than the assertion of the theorem, even this oracle inequality is less natural than those proved in Theorem 1. The next result allows to understand better the properties of the estimator $\widehat{B}_{\mathfrak{h}^{(n)}}$.

Proposition 1. Let $\mathbf{A}^{\text{permute}}$ and $\mathbf{A}^{\text{upper}}$ hold. Suppose additionally that there exists a constant $C_\ell > 0$ such that for any $f \in \mathbb{F}$, $n \geq 1$ and $\mathfrak{h} \in \mathfrak{H}_n$

$$\rho(\Upsilon_{\mathfrak{h}}(f), B(f)) \leq C_\ell \ell(\Lambda_{\mathfrak{h}}(f), A(f)). \quad (3.3)$$

Assume also that there exists a constant $C_\mathcal{E} > 0$ such that for any $n \geq 1$ and $\mathfrak{h} \in \mathfrak{H}_n$

$$\sup_{f \in \mathbb{F}} \mathcal{E}_n(f, \mathfrak{h}) \leq C_\mathcal{E} \Psi_n(\mathfrak{h}). \quad (3.4)$$

Then, for any $f \in \mathbb{F}$ and $n \geq 1$

$$\begin{aligned} \mathcal{R}_B[\widehat{B}_{\mathfrak{h}^{(n)}}(f)] &\leq \inf_{\mathfrak{h} \in \mathfrak{H}_n} \left\{ (2C_\ell + C_\mathcal{E}) \mathcal{B}_A^{(n)}(f, \mathfrak{h}) + (7C_\ell + 2C_\mathcal{E}) \Psi_n(f, \mathfrak{h}) \right\} \\ &\quad + (10C_\ell + 3C_\mathcal{E}) \varepsilon_n. \end{aligned}$$

We remark that the assertions of Theorem 1 and of Proposition 1 differ only by the numerical constant. We would like to emphasize that although Proposition 1 holds under quite restrictive assumption (3.3) it is useful for some problems studied in Section 4. Moreover, it does not require the hypothesis $\mathcal{A}^{\text{upper}}$.

What does make possible the assertion of Theorem 2? The goal of this paragraph is to discuss on informal way which properties of (Ψ_n, ℓ) -selection rule lead to the assertion of Theorem 2. Introduce the “oracle nuisance parameter” h_n^{oracle} which is a minimizer of the right-hand side of oracle-type inequality proved in Theorem 1, that is

$$h_n^{\text{oracle}} = \arg \inf_{h \in \mathfrak{H}_n} \{ \mathcal{B}_A^{(n)}(f, h) + \psi_n(f, h) \}. \tag{3.5}$$

Without loss of generality, we will assume that $h_n^{\text{oracle}} \in \mathfrak{H}_n$.

The first property that we are going to discuss is the following: with “high probability”

$$\psi_n(f, \hat{h}^{(n)}) \leq C_1 \{ \mathcal{B}_A^{(n)}(f, h_n^{\text{oracle}}) + \psi_n(f, h_n^{\text{oracle}}) \}, \quad \forall f \in \mathbb{F},$$

where C_1 is a numerical constant independent of f and n . In many particular problems, one can find a constant C_2 such that

$$\mathcal{B}_A^{(n)}(f, h_n^{\text{oracle}}) + \psi_n(f, h_n^{\text{oracle}}) \leq C_2 \psi_n(f, h_n^{\text{oracle}}), \quad \forall f \in \mathbb{F},$$

and, therefore, with “high probability”

$$\psi_n(f, \hat{h}^{(n)}) \leq C_3 \psi_n(f, h_n^{\text{oracle}}).$$

The latter is well-known in the bandwidth selection problem as overshooting effect. In the most simple case $\mathfrak{H}_n \subset \mathbb{R}_+$ this means that $\hat{h}^{(n)} \geq h_n^{\text{oracle}}$ with probability tending to 1, $n \rightarrow \infty$. It is worth noting that this property is inherent in many adaptive procedures and, therefore, an analog of Theorem 2 for other selection rules seems to be possible.

The second property, which is heavily related to (Ψ_n, ℓ) -selection rule, is the following: with “high probability”

$$\ell(\Lambda_{\hat{h}^{(n)}}, A(f)) \leq C_4 \{ \mathcal{B}_A^{(n)}(f, h_n^{\text{oracle}}) + \psi_n(f, h_n^{\text{oracle}}) \}.$$

In particular if the condition (3.3) of Proposition 1 is fulfilled then

$$\ell(\Upsilon_{\hat{h}^{(n)}}, B(f)) \leq C_5 \{ \mathcal{B}_A^{(n)}(f, h_n^{\text{oracle}}) + \psi_n(f, h_n^{\text{oracle}}) \}, \tag{3.6}$$

and after simple manipulations the assertion of this proposition follows from (3.5) and (3.6).

3.1. Proof of Theorem 2

For the simplicity of notations throughout the proof, we will write \hat{h} instead of $\hat{h}^{(n)}$ and we break the proof on several steps.

1⁰. Let us prove that for any $f \in \mathbb{F}$ and $n \geq 1$

$$\mathbb{P}_{1,f}^{(n)} \{ \hat{h} \notin \mathfrak{U}_n(f) \} \leq (30\varepsilon_n/\delta_n)^q. \tag{3.7}$$

First note that in view of the triangle inequality

$$\begin{aligned} \ell(\Lambda_{\hat{h}}, A(f)) &\leq \ell(\widehat{A}_{\hat{h}}, A(f)) + \ell(\widehat{A}_{\hat{h}}, \Lambda_{\hat{h}}) \\ &\leq \ell(\widehat{A}_{\hat{h}}, A(f)) + \Psi_n(\hat{h}) + \sup_{\eta \in \mathfrak{H}_n} [\ell(\widehat{A}_\eta, \Lambda_\eta) - \Psi_n(\eta)] \\ &= \ell(\widehat{A}_{\hat{h}}, A(f)) + \Psi_n(\hat{h}) + \xi_1. \end{aligned} \tag{3.8}$$

To get the last inequality, we have used that $\hat{h} \in \mathfrak{H}_n$. Next, we get from the definition of \hat{h} and (2.4)

$$\begin{aligned} \Psi_n(\hat{h}) &\leq \widehat{R}_n(\hat{h}) + 2\Psi_n(\hat{h}) \leq \widehat{R}_n(h) + 2\Psi_n(h) + \varepsilon_n \\ &\leq \sup_{\eta \in \mathfrak{H}_n} \ell(\Lambda_{h,\eta}(f), \Lambda_\eta(f)) + 2\Psi_n(h) + \xi_1 + \xi_2 + \varepsilon'_n \\ &\leq \mathcal{B}_A^{(n)}(f, h) + 2\Psi_n(h) + \xi_1 + \xi_2 + \varepsilon'_n, \quad \forall h \in \mathfrak{H}_n. \end{aligned} \tag{3.9}$$

Set $\xi_3 = \sup_{h \in \mathfrak{H}_n} [\Psi_n(h) - C_\psi \psi(f, h)]_+$. We deduce from (2.7), (3.8) and (3.9) that

$$\begin{aligned} \ell(\Lambda_{\hat{h}}, A(f)) &\leq \inf_{h \in \mathfrak{H}} [2\mathcal{B}_A^{(n)}(f, h) + 7\Psi_n(h)] + 5\xi_1 + 3\xi_2 + 2\varepsilon'_n \\ &\leq \inf_{h \in \mathfrak{H}} [2\mathcal{B}_A^{(n)}(f, h) + 7C_\psi \psi(f, h)] + 5\xi_1 + 3\xi_2 + 7\xi_3 + 2\varepsilon'_n. \end{aligned} \tag{3.10}$$

Hence, for any $f \in \mathbb{F}$ and $n \geq 1$

$$\{\hat{h} \notin \mathcal{A}_n(f)\} \subseteq \{5\xi_1 + 3\xi_2 + 7\xi_3 + 2\varepsilon'_n \geq \delta_n\} \subseteq \{10\xi_1 + 6\xi_2 + 14\xi_3 \geq \delta_n\}.$$

To get the last inclusion, we have used that $\varepsilon'_n \leq \delta_n/4$. The statement (3.7) follows now from the Markov inequality and the hypotheses $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$.

2⁰. For any $f \in \mathbb{F}$ and $n \geq 1$ the following is true:

$$\mathbb{P}_{1,f}^{(n)}\{\hat{h} \notin \mathfrak{B}_n(f)\} \leq (5\varepsilon_n/\delta_n)^q. \tag{3.11}$$

Indeed, in view of (3.9)

$$\Psi_n(\hat{h}) \leq 2 \inf_{h \in \mathfrak{H}} [\mathcal{B}_A^{(n)}(f, h) + C_\psi \psi(f, h)] + \xi_1 + \xi_2 + 2\xi_3 + \varepsilon'_n.$$

Hence, putting $\xi_4 = \sup_{h \in \mathfrak{H}_n} [\psi(f, h) - C_\psi \Psi_n(h)]_+$ we obviously have

$$\begin{aligned} \psi(f, \hat{h}) &\leq C_\psi \Psi_n(\hat{h}) + \xi_4 \\ &\leq 2C_\psi \inf_{h \in \mathfrak{H}} [\mathcal{B}_A^{(n)}(f, h) + C_\psi \psi(f, h)] \\ &\quad + C_\psi [\xi_1 + \xi_2 + 2\xi_3 + \varepsilon'_n] + \xi_4. \end{aligned}$$

Taking into account that $C_\Psi \geq 1$ and $\varepsilon'_n \leq \delta_n/4$ we have

$$\begin{aligned} \{\hat{h} \notin \mathfrak{A}_n(f)\} &= \left\{ \psi(f, \hat{h}) \geq 2C_\Psi^2 \inf_{\eta \in \mathfrak{S}_n} [\mathcal{B}_A^{(n)}(f, \eta) + 2\psi_n(f, \eta)] \right\} \\ &\subseteq \{\xi_1 + \xi_2 + 2\xi_3 + \xi_4 + \varepsilon'_n \geq 2\delta_n\} \subseteq \{\xi_1 + \xi_2 + 2\xi_3 + \xi_4 \geq \delta_n\}. \end{aligned}$$

The statement (3.11) follows now from the Markov inequality hypotheses $\mathbf{A}^{\text{upper}}$ and $\mathbf{A}^{\text{upper}}$.

3⁰. Since \hat{h} is $X_1^{(n)}$ -measurable, $\{\widehat{B}_{\hat{h}}, \hat{h} \in \mathfrak{S}\}$ is $X_2^{(n)}$ -measurable and $X_1^{(n)}, X_2^{(n)}$ are independent we get

$$(\mathbb{E}_f^{(n)} \{\rho^q(\widehat{B}_{\hat{h}}, \Upsilon_{\hat{h}}(f))\})^{\frac{1}{q}} = (\mathbb{E}_{1,f}^{(n)} \{\mathcal{E}_n^q(f, \hat{h})\})^{\frac{1}{q}}. \tag{3.12}$$

Next, we obviously have

$$\mathcal{E}_n(f, \hat{h}) = \mathcal{E}_n(f, \hat{h})1_{\hat{h} \in \mathfrak{A}_n(f)} + \mathcal{E}_n(f, \hat{h})1_{\hat{h} \notin \mathfrak{A}_n(f)} \leq \sup_{\mathfrak{h} \in \mathfrak{A}_n(f)} \mathcal{E}_n(f, \mathfrak{h}) + \sup_{\mathfrak{h} \in \mathfrak{S}_n} \mathcal{E}_n(f, \mathfrak{h})1_{\hat{h} \notin \mathfrak{A}_n(f)}.$$

Hence, we get from (3.11)

$$(\mathbb{E}_{1,f}^{(n)} \{\mathcal{E}_n^q(f, \hat{h})\})^{\frac{1}{q}} \leq \sup_{\mathfrak{h} \in \mathfrak{A}_n(f)} \mathcal{E}_n(f, \mathfrak{h}) + \sup_{\mathfrak{h} \in \mathfrak{S}_n} \mathcal{E}_n(f, \mathfrak{h})(5\varepsilon_n/\delta_n) = \tau_n(f). \tag{3.13}$$

4⁰. We have

$$\begin{aligned} \rho(\Upsilon_{\hat{h}}(f), B(f)) &= \rho(\Upsilon_{\hat{h}}(f), B(f))1_{\hat{h} \in \mathfrak{L}(f)} + \rho(\Upsilon_{\hat{h}}(f), B(f))1_{\hat{h} \notin \mathfrak{L}(f)} \\ &\leq \sup_{\mathfrak{h} \in \mathfrak{L}(f)} \rho(\Upsilon_{\mathfrak{h}}(f), B(f)) + \sup_{\mathfrak{h} \in \mathfrak{S}_n} \rho(\Upsilon_{\mathfrak{h}}(f), B(f))1_{\hat{h} \notin \mathfrak{L}(f)}. \end{aligned}$$

Hence, in view of (3.7) we obtain

$$(\mathbb{E}_f^{(n)} \{\rho^q(\Upsilon_{\hat{h}}(f), B(f))\})^{\frac{1}{q}} \leq \nu_n(f). \tag{3.14}$$

It remains to note that in view of triangle inequality

$$\mathcal{R}_B[\widehat{B}_{\hat{h}^{(n)}}, f] \leq (\mathbb{E}_f^{(n)} \{\rho^q(\Upsilon_{\hat{h}}(f), B(f))\})^{\frac{1}{q}} + (\mathbb{E}_f^{(n)} \{\rho^q(\widehat{B}_{\hat{h}}, \Upsilon_{\hat{h}}(f))\})^{\frac{1}{q}}$$

and the assertion of the theorem follows from (3.13) and (3.14).

3.2. Proof of Proposition 1

We have in view of the assumption (3.3) for any $f \in \mathbb{F}$

$$\begin{aligned} \rho(\widehat{B}_{\hat{h}}, B(f)) &\leq \rho(\Upsilon_{\hat{h}}(f), B(f)) + \rho(\widehat{B}_{\hat{h}}, \Upsilon_{\hat{h}}(f)) \\ &\leq C_{\ell\ell}(\Lambda_{\hat{h}}(f), A(f)) + \rho(\widehat{B}_{\hat{h}}, \Upsilon_{\hat{h}}(f)). \end{aligned} \tag{3.15}$$

Since by construction \hat{h} is $X_1^{(n)}$ -measurable and $X_1^{(n)}$ and $X_2^{(n)}$ are independent

$$\left(\mathbb{E}_f^{(n)}\{\ell^q(\Lambda_{\hat{h}}(f), A(f))\}\right)^{\frac{1}{q}} = \left(\mathbb{E}_{1,f}^{(n)}\{\ell^q(\Lambda_{\hat{h}}(f), A(f))\}\right)^{\frac{1}{q}}.$$

We deduce from (3.10) and the hypothesis $\mathbf{A}^{\text{upper}}$ for any $f \in \mathbb{F}$

$$\left(\mathbb{E}_f^{(n)}\{\ell^q(\Lambda_{\hat{h}}(f), A(f))\}\right)^{\frac{1}{q}} \leq \inf_{h \in \mathfrak{H}_n} \{2\mathcal{B}_A^{(n)}(f, h) + 7\psi_n(f, h)\} + 10\varepsilon_n. \tag{3.16}$$

For any $h \in \mathfrak{H}_n$ and $f \in \mathbb{F}$ we have in view of (3.9) and the assumption (3.4),

$$\mathcal{E}_n(f, \hat{h}) \leq C_{\mathcal{E}}\Psi_n(\hat{h}) \leq C_{\mathcal{E}}[\mathcal{B}_A^{(n)}(f, h) + 2\Psi_n(h) + \xi_1 + \xi_2 + \varepsilon_n],$$

and, therefore, we deduce from the hypothesis $\mathbf{A}^{\text{upper}}$ and (3.12)

$$\left(\mathbb{E}_f^{(n)}\{\ell^q(\widehat{\mathcal{B}}_{\hat{h}}, \Upsilon_{\hat{h}}(f))\}\right)^{\frac{1}{q}} \leq C_{\mathcal{E}}[\mathcal{B}_A^{(n)}(f, h) + 2\psi_n(f, h)] + 3C_{\mathcal{E}}\varepsilon_n.$$

Since the left-hand side of the obtained inequality is independent of h we get

$$\left(\mathbb{E}_f^{(n)}\{\ell^q(\widehat{\mathcal{B}}_{\hat{h}}, \Upsilon_{\hat{h}}(f))\}\right)^{\frac{1}{q}} \leq C_{\mathcal{E}} \inf_{h \in \mathfrak{H}_n} \{\mathcal{B}_A^{(n)}(f, h) + 2\psi_n(f, h)\} + 3C_{\mathcal{E}}\varepsilon_n. \tag{3.17}$$

The assertion of the theorem follows now from (3.15), (3.16), (3.17) and the triangle inequality.

3.3. From oracle approach to adaptive estimation

In this section, we discuss how to use the oracle approach in adaptive estimation over the given scale of functional classes. We present several results concerning adaptation which will be directly deduced from Theorems 1, 2 and Proposition 1.

Let $\{\mathbb{F}_\alpha, \alpha \in \mathfrak{A}\}$ be a given collection of subsets of \mathbb{F} and suppose that an abstract oracle inequality (1.1) is established. Define

$$R_n(\mathbb{F}_\alpha) = \sup_{f \in \mathbb{F}_\alpha} \inf_{h \in \mathfrak{H}} \mathcal{A}_\ell^{(n)}(f, h) + cr_n, \quad \alpha \in \mathfrak{A}.$$

We immediately deduce from (1.1) that for any $\alpha \in \mathfrak{A}$

$$\limsup_{n \rightarrow \infty} R_n^{-1}(\mathbb{F}_\alpha) \sup_{f \in \mathbb{F}_\alpha} \mathcal{R}_A[\widehat{A}_{\hat{h}}, f] \leq 1.$$

Using modern statistical language we can state that the estimator $\widehat{A}_{\hat{h}}$ is R_n -adaptive, where $R_n = \{R_n(\mathbb{F}_\alpha), \alpha \in \mathfrak{A}\}$ is the family of normalizations. If additionally one can prove that for any $\alpha \in \mathfrak{A}$

$$\liminf_{n \rightarrow \infty} R_n^{-1}(\mathbb{F}_\alpha) \inf_{\tilde{Q}} \sup_{f \in \mathbb{F}_\alpha} \mathcal{R}_A[\tilde{Q}, f] > 0,$$

where infimum is taken over all $X^{(n)}$ -measurable \mathfrak{S}_1 -valued random mappings we can assert that the estimator $\widehat{A}_{\widehat{h}}$ is *optimally adaptive* over the scale $\{\mathbb{F}_\alpha, \alpha \in \mathfrak{A}\}$. The latter means that this estimator is simultaneously asymptotically minimax on each \mathbb{F}_α .

Consequences of Theorem 1 and Proposition 1. Set $S_n = \{s_n(\mathbb{F}_\alpha), \alpha \in \mathfrak{A}\}$, where

$$s_n(\mathbb{F}_\alpha) = \sup_{f \in \mathbb{F}_\alpha} \inf_{h \in \mathfrak{H}_n} \{ \mathcal{B}_A^{(n)}(f, h) + \psi_n(f, h) \}$$

and let $\widehat{h}^{(n)}$ is obtained by (Ψ_n, ℓ) -selection rule.

The results below follow immediately from Theorem 1 and Proposition 1.

Theorem 3. *Let $\mathbf{A}^{\text{permute}}$ and $\mathbf{A}^{\text{upper}}$ hold and assume that $\varepsilon_n = o(\inf_{\alpha \in \mathfrak{A}} s_n(\mathbb{F}_\alpha))$, $n \rightarrow \infty$.*

(1) *Then, for any $\alpha \in \mathfrak{A}$*

$$\limsup_{n \rightarrow \infty} s_n^{-1}(\mathbb{F}_\alpha) \sup_{f \in \mathbb{F}_\alpha} \mathcal{R}_A[\widehat{A}_{\widehat{h}^{(n)}}, f] \leq 5.$$

(2) *If additionally (3.3) and (3.4) are fulfilled, then for any $\alpha \in \mathfrak{A}$*

$$\limsup_{n \rightarrow \infty} s_n^{-1}(\mathbb{F}_\alpha) \sup_{f \in \mathbb{F}_\alpha} \mathcal{R}_B[\widehat{B}_{\widehat{h}^{(n)}}, f] \leq 7C_\ell + 2C_\mathcal{E}.$$

Thus, the estimators $\widehat{A}_{\widehat{h}^{(n)}}$ and $\widehat{B}_{\widehat{h}^{(n)}}$ are simultaneously S_n -adaptive. Hence, if $\widehat{A}_{\widehat{h}^{(n)}}$ is optimally adaptive in *Problem A* one can expect that $\widehat{B}_{\widehat{h}^{(n)}}$ is optimally adaptive in *Problem B*.

Adaptive analogue of Theorem 2. Define for any $\alpha \in \mathfrak{A}$

$$\psi_n(\mathbb{F}_\alpha, h) = \inf_{f \in \mathbb{F}_\alpha} \psi_n(f, h) \tag{3.18}$$

and introduce the following sets of parameters

$$\begin{aligned} \mathcal{V}_n(\alpha) &= \{h \in \mathfrak{H}_n : \psi_n(\mathbb{F}_\alpha, h) < 4C_\Psi^2 s_n(\mathbb{F}_\alpha)\}, \\ \mathcal{U}_n(\alpha, f) &= \{h \in \mathfrak{H}_n : \ell(\Lambda_h, A(f)) < 8C_\Psi s_n(\mathbb{F}_\alpha)\}. \end{aligned}$$

It is obvious that $\mathfrak{V}_n(f) \subseteq \mathcal{V}_n(\alpha)$ and $\mathfrak{U}_n(f) \subseteq \mathcal{U}_n(\alpha, f)$ for any $f \in \mathbb{F}_\alpha$. Set finally

$$\begin{aligned} \varphi_n(\mathbb{F}_\alpha) &= \sup_{f \in \mathbb{F}_\alpha} \sup_{h \in \mathcal{V}_n(\alpha)} \mathcal{E}_n(f, h) + \sup_{f \in \mathbb{F}_\alpha} \sup_{h \in \mathcal{U}_n(\alpha, f)} \rho(\Upsilon_h, B(f)), \\ \varkappa_n(\mathbb{F}_\alpha) &= (5\varepsilon_n/\delta_n) \sup_{f \in \mathbb{F}_\alpha} \left[\sup_{h \in \mathfrak{H}_n} \mathcal{E}_n(f, h) + 6 \sup_{h \in \mathfrak{H}_n} \rho(\Upsilon_h, B(f)) \right] \end{aligned}$$

and let $\widehat{h}^{(n)}$ is obtained by (Ψ_n, ℓ) -selection rule.

The following statement is the direct consequence of Theorem 2.

Theorem 4. Let $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$ hold.

Assume that $\varkappa_n(\mathbb{F}_\alpha) = o(\varphi_n(\mathbb{F}_\alpha))$, $n \rightarrow \infty$, for any $\alpha \in \mathfrak{A}$. Then, for any $\alpha \in \mathfrak{A}$

$$\limsup_{n \rightarrow \infty} \sup_{f \in \mathbb{F}_\alpha} \varphi_n^{-1}(\mathbb{F}_\alpha) \mathcal{R}_B[\widehat{\mathcal{B}}_{\mathfrak{h}^{(n)}}^{\widehat{\mathfrak{h}}}, f] \leq 1.$$

3.4. Some computations in the density model

Our objective now is to give an example of statistical model and problem in which we are able to check the hypotheses $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$. We also gives some bounds for the quantities $\mathcal{B}_A^{(n)}(f, \mathfrak{h})$ and $\psi_n(f, \mathfrak{h})$ involved in the (Ψ_n, ℓ) -oracle inequality. It allows us, in particular, to compute the rate

$$s_n(\mathbb{F}_\alpha) = \sup_{f \in \mathbb{F}_\alpha} \inf_{\mathfrak{h} \in \mathfrak{H}_n} \{ \mathcal{B}_A^{(n)}(f, \mathfrak{h}) + \psi_n(f, \mathfrak{h}) \}$$

appeared in all adaptive results in the case when $\mathbb{F}_\alpha = \mathbb{N}_{p,d}(\vec{\beta}, L)$, $\alpha = (\vec{\beta}, L)$, where $\mathbb{N}_{p,d}(\vec{\beta}, L)$ anisotropic Nikol'skii class. The results presented in this section form basic tools for the solution of problem studied in Sections 4 and 5.

Let $T^{(m)} = (T_1, \dots, T_m)$, $m \in \mathbb{N}^*$, be i.i.d. random vectors taking values in \mathbb{R}^d and having density \mathfrak{p} with respect to Lebesgue measure. As before $\mathbb{P}_{\mathfrak{p}}^{(m)}$ and $\mathbb{E}_{\mathfrak{p}}^{(m)}$ denote the probability law and mathematical expectation of $T^{(m)}$. The goal is to estimate $A(\mathfrak{p}) = \mathfrak{p}$ and the quality of an estimator procedures is measured by \mathbb{L}_p -risk, $1 \leq p < \infty$, that is

$$\mathcal{R}_A[\widetilde{A}_m, \mathfrak{p}] = (\mathbb{E}_{\mathfrak{p}}^{(m)} [\| \widetilde{A}_m - \mathfrak{p} \|_p^p])^{1/p}, \quad \mathfrak{p} \in \mathfrak{P},$$

where \mathfrak{P} is the set of all probability densities and $\ell^p(g) = \|g\|_p^p = \int_{\mathbb{R}^d} |g(x)|^p dx$.

Remind that $\mathcal{H}^d = \{((2^{-k_1}, \dots, 2^{-k_1}), (k_1, \dots, k_d)) \in \mathbb{N}^d\}$ is the dyadic grid in $(0, 1]^d$ and $K_{\vec{h}}(\cdot)$, $\vec{h} \in \mathcal{H}^d$ be defined in (1.3). Introduce the collection of standard kernel estimators

$$\mathbf{A}(\mathcal{H}_m^d) = \left\{ \widehat{A}_{\vec{h}}(\cdot) = m^{-1} \sum_{i=1}^m K_{\vec{h}}(T_i - \cdot), \vec{h} \in \mathcal{H}_m^d \right\}, \tag{3.19}$$

where $\mathcal{H}_m^d = \{\vec{h} \in \mathcal{H}^d : \ln(m)/m \leq V_{\vec{h}} \leq e^{-\sqrt{\ln(m)}}\}$ and $V_{\vec{h}} = \prod_{j=1}^d h_j$.

From now on we will assume that the kernel K involved in the definition of $K_{\vec{h}}$ is continuous, bounded, symmetric function belonging to $\mathbb{L}_1(\mathbb{R}^d)$ and $\int K = 1$.

3.4.1. Verification of the hypotheses $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$

Recall that “ \star ” denotes the convolution operator and introduce for any $\vec{h}, \vec{\eta} \in \mathcal{H}^d$

$$\widehat{A}_{\vec{h}, \vec{\eta}}(\cdot) = m^{-1} \sum_{i=1}^m [K_{\vec{h}} \star K_{\vec{\eta}}](T_i - \cdot). \tag{3.20}$$

Set also for any $\vec{h} \in \mathcal{H}^d$ and $m \in \mathbb{N}^*$

$$\Delta_m(\vec{h}) = \begin{cases} 128\|K\|_1\|K\|_p(mV_{\vec{h}})^{1/p-1}, & 1 \leq p < 2; \\ 9\|K\|_1\|K\|_2(mV_{\vec{h}})^{-1/2}, & p = 2; \end{cases}$$

and if $p > 2$

$$\Delta_m(\vec{h}) = \frac{480p\|K\|_1}{\ln(p)} \left\{ \frac{1}{\sqrt{m}} \left(\int_{\mathbb{R}^d} \left[\frac{1}{m} \sum_{i=1}^m K_{\vec{h}}^2(Z_i - t) \right]^{p/2} dt \right)^{1/p} + 2\|K\|_p(mV_{\vec{h}})^{-1/2} \right\}.$$

We remark that $\Delta_m(\cdot)$ is not random if $p \leq 2$. It is obvious that the hypothesis $\mathbf{A}^{\text{permute}}$ is fulfilled. Define the empirical process

$$\xi_m(x, \vec{h}) = \frac{1}{m} \sum_{i=1}^m [K_{\vec{h}}(T_i - x) - \mathbb{E}_{\mathbf{p}}^{(m)} K_{\vec{h}}(T_i - x)], \quad x \in \mathbb{R}^d.$$

Auxiliary results. The following statements are the consequences of the results established in Lemmas 1 and 2 in Goldenshluger and Lepski [20] and Theorem 3 in Goldenshluger and Lepski [19]. Set $\mathfrak{P}(D) = \{\mathbf{p} \in \mathfrak{P} : \|\mathbf{p}\|_{\infty} \leq D\}$, $D > 0$.

Lemma 1. For any $1 \leq p < \infty$ and for all $m \in \mathbb{N}^*$

$$\sup_{\mathbf{p} \in \mathfrak{P}} \mathbb{E}_{\mathbf{p}}^{(m)} \left(\sup_{\vec{h} \in \mathcal{H}_m^d} [\|\xi_m(\cdot, \vec{h})\|_p - \|K\|_1^{-1} \Delta_m(\vec{h})]_+^q \right) \leq \varepsilon_m^q(p), \quad p \in [1, 2); \quad (3.21)$$

$$\sup_{\mathbf{p} \in \mathfrak{P}(D)} \mathbb{E}_{\mathbf{p}}^{(m)} \left(\sup_{\vec{h} \in \mathcal{H}_m^d} [\|\xi_m(\cdot, \vec{h})\|_p - \|K\|_1^{-1} \Delta_m(\vec{h})]_+^q \right) \leq \varepsilon_m^q(p), \quad p \in [2, \infty), \quad (3.22)$$

where for any $p \geq 1$ and $D > 0$

$$\limsup_{m \rightarrow \infty} m^d \varepsilon_m(p) = 0, \quad \forall a > 0. \quad (3.23)$$

Moreover for any $p > 2$, there exists $\mathfrak{D} > 0$ completely determined by K , p , D and d such that

$$\sup_{\mathbf{p} \in \mathfrak{P}(D)} \mathbb{E}_{\mathbf{p}}^{(m)} \left(\sup_{\vec{h} \in \mathcal{H}_m^d} [\Delta_m(\vec{h}) - \mathfrak{D}(mV_{\vec{h}})^{-1/2}]_+^q \right) \leq \varepsilon_m^q(p). \quad (3.24)$$

Let us remark the following simple consequence of (3.23) and (3.24).

$$\sup_{\mathbf{p} \in \mathfrak{P}(D)} (\mathbb{E}_{\mathbf{p}}^{(m)} \Delta_m^q(\vec{h}))^{1/q} \leq \mathfrak{D}(mV_{\vec{h}})^{-1/2}, \quad \forall \vec{h} \in \mathcal{H}_m^d, \forall m \in \mathbb{N}^*. \quad (3.25)$$

In order to get this inequality it suffices to note that $(mV_{\vec{h}})^{-1/2} \geq m^{-1/2}$ for all $\vec{h} \in (0, 1]^d$.

Verification of the hypothesis $\mathbf{A}^{\text{upper}}$. Set $\Lambda_{\vec{h}}(\mathfrak{p}, \cdot) = \mathbb{E}_{\mathfrak{p}}^{(m)}\{\widehat{A}_{\vec{h}}(\cdot)\}$, $\vec{h} \in \mathcal{H}_m^d$ and note that

$$\Lambda_{\vec{h}}(\mathfrak{p}, \cdot) = \mathbb{E}_{\mathfrak{p}}^{(m)} K_{\vec{h}}(T_1 - \cdot) = \int_{\mathbb{R}^d} K_{\vec{h}}(z - \cdot) \mathfrak{p}(z) dz. \quad (3.26)$$

In view of the assertions (3.21) and (3.22) of Lemma 1 we can conclude that the hypothesis $\mathbf{A}^{\text{upper}}$ (i) is fulfilled for any $p \geq 1$. In order to verify $\mathbf{A}^{\text{upper}}$ (ii) note first the following obvious identities.

$$\widehat{A}_{\vec{h}, \vec{\eta}} \equiv K_{\vec{h}} \star \widehat{A}_{\vec{\eta}} \equiv K_{\vec{\eta}} \star \widehat{A}_{\vec{h}}, \quad \forall \vec{h}, \vec{\eta} \in \mathcal{H}^d. \quad (3.27)$$

Hence, putting $\Lambda_{\vec{h}, \vec{\eta}}(\mathfrak{p}, \cdot) = \mathbb{E}_{\mathfrak{p}}^{(m)}\{\widehat{A}_{\vec{h}, \vec{\eta}}(\cdot)\}$, $\zeta_{\vec{h}, \vec{\eta}}^{(m)}(\cdot) = \widehat{A}_{\vec{h}, \vec{\eta}}(\cdot) - \Lambda_{\vec{h}, \vec{\eta}}(\mathfrak{p}, \cdot)$ and $\zeta_{\vec{h}}^{(m)}(\cdot) = \widehat{A}_{\vec{h}}(\cdot) - \Lambda_{\vec{h}}(\mathfrak{p}, \cdot)$ we deduce from the Young inequality, Folland [15], Theorem 6.18,

$$\begin{aligned} \|\zeta_{\vec{h}, \vec{\eta}}^{(m)}\|_p &\leq \|K\|_1 \|\zeta_{\vec{\eta}}^{(m)}\|_p, \\ \|\zeta_{\vec{h}, \vec{\eta}}^{(m)}\|_p &\leq \|K\|_1 \|\zeta_{\vec{h}}^{(m)}\|_p \Rightarrow \|\zeta_{\vec{h}, \vec{\eta}}^{(m)}\|_p \leq \|K\|_1 (\|\zeta_{\vec{\eta}}^{(m)}\|_p \wedge \|\zeta_{\vec{h}}^{(m)}\|_p). \end{aligned}$$

It yields for all $\vec{h}, \vec{\eta} \in \mathcal{H}_m^d$ in view of the obvious inequality $a \wedge c \leq b \wedge d + (a - b)_+ \vee (c - d)_+$

$$(\|\zeta_{\vec{h}, \vec{\eta}}^{(m)}\|_p - \Delta_m(\vec{h}) \wedge \Delta_m(\vec{\eta}))_+ \leq \|K\|_1 \sup_{\vec{h} \in \mathcal{H}_m^d} (\|\zeta_{\vec{h}}^{(m)}\|_p - \|K\|_1^{-1} \Delta_m(\vec{h}))_+.$$

In view of the assertions (3.21) and (3.22) of Lemma (1) we can conclude that the hypothesis $\mathbf{A}^{\text{upper}}$ (ii) is fulfilled for any $p \geq 1$.

Verification of the hypothesis $\mathcal{A}^{\text{upper}}$. First, we note that Δ_m is not random if $p \in [1, 2]$ and, therefore, in this case $\mathcal{A}^{\text{upper}}$ is checked with $C_\Psi = 1$.

Next, if $p > 2$ we have in view of the definition of Δ_m and (3.25) for any $\mathfrak{p} \in \mathfrak{P}(D)$

$$\frac{960p \|K\|_1 \|K\|_p}{\ln(p)} (mV_{\vec{h}})^{-1/2} \leq (\mathbb{E}_{\mathfrak{p}}^{(m)} \Delta_m^q(\vec{h}))^{1/q} \leq \mathfrak{D} (mV_{\vec{h}})^{-1/2}, \quad \forall \vec{h} \in \mathcal{H}_m^d. \quad (3.28)$$

Moreover, $\Delta_m \geq \frac{960p \|K\|_1 \|K\|_p}{\ln(p)} (mV_{\vec{h}})^{-1/2}$ and, therefore, $\mathcal{A}^{\text{upper}}$ is fulfilled in view of the assertion (3.22) of Lemma (1) with $C_\Psi = \mathfrak{D} \vee \frac{960p \|K\|_1 \|K\|_p}{\ln(p)}$.

3.4.2. Some bounds for the quantities $\mathcal{B}_A^{(n)}(f, \mathfrak{h})$ and $\psi_n(f, \mathfrak{h})$

Let us find upper estimate for $\mathcal{B}_A^{(n)}(f, \mathfrak{h})$ and lower and upper estimates for $\psi_n(f, \mathfrak{h})$, which both appear in (Ψ_n, ℓ) -oracle inequality. In the considered case $f = \mathfrak{p}$, $n = m$, $\mathfrak{h} = \vec{h}$, $A(f) = A(\mathfrak{p}) = \mathfrak{p}$, $\ell(\cdot) = \|\cdot\|_p$ and $\Psi_n = \Delta_m$.

Note first that (3.27) implies that for any $x \in \mathbb{R}^d$

$$\begin{aligned} \Lambda_{\vec{h}, \vec{\eta}}(\mathfrak{p}, x) - \Lambda_{\vec{\eta}}(\mathfrak{p}, x) &= K_{\vec{\eta}} \star \Lambda_{\vec{h}}(\mathfrak{p}, x) - \Lambda_{\vec{\eta}}(\mathfrak{p}, x) \\ &= \int_{\mathbb{R}^d} K_{\vec{\eta}}(x - y) \left(\int_{\mathbb{R}^d} K_{\vec{h}}(z - y) \mathfrak{p}(z) dz - \mathfrak{p}(y) \right) dy. \end{aligned}$$

Applying the Young inequality we obtain for all $p \geq 1$

$$\mathcal{B}_A^{(m)}(\mathfrak{p}, \vec{h}) \leq (1 + 2\|K\|_1)\|K_{\vec{h}} \star \mathfrak{p} - \mathfrak{p}\|_p, \quad \forall \vec{h} \in \mathcal{H}^d. \tag{3.29}$$

In view of assertion (3.25) of Lemma 1 and the definition of Δ_m we get

$$\underline{\mathbf{C}}(nV_{\vec{h}})^{\frac{1}{p\wedge 2}-1} \leq \psi_n(\mathfrak{p}, \vec{h}) := [\mathbb{E}_{\mathfrak{p}}^{(m)}\{\Delta_m^p(\mathfrak{h})\}]^{\frac{1}{p}} \leq \overline{\mathbf{C}}(nV_{\vec{h}})^{\frac{1}{p\wedge 2}-1}, \quad \forall \mathfrak{p} \in \mathfrak{P}_p, \tag{3.30}$$

where $\mathfrak{P}_p = \mathfrak{P}$ if $p < 2$, $\mathfrak{P}_p = \{\mathfrak{p} \in \mathfrak{P} : \mathfrak{p} \in \mathfrak{P}(D)\}$ if $p \geq 2$ and $\underline{\mathbf{C}} = \overline{\mathbf{C}} = 128\|K\|_1\|K\|_p$ if $p \in [1, 2)$, $\underline{\mathbf{C}} = \overline{\mathbf{C}} = 9\|K\|_1\|K\|_2$ if $p = 2$ and $\underline{\mathbf{C}} = \frac{960p\|K\|_1\|K\|_p}{\ln(p)}$, $\overline{\mathbf{C}} = \mathfrak{D}$ if $p > 2$. We remark that the obtained lower and upper estimates are independent of \mathfrak{p} .

3.4.3. Computation of the rate on anisotropic Nikol'skii class

In view of (3.29) and (3.30), we should bound from above

$$\sup_{\mathfrak{p} \in \mathbb{F}_{\mathfrak{a}}} \inf_{\vec{h} \in H_m^d} \left\{ \|K_{\vec{h}} \star \mathfrak{p} - \mathfrak{p}\|_p + (mV_{\vec{h}})^{\frac{1}{p\wedge 2}-1} \right\}.$$

This quantity with $\mathbb{F}_{\mathfrak{a}} = \mathbb{N}_{p,d}(\vec{\beta}, L)$, $\mathfrak{a} = (\vec{\beta}, L)$, where $\mathbb{N}_{p,d}(\vec{\beta}, L)$ is anisotropic Nikol'skii class was evaluated in Goldenshluger and Lepski [20].

Let $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ denote the canonical basis of \mathbb{R}^d . For function $G : \mathbb{R}^d \rightarrow \mathbb{R}^1$ and real number $u \in \mathbb{R}$ define the first order difference operator with step size u in direction of the variable x_j by

$$\Delta_{u,j}G(x) = G(x + u\mathbf{e}_j) - G(x), \quad j = 1, \dots, d.$$

By induction, the k th order difference operator with step size u in direction of the variable x_j is defined as

$$\Delta_{u,j}^k G(x) = \Delta_{u,j} \Delta_{u,j}^{k-1} G(x) = \sum_{l=1}^k (-1)^{l+k} \binom{k}{l} \Delta_{ul,j} G(x).$$

Definition 1. For given $\vec{\beta} = (\beta_1, \dots, \beta_d) \in (0, \infty)^d$, $L > 0$ and $p \geq 1$ we say that function $G : \mathbb{R}^d \rightarrow \mathbb{R}^1$ belongs to the anisotropic Nikol'skii class $\mathbb{N}_{p,d}(\vec{\beta}, L)$ if $\|G\|_p \leq L$ and for every $j = 1, \dots, d$ there exists natural number $k_j > \beta_j$ such that

$$\|\Delta_{u,j}^{k_j} G\|_p \leq L|u|^{\beta_j}, \quad \forall u \in \mathbb{R}, \forall j = 1, \dots, d.$$

We will use the following specific kernel K in the definition of the family $\mathbf{A}(\mathcal{H}_m^d)$ [see, e.g., Kerkyacharian *et al.* [25] or Goldenshluger and Lepski [20]].

Let $\mathfrak{s} \in \mathbb{N}^*$ be fixed and let $w : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ satisfy $\int w(y) dy = 1$, and $w \in \mathcal{C}^{(\mathfrak{s})}(\mathbb{R}^1)$. Put

$$\mathcal{K}_{\mathfrak{s}}(y) = \sum_{i=1}^{\mathfrak{s}} \binom{\mathfrak{s}}{i} (-1)^{i+1} \frac{1}{i} w\left(\frac{y}{i}\right), \quad K(x) = \prod_{j=1}^d \mathcal{K}_{\mathfrak{s}}(x_j), \quad x = (x_1, \dots, x_d). \tag{3.31}$$

Let $\varepsilon > 0$ be an arbitrary but a priori chosen number and let the kernel K is constructed in accordance with (3.31). Next result can be found for instance, in Goldenshluger and Lepski [20].

For any $\vec{\beta} \in (0, \varepsilon]^d$, $L > 0$ and $p \geq 1$ there exists $\nu > 0$ independent of L such that

$$\begin{aligned} \sup_{\mathfrak{p} \in \mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p} \inf_{\vec{h} \in H_m^d} \{ \|K_{\vec{h}} \star \mathfrak{p} - \mathfrak{p}\|_p + (mV_{\vec{h}})^{\frac{1}{p\wedge 2} - 1} \} &\leq \nu s_m(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p) \\ &=: \nu L^{\frac{1-1/(p\wedge 2)}{\beta+1-1/(p\wedge 2)}} m^{-\frac{1-1/(p\wedge 2)}{1+(1/\beta)(1-1/(p\wedge 2))}}. \end{aligned} \tag{3.32}$$

It is well known that $s_m(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p)$ is the minimax rate of convergence on $\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p$.

Concluding remarks. (Ψ_n, ℓ) -selection rule with $\Psi_n = \Delta_m$, $n = m$, and $\ell(\cdot) = \|\cdot\|_p$ is a particular case of the selection scheme proposed in Goldenshluger and Lepski [20]. Selected in accordance with this rule estimator from the collection (3.19) is, in view of (3.29), (3.30), (3.32) and the first assertion of Theorem 3, optimally-adaptive over the collection of anisotropic Nikol’skii classes. For the first time, it was proved in Goldenshluger and Lepski [20], Theorem 4.

4. Adaptive estimation in the generalized deconvolution model

Consider the following observation scheme introduced in Lepski and Willer [35]:

$$Z_i = X_i + \varepsilon_i Y_i, \quad i = 1, \dots, n, \tag{4.1}$$

where $X_i, i = 1, \dots, n$, are *i.i.d.* d -dimensional random vectors with common density f to be estimated. The noise variables $Y_i, i = 1, \dots, n$, are *i.i.d.* d -dimensional random vectors with known common density g . At last $\varepsilon_i \in \{0, 1\}, i = 1, \dots, n$, are *i.i.d.* Bernoulli random variables with $\mathbb{P}(\varepsilon_1 = 1) = \alpha$, where $\alpha \in [0, 1)$ is supposed to be known. The sequences $\{X_i, i = 1, \dots, n\}$, $\{Y_i, i = 1, \dots, n\}$ and $\{\varepsilon_i, i = 1, \dots, n\}$ are supposed to be mutually independent.

Let us note that the case $\alpha = 1$ corresponds to the pure deconvolution model $Z_i = X_i + Y_i, i = 1, \dots, n$ discussed in Introduction, whereas the case $\alpha = 0$ corresponds to the direct observation scheme $Z_i = X_i, i = 1, \dots, n$. The “intermediate” case $\alpha \in (0, 1)$, considered for the first time in Hesse [23], can be treated as the mathematical modeling of the following situation. One part of the data, namely $(1 - \alpha)n$, is observed without noise. If the indexes corresponding to these observations were known, the density f could be estimated using only this part of the data, with the accuracy corresponding to the direct case. The main question we will address in this intermediate case is whether the same accuracy would be achievable if the latter information is not available?

Thus, $\rho^p(g) = \|g\|_p^p = \int_{\mathbb{R}^d} |g(x)|^p dx$, $\mathbb{F} = \mathfrak{P}$ is the set of all probability densities, $B(f) = f$ and the quality of an estimator procedures is measured by \mathbb{L}_p -risk

$$\mathcal{R}_B[\tilde{B}_n, f] = (\mathbb{E}_f^{(n)} [\|\tilde{B}_n - f\|_p^p])^{1/p}, \quad f \in \mathfrak{P}.$$

At last, let $X_1^{(n)} = (X_1, \dots, X_{[n/2]})$ and $X_2^{(n)} = (X_{[n/2]+1}, \dots, X_n)$, where $[a]$ denotes the integer part of $a \in \mathbb{R}$.

All results presented in this section will be established under the following condition imposed on the distribution of the noise variable Y_1 . In what follows for any $Q \in \mathbb{L}_1(\mathbb{R}^d)$ its Fourier transform will be denoted by \check{Q} .

Assumption 1. (1) if $\alpha \in (0, 1)$ then there exists $\varpi > 0$ such that

$$|1 - \alpha + \alpha \check{g}(t)| \geq \varpi, \quad \forall t \in \mathbb{R}^d;$$

(2) if $\alpha = 1$ then there exists $\vec{\mu} = (\mu_1, \dots, \mu_d) \in (0, \infty)^d$ and $G_1, G_2 > 0$ such that

$$G_1 \prod_{j=1}^d (1 + t_j^2)^{-\frac{\mu_j}{2}} \leq |\check{g}(t)| \leq G_2 \prod_{j=1}^d (1 + t_j^2)^{-\frac{\mu_j}{2}}, \quad \forall t \in \mathbb{R}^d.$$

We remark that Assumption 1 is very weak when $\alpha \in (0, 1)$. It is verified for many distributions, including centered multivariate Laplace and Gaussian ones. Note also that Assumption 1 always holds with $\varpi = 1 - 2\alpha$ if $\alpha < 1/2$. Additionally, it holds with $\varpi = 1 - \alpha$ if \check{g} is a real positive function. The latter is true, in particular, for any probability law obtained by the even number of convolutions of a symmetric distribution with itself. As to the case $\alpha = 1$ Assumption 1 is a well-known in the literature condition referred to *moderately ill-posed* statistical problem. In particular, it is checked for multivariate Laplace and Gamma laws.

Below we introduce families of estimators whose construction involved kernel K . Throughout this section without further mentioning we will additionally assume that \check{K} is compactly supported and $\check{K}(t) = 1$ for all $t \in [-1, 1]^d$.

From now on $\mathbf{c}_1, \mathbf{c}_2, \dots$, denote the constants that may depend only on K, p, D, \mathfrak{D}, d and $\alpha, \varpi, G_1, G_2, \vec{\mu}$ appeared in Assumption 1. In particular they are independent of f and n .

4.1. Idea of estimator construction

Let $K_{\vec{h}}(\cdot), \vec{h} \in (0, 1]^d$, be defined in (1.3) and let $M(\cdot, \vec{h}) : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfy the operator equation

$$K_{\vec{h}}(y) = (1 - \alpha)M(y, \vec{h}) + \alpha \int_{\mathbb{R}^d} g(t - y)M(t, \vec{h}) dt, \quad y \in \mathbb{R}^d. \tag{4.2}$$

Recall that $V_{\vec{h}} = \prod_{j=1}^d h_j, \mathcal{H}_m^d = \{\vec{h} \in \mathcal{H}^d : \ln(m)/m \leq V_{\vec{h}} \leq e^{-\sqrt{\ln(m)}}\}, m \geq 1$, and introduce the following estimator's collection built from $X_2^{(n)}$

$$\mathbf{B}(\mathcal{H}_{[n/2]}^d) = \left\{ \widehat{B}_{\vec{h}}(\cdot) = (n - [n/2])^{-1} \sum_{i=[n/2]+1}^n M(Z_i - \cdot, \vec{h}), \vec{h} \in \mathcal{H}_{[n/2]}^d \right\}. \tag{4.3}$$

Our goal is to select an estimator from this collection and to study its properties. Following our general receipt, we suggest to proceed as follows.

- Consider first the estimation of $A_g(f) = (1 - \alpha)f + \alpha[g \star f]$, where, remind “ \star ” denotes the convolution operator. Note that $A_g(f)$ is the density of Z_1 and, therefore, can be estimated from the observation $X_1^{(n)}$ using standard kernel estimator. Introduce the collection

$$\mathbf{A}(\mathcal{H}_{[n/2]}^d) = \left\{ \widehat{A}_{\vec{h}}(\cdot) = [n/2]^{-1} \sum_{i=1}^{[n/2]} K_{\vec{h}}(Z_i - \cdot), \vec{h} \in \mathcal{H}_{[n/2]}^d \right\}. \tag{4.4}$$

- To each estimator from this collection associate, its \mathbb{L}_p -risk ($\ell(\cdot) = \|\cdot\|_p$)

$$\mathcal{R}_A[\widehat{A}_{\vec{h}}, f] = (\mathbb{E}_f^{(n)}[\|\widehat{A}_{\vec{h}} - A_g(f)\|_p^p])^{1/p}, \quad f \in \mathfrak{F}.$$

- Select an estimator from the collection $\mathbf{A}(\mathcal{H}_{[n/2]}^d)$ in accordance with (Ψ_n, ℓ) -selection rule based on the collection of auxiliary estimators (3.20), where $T^{(m)} = (X_1, \dots, X_{[n/2]})$, $m = [n/2]$ and $\Psi_n = \Delta_{[n/2]}$, $\ell(\cdot) = \|\cdot\|_p$. Thus, the selection rule is

$$\widehat{R}_n(\vec{h}) := \sup_{\vec{\eta} \in \mathcal{H}_{[n/2]}^d} [\|\widehat{A}_{\vec{h}, \vec{\eta}} - \widehat{A}_{\vec{\eta}}\|_p - 2\Delta_{[n/2]}(\vec{\eta})]_+; \tag{4.5}$$

$$\vec{h}_n := \inf_{\vec{h} \in \mathcal{H}_{[n/2]}^d} [\widehat{R}_n(\vec{h}) + 2\Delta_{[n/2]}(\vec{h})]. \tag{4.6}$$

- Choose the estimator $\widehat{B}_{\vec{h}_n}(\cdot)$.

The theoretical properties of the estimator $\widehat{B}_{\vec{h}_n}(\cdot)$ will be then deduced with help of Proposition 1, Theorem 3(2) and Theorem 4.

4.2. Some bounds for the quantity $\mathcal{E}_n(\cdot, \cdot)$

Introduce the following notations. Set for any $x \in \mathbb{R}^d$

$$\chi_{\vec{h}}^{(n)}(x) = \widehat{B}_{\vec{h}} - \Upsilon_{\vec{h}}(f, x), \quad \Upsilon_{\vec{h}}(f, x) = \mathbb{E}_{2,f}^{(n)}\{\widehat{B}_{\vec{h}}(x)\} = \int_{\mathbb{R}^d} M(z - x, \vec{h}) A_g(f, z) dz.$$

Put for brevity $\mathfrak{p} = A_g(f)$. Applying Bahr–Esseen and Rosenthal inequalities, von Bahr and Esseen [46], Rosenthal [42], to the sum of i.i.d. random variables $\chi_{\vec{h}}^{(n)}(x)$, we have for any $x \in \mathbb{R}^d$

$$\begin{aligned} \mathbf{c}_4^{-1} \mathbb{E}_{2,f}^{(n)} |\chi_{\vec{h}}^{(n)}(x)|^p &\leq n^{p-1} \int_{\mathbb{R}^d} |M(z - x, \vec{h})|^p \mathfrak{p}(z) dz, \quad p \leq 2; \\ \mathbf{c}_4^{-1} \mathbb{E}_{2,f}^{(n)} |\chi_{\vec{h}}^{(n)}(x)|^p &\leq \left(n^{-1} \int_{\mathbb{R}^d} M^2(z - x, \vec{h}) \mathfrak{p}(z) dz \right)^{p/2} \\ &\quad + n^{p-1} \int_{\mathbb{R}^d} |M(z - x, \vec{h})|^p \mathfrak{p}(z) dz, \quad p > 2. \end{aligned}$$

Noting that in view of the Fubini theorem $\mathbb{E}_{2,f}^{(n)} \|\chi_{\vec{h}}^{(n)}\|_p^p = \|\mathbb{E}_{2,f}^{(n)} |\chi_{\vec{h}}^{(n)}(\cdot)|^p\|_1$ one has

$$\begin{aligned} (\mathbb{E}_{2,f}^{(n)} \|\chi_{\vec{h}}^{(n)}\|_p^p)^{1/p} &\leq \mathbf{c}_4 n^{1/p-1} \|M(\cdot, \vec{h})\|_p, \quad p \leq 2; \\ (\mathbb{E}_{2,f}^{(n)} \|\chi_{\vec{h}}^{(n)}\|_p^p)^{1/p} &\leq \mathbf{c}_4 n^{-\frac{1}{2}} \left\{ \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} M^2(z-x, \vec{h}) \mathfrak{p}(z) dz \right)^{p/2} dx \right\}^{1/p} + \mathbf{c}_4 n^{\frac{1-p}{p}} \|M(\cdot, \vec{h})\|_p \\ &\leq \mathbf{c}_4 (\|\mathfrak{p}\|_\infty^{1/2-1/p} n^{-\frac{1}{2}} \|M(\cdot, \vec{h})\|_2 + n^{\frac{1-p}{p}} \|M(\cdot, \vec{h})\|_p), \quad p > 2. \end{aligned}$$

To get the last estimate, we have used again the Young inequality. Thus, for any $\vec{h} \in \mathcal{H}^d$ we have the following bound for the quantity $\mathcal{E}(\cdot, \cdot)$ defined in (3.1)

$$\begin{aligned} \mathcal{E}(f, \vec{h}) &:= (\mathbb{E}_{2,f}^{(n)} \|\chi_{\vec{h}}^{(n)}\|_p^p)^{1/p} \\ &\leq \mathbf{c}_5 (n^{-1/2} \|M(\cdot, \vec{h})\|_2 + n^{1/p-1} \|M(\cdot, \vec{h})\|_p), \quad \forall f \in \mathfrak{F}_p. \end{aligned} \tag{4.7}$$

We would like to stress that if $\alpha = 1$ and $\|g\|_\infty < \infty$ then $\mathfrak{F}_p = \mathfrak{F}$ for all $p \geq 1$.

Let $\check{M}(t, \vec{h}), t \in \mathbb{R}^d$, denote the Fourier transform of $M(\cdot, \vec{h})$. Then, we obtain in view of the definition of $M(\cdot, \vec{h})$

$$\check{M}(t, \vec{h}) = \check{K}(t\vec{h})[(1-\alpha) + \alpha \check{g}(-t)]^{-1}, \quad t \in \mathbb{R}^d. \tag{4.8}$$

The conditions imposed on K guarantee that $\check{M}(\cdot, \vec{h}) \in \mathbb{L}_1(\mathbb{R}^d) \cap \mathbb{L}_2(\mathbb{R}^d)$ for any $\vec{h} \in \mathcal{H}^d$ and, hence,

$$\|M(\cdot, \vec{h})\|_\infty \leq (2\pi)^{-d} \|\check{M}(\cdot, \vec{h})\|_1, \quad \|M(\cdot, \vec{h})\|_2 = (2\pi)^{-d} \|\check{M}(\cdot, \vec{h})\|_2.$$

Set $\vec{\mu}(\alpha) = \vec{\mu}, \alpha = 1, \vec{\mu}(\alpha) = (0, \dots, 0), \alpha \in [0, 1)$. We have in view of Assumption 1 for any $\vec{h} \in \mathcal{H}^d$

$$\|M(\cdot, \vec{h})\|_\infty \leq \mathbf{c}_6 \prod_{j=1}^d h_j^{-1-\mu_j(\alpha)}, \quad \|M(\cdot, \vec{h})\|_2 \leq \mathbf{c}_6 \prod_{j=1}^d h_j^{-\frac{1}{2}-\mu_j(\alpha)}. \tag{4.9}$$

Additionally we deduce from (4.9) for any $p > 2$

$$\|M(\cdot, \vec{h})\|_p \leq \|M(\cdot, \vec{h})\|_\infty^{1-\frac{2}{p}} \|M(\cdot, \vec{h})\|_2^{\frac{2}{p}} \leq \mathbf{c}_6 \prod_{j=1}^d h_j^{-1+1/p-\mu_j(\alpha)}, \quad \forall \vec{h} \in \mathcal{H}^d. \tag{4.10}$$

If $p < 2$, we will study only the case $\alpha < 1/2$. Then we have in view of the definition of $M(\cdot, \vec{h})$, applying the Young and triangle inequalities

$$\|K_{\vec{h}}\|_p \geq (1-\alpha) \|M(\cdot, \vec{h})\|_p - \alpha \|M(\cdot, \vec{h})\|_p = (1-2\alpha) \|M(\cdot, \vec{h})\|_p.$$

It yields, for any $p < 2$ and $\alpha < 1/2$

$$\|M(\cdot, \vec{h})\|_p \leq (1 - 2\alpha)^{-1} \|K\|_p \prod_{j=1}^d h_j^{-1+1/p}, \quad \forall \vec{h} \in \mathcal{H}^d. \tag{4.11}$$

Thus, we obtain from (4.7), (4.9), (4.10) and (4.11) for any $\vec{h} \in \mathcal{H}_{[n/2]}^d$

$$\sup_{f \in \mathfrak{F}_p} \mathcal{E}(f, \vec{h}) \leq c_7 \begin{cases} n^{-\frac{1}{2}} \prod_{j=1}^d h_j^{-\frac{1}{2} - \mu_j(\alpha)}, & p \geq 2, \alpha \in [0, 1]; \\ n^{\frac{1}{p}-1} \prod_{j=1}^d h_j^{-1+1/p - \mu_j(\alpha)}, & p < 2, \alpha \in [0, 1/2). \end{cases} \tag{4.12}$$

To get the bound for $p \geq 2$ we have taken into account that $nV_{\vec{h}} \geq \ln(n)$ if $\vec{h} \in \mathcal{H}_{[n/2]}^d$.

In fact (4.7) is true for all $\alpha \in [0, 1]$ but it requires to impose several additional assumption on g and we do not treat the case $p < 2, \alpha > 1/2$ in the present paper. The interested reader can look the paper Rebelles [39], where the corresponding norm is computed when $\alpha = 1$.

4.3. Consequences of Proposition 1 and Theorem 3(2). Case $\alpha \in (0, 1)$

Let us compute the approximation error related to the estimator $\widehat{B}_{\vec{h}}$. Note that for any $\vec{h} \in \mathcal{H}^d$

$$\begin{aligned} \Upsilon_{\vec{h}}(f, x) &:= \int_{\mathbb{R}^d} M(z - x, \vec{h}) A_g(z) dz \\ &= (1 - \alpha) \int_{\mathbb{R}^d} M(z - x, \vec{h}) f(z) dz + \alpha \int_{\mathbb{R}^d} M(z - x, \vec{h}) [f \star g](z) dz \\ &= \int_{\mathbb{R}^d} f(t) \left[(1 - \alpha) M(t - x, \vec{h}) + \alpha \int_{\mathbb{R}^d} M(u, \vec{h}) g(u - [t - x]) du \right] dt \\ &= \int_{\mathbb{R}^d} K_{\vec{h}}(t - x) f(t) dt. \end{aligned}$$

Thus, we have for any $\vec{h} \in \mathcal{H}^d$ and any $p \geq 1$

$$\|\Upsilon_{\vec{h}}(f) - f\|_p = \|K_{\vec{h}} \star f - f\|_p. \tag{4.13}$$

4.3.1. Verification of the condition (3.3) and (3.4) of Proposition 1

We will study two different cases: either $p \neq 2, \alpha < 1/2$ or $p = 2, \alpha \in (0, 1)$.

Set $\Lambda_{\vec{h}}(f, \cdot) = \mathbb{E}_{1,f}^{(n)} \{\widehat{A}_{\vec{h}}(\cdot)\}$, $\vec{h} \in (0, 1]^d$ and note that

$$\Lambda_{\vec{h}}(f, \cdot) = \int_{\mathbb{R}^d} K_{\vec{h}}(z - \cdot) A_g(f, z) dz = (1 - \alpha) [K_{\vec{h}} \star f](\cdot) + \alpha [K_{\vec{h}} \star g \star f](\cdot). \tag{4.14}$$

We have in view of (4.14) and the Young inequality (recall that g is a density)

$$\begin{aligned} \|\Lambda_{\vec{h}}(f) - A_g(f)\|_p &= \|(1 - \alpha)\{K_{\vec{h}} \star f - f\} + \alpha g \star \{K_{\vec{h}} \star f - f\}\|_p \\ &\geq (1 - 2\alpha)\|K_{\vec{h}} \star f - f\|_p. \end{aligned}$$

Thus, if $\alpha < 1/2$ one has for any $p \geq 1$ in view of (4.13)

$$\|\Upsilon_{\vec{h}}(f) - f\|_p \leq (1 - 2\alpha)^{-1} \|\Lambda_{\vec{h}}(f) - A_g(f)\|_p$$

and, therefore, the condition (3.3) of Proposition 1 is fulfilled with $C_\ell = (1 - 2\alpha)^{-1}$.

If $p = 2$ we have in view of (4.14), (4.13), Assumption 1 and the Parseval identity

$$\begin{aligned} \|\Lambda_{\vec{h}}(f) - A_g(f)\|_2 &= \|(1 - \alpha)[\check{K}(\vec{h}\cdot) - 1]\check{f}(\cdot) + \alpha[\check{K}(\vec{h}\cdot) - 1]\check{g}(\cdot)\check{f}(\cdot)\|_2 \\ &= \|\{(1 + \alpha) + \alpha\check{g}(\cdot)\}[\check{K}(\vec{h}\cdot) - 1]\check{f}(\cdot)\|_2 \geq \varpi \|[\check{K}(\vec{h}\cdot) - 1]\check{f}(\cdot)\|_2 \\ &= \varpi \|\Upsilon_{\vec{h}}(f) - f\|_2. \end{aligned}$$

We conclude that the condition (3.3) of Proposition 1 is fulfilled with $C_\ell = \varpi^{-1}$ if $p = 2$.

Let us now check (3.4) of Proposition 1. Note that the definition of $\Delta_{[n/2]}$ implies

$$\Delta_{[n/2]}(\vec{h}) \geq \mathbf{c}_9 \begin{cases} n^{-\frac{1}{2}} \prod_{j=1}^d h_j^{-\frac{1}{2}}, & p \geq 2; \\ n^{\frac{1}{p}-1} \prod_{j=1}^d h_j^{-1+1/p}, & p < 2 \end{cases} \tag{4.15}$$

and therefore, (3.4) is fulfilled with $C_\mathcal{E} = \mathbf{c}_9/\mathbf{c}_7$ in view of (4.12), since recall $\vec{\mu}(\alpha) = 0, \alpha \neq 1$.

4.3.2. Main results

We deduce from Proposition 1 and Lemma 1 the following assertion.

Theorem 5. *Let either $p \neq 2, \alpha < 1/2$ or $p = 2, \alpha \in (0, 1)$. Then for all $n \geq 2$*

$$\mathcal{R}_B[\widehat{B}_{\vec{h}_n}, f] \leq \mathbf{c}_{10} \inf_{\vec{h} \in H_n} \{ \|K_{\vec{h}} \star f - f\|_p + (nV_{\vec{h}})^{\frac{1}{p\wedge 2}-1} \} + \varepsilon_{[n/2]}, \quad \forall f \in \mathfrak{F}_p.$$

We also have the following adaptive result.

Theorem 6. *Let either $p \neq 2, \alpha < 1/2$ or $p = 2, \alpha \in (0, 1)$. Then the estimator $\widehat{B}_{\vec{h}_n}$ is optimally-adaptive over the scale of anisotropic classes $\{\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{F}_p, \vec{\beta} \in (0, \mathfrak{s}]^d, L > 0\}$.*

The fact that $\widehat{B}_{\vec{h}_n}$ is S_n -adaptive with $S_n = \{s_{[n/2]}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{F}_p), \vec{\beta} \in (0, \mathfrak{s}]^d, L > 0\}$ follows from (3.32) with $p = f$ and $m = [n/2]$, the assertion (3.23) of Lemma 1 and the second assertion of Theorem 3.

The lower bound for minimax risk showing that S_n is the family of minimax rates in this problem can be found in Lepski and Willer [35], Theorem 1.

Remark 3. The results presented in Theorems 5 and 6 are new.

4.4. Consequences of Theorem 4. Case $\alpha = 1, p = 2, d = 1$

In this section, we will study the adaptive estimation over the collection of univariate Sobolev classes.

Definition 2. Let $\beta_1 \in (0, \infty)^d$ and $L > 0$ be given. We say that $Q \in \mathbb{L}_1(\mathbb{R}^1)$ belongs to Sobolev class $\mathbb{W}(\beta_1, L)$ if

$$\int_{\mathbb{R}^1} (1 + t^2)^{\beta_1} |\check{Q}(t)|^2 dt \leq L^2.$$

The adaptive estimation over the collection of *anisotropic* Sobolev classes when $\alpha = 1$ was studied in Comte and Lacour [10].

Auxiliary inequalities. The results below follow from the Parseval identity, the properties of the kernel K and Assumption 1. First note, that for any $f \in \mathbb{W}(\beta_1, L)$ and any $h_1 \in (0, 1)$

$$\|K_{h_1} \star A_g(f) - A_g(f)\|_2 \leq G_2 L h_1^{\beta_1 + \mu_1},$$

since obviously $A_g(f) \in \mathbb{W}(\beta_1 + \mu_1, G_2 L)$.

It yields in view of (3.29) and (3.30) with $p = A_g(f)$ and $p = 2$,

$$\begin{aligned} s_n(\mathbb{F}_\alpha) &= \sup_{f \in \mathbb{F}_\alpha} \inf_{\mathfrak{h} \in \mathfrak{S}_n} \{ \mathcal{B}_A^{(n)}(f, \mathfrak{h}) + \psi_n(f, \mathfrak{h}) \} \\ &=: \sup_{f \in \mathbb{W}(\beta_1, L)} \inf_{h_1 \in H_n^1} \{ (1 + 2\|K\|_1) \|K_{h_1} \star A_g(f) - A_g(f)\|_2 + \mathbf{c}_{11} (nh_1)^{-\frac{1}{2}} \} \quad (4.16) \\ &\leq \mathbf{c}_{12} L^{\frac{1}{2\beta_1 + 2\mu_1 + 1}} n^{-\frac{\beta_1 + \mu_1}{2\beta_1 + 2\mu_1 + 1}} =: \mathbf{c}_{12} s_n(\mathbb{W}(\beta_1, L)). \end{aligned}$$

We also have in view of Assumption 1 for any $f \in \mathbb{W}(\beta_1, L), y > 0$ and $h_1 \in (0, 1]$

$$\begin{aligned} &\|K_{h_1} \star f - f\|_2^2 \\ &\leq \int_{-y}^y |\check{K}(h_1 t) - 1|^2 |\check{f}(t)|^2 dt + (1 + \|K\|_1)^2 \int_{\mathbb{R}} 1_{\mathbb{R} \setminus [-y, y]}(t) |\check{f}(t)|^2 dt \\ &\leq \mathbf{c}_{13} \left[(1 + y^{2\mu_1}) \int_{-y}^y |g(t)|^2 |\check{K}(h_1 t) - 1|^2 |\check{f}(t)|^2 dt + L^2 y^{-2\beta_1} \right] \\ &\leq \mathbf{c}_{13} [(1 + y^{2\mu_1}) \|K_{h_1} \star A_g(f) - A_g(f)\|_2^2 + L^2 y^{-2\beta_1}]. \end{aligned}$$

Minimizing the r.h.s. of the latter inequality in y we get for any $f \in \mathbb{W}(\beta_1, L)$ and $h_1 \in (0, 1]$

$$\begin{aligned} \|K_{h_1} \star f - f\|_2 &\leq \mathbf{c}_{13} \left[\|K_{h_1} \star A_g(f) - A_g(f)\|_2 \right. \\ &\quad \left. + L^{\frac{\mu_1}{\mu_1 + \beta_1}} \|K_{h_1} \star A_g(f) - A_g(f)\|_2^{\frac{\beta_1}{\beta_1 + \mu_1}} \right]. \end{aligned} \tag{4.17}$$

Application of Theorem 4. First, we note that all assumptions of Theorem 4 are fulfilled. Indeed, in Section 3.4.1 we have already checked the hypotheses $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$. Moreover, defined in (3.18) quantity in our case is given by

$$\psi_n(\mathbb{W}(\beta_1, L), h_1) = \mathbf{c}_{14} (nh_1)^{-\frac{1}{2}}.$$

It yields together with (4.16) that

$$\mathcal{V}_n(\mathbf{a}) = \left\{ h_1 \in \mathcal{H}_n^1 : h_1 \geq \mathbf{c}_{15} L^{-\frac{2}{2\beta_1 + 2\mu_1 + 1}} n^{-\frac{1}{2\beta_1 + 2\mu_1 + 1}} \right\}, \quad \mathbf{a} = (\beta_1, L),$$

and we have that in view of (4.12)

$$\sup_{f \in \mathbb{F}_{\mathbf{a}}} \sup_{h \in \mathcal{V}_n(\mathbf{a})} \mathcal{E}_n(f, h) \leq \mathbf{c}_{16} L^{\frac{1+2\mu_1}{2\beta_1 + 2\mu_1 + 1}} n^{-\frac{\beta_1}{2\beta_1 + 2\mu_1 + 1}}. \tag{4.18}$$

Since $\mathcal{U}_n(\mathbf{a}, f) = \{h_1 \in \mathcal{H}_n^1 : \|K_{h_1} \star A_g(f) - A_g(f)\|_2 < 8s_n(\mathbb{W}(\beta_1, L))\}$ we obtain from (4.17)

$$\sup_{f \in \mathbb{F}_{\mathbf{a}}} \sup_{h \in \mathcal{U}_n(\mathbf{a}, f)} \|K_{h_1} \star f - f\|_2 \leq \mathbf{c}_{17} L^{\frac{\mu_1}{\mu_1 + \beta_1}} [s_n(\mathbb{W}(\beta_1, L))]^{\frac{\beta_1}{\beta_1 + \mu_1}} = \mathbf{c}_{17} L^{\frac{1+2\mu_1}{2\beta_1 + 2\mu_1 + 1}} n^{-\frac{\beta_1}{2\beta_1 + 2\mu_1 + 1}}.$$

It yields together with (4.18)

$$\varphi_n(\mathbb{W}(\beta_1, L)) = \mathbf{c}_{18} L^{\frac{1+2\mu_1}{2\beta_1 + 2\mu_1 + 1}} n^{-\frac{\beta_1}{2\beta_1 + 2\mu_1 + 1}}, \tag{4.19}$$

where \mathbf{c}_{18} independent of L . Putting $\mu^* = \max_{j=1, \dots, d} \mu_j$ we note that

$$\delta_n \geq \mathbf{c}_{19} n^{-\frac{1}{2}}, \quad \sup_{f \in \mathfrak{F}} \sup_{h_1 \in \mathcal{H}_n^1} \mathcal{E}_n(f, h_1) \leq \mathbf{c}_{20} n^{\mu^*},$$

$$\sup_{f \in \mathbb{W}(\beta_1, L)} \sup_{h_1 \in (0, 1]} \|K_{h_1} \star f - f\|_2 \leq 2\|K\|_1 L.$$

Hence, in view of the assertion (3.23) of Lemma 1 we can state that

$$\limsup_{n \rightarrow \infty} n^a \varkappa_n(\mathbb{W}(\beta_1, L)) = 0, \quad \forall a > 0. \tag{4.20}$$

Thus, we deduce from (4.17), (4.19), (4.20) and Theorem 4 the following result.

Assertion 1. Let \mathbf{h}_n comes from (4.5)–(4.6). Then for any $\mu_1 > 0, \beta_1 > 0$ and $L > 0$

$$\limsup_{n \rightarrow \infty} \varphi_n^{-1}(\mathbb{W}(\beta_1, L)) \sup_{f \in \mathbb{W}(\beta_1, L)} \{ \mathbb{E}_f^{(n)} \|\widehat{B}_{\mathbf{h}_n} - f\|_2^2 \}^{\frac{1}{2}} < \mathbf{c}_{21},$$

where \mathbf{c}_{21} is independent of L .

The estimator $\widehat{B}_{\mathbf{h}_n}$ is optimally-adaptive over the scale $\{\mathbb{W}(\beta_1, L), \beta_1 > 0, L > 0\}$.

Assertion 1 is the particular case of the results obtained in Comte and Lacour [10], which were established by the use of completely different selection scheme.

5. Simultaneous adaptive estimation of partial derivatives in the density model

Let as previously $X_i, i = 1, \dots, n$, be *i.i.d.* d -dimensional random vectors with unknown density f . This is the particular case of the model considered above corresponding to $\alpha = 0$.

For any multi-index $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d$ let $f^{(\mathbf{m})} = \frac{\partial^{m_1 + \dots + m_d} f}{\partial x_1^{m_1} \dots \partial x_d^{m_d}}$ be the partial derivative of f of order \mathbf{m} . The goal is to estimate $B(f) = f^{(\mathbf{m})}$ under \mathbb{L}_p -loss, that is

$$\mathcal{R}_B[\widetilde{B}_n, f] = (\mathbb{E}_f^{(n)} [\|\widetilde{B}_n - f^{(\mathbf{m})}\|_p^p])^{1/p}, \quad f \in \mathfrak{F}.$$

Remark 4. We will keep all previous notations with only one change related to the following simple observation. All oracle results established in the present paper remain valid if one replace in all formulas the set \mathfrak{H}_n (abstract model) or \mathcal{H}_m^d (density model) by its arbitrary subset. In particular we will use below the selection rule (4.5)–(4.6) from the collection $\mathbf{A}(\cdot)$, which is parameterized by either $\mathcal{H}_{[n/2]}^d$ as it was before or by its subset $\mathcal{H}_{[n/2]}^{\text{isotrop}} := \{\vec{h} \in \mathcal{H}_{[n/2]}^d : h_1 = \dots = h_d\}$. The latter set will be used when the adaptation over the scale of isotropic, ($\beta_1 = \dots = \beta_d$) Nikol’skii classes is studied.

Thus, let H_n denote either $\mathcal{H}_{[n/2]}^d$ or $\mathcal{H}_n^{\text{isotrop}}$ and let the collection $\mathbf{A}(H_n)$ be given by (4.4) (where $\mathcal{H}_{[n/2]}^d$ is replaced by H_n) and $Z_i = X_i, i = 1, \dots, [n/2]$. Introduce the family of estimators

$$\mathbf{B}(H_n) = \left\{ \widehat{B}_{\vec{h}}(\cdot) = (n - [n/2])^{-1} \sum_{i=[n/2]+1}^n \prod_{j=1}^d (-1)^{m_j} h_j^{-m_j} K_{\vec{h}}^{(\mathbf{m})}(X_i - \cdot), \vec{h} \in H_n \right\},$$

where $K^{(\mathbf{m})}$ denotes the partial derivative of K of the order \mathbf{m} .

It is important to note that the estimator $\widehat{B}_{\mathbf{h}_n}$ would coincide with $\widehat{A}_{\mathbf{h}_n}^{(\mathbf{m})}$ if we would construct it using observations $X_i, i = 1, \dots, [n/2]$ instead of $X_i, i = [n/2] + 1, \dots, n$. The interest to this

collection is dictated by the following minimax result. For any $\vec{\beta} \in (0, \infty)^d$ and $\mathbf{m} \in \mathbb{N}^d$ put

$$\frac{1}{\vec{\beta}} = \sum_{j=1}^d \frac{1}{\beta_j}, \quad \frac{1}{\omega} = \sum_{j=1}^d \frac{m_j}{\beta_j}.$$

Proposition 2. For any $p > 1, L > 0$, any $\vec{\beta} \in (0, \infty)^d$ and $\mathbf{m} \in \mathbb{N}^d$ provided $\omega > 1$ one can find a kernel K and $\vec{h}_n \in (0, 1]^d$ such that the estimator $\widehat{A}_{\vec{h}_n}^{(\mathbf{m})}$ is rate-minimax for $f^{(\mathbf{m})}$ on $\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p$.

The minimax rate of convergence is given by

$$\varphi_n^{(\mathbf{m})}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p) = L^{\frac{(1/\vec{\beta})(1-1/p \wedge 2)+1/\omega}{1+(1/\vec{\beta})(1-1/p \wedge 2)}} n^{-\frac{(1-1/\omega)(1-1/p \wedge 2)}{1+(1/\vec{\beta})(1-1/p \wedge 2)}}.$$

The conditions $\omega > 1$ and $p > 1$ are necessary for the existence of uniformly consistent on $\mathbb{N}_{p,d}(\vec{\beta}, L)$ estimators.

The assertions of the proposition seems to be new.

Now, let us formulate the main result of this section. As previously let $\mathfrak{s} \in \mathbb{N}^*$ be an arbitrary but a priory chosen number. Let the kernel K , involved to the description of estimators from the collections $\mathbf{A}(H_n)$ and $\mathbf{B}(H_n)$, be constructed in accordance with (3.31).

Set at last $\Pi_{\mathfrak{s}} = \{(\vec{\beta}, \mathbf{m}) \in (0, \mathfrak{s}]^d \times \mathbb{N}^d : \omega > 1\}$.

Theorem 7. Let \vec{h}_n be the issue of (4.5)–(4.6) with $H_n = \mathcal{H}_{[n/2]}^d$ if $m_1 = \dots = m_d$ or with $H_n = \mathcal{H}_n^{\text{isotrop}}$ if $\beta_1 = \dots = \beta_d$. Then for any $p \geq 1, L > 0$, and $(\vec{\beta}, \mathbf{m}) \in \Pi_{\mathfrak{s}}$

$$\limsup_{n \rightarrow \infty} [\varphi_n^{(\mathbf{m})}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p)]^{-1} \sup_{f \in \mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p} \left\{ \mathbb{E}_f^{(n)} \left\| \widehat{B}_{\vec{h}_n} - f^{(\mathbf{m})} \right\|_p^p \right\}^{\frac{1}{p}} < \mathbf{c}_{21},$$

where \mathbf{c}_{21} is independent of L .

Theorem 7 together with second statement of Proposition 2 allows us to assert that the estimator $\widehat{B}_{\vec{h}_n}$ is optimally-adaptive over the scale of either anisotropic Nikol'skii classes (if $m_1 = \dots = m_d$) or isotropic ones (without any restriction on the order of the considered partial derivative). Note that the problem is completely solved in the dimension 1. Also we conclude that the differentiation of an optimally-adaptive estimator leads to the optimally-adaptive estimator of the corresponding partial derivative.

5.1. Proof of Proposition 2

Lemma 2. For any $p \geq 1$, any $\vec{\beta}$ and \mathbf{m} , provided $\omega > 1$, and any $f \in \mathbb{N}_{p,d}(\vec{\beta}, L)$ one has

$$\|f^{(\mathbf{m})}\|_p \leq \mathbf{c}_{23} L^{1/\omega} \|f\|_p^{1-1/\omega}, \tag{5.1}$$

where \mathbf{c}_{23} is independent of L .

The author is sure that this *Kolmogorov-type* inequality should be known but he was unable to find the exact reference.

Proof of Lemma 2. The following inequality was proved in Goldenshluger and Lepski [21]. For any $p \geq 1$ and $\mathfrak{s} \in \mathbb{N}^d$ there exists \mathbf{c}_{22} is independent of L such that

$$\sup_{F \in \mathbb{N}_{p,d}(\vec{\beta}, L)} \|K_{\vec{h}} \star F - F\|_p \leq \mathbf{c}_{22} L \sum_{j=1}^d h_j^{\beta_j}, \quad \forall \vec{h} \in (0, 1]^d, \forall \vec{\beta} \in (0, \mathfrak{s}]^d. \tag{5.2}$$

The following statement can be found in Nikol'skii [38], Chapter 5, Theorem 5.6.3.

If $\omega > 1$ and $f \in \mathbb{N}_{p,d}(\vec{\beta}, L)$, then $f^{(m)}$ exists and

$$f^{(m)} \in \mathbb{N}_{p,d}(\vec{\gamma}, \mathbf{c}L), \quad \gamma_j = \beta_j(1 - 1/\omega), \quad j = 1, \dots, d, \tag{5.3}$$

and $\mathbf{c} > 0$ is independent of L .

Since $\vec{\beta}$ is fixed we can always choose $\mathfrak{s} \in \mathbb{N}^d$ in order to have $\vec{\beta} \in (0, \mathfrak{s}]^d$ and, then, (5.2) will be fulfilled. We obviously have by the triangle inequality

$$\|f^{(m)}\|_p \leq \|(K_{\vec{h}})^{(m)} \star f - f^{(m)}\|_p + \|(K_{\vec{h}})^{(m)} \star f\|_p.$$

It is easy to see that $(K_{\vec{h}})^{(m)} = \prod_{j=1}^d (-1)^{m_j} h_j^{-m_j} K_{\vec{h}_n}^{(m)}$ for any $\vec{h} \in (0, 1]^d$ and we have first by integrating by parts

$$(K_{\vec{h}})^{(m)} \star f - f^{(m)} = K_{\vec{h}} \star f^{(m)} - f^{(m)}. \tag{5.4}$$

Hence, applying (5.2) with $F = f^{(m)}$ and $\vec{\beta}$ replaced by $\vec{\gamma}$ we obtain in view of (5.3)

$$\|(K_{\vec{h}})^{(m)} \star f - f^{(m)}\|_p \leq \mathbf{c}_{24} L \sum_{j=1}^d h_j^{\gamma_j}, \quad \forall \vec{h} \in (0, 1]^d. \tag{5.5}$$

Moreover, we obtain applying the Young inequality

$$\|(K_{\vec{h}})^{(m)} \star f\|_p \leq \|(K_{\vec{h}})^{(m)}\|_1 \|f\|_p = \|f\|_p \|K^{(m)}\|_1 \prod_{j=1}^d h_j^{-m_j},$$

that yields together with (5.5) for any $f \in \mathbb{N}_{p,d}(\vec{\beta}, L)$

$$\|f^{(m)}\|_p \leq \mathbf{c}_{25} \left\{ L \sum_{j=1}^d h_j^{\beta_j(1-1/\omega)} + \|f\|_p \prod_{j=1}^d h_j^{-m_j} \right\}, \quad \forall \vec{h} \in (0, 1]^d.$$

Choosing $h_j = (\|f\|_p/L)^{1/\beta_j}$ (that is possible since $\|f\|_p \leq L$ in view of the definition of the Nikol'skii class) we come to the assertion of the lemma. □

Proof of the proposition. Since

$$\widehat{A}_{\vec{h}_n}^{(m)}(\cdot) = [n/2]^{-1} \sum_{i=1}^{[n/2]} \prod_{j=1}^d (-1)^{m_j} h_j^{-m_j} K_{\vec{h}_n}^{(m)}(X_i - \cdot)$$

we have in view of (5.5) for any $p \geq 1$, $f \in \mathbb{N}_{p,d}(\vec{\beta}, L)$ and $\vec{h} \in (0, 1]^d$

$$\|\mathbb{E}_{1,f}^{(n)}(\widehat{A}_{\vec{h}}^{(m)}) - f^{(m)}\|_p = \|(K_{\vec{h}}^{(m)}) \star f - f^{(m)}\|_p \leq \mathbf{c}_{24} L \sum_{j=1}^d h_j^{\beta_j(1-1/\omega)}. \tag{5.6}$$

Next, repeating the computations led to (4.12) (remind that our considerations here correspond to the case $\alpha = 0$) we get for any $p \geq 1$, $f \in \mathfrak{F}$ and $\vec{h} \in (0, 1]^d$

$$\left[\mathbb{E}_{1,f}^{(n)} \|\widehat{A}_{\vec{h}}^{(m)} - \mathbb{E}_{1,f}^{(n)}(\widehat{A}_{\vec{h}}^{(m)})\|_p^p \right]^{\frac{1}{p}} \leq \mathbf{c}_{26} \prod_{j=1}^d h_j^{-m_j} \left(n \prod_{j=1}^d h_j \right)^{\frac{1}{p\wedge 2} - 1}. \tag{5.7}$$

We deduce from (5.6) and (5.7) that for any $p \geq 1$, $f \in \mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{F}_p$ and $\vec{h} \in (0, 1]^d$

$$\begin{aligned} \mathcal{R}_A[\widehat{A}_{\vec{h}}, f^{(m)}] &:= \left[\mathbb{E}_{1,f}^{(n)} \|\widehat{A}_{\vec{h}}^{(m)} - f^{(m)}\|_p^p \right]^{\frac{1}{p}} \\ &\leq \mathbf{c}_{27} \left\{ L \sum_{j=1}^d h_j^{\beta_j(1-1/\omega)} + \prod_{j=1}^d h_j^{-m_j} \left(n \prod_{j=1}^d h_j \right)^{\frac{1}{p\wedge 2} - 1} \right\}. \end{aligned}$$

Noting that the right-hand side of the obtained inequality is independent of f and minimizing it with respect to \vec{h} we come to the following bound.

$$\sup_{f \in \mathbb{N}_{p,d}(\vec{\beta}, L)} \mathcal{R}_A[\widehat{A}_{\vec{h}_n}, f^{(m)}] \leq \mathbf{c}_{28} \varphi_n^{(m)}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{F}_p),$$

where \vec{h}_n is the minimizer of the right-hand side of penultimate inequality.

Thus, we have proved that the maximal risk is upper-bounded by $\varphi_n^{(m)}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{F}_p)$. The proof of the corresponding lower bound estimate (including the third assertion of the proposition) follows immediately from general lower bound construction established in Goldenshluger and Lepski [21], Theorem 3 (tail and dense zones) and it can be omitted. \square

5.2. Proof of Theorem 7

The proof consists in the application of Theorem 4. Since in Section 3.4.1 we have already checked the hypothesis $\mathbf{A}^{\text{permute}}$, $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$, it remains to compute the quantities

$$\varphi_n(\mathbb{F}_a) = \sup_{f \in \mathbb{F}_a} \sup_{\mathfrak{h} \in \mathcal{V}_n(a)} \mathcal{E}_n(f, \mathfrak{h}) + \sup_{f \in \mathbb{F}_a} \sup_{\mathfrak{h} \in \mathcal{U}_n(a, f)} \rho(\Upsilon_{\mathfrak{h}}, B(f)),$$

$$\varkappa_n(\mathbb{F}_\alpha) = (5\varepsilon_n/\delta_n) \sup_{f \in \mathbb{F}_\alpha} \left[\sup_{\mathfrak{h} \in \mathfrak{H}_n} \mathcal{E}_n(f, \mathfrak{h}) + 6 \sup_{\mathfrak{h} \in \mathfrak{H}_n} \rho(\Upsilon_{\mathfrak{h}}, B(f)) \right],$$

where $\mathcal{U}_n(\mathfrak{a}, f) = \{\mathfrak{h} : \ell(\Lambda_{\mathfrak{h}}, A(f)) < 8C_{\Psi} s_n(\mathbb{F}_\alpha)\}$ and $\mathcal{V}_n(\mathfrak{a}) = \{\mathfrak{h} : \psi_n(\mathbb{F}_\alpha, \mathfrak{h}) < 4C_{\Psi}^2 s_n(\mathbb{F}_\alpha)\}$.

In the considered problem $\mathbb{F}_\alpha = \mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p$, $\mathfrak{a} = (\vec{\beta}, L)$, $\mathfrak{h} = \vec{h} \in H_n$ and

$$\begin{aligned} \ell(\Lambda_{\mathfrak{h}}, A(f)) &= \|K_{\vec{h}} \star f - f\|_p, \\ \rho(\Upsilon_{\mathfrak{h}}, B(f)) &= \|(K_{\vec{h}})^{(m)} \star f - f^{(m)}\|_p. \end{aligned} \tag{5.8}$$

We also have from (3.30) and (5.7) that for any $f \in \mathfrak{P}_p$ and independently of $\vec{\beta}$ and L

$$\psi_n(\mathbb{F}_\alpha, \mathfrak{h}) = \underline{C}(nV_{\vec{h}})^{\frac{1}{p\wedge 2}-1}, \quad \mathcal{E}_n(f, \vec{h}) \leq \mathbf{c}_{26} \prod_{j=1}^d h_j^{-m_j} \left(n \prod_{j=1}^d h_j \right)^{\frac{1}{p\wedge 2}-1}. \tag{5.9}$$

To get the second inequality we have taken into account that the estimators $\widehat{B}_{\vec{h}}$ and $\widehat{A}_{\vec{h}}^{(m)}$ have the same distribution. Moreover,

$$s_{[n/2]}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p) = \mathbf{c}_{29} L^{\frac{1-1/(p\wedge 2)}{\vec{\beta}+1-1/(p\wedge 2)}} n^{-\frac{1-1/(p\wedge 2)}{1+(1/\vec{\beta})(1-1/(p\wedge 2))}}$$

as it was found in (3.32). It remains to note that $\delta_n \geq \mathbf{c}_{30} n^{-\frac{1}{2}}$ in view of (3.30) and the definition of $\mathcal{H}_{[n/2]}^d$ and moreover in view of (5.8), (5.9) and (5.3) and (5.4)

$$\sup_{\vec{h} \in \mathcal{H}_{[n/2]}^d} \sup_{f \in \mathfrak{P}_p} \mathcal{E}_n(f, \mathfrak{h}) \leq n^{m^*}, \quad \sup_{f \in \mathbb{N}_{p,d}(\vec{\beta}, L)} \|(K_{\vec{h}})^{(m)} \star f - f^{(m)}\|_p \leq \mathbf{c}_{31} L,$$

where $m^* = \max_{j=1, \dots, d} m_j$. To get the last bound, we have also used the Young inequality.

Since the hypotheses $\mathbf{A}^{\text{upper}}$ and $\mathcal{A}^{\text{upper}}$ are checked with $\varepsilon_{[n/2]}(p)$ verifying the assertion (3.23) of Lemma 1 we can first assert that

$$\limsup_{n \rightarrow \infty} n^a \varkappa_n(\mathbb{F}_\alpha) = 0, \quad \forall a > 0. \tag{5.10}$$

Next, in view of (5.4) we have for any $\vec{h} \in (0, 1]^d$

$$(K_{\vec{h}})^{(m)} \star f - f^{(m)} = K_{\vec{h}} \star f^{(m)} - f^{(m)} = (K_{\vec{h}} \star f - f)^{(m)}$$

and, applying Lemma 2, we obtain for any $f \in \mathbb{N}_{p,d}(\vec{\beta}, L)$

$$\begin{aligned} \rho(\Upsilon_{\mathfrak{h}}, B(f)) &:= \|(K_{\vec{h}})^{(m)} \star f - f^{(m)}\|_p \leq \mathbf{c}_{23} L^{1/\omega} \|K_{\vec{h}} \star f - f\|_p^{1-1/\omega} \\ &=: \mathbf{c}_{23} L^{1/\omega} \ell^{1-1/\omega}(\Lambda_{\mathfrak{h}}, A(f)). \end{aligned}$$

It yields for any $(\vec{\beta}, \mathbf{m}) \in \Pi_{\mathfrak{s}}$ and any $p \geq 1$

$$\begin{aligned} \sup_{f \in \mathbb{F}_{\mathfrak{a}}} \sup_{\mathfrak{h} \in \mathcal{U}_n(\mathfrak{a}, f)} \rho(\Upsilon_{\mathfrak{h}}, B(f)) &\leq \mathbf{c}_{32} L^{1/\omega} [s_n(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p)]^{1-1/\omega} \\ &= \mathbf{c}_{33} L^{\frac{(1/\beta)(1-1/p \wedge 2)+1/\omega}{1+(1/\beta)(1-1/p \wedge 2)}} n^{-\frac{(1-1/\omega)(1-1/(p \wedge 2))}{1+(1/\beta)(1-1/(p \wedge 2))}} \\ &= \mathbf{c}_{33} \varphi_n^{(\mathbf{m})}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p). \end{aligned} \tag{5.11}$$

Put for brevity $s_n = s_n(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p)$, $r = p \wedge 2$ and remark that

$$\mathcal{V}_n(\mathfrak{a}) = \{ \vec{h} \in H_n : V_{\vec{h}}^{1-\frac{1}{r}} > n^{\frac{1}{r}-1} (\mathbf{c}_{33} s_n)^{-1} \} = \{ \vec{h} \in H_n : V_{\vec{h}} > \mathbf{c}_{34} (L n^{1-1/r})^{-\frac{1}{\beta+1-1/r}} \}.$$

We obtain in view of (5.9)

$$\mathcal{E}_n(f, \vec{h}) \leq \mathbf{c}_{26} \begin{cases} n^{1/r-1} V_{\vec{h}}^{1/r-1-m_1}, & m_1 = \dots = m_d; \\ n^{1/r-1} h_1^{d(1/r-1-d^{-1} \sum_{j=1}^d m_j)}, & \vec{h} \in \mathcal{H}_n^{\text{isotrop}}. \end{cases}$$

Note that both bounds coincide since $V_{\vec{h}} = h_1^d$ if $\vec{h} \in \mathcal{H}_n^{\text{isotrop}}$ and $m_1 = d^{-1} \sum_{j=1}^d m_j$ in the first case. Simple algebra shows that

$$\sup_{f \in \mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p} \sup_{\vec{h} \in \mathcal{V}_n(\mathfrak{a})} \mathcal{E}_n(f, \mathfrak{h}) \leq \mathbf{c}_{35} \varphi_n^{(\mathbf{m})}(\mathbb{N}_{p,d}(\vec{\beta}, L) \cap \mathfrak{P}_p). \tag{5.12}$$

The assertion of the theorem follows now from (5.10), (5.11), (5.12) and Theorem 4.

Acknowledgements

The author is grateful to A. Goldenshluger and Yu. Golubev for fruitful discussions.

This work has been carried out in the framework of the Labex Archimède (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the ‘‘Investissements d’Avenir’’ French Government program managed by the French National Research Agency (ANR).

References

- [1] Abramovich, F., Grinshtein, V., Petsa, A. and Sapatinas, T. (2010). On Bayesian testimation and its application to wavelet thresholding. *Biometrika* **97** 181–198. [MR2594426](#)
- [2] Baraud, Y., Birgé, L. and Sart, M. (2017). A new method for estimation and model selection: ρ -estimation. *Invent. Math.* **207** 425–517. [MR3595933](#)
- [3] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. [MR1679028](#)
- [4] Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)* **3** 203–268. [MR1848946](#)

- [5] Bunea, F., Tsybakov, A.B. and Wegkamp, M.H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. [MR2351101](#)
- [6] Cai, T.T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27** 898–924. [MR1724035](#)
- [7] Cavalier, L. and Golubev, Yu. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. [MR2283712](#)
- [8] Cavalier, L. and Tsybakov, A.B. (2001). Penalized blockwise Stein’s method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.* **10** 247–282. [MR1867161](#)
- [9] Chernousova, E. and Golubev, G.K. (2014). Pointwise adaptive estimation of a multivariate function. *Math. Methods Statist.* **23** 1–16.
- [10] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 569–609. [MR3088382](#)
- [11] Dalalyan, A. and Tsybakov, A.B. (2008). Aggregation by exponential weighting, sharp PAC-Bayesian bounds and sparsity. *Mach. Learn.* **72** 39–61.
- [12] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25** 2626–2637. [MR1604428](#)
- [13] Efromovich, S. (1998). Simultaneous sharp estimation of functions and their derivatives. *Ann. Statist.* **26** 273–278. [MR1611776](#)
- [14] Egishyants, S.A. and Ostrovskii, E.I. (1996). Local and global upper functions for random fields. *Theory Probab. Appl.* **41** 657–665.
- [15] Folland, G.B. (1999). *Real Analysis. Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics (New York)*. New York: Wiley. [MR1681462](#)
- [16] Goldenshluger, A. (2009). A universal procedure for aggregating estimators. *Ann. Statist.* **37** 542–568. [MR2488362](#)
- [17] Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. [MR2543590](#)
- [18] Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via \mathbb{L}_p -norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. [MR2449122](#)
- [19] Goldenshluger, A. and Lepski, O. (2011). Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* **39** 2318–2384. [MR2932670](#)
- [20] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. [MR2850214](#)
- [21] Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on R^d . *Probab. Theory Related Fields* **159** 479–543. [MR3230001](#)
- [22] Goldenshluger, A.V. and Lepski, O.V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl.* **57** 209–226. [MR3201652](#)
- [23] Hesse, C.H. (1995). Deconvolving a density from partially contaminated observations. *J. Multivariate Anal.* **55** 246–260. [MR1370403](#)
- [24] Juditsky, A.B., Lepski, O.V. and Tsybakov, A.B. (2009). Nonparametric estimation of composite functions. *Ann. Statist.* **37** 1360–1404. [MR2509077](#)
- [25] Kerkycharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. [MR1863916](#)
- [26] Kerkycharian, G., Lepski, O. and Picard, D. (2008). Nonlinear estimation in anisotropic multiindex denoising. Sparse case. *Theory Probab. Appl.* **52** 58–77.
- [27] Knapik, B. and Solomond, J.-B. (2015). A general approach to posterior contraction in nonparametric inverse problems. Available at [arXiv:1407.0335v2](#).
- [28] Lepski, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. [MR3099129](#)

- [29] Lepski, O. (2013). Upper functions for positive random functionals. I. General setting and Gaussian random functions. *Math. Methods Statist.* **22** 1–27. [MR3040410](#)
- [30] Lepski, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 1. *Math. Methods Statist.* **22** 83–99. [MR3071956](#)
- [31] Lepski, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 2. *Math. Methods Statist.* **22** 193–212. [MR3107668](#)
- [32] Lepski, O. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43** 1178–1242. [MR3346701](#)
- [33] Lepski, O. (2016). Upper functions for \mathbb{L}_p -norms of Gaussian random fields. *Bernoulli* **22** 732–773. [MR3449799](#)
- [34] Lepski, O.V. and Levit, B.Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.* **7** 123–156. [MR1643256](#)
- [35] Lepski, O.V. and Willer, T. (2017). Lower bounds in the convolution structure density model. *Bernoulli* **23** 884–926. [MR3606754](#)
- [36] Massart, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math.* **1896**. Berlin: Springer. [MR2319879](#)
- [37] Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1998)*. *Lecture Notes in Math.* **1738** 85–277. Berlin: Springer. [MR1775640](#)
- [38] Nikol'skiĭ, S.M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya*. Moscow: “Nauka”. [Approximation of functions of several variables and imbedding theorems.] Second edition, revised and supplemented. [MR0506247](#)
- [39] Rebelles, G. (2016). Structural adaptive deconvolution under \mathbb{L}_p -losses. *Math. Methods Statist.* **25** 26–53. [MR3480609](#)
- [40] Reynaud-Bouret, P., Rivoirard, V., Grammont, F. and Tuleau-Malot, C. (2014). Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neurosci.* **4** Art. 3, 41. [MR3197017](#)
- [41] Rigollet, P. and Tsybakov, A.B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. [MR2356821](#)
- [42] Rosenthal, H.P. (1970). On the subspaces of \mathbb{L}_p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303. [MR0271721](#)
- [43] Tsybakov, A. (2003). Optimal rate of aggregation. In *Proc. COLT. Lecture Notes in Artificial Intelligence* **2777** 303–313.
- [44] van de Geer, S.A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge: Cambridge Univ. Press. [MR1739079](#)
- [45] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. [MR1385671](#)
- [46] von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.* **36** 299–303. [MR0170407](#)
- [47] Wegkamp, M. (2003). Model selection in nonparametric regression. *Ann. Statist.* **31** 252–273. [MR1962506](#)

Received October 2016 and revised March 2017