Oracle inequalities and adaptive estimation in the convolution structure density model

O. V Lepski, T. Willer

Submitted to the Annals of Statistics

ORACLE INEQUALITIES AND ADAPTIVE ESTIMATION IN THE CONVOLUTION STRUCTURE DENSITY MODEL

By O.V. Lepski * T. Willer *

Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

We study the problem of nonparametric estimation under $\mathbb{L}_p$-loss, $p \in [1, \infty)$, in the framework of the convolution structure density model on $\mathbb{R}^d$. This observation scheme is a generalization of two classical statistical models, namely density estimation under direct and indirect observations. The original pointwise selection rule from a family of "kernel-type" estimators is proposed. For the selected estimator, we prove an $\mathbb{L}_p$-norm oracle inequality and several of its consequences. Next, the problem of adaptive minimax estimation under $\mathbb{L}_p$–loss over the scale of anisotropic Nikol'skii classes is addressed. We fully characterize the behavior of the minimax risk for different relationships between regularity parameters and norm indexes in the definitions of the functional class and of the risk. We prove that the proposed selection rule leads to the construction of an optimally or nearly optimally (up to logarithmic factor) adaptive estimator.

1. Introduction. In the present paper we will investigate the following observation scheme introduced in Lepski and Willer (2017). Suppose that we observe i.i.d. vectors $Z_i \in \mathbb{R}^d, i = 1, \dots, n$, with a common probability density $\mathfrak{p}$ satisfying the following structural assumption

$$\mathfrak{p} = (1-\alpha)f + \alpha[f \star g], \quad f \in \mathbb{F}_g(R), \;\; \alpha \in [0,1], \tag{1.1}$$

where $\alpha \in [0,1]$ and $g : \mathbb{R}^d \to \mathbb{R}$ are supposed to be known and $f : \mathbb{R}^d \to \mathbb{R}$ is the function to be estimated. We will call the observation scheme (1.1) *convolution structure density model*.

Here and later, for two functions $f, g \in \mathbb{L}_1\big(\mathbb{R}^d\big)$

$$\big[f \star g\big](x) = \int_{\mathbb{R}^d} f(x-z)g(z)\nu_d(\mathrm{d}z), \;\; x \in \mathbb{R}^d,$$

*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).

AMS 2000 subject classifications: 62G05, 62G20

Keywords and phrases: deconvolution model, density estimation, oracle inequality, adaptive estimation, kernel estimators, $\mathbb{L}_p$–risk, anisotropic Nikol'skii class

and for any $\alpha \in [0,1]$, $g \in \mathbb{L}_1(\mathbb{R}^d)$ and $R>1$,

$$\mathbb{F}_g(R) = \Big\{ f \in \mathbb{B}_{1,d}(R) : \ (1-\alpha)f + \alpha[f \star g] \in \mathfrak{P}(\mathbb{R}^d) \Big\}.$$

Here $\mathfrak{P}(\mathbb{R}^d)$ denotes the set of probability densities on $\mathbb{R}^d$, $\mathbb{B}_{s,d}(R)$ is the ball of radius $R>0$ in $\mathbb{L}_s(\mathbb{R}^d) := \mathbb{L}_s(\mathbb{R}^d, \nu_d), 1 \le s \le \infty$ and ν_d is the Lebesgue measure on $\mathbb{R}^d$. The convolution structure density model (1.1) will be studied for an arbitrary $g \in \mathbb{L}_1(\mathbb{R}^d)$ and $f \in \mathbb{F}_g(R)$. Then, except in the case $\alpha = 0$, the function f is not necessarily a probability density.

We remark that if one assumes additionally that $f, g \in \mathfrak{P}(\mathbb{R}^d)$, this model can be interpreted as follows. The observations $Z_i \in \mathbb{R}^d, i = 1, \ldots, n$, can be written as a sum of two independent random vectors, that is,

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

where $X_i, i = 1, \ldots, n$, are *i.i.d.* d-dimensional random vectors with a common density f, to be estimated. The noise variables $Y_i, i = 1, \ldots, n$, are *i.i.d.* d-dimensional random vectors with a known common density g. At last $\varepsilon_i \in \{0,1\}, i = 1, \ldots, n$, 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 $\{X_i, i = 1, \ldots, n\}$, $\{Y_i, i = 1, \ldots, n\}$ and $\{\epsilon_i, i = 1, \ldots, n\}$ are supposed to be mutually independent.

The observation scheme (1.2) can be viewed as the generalization of two classical statistical models. Indeed, the case $\alpha = 1$ corresponds to the standard deconvolution model $Z_i = X_i + Y_i,\ i = 1, \ldots, n$. Another "extreme" case $\alpha = 0$ corresponds to the direct observation scheme $Z_i = X_i,\ i = 1, \ldots, n$. The "intermediate" case $\alpha \in (0,1)$, considered for the first time in Hesse (1995), can be treated as the mathematical modeling of the following situation. One part of the data, namely $(1-\alpha)n$, is observed without noise, while the other part is contaminated by additional noise. If the indexes corresponding to that first part were known, the density f could be estimated using only this part of the data, with the accuracy corresponding to the direct case. The question we address now is: can one obtain the same accuracy if the latter information is not available? We will see that the answer to the aforementioned question is positive, but the construction of optimal estimation procedures is based upon ideas corresponding to the "pure" deconvolution model.

We want to estimate f using the observations $Z^{(n)} = (Z_1, \ldots, Z_n)$. By estimator, we mean any $Z^{(n)}$-measurable map $\hat{f} : \mathbb{R}^n \to \mathbb{L}_p(\mathbb{R}^d)$. The accuracy of an estimator $\hat{f}$ is measured by the $\mathbb{L}_p$–risk

$$\mathcal{R}_n^{(p)}[\hat{f}, f] := \Big(\mathbb{E}_f \| \hat{f} - f \|_p^p \Big)^{1/p}, \ p \in [1, \infty),$$

where $\mathbb{E}_f$ denotes the expectation with respect to the probability measure $\mathbb{P}_f$ of the observations $Z^{(n)} = (Z_1, \dots, Z_n)$. Also, $\|\cdot\|_p$, $p \in [1, \infty)$, is the $\mathbb{L}_p$-norm on $\mathbb{R}^d$ and without further mentioning we will assume that $f \in \mathbb{L}_p\big(\mathbb{R}^d\big)$.

1.1. *Oracle approach via local selection.* Let $\mathcal{F}\big(\mathbb{H}\big) = \big\{\hat{f}_{\vec{h}}, \vec{h} \in \mathbb{H}\big\}$ be a family of estimators of "kernel-type" estimators, see Section 2.1, parameterized by a collection of multi-bandwidths $\mathbb{H}$ built from the observation $Z^{(n)}$. We want to construct a $Z^{(n)}$-measurable random map $\vec{\mathbf{h}} : \mathbb{R}^d \to \mathbb{H}$ and for any $p \in [1, \infty)$ and $n \geq 1$ to bound from above the $\mathbb{L}_p$-risk of the selected estimator $\hat{f}_{\vec{\mathbf{h}}(\cdot)}$. Our selection rule presented in Section 2.1 can be viewed as a generalization and modification of statistical procedures proposed in Kerkyacharian et al. (2001) and Goldenshluger and Lepski (2014). In Section 2.2, the following risk bound will be established

$$\mathcal{R}_n^{(p)}\big[\hat{f}_{\vec{\mathbf{h}}(\cdot)}; f\big] \leq C_1 \Big\| \inf_{\vec{h} \in \mathbb{H}} A_n\left(f, \vec{h}, \cdot\right) \Big\|_p + C_2 n^{-\frac{1}{2}}, \quad \forall f \in \mathbb{F}_g(R). \tag{1.3}$$

Here C_1 and C_2 are numerical constants which depend on d and p only, and $A_n(\cdot, \cdot, x), x \in \mathbb{R}^d$, is an explicitly known functional. We call (1.3) an $\mathbb{L}_p$-*norm oracle inequality obtained by local selection.* Since the selection rule from the considered family is done pointwisely, i.e. for any $x \in \mathbb{R}^d$, this allows to take into account the "local structure" of the function to be estimated. The $\mathbb{L}_p$-norm oracle inequality is then obtained by the integration of the pointwise risk of the proposed estimator, which is a kernel estimator with the bandwidth being a multivariate random function. This, in its turn, allows us to derive different minimax adaptive results thanks to an unique $\mathbb{L}_p$-norm oracle inequality. It is worth noting in this context that estimation procedures based on a local selection scheme can be applied to the estimation of functions belonging to much more general functional classes than whose based on global selection schemes, see for instance Goldenshluger and Lepski (2011) and Goldenshluger and Lepski (2014) for comparison. We will see however that although $A_n(\cdot, \cdot, x), x \in \mathbb{R}^d$, is known explicitly, its computation in particular problems is not a simple task. The main difficulty here is mostly related to the fact that (1.3) is proved without any assumption (except for the model requirements) imposed on the underlying function f. It turns out that under some nonrestrictive assumptions imposed on f, the obtained bound can be considerably simplified, see Section 3.

1.2. *Adaptive estimation.* Let Σ be a given subset of $\mathbb{L}_p\big(\mathbb{R}^d\big)$. For any estimator $\tilde{f}_n$, define its *maximal risk* by $\mathcal{R}_n^{(p)}\big[\tilde{f}_n; \Sigma\big] = \sup_{f \in \Sigma} \mathcal{R}_n^{(p)}\big[\tilde{f}_n; f\big]$ and its *minimax risk* on Σ is given by

$$\phi_n(\Sigma) := \inf_{\tilde{f}_n} \mathcal{R}_n^{(p)}\big[\tilde{f}_n; \Sigma\big].$$

Here, the infimum is taken over all possible estimators. An estimator whose maximal risk is bounded, up to some constant factor, by $\phi_n(\Sigma)$, is called minimax on Σ.

Let $\big\{\Sigma_\vartheta, \vartheta \in \Theta\big\}$ be a collection of subsets of $\mathbb{L}_p\big(\mathbb{R}^d, \nu_d\big)$, where ϑ is a nuisance parameter which may have a very complicated structure.

The problem of adaptive estimation can be formulated as follows: *is it possible to construct a single estimator $\hat{f}_n$ which would be simultaneously minimax on each class Σ_ϑ, $\vartheta \in \Theta$, i.e.*

$$\limsup_{n\to\infty} \phi_n^{-1}(\Sigma_\vartheta)\mathcal{R}_n^{(p)}\big[\hat{f}_n; \Sigma_\vartheta\big] < \infty, \quad \forall \vartheta \in \Theta?$$

We refer to this question as *the problem of minimax adaptive estimation over the scale of* $\{\Sigma_\vartheta,\ \vartheta \in \Theta\}$. If such an estimator exists, we will call it optimally adaptive.

From oracle approach to adaptation. Let the oracle inequality (1.3) be established. Define

$$R_n\big(\Sigma_\vartheta\big) = \sup_{f\in\Sigma_\vartheta}\bigg\| \inf_{\vec{h}\in\mathbb{H}} A_n\Big(f, \vec{h}, \cdot\Big)\bigg\|_p + n^{-\frac{1}{2}}, \quad \vartheta \in \Theta.$$

We immediately deduce from (1.3) that for any $\vartheta \in \Theta$

$$\limsup_{n\to\infty} R_n^{-1}\big(\Sigma_\vartheta\big)\mathcal{R}_n^{(p)}\big[\hat{f}_{\vec{\mathbf{h}}(\cdot)}; \Sigma_\vartheta\big] < \infty.$$

Hence, the minimax adaptive optimality of the estimator $\hat{f}_{\vec{\mathbf{h}}(\cdot)}$ is reduced to the comparison of the normalization $R_n\big(\Sigma_\vartheta\big)$ with the minimax risk $\phi_n(\Sigma_\vartheta)$. Indeed, if one proves that for any $\vartheta \in \Theta$

$$\liminf_{n\to\infty} R_n\big(\Sigma_\vartheta\big)\phi_n^{-1}(\Sigma_\vartheta) < \infty,$$

then the estimator $\hat{f}_{\vec{\mathbf{h}}(\cdot)}$ is *optimally adaptive* over the scale $\big\{\Sigma_\vartheta, \vartheta \in \Theta\big\}$. Using the modern statistical language we call the estimator $\hat{f}_n$ *nearly optimally adaptive* if

$$\limsup_{n\to\infty} \phi^{-1}_{\frac{n}{\ln n}}(\Sigma_\vartheta)\mathcal{R}_n^{(p)}\big[\hat{f}_n; \Sigma_\vartheta\big] < \infty, \quad \forall \vartheta \in \Theta.$$

Objectives. In the framework of the *convolution structure density model*, we will be interested in adaptive estimation over the scale

$$\Sigma_\vartheta = \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big) \cap \mathbb{F}_{g,\infty}(R, Q), \quad \vartheta = \big(\vec{\beta}, \vec{r}, \vec{L}, R, Q\big),$$

where $\mathbb{F}_{g,\mathbf{u}}(R,Q) := \Big\{f \in \mathbb{F}_g(R) : \ (1-\alpha)f + \alpha[f \star g] \in \mathbb{B}_{\infty,d}(Q)\Big\}$ and $\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$ is the anisotropic Nikolskii class, see Definition 1 below. Here we only mention that the adaptive estimation over the scale $\big\{\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$, $\big(\vec{\beta}, \vec{r}, \vec{L}\big) \in (0,\infty)^d \times [1,\infty]^d \times (0,\infty)^d\big\}$ is usually viewed as the adaptation to anisotropy and inhomogeneity of the function to be estimated. As to the assumption $f \in \mathbb{F}_{g,\infty}(R,Q)$ it simply means that the common density of observations $\mathfrak{p}$ is uniformly bounded by Q. In particular, this is always the case if $\alpha = 1$ and $\|g\|_\infty < \infty$.

Additionally, we will study the adaptive estimation over the collection

$$\Sigma_\vartheta = \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big) \cap \mathbb{F}_g(R) \cap \mathbb{B}_{\infty,d}(Q),\ \vartheta = \big(\vec{\beta}, \vec{r}, \vec{L}, R, Q\big).$$

We will show that the boundedness of the underlying function allows to improve considerably the accuracy of estimation.

Historical notes. The minimax adaptive estimation is a very active area of mathematical statistics, and the interested reader can find a very detailed overview as well as several open problems in adaptive estimation in Lepski (2015). Below we will discuss only the articles whose results are relevant to our consideration, i.e. the density setting under $\mathbb{L}_p$-loss, from a minimax perspective. As already said, the convolution structure density model includes itself the density estimation under direct and indirect observations.

Direct case, $\alpha = 0$. There is a vast literature dealing with minimax and minimax adaptive density estimation, see for example, Efroimovich (1986), Hasminskii and Ibragimov (1990), Golubev (1992), Donoho et al. (1996), Devroye and Lugosi (1997), Rigollet (2006), Rigollet and Tsybakov (2007), Samarov and Tsybakov (2007), Birgé (2008), Giné and Nickl (2009), Akakpo (2012), Gach et al. (2013), Lepski (2013), among many others. Special attention was paid to the estimation of densities with unbounded support, see Juditsky and Lambert–Lacroix (2004), Reynaud–Bouret et al. (2011). The most developed results can be found in Goldenshluger and Lepski (2011), Goldenshluger and Lepski (2014) and in Section 4 we will compare in detail our results with those obtained in these papers.

Intermediate case, $\alpha \in (0,1)$. To the best of our knowledge, adaptive estimation in the case of partially contaminated observations has not been studied yet. We were able to find only two papers dealing with minimax estimation. The first one is Hesse (1995) (where the discussed model was introduced in dimension 1) in which the author evaluated the $\mathbb{L}_\infty$-risk of the proposed estimator over a functional class formally corresponding to the Nikol'skii class $\mathbb{N}_{\infty,1}(2,1)$. In Yuana and Chenb (2002) the latter result was developed to the multidimensional setting, i.e. to the minimax estimation on $\mathbb{N}_{\infty,d}\big(\vec{2},1\big)$. The most intriguing fact is that the accuracy of estimation in partially contaminated noise is the same as in the direct observation scheme. However none of these articles studied the optimality of the proposed estimators. We will come back to the aforementioned papers in Section 1.3 in order to compare the assumptions imposed on the noise density g.

Deconvolution case, $\alpha = 1$. First let us remark that the behavior of the Fourier transform of the function g plays an important role in all the works dealing with deconvolution. Indeed ill-posed problems correspond to Fourier transforms decaying towards zero. Our results will be established for "moderately" ill posed problems, so we detail only results in papers studying that type of operators. This assumption means that there exist $\vec{\mu} = (\mu_1, \dots, \mu_d) \in (0,\infty)^d$ and $\Upsilon_1 > 0, \Upsilon_2 > 0$ such that the Fourier transform $\check{g}$ of g satisfies:

$$(1.4)\quad \Upsilon_1 \prod_{j=1}^{d} (1+t_j^2)^{-\frac{\mu_j}{2}} \le \big|\check{g}(t)\big| \le \Upsilon_2 \prod_{j=1}^{d} (1+t_j^2)^{-\frac{\mu_j}{2}},\ \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Some minimax and minimax adaptive results in dimension 1 over different classes of smooth functions can be found in particular in Stefanski and Carroll (1990), Fan (1991), Fan (1993), Pensky and Vidakovic (1999), Fan and Koo (2002), Comte and al. (2006), Butucea and Tsybakov (2008), Hall and Meister (2007), Meister (2009), Lounici and Nickl (2011), Kerkyacharian et al. (2011).

There are very few results in the multidimensional setting. It seems that Masry (1993) was the first paper where the deconvolution problem was studied for multivariate densities. It is worth noting that Masry (1993) considered more general weakly dependent observations and this paper formally does not deal with the minimax setting. However the results obtained in this paper could be formally compared with the estimation under $\mathbb{L}_\infty$-loss over the isotropic Hölder class of regularity 2, i.e. $\mathbb{N}_{\infty,d}\big(\vec{2}, 1\big)$ which is exactly the same setting as in Yuana and Chenb (2002) in the case of partially contaminated observations. Let us also remark that there is no lower bound result in Masry (1993). The most general results in the deconvolution model were obtained in Comte and Lacour (2013) and Rebelles (2016) and in Section 4 we will compare in detail our results with those obtained in these papers.

1.3. *Assumption on the function g.* Later on for any $U \in \mathbb{L}_1\big(\mathbb{R}^d\big)$, let $\check{U}$ denote its Fourier transform. All our results will be established under the following condition.

Assumption 1. *(1) if $\alpha \neq 1$ then there exists $\varepsilon > 0$ such that*

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

(2) if $\alpha = 1$ then there exists $\vec{\mu} = (\mu_1, \dots, \mu_d) \in (0,\infty)^d$ and $\Upsilon_0 > 0$ s.t.

$$|\check{g}(t)| \ge \Upsilon_0 \textstyle\prod_{j=1}^{d} (1+t_j^2)^{-\frac{\mu_j}{2}}, \quad \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Remind that the following assumption is well-known in the literature:

$$\Upsilon_0 \textstyle\prod_{j=1}^{d} (1+t_j^2)^{-\frac{\mu_j}{2}} \le |\check{g}(t)| \le \Upsilon \prod_{j=1}^{d} (1+t_j^2)^{-\frac{\mu_j}{2}}, \quad \forall t \in \mathbb{R}^d.$$

It is referred to as a *moderately ill-posed* statistical problem. In particular, the assumption is satisfied for the centered multivariate Laplace law.

Note that Assumption 1 (1) is very weak and it is verified for many distributions, including centered multivariate Laplace and Gaussian ones. Note also that this assumption always holds with $\varepsilon = 1-2\alpha$ if $\alpha < 1/2$. Additionally, it holds with $\varepsilon = 1-\alpha$ if $\check{g}$ is a real positive function. The latter is true, in particular, for any probability law obtained by an even number of convolutions of a symmetric distribution with itself.

At last, our Assumption 1 is weaker than the conditions imposed in Hesse (1995) and Yuana and Chenb (2002). In these papers $\check{g} \in \mathbb{C}^{(2)}\big(\mathbb{R}^d\big)$, $\check{g}(t) \neq 0$ for any $t \in \mathbb{R}^d$ and $\big|1-\alpha+\alpha\check{g}(t)\big| \geq 1-\alpha$ for any $t \in \mathbb{R}^d$.

2. Pointwise selection rule and $\mathbb{L}_p$-norm oracle inequality. To present our results in an unified way, let us define $\vec{\boldsymbol{\mu}}(\alpha) = \vec{\mu}$, $\alpha = 1$, $\vec{\boldsymbol{\mu}}(\alpha) = (0,\ldots,0)$, $\alpha \in [0,1)$. Let $K : \mathbb{R}^d \to \mathbb{R}$ be a continuous function belonging to $\mathbb{L}_1\big(\mathbb{R}^d\big)$, $\int_{\mathbb{R}} K = 1$, and such that its Fourier transform $\check{K}$ satisfies the following condition.

Assumption 2. *There exist* $\mathbf{k}_1 > 0$ *and* $\mathbf{k}_2 > 0$ *such that*

$$\int_{\mathbb{R}^d} \big|\check{K}(t)\big| \prod_{j=1}^{d} (1+t_j^2)^{\frac{\boldsymbol{\mu}_j(\alpha)}{2}} \mathrm{d}t \leq \mathbf{k}_1, \; \int_{\mathbb{R}^d} \big|\check{K}(t)\big|^2 \prod_{j=1}^{d} (1+t_j^2)^{\boldsymbol{\mu}_j(\alpha)} \mathrm{d}t \leq \mathbf{k}_2^2.$$

Set $\mathcal{H} = \big\{e^k,\ k \in \mathbb{Z}\big\}$ and let $\mathcal{H}^d = \big\{\vec{h} = (h_1,\ldots,h_d) :\ h_j \in \mathcal{H}, j = 1,\ldots,d\big\}$. Define for any $\vec{h} = (h_1,\ldots,h_d) \in \mathcal{H}^d$

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

Later on for any $u,v \in \mathbb{R}^d$ the operations and relations u/v, uv, $u \vee v$,$u \wedge v$, $u \geq v$, $au, a \in \mathbb{R}$, are understood in coordinate-wise sense. In particular $u \geq v$ means that $u_j \geq v_j$ for any $j = 1,\ldots,d$.

2.1. *Pointwise selection rule from the family of kernel estimators.* For any $\vec{h} \in (0,\infty)^d$ let $M\big(\cdot,\vec{h}\big)$ satisfy the operator equation

$$K_{\vec{h}}(y) = (1-\alpha)M\big(y,\vec{h}\big) + \alpha \int_{\mathbb{R}^d} g(t-y) M\big(t,\vec{h}\big)\mathrm{d}t, \quad y \in \mathbb{R}^d. \tag{2.1}$$

Note that although the explicit expression of $M\big(\cdot,\vec{h}\big)$ is not available its Fourier transform can be easily deduced from (2.1), see Section 5.1.2.

For any $\vec{\mathrm{h}} \in \mathcal{H}^d$ introduce $\widehat{f}_{\vec{\mathrm{h}}}(x) = n^{-1}\sum_{i=1}^n M\big(Z_i - x, \vec{\mathrm{h}}\big), x \in \mathbb{R}^d$. Our first goal is to propose for any given $x \in \mathbb{R}^d$ a data-driven selection rule from the family of estimators $\mathcal{F}\big(\mathcal{H}^d\big) = \big\{\widehat{f}_{\vec{\mathrm{h}}}(x),\ \vec{\mathrm{h}} \in \mathcal{H}^d\big\}$. Define for any $\vec{\mathrm{h}} \in \mathcal{H}^d$

$\widehat{U}_n\big(x,\vec{\mathrm{h}}\big) = \sqrt{2n^{-1}\lambda_n\big(\vec{\mathrm{h}}\big)\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big)} + \frac{4}{3}n^{-1}M_\infty\lambda_n\big(\vec{\mathrm{h}}\big)\prod_{j=1}^d \mathrm{h}_j^{-1}(\mathrm{h}_j\wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}$,

where we have put $\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big) = \frac{1}{n}\sum_{i=1}^n M^2\big(Z_i - x,\vec{\mathrm{h}}\big)$ and

$$\lambda_n\big(\vec{\mathrm{h}}\big) = 4\ln(M_\infty) + 6\ln(n) + (8p+26)\sum_{j=1}^d \big[1+\boldsymbol{\mu}_j(\alpha)\big]\big|\ln(\mathrm{h}_j)\big|;$$

$$M_\infty = \big[(2\pi)^{-d}\big\{\varepsilon^{-1}\big\|\check{K}\big\|_1 1_{\alpha\neq 1} + \Upsilon_0^{-1}\mathbf{k}_1 1_{\alpha=1}\big\}\big]\vee 1.$$

Pointwise selection rule. Let $\mathbb{H}$ be an arbitrary subset of $\mathcal{H}^d$. For any $\vec{h}\in\mathbb{H}$ and $x\in\mathbb{R}^d$ introduce

$$\widehat{\mathcal{R}}_{\vec{h}}(x) = \sup_{\vec{\eta}\in\mathbb{H}}\Big[\big|\widehat{f}_{\vec{h}\vee\vec{\eta}}(x) - \widehat{f}_{\vec{\eta}}(x)\big| - 4\widehat{U}_n\big(x,\vec{h}\vee\vec{\eta}\big) - 4\widehat{U}_n\big(x,\vec{\eta}\big)\Big]_+;$$

where $\widehat{U}_n^*\big(x,\vec{h}\big) = \sup_{\vec{\eta}\in\mathbb{H}:\vec{\eta}\geq\vec{h}}\widehat{U}_n\big(x,\vec{\eta}\big)$, and define

$$\vec{\mathbf{h}}(x) = \arg\inf_{\vec{h}\in\mathbb{H}}\Big[\widehat{\mathcal{R}}_{\vec{h}}(x) + 8\widehat{U}_n^*\big(x,\vec{h}\big)\Big]. \tag{2.2}$$

Our final estimator is $\widehat{f}_{\vec{\mathbf{h}}(x)}(x),\ x\in\mathbb{R}^d$ and we will call (2.2) the *pointwise selection rule*. Note that the estimator $\widehat{f}_{\vec{\mathbf{h}}(\cdot)}(\cdot)$ does not necessarily belong to the collection $\big\{\widehat{f}_{\vec{\mathrm{h}}}(\cdot),\ \vec{\mathrm{h}}\in\mathcal{H}^d\big\}$ since the multi-bandwidth $\vec{\mathbf{h}}(\cdot)$ is a d-variate function, which is not necessarily constant on $\mathbb{R}^d$. The latter fact allows to take into account the "local structure" of the function to be estimated. Moreover, $\vec{\mathbf{h}}(\cdot)$ is chosen with respect to the observations, and therefore it is a random vector-function.

2.2. *$\mathbb{L}_p$-norm oracle inequality.* Introduce for any $x\in\mathbb{R}^d$ and $\vec{h}\in\mathcal{H}^d$

$$U_n^*\big(x,\vec{h}\big) = \sup_{\vec{\eta}\in\mathcal{H}^d:\ \vec{\eta}\geq\vec{h}} U_n\big(x,\vec{\eta}\big),\quad S_{\vec{h}}(x,f) = \int_{\mathbb{R}^d} K_{\vec{h}}(t-x)f(t)\nu_d(\mathrm{d}t);$$

where we have put

$U_n\big(x,\vec{\eta}\big) = \sqrt{2n^{-1}\lambda_n\big(\vec{\eta}\big)\sigma^2\big(x,\vec{\eta}\big)} + \frac{4}{3}n^{-1}M_\infty\lambda_n\big(\vec{\eta}\big)\prod_{j=1}^d \eta_j^{-1}(\eta_j\wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}$,

and $\sigma^2\big(x,\vec{\eta}\big) = \int_{\mathbb{R}^d} M^2\big(t-x,\vec{\eta}\big)\mathfrak{p}(t)\nu_d(\mathrm{d}t)$.

For any $\mathbb{H}\subseteq\mathcal{H}^d$, $\vec{h}\in\mathbb{H}$ and $x\in\mathbb{R}^d$ introduce also

$$B_{\vec{h}}^*(x,f) = \sup_{\vec{\eta}\in\mathbb{H}}\big|S_{\vec{h}\vee\vec{\eta}}(x,f) - S_{\vec{\eta}}(x,f)\big|,\quad B_{\vec{h}}(x,f) = \big|S_{\vec{h}}(x,f) - f(x)\big|. \tag{2.3}$$

THEOREM 1. *Let Assumptions 1 and 2 be fulfilled. Then for any $\mathbb{H}\subseteq\mathcal{H}^d$, $n\geq 3$, $p\in[1,\infty)$ and any $f\in\mathbb{F}_g(R)$*

$$\mathcal{R}_n^{(p)}[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}, f] \leq \Big\|\inf_{\vec{h}\in\mathbb{H}}\Big\{2B_{\vec{h}}^*(\cdot,f) + B_{\vec{h}}(\cdot,f) + 49U_n^*\big(\cdot,\vec{h}\big)\Big\}\Big\|_p + \mathbf{C}_p n^{-\frac{1}{2}}.$$

The explicit expression for the constant $\mathbf{C}_p$ (independent of f, n and $\mathbb{H}$) can be found in the proof of the theorem. Later on we will pay attention to a special choice for the collection of multi-bandwidths, namely

$$\mathcal{H}^d_{\text{isotr}} := \big\{\vec{h} \in \mathcal{H}^d : \ \vec{h} = (h, \dots, h), \ \ h \in \mathcal{H}\big\}.$$

More precisely, in Part II, the selection from the corresponding family of kernel estimators will be used for the adaptive estimation over the collection of isotropic Nikolskii classes. Note also that if $\mathbb{H} = \mathcal{H}^d_{\text{isotr}}$ then obviously $B^*_{\vec{h}}(\cdot, f) \leq 2\sup_{\vec{\eta}\in\mathcal{H}^d_{\text{isotr}}:\, \eta\leq h} B_{\vec{\eta}}(\cdot, f)$ for any $\vec{h} = (h, \dots, h) \in \mathcal{H}^d_{\text{isotr}}$ and we come to the following corollary of Theorem 1.

Corollary 1. *Let Assumptions 1 and 2 be fulfilled. Then for any $n \geq 3$, $p \in [1, \infty)$ and any $f \in \mathbb{F}_g(R)$*

$$\mathcal{R}^{(p)}_n[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}, f] \leq \bigg\| \inf_{\vec{h}\in\mathcal{H}^d_{isotr}} \bigg\{ 5 \sup_{\vec{\eta}\in\mathcal{H}^d_{isotr}\,:\eta\leq h} B_{\vec{\eta}}(\cdot, f) + 49U^*_n\big(\cdot, \vec{h}\big)\bigg\}\bigg\|_p + \mathbf{C}_p n^{-\frac{1}{2}}.$$

The oracle inequality proved in Theorem 1 is particularly useful since it does not require any assumption on the underlying function f (except for the restrictions ensuring the existence of the model and of the risk). However, the quantity appearing in the right hand side of this inequality, namely

$$\Big\| \inf_{\vec{h}\in\mathbb{H}} \Big\{ 2B^*_{\vec{h}}(\cdot, f) + B_{\vec{h}}(\cdot, f) + 49U^*_n\big(\cdot, \vec{h}\big)\Big\}\Big\|_p$$

is not easy to analyze. In particular, in order to use the result of Theorem 1 for adaptive estimation, one has to be able to compute

$$\sup_{f\in\mathbb{F}} \Big\| \inf_{\vec{h}\in\mathbb{H}} \Big\{ 2B^*_{\vec{h}}(\cdot, f) + B_{\vec{h}}(\cdot, f) + 49U^*_n\big(\cdot, \vec{h}\big)\Big\}\Big\|_p$$

for a given class $\mathbb{F} \subset \mathbb{L}_p\big(\mathbb{R}^d\big) \cap \mathbb{F}_g(R)$ with either $\mathbb{H} = \mathcal{H}^d$ or $\mathbb{H} = \mathcal{H}^d_{\text{isotr}}$. It turns out that under some nonrestrictive assumptions imposed on f, the obtained bounds can be considerably simplified. Moreover, new inequalities will allow us to better understand the way for proving adaptive results.

3. Abstract upper bound theorem.

Define $\forall \mathbf{q}, \mathbf{u} \in [1, \infty], D > 0$,

$$\mathbb{F}_{g,\mathbf{u}}(R, D) := \Big\{ f \in \mathbb{F}_g(R) : \ (1-\alpha)f + \alpha[f \star g] \in \mathbb{B}^{(\infty)}_{\mathbf{u},d}(D)\Big\},$$

where $\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D)$ is the ball of radius D in the weak-type space $\mathbb{L}_{\mathbf{u},\infty}\big(\mathbb{R}^d\big)$, i.e. $\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D) = \big\{\lambda : \mathbb{R}^d \to \mathbb{R} : \ \|\lambda\|_{\mathbf{u},\infty} < D\big\}$, where

$$\|\lambda\|_{\mathbf{u},\infty} = \inf\big\{C : \ \nu_d\big(x : |\lambda(x)| > \mathfrak{z}\big) \leq C^{\mathbf{u}}\mathfrak{z}^{-\mathbf{u}}, \ \forall \mathfrak{z} > 0\big\}.$$

As usual $\mathbb{B}^{(\infty)}_{\infty,d}(D) = \mathbb{B}_{\infty,d}(D)$ and obviously $\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D) \supset \mathbb{B}_{\mathbf{u},d}(D)$.

It is worth noting that the assumption $f\in\mathbb{F}_{g,\mathbf{u}}(R,D)$ simply means that the common density of the observations $\mathfrak{p}$ belongs to $\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D)$.

Our objective is to bound from above $\sup_{f\in\mathbb{F}}\mathcal{R}^{(p)}_n[\widehat{f}_{\vec{\mathbf{h}}(\cdot)},f]$ for any $\mathbb{F}\subset\mathbb{F}_{g,\mathbf{u}}(R,D)\cap\mathbb{B}_{\mathbf{q},d}(D)$. Since $\mathbb{F}$ is an arbitrary set this bound can be applied to the adaptation over different scales of functional classes. In particular, the results below form the basis for our consideration in Section 4.

Remark 1. *Note that $\mathbb{F}_{g,\mathbf{1}}(R,D)=\mathbb{F}_g(R)$ for any $D\geq 1$. Moreover, it is easily seen that $\mathbb{F}_{g,\infty}\big(R,R\|g\|_\infty\big)=\mathbb{F}_g(R)$ if $\alpha=1$ and $\|g\|_\infty<\infty$. At last $\mathbb{F}_{g,\infty}(R,Q\|g\|_1)\supset\mathbb{F}_g(R)\cap\mathbb{B}_{\infty,d}(Q)$ for any $\alpha\in[0,1]$ and $Q>0$.*

Introduce for any $\vec{h}\in\mathcal{H}^d$

$$F_n\big(\vec{h}\big)=\frac{\sqrt{\ln n+\sum_{j=1}^d|\ln h_j|}}{\sqrt{n}\prod_{j=1}^d h_j^{\frac{1}{2}}(h_j\wedge 1)^{\boldsymbol{\mu}_j(\alpha)}},\quad G_n\big(\vec{h}\big)=\frac{\ln n+\sum_{j=1}^d|\ln h_j|}{n\prod_{j=1}^d h_j(h_j\wedge 1)^{\boldsymbol{\mu}_j(\alpha)}}.$$

Furthermore let $\mathbb{H}$ be either $\mathcal{H}^d$ or $\mathcal{H}^d_{\text{isotr}}$ and for any $v,z>0$ define

$$\mathfrak{H}(v)=\big\{\vec{h}\in\mathbb{H}:\ G_n\big(\vec{h}\big)\leq av\big\},\ \ \mathfrak{H}(v,z)=\big\{\vec{h}\in\mathfrak{H}(v):\ F_n\big(\vec{h}\big)\leq avz^{-\frac{1}{2}}\big\}.$$

Here $a>0$ is a numerical constant whose explicit expression is given in the beginning of Section 5.2. Put also for any $v>0$, $l_{\mathbb{H}}(v)=v^{p-1}(1+|\ln(v)|)^{t(\mathbb{H})}$, where $t(\mathbb{H})=d-1$ if $\mathbb{H}=\mathcal{H}^d$ and $t(\mathbb{H})=0$ if $\mathbb{H}=\mathcal{H}^d_{\text{isotr}}$.

Remark 2. *Note that $\mathfrak{H}(v)\neq\emptyset$ and $\mathfrak{H}(v,z)\neq\emptyset$ whatever the values of $v>0$ and $z\geq 2$. Indeed, for any $v>0$, $z>2$ one can find $b>1$ such that*

$$\big(\ln n+d\ln b\big)(nb^d)^{-1}\leq\big[a^2v^2z^{-1}\big]\wedge av.$$

The latter means that $\vec{b}=(b,\dots,b)\in\mathfrak{H}(v,z)\cap\mathfrak{H}(v)$.

All the results in this section will be proved under an additional condition imposed on the kernel K.

Assumption 3. *Let $\mathcal{K}:\mathbb{R}\to\mathbb{R}$ be a compactly supported, bounded function and $\int\mathcal{K}=1$. Then*

$$K(x)=\prod_{j=1}^d\mathcal{K}(x_j),\ \forall x\in\mathbb{R}^d.$$

Without loss of generality we will assume that $\|\mathcal{K}\|_\infty\geq 1$ and supp$(\mathcal{K})\subset[-c_{\mathcal{K}},c_{\mathcal{K}}]$ with $c_{\mathcal{K}}\geq 1$.

Introduce the following notations. Set for any $h\in\mathcal{H}$ and $j=1,\dots,d$

$$b^*_{h,f,j}(\cdot)=\bigg|\int_{\mathbb{R}}\mathcal{K}(u)f\big(\,\cdot+uh\mathbf{e}_j\big)\nu_1(\mathrm{d}u)-f(\cdot)\bigg|,\ b_{h,f,j}(\cdot)=\sup_{\eta\in\mathcal{H}:\eta\leq h}b^*_{\eta,f,j}(\cdot),$$

where $(\mathbf{e}_1,\dots,\mathbf{e}_d)$ denotes the canonical basis of $\mathbb{R}^d$. Introduce $\forall s\in[1,\infty]$

$$\mathbf{B}^*_{j,s,\mathbb{F}}(\mathbf{h})=\sup_{f\in\mathbb{F}}\sum_{h\in\mathcal{H}:\,h\le\mathbf{h}}\big\|b^*_{h,f,j}\big\|_s,\quad \mathbf{B}_{j,s,\mathbb{F}}(\mathbf{h})=\sup_{f\in\mathbb{F}}\big\|b_{\mathbf{h},f,j}\big\|_s,\ j=1,\dots,d.$$

For any $v>0$ and $j=1,\dots,d$, set $\mathbf{V}_j(v)=\big\{\mathbf{h}\in\mathcal{H}:\mathbf{B}_{j,\infty,\mathbb{F}}(\mathbf{h})\le\mathbf{c}v\big\}$ and

$$J\big(\vec{h},v\big)=\big\{j\in\{1,\dots,d\}:h_j\in\mathbf{V}_j(v)\big\},\quad \vec{h}\in\mathcal{H}^d.$$

where $\mathbf{c}=(20d)^{-1}\big[\max(2c_{\mathcal{K}}\|\mathcal{K}\|_\infty,\|\mathcal{K}\|_1)\big]^{-d}$. As usual the complement of $J\big(\vec{h},v\big)$ will be denoted by $\bar{J}\big(\vec{h},v\big)$. Furthermore, the summation over the empty set is supposed to be zero.

For any $\vec{s}=(s_1,\dots,s_d)\in[1,\infty)^d$, $\mathbf{u}\ge1$ and $v>0$ introduce

$$\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F},\mathbf{u})=\inf_{z\ge2}\inf_{\vec{h}\in\mathfrak{H}(v,z)}\Bigg[\sum_{j\in\bar{J}(\vec{h},v)}v^{-s_j}\big[\mathbf{B}_{j,s_j,\mathbb{F}}\big(h_j\big)\big]^{s_j}+z^{-\mathbf{u}}\Bigg];\tag{3.1}$$

$$\boldsymbol{\Lambda}_{\vec{s}}\big(v,\mathbb{F}\big)=\inf_{\vec{h}\in\mathfrak{H}(v)}\Bigg[\sum_{j\in\bar{J}(\vec{h},v)}v^{-s_j}\big[\mathbf{B}_{j,s_j,\mathbb{F}}\big(h_j\big)\big]^{s_j}+v^{-2}F_n^2\big(\vec{h}\big)\Bigg].\tag{3.2}$$

Furthermore, we assume that any quantity depending on $\overline{\boldsymbol{v}}$ is equal to zero when $\overline{\boldsymbol{v}}=\infty$.

Theorem 2. *Let assumptions of Theorem 1 be fulfilled and suppose additionally that K satisfies Assumption 3. Then for any $n\ge3$, $p>1,\mathbf{q}>1,R>1,D>0,0<\underline{\boldsymbol{v}}\le\overline{\boldsymbol{v}}\le\infty,\mathbf{u}\in(p/2,\infty],\mathbf{u}\ge\mathbf{q}$, $\vec{s}\in(1,\infty)^d$, $\vec{q}\in[p,\infty)^d$ and any $\mathbb{F}\subset\mathbb{B}_{\mathbf{q},d}(D)\cap\mathbb{F}_{g,\mathbf{u}}(R,D)$*

$$\begin{aligned}\sup_{f\in\mathbb{F}}\mathcal{R}_n^{(p)}\big[\widehat{f}_{\vec{\mathbf{h}}(\cdot)},f\big]\;&\le\;C^{(2)}\Bigg[l_{\mathbb{H}}(\underline{\boldsymbol{v}})+\int_{\underline{\boldsymbol{v}}}^{\overline{\boldsymbol{v}}}v^{p-1}\big[\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F},\mathbf{u})\wedge\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F})\big]\mathrm{d}v\\&\qquad+\overline{\boldsymbol{v}}^p\boldsymbol{\Lambda}_{\vec{q}}(\overline{\boldsymbol{v}},\mathbb{F},\mathbf{u})\Bigg]^{\frac{1}{p}}+\mathbf{C}_pn^{-\frac12}.\end{aligned}$$

If additionally $\mathbf{q}\in(p,\infty)$ one has also

$$\begin{aligned}\sup_{f\in\mathbb{F}}\mathcal{R}_n^{(p)}\big[\widehat{f}_{\vec{\mathbf{h}}(\cdot)},f\big]\;&\le\;C^{(2)}\Bigg[l_{\mathbb{H}}(\underline{\boldsymbol{v}})+\int_{\underline{\boldsymbol{v}}}^{\overline{\boldsymbol{v}}}v^{p-1}\big[\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F},\mathbf{u})\wedge\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F})\big]\mathrm{d}v\\&\qquad+\overline{\boldsymbol{v}}^{p-\mathbf{q}}\Bigg]^{\frac{1}{p}}+\mathbf{C}_pn^{-\frac12}.\end{aligned}$$

Moreover, if $\mathbf{q}=\infty$ one has

$$\begin{aligned}\sup_{f\in\mathbb{F}}\mathcal{R}_n^{(p)}\big[\widehat{f}_{\vec{\mathbf{h}}(\cdot)},f\big]\;&\le\;C^{(2)}\Bigg[l_{\mathbb{H}}(\underline{\boldsymbol{v}})+\int_{\underline{\boldsymbol{v}}}^{\overline{\boldsymbol{v}}}v^{p-1}\big[\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F},\mathbf{u})\wedge\boldsymbol{\Lambda}_{\vec{s}}(v,\mathbb{F})\big]\mathrm{d}v\\&\qquad+\boldsymbol{\Lambda}_{\vec{s}}(\overline{\boldsymbol{v}},\mathbb{F},\mathbf{u})\Bigg]^{\frac{1}{p}}+\mathbf{C}_pn^{-\frac12}.\end{aligned}$$

Finally, if $\mathbb{H} = \mathcal{H}^d_{isotr}$ *all the assertions above remain true for any* $\vec{s} \in [1,\infty)^d$ *if one replaces in (3.1)–(3.2)* $\mathbf{B}_{j,s_j,\mathbb{F}}(\cdot)$ *by* $\mathbf{B}^*_{j,s_j,\mathbb{F}}(\cdot)$*.*

$\mathbf{1^0}$. It is important to emphasize that $C^{(2)}$ depends only on $\vec{s},\vec{q},g,\mathcal{K},d$, $R,D,\mathbf{u}$ and $\mathbf{q}$. Note also that the assertions of the theorem remain true if we minimize right hand sides of obtained inequalities w.r.t $\vec{s},\vec{q}$ since their left hand sides are independent of $\vec{s}$ and $\vec{q}$. In this context it is important to realize that $C^{(2)} = C^{(2)}(\vec{s},\cdots)$ is bounded for any $\vec{s} \in (1,\infty)^d$ but $C^{(2)}(\vec{s},\cdots) = \infty$ if there exists $j = 1,\dots,d$ such that $s_j = 1$. Contrary to that $C^{(2)}(\vec{s},\cdots) < \infty$ for any $\vec{s} \in [1,\infty)^d$ if $\mathbb{H} = \mathcal{H}^d_{\text{isotr}}$ and it explains in particular the fourth assertion of the theorem.

$\mathbf{2^0}$. It is worth noting that all bounds presented in the theorem are heavily based on the result given in (5.39) of Section 5.2. This is $\mathbb{L}_p$-norm oracle inequality on $\mathbb{F}_{g,\mathbf{u}}(R,D) \cap \mathbb{B}_{\mathbf{q},d}(D)$. In particular, it does not require Assumption 3 and it is established for any compactly supported K satisfying Assumption 2.

$\mathbf{3^0}$. Note also that $D,R,\mathbf{u},\mathbf{q}$ are not involved in the construction of our pointwise selection rule. That means that one and the same estimator can be actually applied on any $\mathbb{F} \subset \bigcup_{R,D,\mathbf{u},\mathbf{q}} \mathbb{B}_{\mathbf{q},d}(D) \cap \mathbb{F}_{g,\mathbf{u}}(R,D)$. Moreover, the assertion of the theorem has a non-asymptotical nature; we do not suppose that the number of observations n is large.

$\mathbf{4^0}$. As we see, the application of our results to some functional class is mainly reduced to the computation of the functions $\mathbf{B}^*_{j,s,\mathbb{F}}(\cdot)$ $j = 1,\dots,d$, for some properly chosen s. Note however that this task is not necessary for many functional classes at least for the classes defined by the help of kernel approximation. Indeed, a typical description of $\mathbb{F}$ can be summarized as follows. Let $\lambda_j : \mathbb{R}_+ \to \mathbb{R}_+$, be such that $\lambda_j(0) = 0, \lambda_j \uparrow$ for any $j = 1,\dots,d$. Then, the functional class is defined as a collection of functions satisfying

$$\big\|b_{\mathbf{h},f,j}\big\|_{r_j} \le \lambda_j(\mathbf{h}), \quad \forall \mathbf{h} \in \mathcal{H}, \tag{3.3}$$

for some $\vec{r} \in [1,\infty]$. It yields obviously $\mathbf{B}_{j,r_j,\mathbb{F}}(\cdot) \le \lambda_j(\cdot)$ for any $j = 1,\dots,d$, and the result of Theorem 2 remains valid if we replace formally $\mathbf{B}_{j,r_j,\mathbb{F}}(\cdot)$ by $\lambda_j(\cdot)$ in all the expressions appearing in this theorem.

In Section 6 (Proposition 2) we show that for some particular kernel K^*, the anisotropic Nikol'skii class $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ is included into the class defined by (3.3) with $\lambda_j(\mathbf{h}) = L_j \mathbf{h}^{\beta_j}$, whatever the values of $\vec{\beta},\vec{L}$ and $\vec{r}$.

4. Adaptive estimation over the scale of anisotropic classes. Let $(\mathbf{e}_1,\dots,\mathbf{e}_d)$ denote the canonical basis of $\mathbb{R}^d$. For some function $G : \mathbb{R}^d \to \mathbb{R}^1$ 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 $\Delta_{u,j}G(x)=G(x+u\mathbf{e}_j)-G(x),\ j=1,\dots,d$.

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

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

Definition 1. *For given vectors $\vec{\beta}=(\beta_1,\dots,\beta_d)\in(0,\infty)^d$, $\vec{r}=(r_1,\dots,r_d)\in[1,\infty]^d$, and $\vec{L}=(L_1,\dots,L_d)\in(0,\infty)^d$ we say that a function $G:\mathbb{R}^d\to\mathbb{R}^1$ belongs to the anisotropic Nikolskii class $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ if $\|G\|_{r_j}\le L_j$ for all $j=1,\dots,d$ and there exists natural number $k_j>\beta_j$ such that*

$$\big\|\Delta_{u,j}^{k_j}G\big\|_{r_j}\le L_j|u|^{\beta_j},\quad \forall u\in\mathbb{R},\quad \forall j=1,\dots,d.$$

If $\beta_j=\boldsymbol{\beta}\in(0,\infty), r_j=\mathbf{r}\in[1,\infty]$ and $L_j=\mathbf{L}\in(0,\infty)$ for any $j=1,\dots,d$ the corresponding Nikolskii class, denoted furthermore $\mathbb{N}_{\mathbf{r},d}(\boldsymbol{\beta},\mathbf{L})$, is called isotropic. The following quantities related to the parameters of the Nikol'skii class will be of the great importance:

$$\tfrac{1}{\beta(\alpha)}=\textstyle\sum_{j=1}^{d}\frac{2\boldsymbol{\mu}_j(\alpha)+1}{\beta_j},\quad \tfrac{1}{\omega(\alpha)}=\sum_{j=1}^{d}\frac{2\boldsymbol{\mu}_j(\alpha)+1}{\beta_j r_j},\quad L(\alpha)=\prod_{j=1}^{d}L_j^{\frac{2\boldsymbol{\mu}_j(\alpha)+1}{\beta_j}}.$$

Define also for any $1\le s\le\infty$ and $\alpha\in[0,1]$

$$\varkappa_\alpha(s)=\omega(\alpha)(2+1/\beta(\alpha))-s,\quad \tau(s)=1-1/\omega(0)+1/(s\beta(0)).$$

4.1. *Construction of kernel K.* We keep Assumption 2 and enforced Assumption 3 by Assumption 4 below related to the following specific construction of kernel K used in the definition of the estimator's family $\big\{\widehat{f}_{\vec{\mathrm{h}}}(\cdot),\ \vec{\mathrm{h}}\in\mathcal{H}^d\big\}$ [see, e.g., Kerkyacharian et al. (2001) or Goldenshluger and Lepski (2014)]. Let ℓ be an integer number, $\mathcal{K}:\mathbb{R}^1\to\mathbb{R}^1$ be a compactly supported continuous function satisfying $\int_{\mathbb{R}^1}\mathcal{K}(y)\mathrm{d}y=1$, and $\mathcal{K}\in\mathbb{C}(\mathbb{R}^1)$. Put

$$\mathcal{K}_\ell(y)=\textstyle\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+1}\frac{1}{i}\mathcal{K}\Big(\frac{y}{i}\Big),$$

and add the following structural condition to Assumption 2.

Assumption 4. $K(x)=\prod_{j=1}^{d}\mathcal{K}_\ell(x_j),\ \forall x\in\mathbb{R}^d$.

4.2. *Main results.* Set $\delta_n=L(\alpha)n^{-1}\ln(n)$, $t(\mathbb{H})=d-1$ if $\mathbb{H}=\mathcal{H}^d$ and $t(\mathbb{H})=0$ if $\mathbb{H}=\mathcal{H}^d_{\mathrm{isotr}}$ and let

$$\mathfrak{b}_n(\mathbb{H})=\begin{cases}[\ln(n)]^{t(\mathbb{H})}, & \varkappa_\alpha(p)>p\omega(\alpha);\\ \ln^{\frac{1}{p}}(n)\vee[\ln(n)]^{t(\mathbb{H})}, & \varkappa_\alpha(p)=p\omega(\alpha);\\ \ln^{\frac{1}{p}}(n), & \varkappa_\alpha(p)=0;\\ 1, & \text{otherwise},\end{cases}$$

4.2.1. *Bounded case.* The first problem we address is the adaptive estimation over the collection of the functional classes $\big\{\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\cap\mathbb{F}_g(R)\cap\mathbb{B}_{\infty,d}(Q)\big\}_{\vec{\beta},\vec{r},\vec{L},R,Q}$. As it was conjectured in Lepski and Willer (2017), the boundedness of the function belonging to $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\cap\mathbb{F}_g(R)$ is a minimal condition allowing to eliminate the inconsistency zone. The results obtained in Theorem 3 together with those from Theorem 2 in Lepski and Willer (2017) confirm this conjecture. Define

$$\rho(\alpha)\;=\;\begin{cases}\frac{1-1/p}{1-1/\omega(\alpha)+1/\beta(\alpha)}, & \varkappa_\alpha(p)>p\omega(\alpha);\\ \frac{\beta(\alpha)}{2\beta(\alpha)+1}, & 0<\varkappa_\alpha(p)\leq p\omega(\alpha);\\ \frac{\tau(p)\omega(\alpha)\beta(0)}{z(\alpha)}, & \varkappa_\alpha(p)\leq 0,\;\;\tau(\infty)>0;\\ \frac{\omega(\alpha)}{p}, & \varkappa_\alpha(p)\leq 0,\;\tau(\infty)\leq 0.\end{cases}\tag{4.1}$$

THEOREM 3. *Let $\alpha\in[0,1]$, $\ell\in\mathbb{N}^*$ and $g\in\mathbb{L}_1\big(\mathbb{R}^d\big)$, satisfying Assumption 1, be fixed. Let K satisfy Assumptions 2 and 4.*

1) Then for any $p\in(1,\infty)$, $Q>0$, $R>0$, $L_0>0$, $\vec{\beta}\in(0,\ell]^d$, $\vec{r}\in(1,\infty]^d$ and $\vec{L}\in[L_0,\infty)^d$ there exists $C<\infty$, independent of $\vec{L}$, such that:

$$\limsup_{n\to\infty}\sup_{f\in\mathbb{N}_{\vec{r},d}\left(\vec{\beta},\vec{L}\right)\cap\mathbb{F}_g(R)\cap\mathbb{B}_{\infty,d}(Q)}\mathfrak{b}_n\big(\mathcal{H}^d\big)^{-1}\delta_n^{-\rho(\alpha)}\mathcal{R}_p^{(n)}\big[\widehat{f}_{\vec{\mathbf{h}},\mathbb{H}};f\big]\leq C,$$

where $\rho(\alpha)$ is defined in (4.1).

2) For any $p\in(1,\infty)$, $Q>0$, $R>0, L_0>0$, $\boldsymbol{\beta}\in(0,\ell]$, $\mathbf{r}\in[1,\infty]$ and $\mathbf{L}\in[L_0,\infty)$ there exists $C<\infty$, independent of $\mathbf{L}$, such that:

$$\limsup_{n\to\infty}\sup_{f\in\mathbb{N}_{\mathbf{r},d}\left(\boldsymbol{\beta},\mathbf{L}\right)\cap\mathbb{F}_g(R)\cap\mathbb{B}_{\infty,d}(Q)}\mathfrak{b}_n\big(\mathcal{H}^d_{isotr}\big)^{-1}\delta_n^{-\rho(\alpha)}\mathcal{R}_p^{(n)}\big[\widehat{f}_{\vec{\mathbf{h}},\mathcal{H}^d_{isotr}};f\big]\leq C.$$

Some remarks are in order. $\mathbf{1^0}$. Our estimation procedure is completely data-driven, i.e. independent of $\vec{\beta},\vec{r},\vec{L},R$, Q, and the assertions of Theorem 3 are completely new if $\alpha\neq 0$. Comparing the results obtained in Theorems 3 and 2 in Lepski and Willer (2017) we can assert that our estimator is optimally-adaptive if $\varkappa_\alpha(p)<0$ and nearly optimally adaptive if $0<\varkappa_\alpha(p)<p\omega(\alpha)$. The construction of an estimation procedure which

would be optimally-adaptive when $\varkappa_\alpha(p) \geq 0$ is an open problem, and we conjecture that the lower bounds for the asymptotics of the minimax risk found in Theorem 2 in Lepski and Willer (2017) are sharp in order. This conjecture in the case $\alpha = 1$ is partially confirmed by the results obtained in Comte and Lacour (2013) and Rebelles (2016). Since both articles deal with the estimation of unbounded functions we will discuss them in the next section.

$\mathbf{2^0}$. We note that the asymptotic of the minimax risk under partially contaminated observations, $\alpha \in (0,1)$, is independent of α and coincides with the asymptotic of the risk in the direct observation model, $\alpha = 0$. For the first time this phenomenon was discovered in Hesse (1995) and Yuana and Chenb (2002). In the very recent paper Lepski (2017), in the particular case $\vec{r} = (p, \dots, p)$, $p \in (1, \infty)$ the optimally adaptive estimator was built. It is easy to check that independently of the value of $\vec{\beta}$ and $\vec{r}$, the corresponding set of parameters belongs to the dense zone. Note however that our estimator is only nearly optimally-adaptive in this zone, but it is applied to a much more general collection of functional classes. It is worth noting that the estimator procedure, used in Lepski (2017), has nothing in common with our pointwise selection rule.

$\mathbf{3^0}$. As to the direct observation scheme, $\alpha = 0$, our results coincide with those obtained recently in Goldenshluger and Lepski (2014), when $p\omega(0) > \varkappa_0(p)$. However, for the tail zone $p\omega(0) \leq \varkappa_0(p)$, our bound is slightly better since the bound obtained in the latter paper contains an additional factor $\ln^{\frac{d}{p}}(n)$. It is interesting to note that although both estimator constructions are based upon local selections from the family of kernel estimators, the selection rules are different.

$\mathbf{4^0}$. Let us finally discuss the results corresponding to the tail zone, $\varkappa_\alpha(p) > p\omega(\alpha)$. First, the lower bound for the minimax risk is given by $[L(\alpha)n^{-1}]^{\rho(\alpha)}$ while the accuracy provided by our estimator is

$$\ln^{\frac{d-1}{p}}(n)[L(\alpha)n^{-1}\ln(n)]^{\rho(\alpha)}.$$

As it was mentioned , the passage from $[L(\alpha)n^{-1}]^{\rho(\alpha)}$ to $[L(\alpha)n^{-1}\ln(n)]^{\rho(\alpha)}$ seems to be an unavoidable payment for the application of a local selection scheme. It is interesting to note that the additional factor $\ln^{\frac{d-1}{p}}(n)$ disappears in the dimension $d = 1$. First, note that if $\alpha = 0$ the one-dimensional setting was considered in Juditsky and Lambert–Lacroix (2004) and Reynaud–Bouret et al. (2011). The setting of Juditsky and Lambert–Lacroix (2004) corresponds to $r = \infty$, while Reynaud–Bouret et al. (2011) deal with the case of $p = 2$ and $\tau(2) > 0$. Both settings rule out the sparse zone. The rates of convergence found in these papers are easily recovered from our results corresponding to the tail and dense zones.

Next, we remark that the aforementioned factor appears only when anisotropic functional classes are considered. Indeed, in view of the second assertion of Theorem 3 our estimator is nearly optimally adaptive on the tail zone in the isotropic case. The natural question arising in this context, is whether the $\ln^{\frac{d-1}{p}}(n)$-factor is an unavoidable payment for anisotropy of the underlying function or not?

$\mathbf{5^0}$. We finish our discussion with the following remark. If $\alpha\neq 1$ the assumption $f\in\mathbb{F}_{g,\infty}(R,Q)$ implies in many cases that f is uniformly bounded and, therefore, Theorem 3 is applicable. In particular it is always the case if the model (1.2) is considered. Indeed $f,g\in\mathfrak{P}\big(\mathbb{R}^d\big)$ in this case, which implies $\|f\|_\infty\leq(1-\alpha)^{-1}\|\mathfrak{p}\|_\infty\leq(1-\alpha)^{-1}Q$. Another case is $\|g\|_\infty<\infty$ and recall that this assumption was used in the proofs of Theorems 1 and 2 in Lepski and Willer (2017). We obviously have that

$$\|f\|_\infty\leq(1-\alpha)^{-1}\big[Q+\alpha R\|g\|_\infty\big].$$

More generally $\|f\|_\infty\leq(1-\alpha)^{-1}(Q+\alpha D)$ if $f\in\mathbb{F}_{g,\infty}(R,Q)$ and $\|f\star g\|_\infty\leq D$. Since the definition of the Nikol'skii class implies that $\|f\|_{r^*}\leq L^*$, where $r^*=\sup_{j=1,\dots,d}r_j$ and $L^*=\sup_{j=1,\dots,d}L_j$, the latter condition can be verified in particular if $\|g\|_q<\infty, 1/q=1-1/r^*$. All saying above explains why we study the estimation of unbounded functions only in the case $\alpha=1$.

4.2.2. *Unbounded case,* $\alpha=1$. The problem we address now is the adaptive estimation over the collection of functional classes $\big\{\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\cap\mathbb{F}_{g,\infty}(R,Q)\big\}_{\vec{\beta},\vec{r},\vec{L},R,Q}$. As we already mentioned, if additionally $\|g\|_\infty<\infty$ then $\mathbb{F}_{g,\infty}(R,Q)=\mathbb{F}_g(R)$ for any $Q\geq R\|g\|_\infty$ and, therefore, in view of Theorem 1 in Lepski and Willer (2017) there is no consistent estimator if either $p=1$ or $\varkappa_\alpha(p)\leq 0,\ \tau(p)\leq 0,\ p^*=p$. For this reason, later on we will only consider the parameters $\vec{\beta},\vec{r}$ belonging to the set $\mathcal{P}_{p,\vec{\mu}}$ defined below.

$$\mathcal{P}_{p,\vec{\mu}}=(0,\infty)^d\times[1,\infty]^d\setminus\Big\{\vec{\beta},\vec{r}:\ \varkappa_\alpha(p)\leq 0,\ \tau(p)\leq 0,\ \max\nolimits_{j=1,\dots,d}r_j\leq p\Big\}.$$

Set $z(\alpha)=\omega(\alpha)(2+1/\beta(\alpha))\beta(0)\tau(\infty)+1$, $p^*=\big[\max_{l=1,\dots,d}r_l\big]\vee p$ and let

$$\varrho(\alpha)=\begin{cases}\frac{1-1/p}{1-1/\omega(\alpha)+1/\beta(\alpha)}, & \varkappa_\alpha(p)>p\omega(\alpha);\\ \frac{\beta(\alpha)}{2\beta(\alpha)+1}, & 0<\varkappa_\alpha(p)\leq p\omega(\alpha);\\ \frac{\tau(p)\omega(\alpha)\beta(0)}{z(\alpha)}, & \varkappa_\alpha(p)\leq 0,\ \ \tau(p^*)>0;\\ \frac{\omega(\alpha)(1-p^*/p)}{\varkappa_\alpha(p^*)}, & \varkappa_\alpha(p)\leq 0,\ \tau(p^*)\leq 0.\end{cases}\tag{4.2}$$

We will assume $0/0 = 0$, which implies in particular $\frac{1-p^*/p}{\varkappa_\alpha(p^*)} = 0$ if $p^* = p$ and $\varkappa_\alpha(p) = 0$. Note also that $\varkappa_\alpha(p^*)/p^* = -1$ if $p^* = \infty$.

THEOREM 4. *Let $\ell \in \mathbb{N}^*$ and $g \in \mathbb{L}_1\big(\mathbb{R}^d\big)$, satisfying Assumption 1 be fixed and let K satisfy Assumptions 2 and 4.*

1) Then for any $p > [\min_{j=1,\dots} \mu_j]^{-1}$, $R, Q > 0$, $0 < L_0 \leq L_\infty < \infty$, $\big(\vec{\beta}, \vec{r}\big) \in \mathcal{P}_{p,\vec{\mu}} \cap \big\{(0,\ell]^d \times (1,\infty]^d\big\}$ and $\vec{L} \in [L_0, L_\infty]^d$ there exists $C < \infty$, independent of $\vec{L}$, such that:

$$\limsup_{n\to\infty} \sup_{f \in \mathbb{N}_{\vec{r},d}\left(\vec{\beta},\vec{L}\right) \cap \mathbb{F}_{g,\infty}(R,Q)} \mathfrak{b}_n\big(\mathcal{H}^d\big)^{-1} \delta_n^{-\varrho(1)} \mathcal{R}_p^{(n)}\big[\widehat{f}_{\vec{\mathbf{h}},\mathcal{H}^d}; f\big] \leq C,$$

where $\varrho(\cdot)$ is defined in (4.2).

2) For any $p > [\min_{j=1,\dots} \mu_j]^{-1}$, $R, Q > 0$, $0 < L_0 \leq L_\infty < \infty$, $(\boldsymbol{\beta}, \mathbf{r}) \in \mathcal{P}_{p,\vec{\mu}} \cap \big\{(0,\ell] \times (1,\infty]\big\}$ and $\mathbf{L} \in [L_0, L_\infty]$ there exists $C < \infty$, independent of $\mathbf{L}$, such that:

$$\limsup_{n\to\infty} \sup_{f \in \mathbb{N}_{\mathbf{r},d}\left(\boldsymbol{\beta},\mathbf{L}\right) \cap \mathbb{F}_{g,\infty}(R,Q)} \mathfrak{b}_n\big(\mathcal{H}^d_{isotr}\big)^{-1} \delta_n^{-\varrho(1)} \mathcal{R}_p^{(n)}\big[\widehat{f}_{\vec{\mathbf{h}},\mathcal{H}^d_{isotr}}; f\big] \leq C.$$

Some remarks are in order.

$\mathbf{1^0}$. Note that $\|g\|_1 < \infty$, $\|g\|_\infty < \infty$ implies that $\|g\|_2 < \infty$ and, therefore the Parseval identity together with Assumption 1 allows us to assert that

$$\|g\|_\infty < \infty \quad \Rightarrow \quad \mu_j > 1/2, \quad \forall j = 1, \dots, d.$$

Hence, the condition $p > [\min_{j=1,\dots} \mu_j]^{-1}$ is automatically checked if $p \geq 2$ and $\|g\|_\infty < \infty$. Also, it is worth noting that considering the adaptation over the collection of isotropic classes, we do not require that the coordinates of $\vec{\mu}$ would be the same. The latter is true for the second assertion of Theorem 3 as well. At last, analyzing the proof of the theorem, we can assert that the second assertion remains true under the slightly weaker assumption $p > d(\mu_1 + \dots + \mu_d)^{-1}$.

$\mathbf{2^0}$. The assertion of Theorem 4 has no analogue in the existing literature except the results obtained in Comte and Lacour (2013) and Rebelles (2016). Comte and Lacour (2013) deals with the particular case $p = 2$, $\vec{r} = (2, \dots, 2)$ while Rebelles (2016) studied the case $\vec{r} = (p, \dots, p)$, $p \in (1, \infty)$. It is easy to check that in both papers whatever the value of $\vec{\beta}$ and $\vec{\mu}$, the corresponding set of parameters belongs to the dense zone. Note also that the estimation procedures used in Comte and Lacour (2013) as well as in Rebelles (2016), if $p \geq 2$, (both based on a global version of the Goldenshluger-Lepski method) are optimally-adaptive. They attain the asymptotic of minimax risks corresponding to the dense zone found in Theorem 1 in Lepski and Willer

(2017), while our method is only nearly optimally adaptive. However, it is well-known that the global selection from the family of standard kernel estimators leads to correct results only if $\vec{r} = (p, \dots, p)$ when the $\mathbb{L}_p$-risk is considered, see, for instance Goldenshluger and Lepski (2011). On the other hand, estimation procedures based on a local selection scheme, which can be applied to the estimation of functions belonging to much more general functional classes, often do not lead to an optimally adaptive method. Fortunately, the loss of accuracy inherent to local procedures is logarithmic w.r.t. the number of observations.

3^0. Together with Theorems 1 and 2 in Lepski and Willer (2017), Theorems 3 and 4 provide the full classification of the asymptotics of the minimax risks over anisotropic/isotropic Nikolskii classes for the class parameters belonging to the sparse zone and, up to some logarithmic factor, belonging to the tail and dense zones as well as the boundaries. We mean that the results of these theorems are valid for any `fixed` $\vec{\beta} \in (0,\infty)^d, \vec{r} \in (1,\infty]^d$ and $\vec{L} \in (0,\infty)^d$. Indeed, for given $\vec{\beta}$ and $\vec{L}$ one can choose $L_0 = \min_{j=1,\dots d} L_j$, $L_\infty = \max_{j=1,\dots d} L_j$ and the number ℓ, used in the construction of kernel $\mathcal{K}_\ell$, as any integer strictly larger than $\max_{j=1,\dots d} \beta_j$.

5. Proofs of Theorems 1–2.

5.1. *Proof of Theorem 1.* The main ingredients of the proof of the theorem are given in Proposition 1. Their proofs are postponed to Section 5.1.2. Introduce for any $\vec{h} \in \mathcal{H}^d$

$$\xi_n(x,\vec{h}) = \tfrac{1}{n}\textstyle\sum_{i=1}^n \big[M\big(Z_i - x, \vec{h}\big) - \mathbb{E}_f M\big(Z_i - x, \vec{h}\big)\big], \quad x \in \mathbb{R}^d.$$

Proposition 1. *Let Assumptions 1 and 2 be fulfilled. Then for any $n \geq 3$ and any $p > 1$*

(i) $\int_{\mathbb{R}^d} \mathbb{E}_f \Big\{ \sup_{\vec{h}\in\mathcal{H}^d} \big[\big|\xi_n\big(x,\vec{h}\big)\big| - U_n\big(x,\vec{h}\big)\big]_+^p \Big\} \nu_d(\mathrm{d}x) \leq C_p n^{-\frac{p}{2}}$;

(ii) $\int_{\mathbb{R}^d} \mathbb{E}_f \Big\{ \sup_{\vec{h}\in\mathcal{H}^d} \big[\widehat{U}_n\big(x,\vec{h}\big) - 3U_n\big(x,\vec{h}\big)\big]_+^p \Big\} \nu_d(\mathrm{d}x) \leq C'_p n^{-\frac{p}{2}}$;

(iii) $\int_{\mathbb{R}^d} \mathbb{E}_f \Big\{ \sup_{\vec{h}\in\mathcal{H}^d} \big[U_n\big(x,\vec{h}\big) - 4\widehat{U}_n\big(x,\vec{h}\big)\big]_+^p \Big\} \nu_d(\mathrm{d}x) \leq C'_p n^{-\frac{p}{2}}$.

The explicit expression of constant C_p and C'_p can be found in the proof.

5.1.1. *Proof of the theorem.* We start by proving the so-called pointwise oracle inequality for losses.

Pointwise oracle inequality for losses. Let $\vec{h} \in \mathbb{H}$ and $x \in \mathbb{R}^d$ be fixed. We have in view of the triangle inequality

$$\begin{aligned}\left|\widehat{f}_{\vec{\mathbf{h}}(x)}(x) - f(x)\right| &\leq \left|\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) - \widehat{f}_{\vec{\mathbf{h}}(x)}(x)\right| + \left|\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) - \widehat{f}_{\vec{h}}(x)\right| \\ &\quad + \left|\widehat{f}_{\vec{h}}(x) - f(x)\right|. \end{aligned}\tag{5.1}$$

1^0. First, note that obviously $\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) = \widehat{f}_{\vec{h}\vee\vec{\mathbf{h}}(x)}(x)$ and, therefore,

$$\begin{aligned}\left|\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) - \widehat{f}_{\vec{\mathbf{h}}(x)}(x)\right| &= \left|\widehat{f}_{\vec{h}\vee\vec{\mathbf{h}}(x)}(x) - \widehat{f}_{\vec{\mathbf{h}}(x)}(x)\right| \leq \widehat{\mathcal{R}}_{\vec{h}}(x) \\ &\quad + 4\widehat{U}_n\big(x, \vec{\mathbf{h}}(x)\vee\vec{h}\big) + 4\widehat{U}_n\big(x, \vec{\mathbf{h}}(x)\big).\end{aligned}$$

Moreover by definition, $\widehat{U}_n\big(x,\vec{\eta}\big) \leq \widehat{U}^*_n\big(x,\vec{\eta}\big)$ for any $\vec{\eta} \in \mathcal{H}^d$. Next, for any $\vec{h}, \vec{\eta} \in \mathcal{H}^d$ we have obviously $\widehat{U}_n\big(x, \vec{h}\vee\vec{\eta}\big) \leq \widehat{U}^*_n\big(x,\vec{h}\big) \wedge \widehat{U}^*_n\big(x,\vec{\eta}\big)$. Thus,

$$\left|\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) - \widehat{f}_{\vec{\mathbf{h}}(x)}(x)\right| \leq \widehat{\mathcal{R}}_{\vec{h}}(x) + 8\widehat{U}^*_n\big(x, \vec{\mathbf{h}}(x)\big).\tag{5.2}$$

Similarly we have

$$\left|\widehat{f}_{\vec{\mathbf{h}}(x)\vee\vec{h}}(x) - \widehat{f}_{\vec{h}}(x)\right| \leq \widehat{\mathcal{R}}_{\vec{\mathbf{h}}(x)}(x) + 8\widehat{U}^*_n\big(x, \vec{h}\big).\tag{5.3}$$

The definition of $\vec{\mathbf{h}}(x)$ implies that for any $\vec{h} \in \mathbb{H}$

$$\widehat{\mathcal{R}}_{\vec{\mathbf{h}}(x)}(x) + 8\widehat{U}^*_n\big(x,\vec{\mathbf{h}}(x)\big) + \widehat{\mathcal{R}}_{\vec{h}}(x) + 8\widehat{U}^*_n\big(x,\vec{h}\big) \leq 2\widehat{\mathcal{R}}_{\vec{h}}(x) + 16\widehat{U}^*_n\big(x,\vec{h}\big)$$

and we get from (5.1), (5.2) and (5.3) for any $\vec{h} \in \mathbb{H}$

$$\left|\widehat{f}_{\vec{\mathbf{h}}(x)}(x) - f(x)\right| \leq 2\widehat{\mathcal{R}}_{\vec{h}}(x) + 16\widehat{U}^*_n\big(x,\vec{h}\big) + \left|\widehat{f}_{\vec{h}}(x) - f(x)\right|.\tag{5.4}$$

2^0. We obviously have for any $\vec{h}, \vec{\eta} \in \mathcal{H}^d$

$$\begin{aligned}\left|\widehat{f}_{\vec{h}\vee\vec{\eta}}(x) - \widehat{f}_{\vec{\eta}}(x)\right| &\leq \left|\mathbb{E}_f M\big(Z_1 - x, \vec{h}\vee\vec{\eta}\big) - \mathbb{E}_f M\big(Z_1 - x, \vec{\eta}\big)\right| \\ &\quad + \left|\xi_n\big(x, \vec{h}\vee\vec{\eta}\big)\right| + \left|\xi_n\big(x,\vec{\eta}\big)\right|.\end{aligned}$$

Note that for any $\mathrm{h} \in \mathcal{H}^d$

$$\begin{aligned}\mathbb{E}_f M\big(Z_1 - x, \vec{\mathrm{h}}\big) &:= \int_{\mathbb{R}^d} M\big(t - x, \vec{\mathrm{h}}\big)\mathfrak{p}(t)\nu_d(\mathrm{d}t) \\ &= (1-\alpha)\int_{\mathbb{R}^d} M\big(t - x, \vec{\mathrm{h}}\big) f(t)\nu_d(\mathrm{d}t) + \alpha\int_{\mathbb{R}^d} M\big(t - x, \vec{\mathrm{h}}\big)\big[f\star g\big](t)\nu_d(\mathrm{d}t),\end{aligned}$$

in view of the assumption (1.1) imposed on the density $\mathfrak{p}$. Note that

$$\begin{aligned}&(1-\alpha)\int_{\mathbb{R}^d} M\big(t-x,\vec{\mathrm{h}}\big)f(t)\nu_d(\mathrm{d}t)+\alpha\int_{\mathbb{R}^d} M\big(t-x,\vec{\mathrm{h}}\big)\big[f\star g\big](t)\nu_d(\mathrm{d}t)\\ &=\int_{\mathbb{R}^d} f(z)\bigg[(1-\alpha)M\big(z-x,\vec{\mathrm{h}}\big)+\alpha\int_{\mathbb{R}^d} M\big(u,\vec{\mathrm{h}}\big)g(u-z+x)\nu_d(\mathrm{d}u)\bigg]\nu_d(\mathrm{d}z)\end{aligned}$$

and, therefore, in view of the definition of $M\big(\cdot,\vec{h}\big)$, c.f. (2.1), we obtain

$$\mathbb{E}_f M\big(Z_1-x,\vec{\mathrm{h}}\big)=\int_{\mathbb{R}^d} K_{\vec{\mathrm{h}}}(z-x)f(z)\nu_d(\mathrm{d}z)=:S_{\vec{\mathrm{h}}}(x,f),\quad \forall \mathrm{h}\in\mathcal{H}^d. \tag{5.5}$$

We deduce from (5.5) that

$$\big|\mathbb{E}_f M\big(Z_1-x,\vec{h}\vee\vec{\eta}\big)-\mathbb{E}_f M\big(Z_1-x,\vec{\eta}\big)\big|=\big|S_{\vec{h}\vee\vec{\eta}}(x,f)-S_{\vec{\eta}}(x,f)\big|$$

and, therefore, for any $\vec{h},\vec{\eta}\in\mathcal{H}^d$

$$\begin{aligned}\Big|\widehat{f}_{\vec{h}\vee\vec{\eta}}(x)-\widehat{f}_{\vec{\eta}}(x)\Big| \;&\leq\; \big|S_{\vec{h}\vee\vec{\eta}}(x,f)-S_{\vec{\eta}}(x,f)\big|\\ &\quad+\big|\xi_n\big(x,\vec{h}\vee\vec{\eta}\big)\big|+\big|\xi_n\big(x,\vec{\eta}\big)\big|.\end{aligned} \tag{5.6}$$

3^0. Set for any $\vec{h}\in\mathcal{H}^d$ and any $x\in\mathbb{R}^d$

$$\upsilon(x)=\sup_{\vec{\eta}\in\mathcal{H}^d}\big[\big|\xi_n\big(x,\vec{\eta}\big)\big|-U_n\big(x,\vec{\eta}\big)\big]_+$$

$$\varpi_1(x)=\sup_{\vec{h}\in\mathcal{H}^d}\big[U_n\big(x,\vec{h}\big)-4\widehat{U}_n\big(x,\vec{h}\big)\big]_+,\;\varpi_2(x)=\sup_{\vec{h}\in\mathcal{H}^d}\big[\widehat{U}_n\big(x,\vec{h}\big)-3U_n\big(x,\vec{h}\big)\big]_+$$

We obtain in view of (5.6) that for any $\vec{h}\in\mathbb{H}$ (since obviously $\vec{h}\vee\vec{\eta}\in\mathcal{H}^d$ for any $\vec{h},\vec{\eta}\in\mathcal{H}^d$)

$$\widehat{\mathcal{R}}_{\vec{h}}(x)\leq B^*_{\vec{h}}(x,f)+2\upsilon(x)+2\varpi_1(x). \tag{5.7}$$

Using the obvious inequality $(\sup_\alpha F_\alpha-\sup_\alpha G_\alpha)_+\leq\sup_\alpha(F_\alpha-G_\alpha)_+$ get

$$\big[\widehat{U}^*_n\big(x,\vec{h}\big)-3U^*_n\big(x,\vec{h}\big)\big]_+\leq\sup_{\vec{\eta}\in\mathcal{H}^d}\big[\widehat{U}_n\big(x,\vec{\eta}\big)-3U_n\big(x,\vec{\eta}\big)\big]_+=:\varpi_2(x) \tag{5.8}$$

We get from (5.4), (5.7) and (5.8)

$$\begin{aligned}\Big|\widehat{f}_{\vec{\mathbf{h}}(x)}(x)-f(x)\Big| \;&\leq\; 2B^*_{\vec{h}}(x,f)+4\upsilon(x)+4\varpi_1(x)+48U^*_n\big(x,\vec{h}\big)\\ &\quad+16\varpi_2(x)+\big|\widehat{f}_{\vec{h}}(x)-f(x)\big|.\end{aligned}$$

It remains to note that

$$\big|\widehat{f}_{\vec{h}}(x)-f(x)\big|\leq B_{\vec{h}}(x,f)+\big|\xi_n\big(x,\vec{h}\big)\big|\leq B_{\vec{h}}(x,f)+U_n\big(x,\vec{h}\big)+\upsilon(x),$$

and we obtain for any $x \in \mathbb{R}^d$, putting $\mathbf{z}(x) = 5\upsilon(x) + 4\varpi_1(x) + 16\varpi_2(x)$,

$$\big|\widehat{f}_{\vec{\mathbf{h}}(x)}(x) - f(x)\big| \leq 2B^*_{\vec{h}}(x,f) + B_{\vec{h}}(x,f) + 49U^*_n\big(x,\vec{h}\big) + \mathbf{z}(x), \quad \forall \vec{h} \in \mathbb{H}.$$

Noting that the left hand side of the latter inequality is independent of $\vec{h}$ we obtain for any $x \in \mathbb{R}^d$

$$\Big|\widehat{f}_{\vec{\mathbf{h}}(x)}(x) - f(x)\Big| \leq \inf_{\vec{h}\in\mathbb{H}} \Big\{2B^*_{\vec{h}}(x,f) + B_{\vec{h}}(x,f) + 49U^*_n\big(x,\vec{h}\big)\Big\} + \mathbf{z}(x). \tag{5.9}$$

This is the pointwise oracle inequality.

Application of Proposition 1. Set for any $x \in \mathbb{R}^d$

$$R_n(x) = \inf_{\vec{h}\in\mathbb{H}} \Big\{2B^*_{\vec{h}}(x,f) + B_{\vec{h}}(x,f) + 49U^*_n\big(x,\vec{h}\big)\Big\}$$

Applying Proposition 1 we get from (5.9) and the triangle inequality

$$\begin{aligned}\mathcal{R}^{(p)}_n[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}, f] \quad \leq \quad & \big\|R_n\big\|_p + 5\bigg[\int_{\mathbb{R}^d} \mathbb{E}_f\big\{\upsilon(x)\big\}^p\bigg]^{\frac{1}{p}} + 4\bigg[\int_{\mathbb{R}^d} \mathbb{E}_f\big\{\varpi_1(x)\big\}^p\bigg]^{\frac{1}{p}} \\ & +16\bigg[\int_{\mathbb{R}^d} \mathbb{E}_f\big\{\varpi_2(x)\big\}^p\bigg]^{\frac{1}{p}} \leq \big\|R_n\big\|_p + \mathbf{C}_p n^{-\frac{1}{2}},\end{aligned}$$

where $\mathbf{C}_p = 5(C_p)^{\frac{1}{p}} + 20(C'_p)^{\frac{1}{p}}$. The theorem is proved. ∎

5.1.2. *Proof of Proposition 1.* Since the proof of the proposition is quite long and technical, we divide it into several steps.

Preliminaries. We start the proof with the following simple remark. Let $\check{M}\big(t,\vec{h}\big), t \in \mathbb{R}^d$, denote the Fourier transform of $M\big(\cdot,\vec{h}\big)$. Then, we obtain in view of the definition of $M\big(\cdot,\vec{h}\big)$

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

1^0. Note that Assumptions 1 and 2 guarantee that $\check{M}\big(\cdot,\vec{h}\big) \in \mathbb{L}_1\big(\mathbb{R}^d\big) \cap \mathbb{L}_2\big(\mathbb{R}^d\big)$ for any $\vec{h} \in \mathcal{H}^d$ and, therefore,

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

Thus, putting $\mathcal{M}_\infty(\vec{h}) = M_\infty \prod_{j=1}^d h_j^{-1}(h_j \wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}$, we obtain $\forall \vec{h} \in \mathcal{H}^d$

$$\big\|M\big(\cdot,\vec{h}\big)\big\|_\infty \leq \mathcal{M}_\infty(\vec{h}), \quad \big\|M\big(\cdot,\vec{h}\big)\big\|_2 \leq M_2 \prod_{j=1}^d h_j^{-\frac{1}{2}}(h_j \wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}, \tag{5.10}$$

in view of Assumptions 1 and 2. Here we have put $M_2 = \big[(2\pi)^{-d}\big\{\varepsilon^{-1}\big\|\check{K}\big\|_2 1_{\alpha\neq 1} + \Upsilon_0^{-1}\mathbf{k}_2 1_{\alpha=1}\big\}\big] \vee 1$. Additionally we deduce from (5.10) for any $\vec{h} \in \mathcal{H}^d$

$$\big\|M\big(\cdot,\vec{h}\big)\big\|_4^4 \leq M_2^2 M_\infty^2 \prod_{j=1}^d h_j^{-3}(h_j \wedge 1)^{-4\boldsymbol{\mu}_j(\alpha)}. \tag{5.11}$$

Let $\mathcal{L}\big(\cdot,\vec{h}\big)$ be either $M\big(\cdot,\vec{h}\big)$ or $M^2\big(\cdot,\vec{h}\big)$ and let $\mathcal{L}_\infty\big(\vec{h}\big)$ denote either $\mathcal{M}_\infty\big(\vec{h}\big)$ or $\mathcal{M}^2_\infty\big(\vec{h}\big)$.
We have in view of (5.10), denoting $T(\vec{h})=\sum_{j=1}^d[1+\boldsymbol{\mu}_j(\alpha)]|\ln(h_j)|$,

$$\mathcal{L}^{-1}_\infty\big(\vec{h}\big)\vee\mathcal{L}_\infty\big(\vec{h}\big)\;\;\leq\;\;M^2_\infty e^{2T(\vec{h})},\quad\forall\vec{h}\in\mathcal{H}^d. \tag{5.12}$$

Additionally, we get from (5.10) and (5.11)

$$\big\|\mathcal{L}\big(\cdot,\vec{h}\big)\big\|_2^2\leq M_2^2M_\infty^2e^{4T(\vec{h})},\quad\forall\vec{h}\in\mathcal{H}^d. \tag{5.13}$$

Set $\sigma^{\mathcal{L}}\big(x,\vec{h}\big)=\sqrt{\int_{\mathbb{R}^d}\mathcal{L}^2\big(t-x,\vec{h}\big)\mathfrak{p}(t)\nu_d(\mathrm{d}t)}$ and note that in view of (5.13)

$$\int_{\mathbb{R}^d}\big[\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big]^2\nu_d(\mathrm{d}x)=\big\|\mathcal{L}\big(\cdot,\vec{h}\big)\big\|_2^2\leq M_2^2M_\infty^2e^{4T(\vec{h})},\;\;\forall\vec{h}\in\mathcal{H}^d. \tag{5.14}$$

Next, we have in view of (5.12)

$$\big\|\sigma^{\mathcal{L}}\big(\cdot,\vec{h}\big)\big\|_\infty\leq\mathcal{L}_\infty\big(\vec{h}\big)\leq M^2_\infty e^{2T(\vec{h})}. \tag{5.15}$$

2^0. Define for any $x\in\mathbb{R}^d$ and $\vec{h}\in\mathcal{H}^d$

$$\zeta^{\mathcal{L}}\big(x,\vec{h}\big)=n^{-1}\textstyle\sum_{i=1}^n\big[\mathcal{L}\big(Z_i-x,\vec{h}\big)-\mathbb{E}\mathcal{L}\big(Z_i-x,\vec{h}\big)\big];$$
$$z_n\big(x,\vec{h}\big)=3\ln(n)+(8p+22)T(\vec{h})+2\big|\ln\big(\big\{\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big\}\vee\big\{n^{-3/2}\mathcal{L}_\infty\big(\vec{h}\big)\big\}\big)\big|;$$
$$V^{\mathcal{L}}\big(x,\vec{h}\big)=\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\sqrt{2n^{-1}z_n\big(x,\vec{h}\big)}+(4/3)n^{-1}z_n\big(x,\vec{h}\big)\mathcal{L}_\infty\big(\vec{h}\big);$$
$$U^{\mathcal{L}}\big(x,\vec{h}\big)=\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\sqrt{2n^{-1}\lambda_n\big(\vec{h}\big)}+(4/3)n^{-1}\lambda_n\big(\vec{h}\big)\mathcal{L}_\infty\big(\vec{h}\big),$$

where remind $\lambda_n\big(\vec{h}\big)=4\ln(M_\infty)+6\ln(n)+(8p+26)T(\vec{h})$.

Noting that $\sup_{z\in[a,b]}|\ln z|\leq|\ln a|\vee|\ln b|$ for any $0<a<b<\infty$ we deduce from (5.15) $z_n\big(x,\vec{h}\big)\leq\lambda_n\big(\vec{h}\big)$ for any $x\in\mathbb{R}^d$ and, therefore,

$$V^{\mathcal{L}}\big(x,\vec{h}\big)\leq U^{\mathcal{L}}\big(x,\vec{h}\big),\quad\forall\vec{h}\in\mathcal{H}^d. \tag{5.16}$$

First step. Let $x\in\mathbb{R}^d$ and $\vec{h}\in\mathcal{H}^d$ be fixed. Put $b=8p+22$. We obtain for any $z\geq1$ and $q\geq1$ by the integration of the Bernstein inequality

$$\begin{aligned}&\mathbb{E}_f\Big\{\big|\zeta^{\mathcal{L}}\big(x,\vec{h}\big)\big|-\sqrt{2n^{-1}z}\sigma^{\mathcal{L}}\big(x,\vec{h}\big)-(4/3)n^{-1}z\mathcal{L}_\infty\big(\vec{h}\big)\Big\}_+^q\\&\leq2\Gamma(q+1)\big[\sqrt{2n^{-1}}\sigma^{\mathcal{L}}\big(x,\vec{h}\big)+(4/3)n^{-1}\mathcal{L}_\infty\big(\vec{h}\big)\big]^q\exp\{-z\},\end{aligned}$$

where Γ is the Gamma-function.

1^0. Choose $z = z_n(x,\vec{h})$. Noting that for any $n \in \mathbb{N}^*$ and $x \in \mathbb{R}^d$

$$\sqrt{2n^{-1}z}\sigma^{\mathcal{L}}(x,\vec{h}) + (4/3)n^{-1}\mathcal{L}_\infty(\vec{h}) \leq 3\mathcal{L}_\infty(\vec{h})n^{-\frac{1}{2}}$$

and taking into account that $\exp\{-|\ln(y)|\} \leq y$ for any $y > 0$, we get

$$\begin{aligned}
&\mathbb{E}_f\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - V^{\mathcal{L}}(x,\vec{h})\big\}_+^q \\
&\leq 2\times 3^q\Gamma(q+1)n^{-\frac{q}{2}-3}\mathcal{L}_\infty^q(\vec{h})e^{bT(\vec{h})}\big(\{\sigma^{\mathcal{L}}(x,\vec{h})\} \vee \{n^{-3/2}\mathcal{L}_\infty(\vec{h})\}\big)^2 \\
&\leq C_q^{(1)}n^{-\frac{q}{2}-3}e^{(2q-b)T(\vec{h})}\big(\{\sigma^{\mathcal{L}}(x,\vec{h})\} \vee \{n^{-3/2}\mathcal{L}_\infty(\vec{h})\}\big)^2. \qquad (5.17)
\end{aligned}$$

Here to get the second inequality we have used (5.12) and put $C_q^{(1)} = 2M_\infty^{2q}3^q\Gamma(q+1)$.

Set $\mathcal{X}(\vec{h}) = \{x \in \mathbb{R}^d : \sigma^{\mathcal{L}}(x,\vec{h}) \geq n^{-3/2}\mathcal{L}_\infty(\vec{h})\}$, $\bar{\mathcal{X}}(\vec{h}) = \mathbb{R}^d \setminus \mathcal{X}(\vec{h})$ and later on the integration over the empty set is supposed to be zero.

Putting $C_p^{(2)} = C_p^{(1)}M_2^2M_\infty^2$, we have in view of (5.16), (5.14) and (5.17) applied with $q = p$ that for any $\vec{h} \in \mathcal{H}^d$

$$\int_{\mathcal{X}(\vec{h})}\mathbb{E}_f\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - U^{\mathcal{L}}(x,\vec{h})\big\}_+^p\nu_d(\mathrm{d}x) \leq C_p^{(2)}n^{-\frac{p}{2}}e^{(2p+4-b)T(\vec{h})}. \qquad (5.18)$$

2^0. Introduce the following notations. For any $i = 1,\dots,n$ set

$$\Psi_i(x,\vec{h}) = 1\Big\{\big|\mathcal{L}(Z_i - x,\vec{h}) - \mathbb{E}\mathcal{L}(Z_i - x,\vec{h})\big| \geq n^{-1}\mathcal{L}_\infty(\vec{h})\Big\},$$

and define the random event $D(x,\vec{h}) = \Big\{\sum_{i=1}^n\Psi_i(x,\vec{h}) \geq 2\Big\}$. As usual, the complimentary event will be denoted by $\bar{D}(x,\vec{h})$. Set finally $\pi(x,\vec{h}) = \mathbb{P}_f\{\Psi_1(x,\vec{h}) = 1\}$.
We obviously have $|\zeta^{\mathcal{L}}(x,\vec{h})|1_{\bar{D}(x,\vec{h})} \leq 3n^{-1}\mathcal{L}_\infty(\vec{h}) < U^{\mathcal{L}}(\vec{h})$ and, therefore,

$$1_{\bar{D}(x,\vec{h})}\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - U^{\mathcal{L}}(x,\vec{h})\big\}_+^p = 0. \qquad (5.19)$$

Applying Cauchy-Schwartz inequality, we deduce from (5.19) that

$$\mathbb{E}_f\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - U^{\mathcal{L}}(x,\vec{h})\big\}_+^p \leq \mathbb{E}_f^{\frac{1}{2}}\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - U^{\mathcal{L}}(x,\vec{h})\big\}_+^{2p}\,\mathbb{P}_f\{D(x,\vec{h})\}.$$

Using (5.17) with $q = 2p$ and (5.12) we obtain for any $x \in \bar{\mathcal{X}}(\vec{h})$

$$\mathbb{E}_f\big\{|\zeta^{\mathcal{L}}(x,\vec{h})| - U^{\mathcal{L}}(x,\vec{h})\big\}_+^p \leq C_p^{(3)}n^{-\frac{p+6}{2}}e^{\mathbf{c}_pT(\vec{h})}\Big[\mathbb{P}_f\{D(x,\vec{h})\}\Big]^{\frac{1}{2}}, \qquad (5.20)$$

where we have put $C_p^{(3)} = \big[C_{2p}^{(1)}\big]^{\frac{1}{2}}M_\infty^2$ and $\mathbf{c}_p = 2p + 2 - b/2$.

For any $\lambda > 0$ we have in view of the exponential Markov inequality

$$\mathbb{P}_f\big\{D\big(x,\vec{h}\big)\big\} = \mathbb{P}_f\Big\{\textstyle\sum_{i=1}^n \Psi_i\big(x,\vec{h}\big) \geq 2\Big\}$$
$$\leq e^{-2\lambda}\big[e^{\lambda}\pi\big(x,\vec{h}\big) + 1 - \pi\big(x,\vec{h}\big)\big]^n = e^{-2\lambda}\big[(e^{\lambda}-1)\pi\big(x,\vec{h}\big)+1\big]^n$$
$$\leq \exp\{-2\lambda + n(e^{\lambda}-1)\pi\big(x,\vec{h}\big)\}.$$

Tchebychev inequality yields $\pi\big(x,\vec{h}\big) \leq n^2\mathcal{L}_\infty^{-2}\big(\vec{h}\big)\big[\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big]^2$ and we get

$$\mathbb{P}_f\big\{D\big(x,\vec{h}\big)\big\} \leq \exp\big\{-2\lambda + n^3\mathcal{L}_\infty^{-2}\big(\vec{h}\big)\big[\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big]^2(e^{\lambda}-1)\big\},\quad \forall \vec{h}\in\mathcal{H}^d.$$

Note that the definition of $\bar{\mathcal{X}}\big(\vec{h}\big)$ implies $n^3\mathcal{L}_\infty^{-2}\big(\vec{h}\big)\big[\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big]^2 < 1$ for any $x\in\bar{\mathcal{X}}\big(\vec{h}\big)$. Hence, choosing $\lambda = \ln 2 - 2\ln\big\{n^{3/2}\mathcal{L}_\infty^{-1}\big(\vec{h}\big)\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big\}$ we have

$$\mathbb{P}_f\big\{D\big(x,\vec{h}\big)\big\} \leq (e^2/4)n^6\mathcal{L}_\infty^{-4}\big(\vec{h}\big)\big[\sigma^{\mathcal{L}}\big(x,\vec{h}\big)\big]^4,\quad \forall x\in\bar{\mathcal{X}}\big(\vec{h}\big).$$

It yields, together with (5.12), (5.14) and (5.20) and for any $\vec{h}\in\mathcal{H}^d$

$$\text{(5.21)}\quad \int_{\bar{\mathcal{X}}(\vec{h})}\mathbb{E}_f\big\{\big|\zeta^{\mathcal{L}}\big(x,\vec{h}\big)\big| - U^{\mathcal{L}}\big(x,\vec{h}\big)\big\}_+^p\nu_d(\mathrm{d}x) \leq C_p^{(4)}n^{-\frac{p}{2}}e^{(2p+10-b/2)T(\vec{h})},$$

where $C_p^{(4)} = C_p^{(3)}(e/2)M_\infty^6M_2^2$. Putting $C_p^{(5)} = C_p^{(2)} + C_p^{(4)}$ and noting that $2p+10-b/2<0$ we obtain from (5.18) and (5.21) for any $\vec{h}\in\mathcal{H}^d$

$$\text{(5.22)}\quad \int_{\mathbb{R}^d}\mathbb{E}_f\big\{\big|\zeta^{\mathcal{L}}\big(x,\vec{h}\big)\big| - U^{\mathcal{L}}\big(x,\vec{h}\big)\big\}_+^p\nu_d(\mathrm{d}x) \leq C_p^{(5)}n^{-\frac{p}{2}}e^{(2p+10-b/2)T(\vec{h})}.$$

3^0. Choosing $\mathcal{L}=M$ and $\mathcal{L}_\infty = \mathcal{M}_\infty$ we get from (5.22) and the definition of b for any $\vec{h}\in\mathcal{H}^d$

$$\text{(5.23)}\qquad \int_{\mathbb{R}^d}\mathbb{E}_f\big\{\big|\xi_n\big(x,\vec{h}\big)\big| - U_n\big(x,\vec{h}\big)\big\}_+^p\nu_d(\mathrm{d}x) \leq C_p^{(5)}n^{-\frac{p}{2}}e^{-T(\vec{h})}.$$

The first assertion follows from (5.23) with $C_p = C_p^{(5)}\sum_{k\in\mathbb{Z}^d}e^{-\sum_{j=1}^d|k_j|}$.

Second step. Denoting $\chi\big(x,\vec{h}\big) = \big\{\big|\widehat{\sigma}^2\big(x,\vec{h}\big) - \sigma^2\big(x,\vec{h}\big)\big| - \mathfrak{U}_n\big(x,\vec{h}\big)\big\}_+$,

$$\mathfrak{U}_n\big(x,\vec{h}\big) = \sigma^{M^2}\big(x,\vec{h}\big)\sqrt{2n^{-1}\lambda_n\big(\vec{h}\big)} + (4/3)n^{-1}\lambda_n\big(\vec{h}\big)\mathcal{M}_\infty^2\big(\vec{h}\big),$$

and choosing $\mathcal{L}=M^2$ and $\mathcal{L}_\infty = \mathcal{M}_\infty^2$, we get from (5.22) for any $\vec{h}\in\mathcal{H}^d$

$$\text{(5.24)}\qquad \int_{\mathbb{R}^d}\mathbb{E}_f\big\{\chi^p\big(x,\vec{h}\big)\big\}\nu_d(\mathrm{d}x) \leq C_p^{(5)}n^{-\frac{p}{2}}e^{(2p+10-b/2)T(\vec{h})}.$$

Note that $\sigma^{M^2}\big(x,\vec{h}\big) \leq \mathcal{M}_\infty\big(\vec{h}\big)\sigma\big(x,\vec{h}\big)$ and, therefore, for any $x\in\mathbb{R}^d$

$$\mathfrak{U}_n\big(x,\vec{h}\big) \leq \mathcal{M}_\infty\big(\vec{h}\big)U_n\big(x,\vec{h}\big),\quad \forall\vec{h}\in\mathcal{H}^d.$$

This implies,

$$
\begin{aligned}
2n^{-1}\lambda_n(\vec{h})\widehat{\sigma}^2(x,\vec{h}) \quad &\leq \quad 2n^{-1}\lambda_n(\vec{h})\sigma^2(x,\vec{h}) + 2n^{-1}\lambda_n(\vec{h})\mathcal{M}_\infty(\vec{h})U_n(x,\vec{h}) \\
&\quad\; +2n^{-1}\lambda_n(\vec{h})\mathcal{M}_\infty(\vec{h})\chi^*(x,\vec{h}),
\end{aligned}
$$

where we have denoted $\chi^*(x,\vec{h}) = \mathcal{M}_\infty^{-1}(\vec{h})\chi(x,\vec{h})$. Hence

$$
\widehat{U}_n(x,\vec{h}) \leq U_n(x,\vec{h}) + \sqrt{2n^{-1}\lambda_n(\vec{h})\mathcal{M}_\infty(\vec{h})\big[U_n(x,\vec{h}) + \chi^*(x,\vec{h})\big]}. \tag{5.25}
$$

By the same reason

$$
U_n(x,\vec{h}) \leq \widehat{U}_n(x,\vec{h}) + \sqrt{2n^{-1}\lambda_n(\vec{h})\mathcal{M}_\infty(\vec{h})\big[U_n(x,\vec{h}) + \chi^*(x,\vec{h})\big]}. \tag{5.26}
$$

Note that the definition of $\widehat{U}_n(x,\vec{h})$ and $U_n(x,\vec{h})$ implies that

$$
2n^{-1}\lambda_n(\vec{h})\mathcal{M}_\infty(\vec{h}) \leq (3/2)\min\big[\widehat{U}_n(x,\vec{h}), U_n(x,\vec{h})\big]. \tag{5.27}
$$

Using the inequality $\sqrt{|ab|} \leq 2^{-1}(|ay| + |b/y|)$, $y > 0$ we get from (5.25), (5.26) and (5.27)

$$
\begin{aligned}
\widehat{U}_n(x,\vec{h}) \quad &\leq \quad \big(1 + \sqrt{3/2} + (3/4)y\big)U_n(x,\vec{h}) + (2y)^{-1}\chi^*(x,\vec{h}); \\
U_n(x,\vec{h}) \quad &\leq \quad \big(1 + (3/4)y\big)\widehat{U}_n(x,\vec{h}) + (2y)^{-1}U_n(x,\vec{h}) + (2y)^{-1}\chi^*(x,\vec{h}).
\end{aligned}
$$

Choosing $y = 1/2$ in the first inequality and $y = 1$ in the second we get for any $x \in \mathbb{R}^d$ and $\vec{h} \in \mathcal{H}^d$

$$
\big[\widehat{U}_n(x,\vec{h}) - 3U_n(x,\vec{h})\big]_+ \quad \leq \quad \chi^*(x,\vec{h}); \tag{5.28}
$$

$$
\big[U_n(x,\vec{h}) - 4\widehat{U}_n(x,\vec{h})\big]_+ \quad \leq \quad \chi^*(x,\vec{h}). \tag{5.29}
$$

Remembering that $b = 8p + 22$ we obtain from (5.28), (5.29), (5.24) and (5.12) for any $\vec{h} \in \mathcal{H}^d$, denoting $C'_p = M_\infty^{2p} C_p^{(5)}$.

$$
\int_{\mathbb{R}^d} \mathbb{E}_f\big[\widehat{U}_n(x,\vec{h}) - 3U_n(x,\vec{h})\big]_+^p \nu_d(\mathrm{d}x) \leq C'_p n^{-\frac{p}{2}} e^{-T(\vec{h})}; \tag{5.30}
$$

$$
\int_{\mathbb{R}^d} \mathbb{E}_f\big[U_n(x,\vec{h}) - 4\widehat{U}_n(x,\vec{h})\big]_+^p \nu_d(\mathrm{d}x) \leq C'_p n^{-\frac{p}{2}} e^{-T(\vec{h})}. \tag{5.31}
$$

The second and third assertions follow from (5.30) and (5.31). ∎

5.2. *Proof of Theorem 2.* The proof of the theorem is very long and technical and we break it on two parts, which in its turn are divided on several steps. Introduce the following notations: $c_1 = M_2\sqrt{2D}$, $c_2 = \frac{4M_\infty}{3}$,

$$a = \Big\{196\big[(c_1\sqrt{c_3}) \vee (c_2c_3)\big]\Big\}^{-1},$$

where $c_3 = 2\max\big\{4\ln(M_\infty), (8p+26)\max_{j=1,\dots,d}[1+\boldsymbol{\mu}_j(\alpha)]\big\}$.

5.2.1. *Preliminaries.* Recall that for any locally integrable function $\lambda : \mathbb{R}^d \to \mathbb{R}$ its strong maximal function is defined as

$$\mathfrak{M}[\lambda](x) := \sup_H \frac{1}{\nu_d(H)} \int_H \lambda(t)\mathrm{d}t, \quad x \in \mathbb{R}^d, \tag{5.32}$$

where the supremum is taken over all possible rectangles H in $\mathbb{R}^d$ with sides parallel to the coordinate axes, containing point x. It is well known that the strong maximal operator $\lambda \mapsto \mathfrak{M}[\lambda]$ is of the strong $(\mathbf{t},\mathbf{t})$–type for all $1 < \mathbf{t} \le \infty$, i.e., if $\lambda \in \mathbb{L}_{\mathbf{t}}(\mathbb{R}^d)$ then $\mathfrak{M}[\lambda] \in \mathbb{L}_{\mathbf{t}}(\mathbb{R}^d)$ and there exists a constant $C_{\mathbf{t}}$ depending on $\mathbf{t}$ only such that

$$\big\|\mathfrak{M}[\lambda]\big\|_{\mathbf{t}} \le C_{\mathbf{t}}\|\lambda\|_{\mathbf{t}}, \quad \mathbf{t} \in (1,\infty]. \tag{5.33}$$

Let $\mathfrak{m}[\lambda]$ be defined by (5.32), where, instead of rectangles, the supremum is taken over all possible cubes H in $\mathbb{R}^d$ with sides parallel to the coordinate axes, containing point x. Then, it is known that $\lambda \mapsto \mathfrak{m}[\lambda]$ is of the weak $(1,1)$-type, i.e. there exists $C_{\mathbf{1}}$ depending on d only s.t. for any $\lambda \in \mathbb{L}_1(\mathbb{R}^d)$

$$\nu_d\Big\{x : \big|\mathfrak{m}[\lambda](x)\big| \ge \mathfrak{z}\Big\} \le C_{\mathbf{1}}\mathfrak{z}^{-1}\|\lambda\|_1, \quad \forall \mathfrak{z} > 0. \tag{5.34}$$

The results presented below deal with the weak property of the strong maximal function. The following inequality can be found in Guzman (1975). There exists a constant $\mathbf{C} > 0$ depending on d only such that

$$\nu_d\Big\{x : \big|\mathfrak{M}[\lambda](x)\big| \ge \mathfrak{z}\Big\} \le \mathbf{C}\int_{\mathbb{R}^d} \tfrac{|\lambda(x)|}{\mathfrak{z}}\Bigg\{1 + \bigg(\ln_+ \tfrac{|\lambda(x)|}{\mathfrak{z}}\bigg)^{d-1}\Bigg\}\mathrm{d}x, \ \mathfrak{z} > 0,$$

where for all $z \in \mathbb{R}$, $\ln_+(z) := \max\{\ln(z), 0\}$.

Lemma 1. *For any given $d \ge 1, R > 0$, $Q > 0$ and $\mathbf{q} \in (1,\infty]$ there exists $C(d,\mathbf{q},R,Q)$ such that for any $\lambda \in \mathbb{B}_{1,d}(R) \cap \mathbb{B}_{\mathbf{q},d}(Q)$*

$$\nu_d\Big\{x : \big|\mathfrak{M}[\lambda](x)\big| \ge \mathfrak{z}\Big\} \le C(d,\mathbf{q},R,Q)\mathfrak{z}^{-1}\big(1 + |\ln(\mathfrak{z})|\big)^{d-1}, \quad \forall \mathfrak{z} > 0.$$

The proof of the lemma is an elementary consequence of the aforementioned result and can be omitted.

Recall also the particular case of the Young inequality for weak-type spaces, see Grafakos (2008), Theorem 1.2.13. For any $\mathbf{u} \in (1,\infty]$ there exists $C_{\mathbf{u}} > 0$ such that for any $\lambda_1 \in \mathbb{L}_1(\mathbb{R}^d)$ and $\lambda_2 \in \mathbb{L}_{\mathbf{u},\infty}(\mathbb{R}^d)$ one has

$$\|\lambda_1 \star \lambda_2\|_{\mathbf{u},\infty} \le C_{\mathbf{u}} \|\lambda_1\|_1 \|\lambda_2\|_{\mathbf{u},\infty}. \tag{5.35}$$

Let $\mathfrak{J}$ denote the set of all the subsets of $\{1,\dots d\}$ endowed with the empty set $\emptyset$. For any $J \in \mathfrak{J}$ and $y \in \mathbb{R}^d$ set $y_J = \{y_j,\ j \in J\} \in \mathbb{R}^{|J|}$ and we will write $y = (y_J, y_{\bar{J}})$, where as usual $\bar{J} = \{1,\dots d\} \setminus J$.

For any $j = 1,\dots,d$ introduce the $d \times d$ matrix $\mathbf{E}_j = (\mathbf{0},\dots,\mathbf{e}_j,\dots,\mathbf{0})$ where, recall, $(\mathbf{e}_1,\dots,\mathbf{e}_d)$ denotes the canonical basis of $\mathbb{R}^d$. Set also $\mathbf{E}[J] = \sum_{j\in J} \mathbf{E}_j$. Later on $\mathbf{E}_0 = \mathbf{E}[\emptyset]$ denotes the matrix with zero entries.

To any $J \in \mathfrak{J}$ and any $\lambda : \mathbb{R}^d \to \mathbb{R}$ associate the function

$$\lambda_J(y_J, z_{\bar{J}}) = \lambda\big(z + \mathbf{E}[J](y-z)\big), \quad y, z \in \mathbb{R}^d,$$

with the obvious agreement $\lambda_J \equiv \lambda$ if $J = \{1,\dots d\}$, which is always the case if $d = 1$. For any $\vec{h} \in \mathcal{H}^d$ and $J \subseteq \{1,\dots d\}$ set $K_{\vec{h},J}(u_J) = \prod_{j\in J} h_j^{-1} \mathcal{K}(u_j/h_j)$ and define for any $y \in \mathbb{R}^d$

$$\big[K_{\vec{h}} \circ \lambda\big]_J(y) = \int_{\mathbb{R}^{|\bar{J}|}} K_{\vec{h},\bar{J}}(u_{\bar{J}} - y_{\bar{J}}) \lambda(y_J, u_{\bar{J}}) \nu_{|\bar{J}|}(\mathrm{d}u_{\bar{J}}),$$

where $\nu_{|\bar{J}|}$ is the Lebesgue measure on $\mathbb{R}^{|\bar{J}|}$. For any $\vec{h}, \vec{\eta} \in \mathcal{H}^d$ set

$$B_{\vec{h},\vec{\eta}}(x,f) = |S_{\vec{h}\vee\vec{\eta}}(x,f) - S_{\vec{\eta}}(x,f)|.$$

LEMMA 2. *Let Assumption 3 hold. One can find $k \in \{1,\dots d\}$ and a collection of indexes $\{j_1 < j_2 < \cdots < j_k\} \in \{1,\dots,d\}$ such that for any $x \in \mathbb{R}^d$ and any $f : \mathbb{R}^d \to \mathbb{R}$*

$$B_{\vec{h},\vec{\eta}}(x,f) \le \sum_{l=1}^k \Big(\Big[\big|K_{\vec{h}\vee\vec{\eta}}\big| \circ b_{h_{j_l},f,j_l} \Big]_{J_l}(x) + \Big[\big|K_{\vec{\eta}}\big| \circ b_{h_{j_l},f,j_l} \Big]_{J_l}(x) \Big);$$

$$B_{\vec{h}}(x,f) \le \sum_{l=1}^k \Big[\big|K_{\vec{h}}\big| \circ b_{h_{j_l},f,j_l} \Big]_{J_l}(x), \qquad J_l = \{j_1,\dots,j_l\}.$$

The proof of the lemma can be found in Lepski (2015), Lemma 2. Also, let us mention the following bound which is a trivial consequence of the Young inequality and the Fubini theorem. If $\lambda \in \mathbb{L}_{\mathbf{t}}(\mathbb{R}^d)$ then for any $\mathbf{t} \in [1,\infty]$

$$\sup_{J\in\mathfrak{J}} \big\| \big[K_{\vec{h}} \circ \lambda\big]_J \big\|_{\mathbf{t}} \le \|\mathcal{K}\|_1^d \|\lambda\|_{\mathbf{t}}, \quad \forall \vec{h} \in \mathcal{H}^d. \tag{5.36}$$

To any $J \in \mathfrak{J}$ and any locally integrable function $\lambda : \mathbb{R}^d \to \mathbb{R}_+$ we associate the operator

$$\mathfrak{M}_J[\lambda](x) = \sup_{H_{|\bar{J}|}} \frac{1}{\nu_{|\bar{J}|}(H_{|\bar{J}|})} \int_{H_{|\bar{J}|}} \lambda\big(t + \mathbf{E}[J][x-t]\big) \nu_{|\bar{J}|}(\mathrm{d}t_{\bar{J}}) \tag{5.37}$$

where the supremum is taken over all hyper-rectangles in $\mathbb{R}^{|\bar J|}$ containing $x_{\bar J}=(x_j, j\in\bar J)$ and with sides parallel to the axis.

As we see $\mathfrak{M}_J[\lambda]$ is the strong maximal operator applied to the function obtained from λ by fixing the coordinates whose indices belong to J. It is obvious that $\mathfrak{M}_\emptyset[\lambda]\equiv\mathfrak{M}[\lambda]$ and $\mathfrak{M}_{\{1,\dots,d\}}[\lambda]\equiv\lambda$.

The following result is a direct consequence of (5.33) and of the Fubini theorem. For any $\mathbf{t}\in(1,\infty]$ there exists $\mathbf{C_t}$ such that for any $\lambda\in\mathbb{L}_{\mathbf{t}}\big(\mathbb{R}^d\big)$

$$\sup_{J\in\mathfrak{J}}\big\|\mathfrak{M}_J[\lambda]\big\|_{\mathbf{t}}\leq\mathbf{C_t}\|\lambda\|_{\mathbf{t}}. \tag{5.38}$$

Obviously this inequality holds if $\mathbf{t}=\infty$ with $\mathbf{C}_\infty=1$.

5.2.2. **Part I.** For any $\vec h\in\mathcal{H}^d$ and any $v>0$, let

$$\mathcal{B}_{\vec h}(\cdot,f)=2B^*_{\vec h}(\cdot,f)+B_{\vec h}(\cdot,f),\quad \mathcal{A}(\vec h,f,v)=\big\{x\in\mathbb{R}^d:\ \mathcal{B}_{\vec h}(x,f)\geq 2^{-1}v\big\},$$

Introduce for any $v>0$ and $f\in\mathbb{F}_{g,\mathbf{u}}(R,D)$

$$\Lambda(v,f)=\inf_{\vec h\in\mathfrak{H}(v)}\Big[\nu_d\big(\mathcal{A}(\vec h,f,v)\big)+v^{-2}F_n^2\big(\vec h\big)\Big];$$

$$\Lambda(v,f,\mathbf{u})=\inf_{z\geq 2}\ \inf_{\vec h\in\mathfrak{H}(v,z)}\big[\nu_d\big(\mathcal{A}(\vec h,f,v)\big)+z^{-\mathbf{u}}\big];$$

$$\Lambda_p(v,f,\mathbf{u})=\inf_{z\geq 2}\ \inf_{\vec h\in\mathfrak{H}(v,z)}\Big[\int_{\mathcal{A}(\vec h,f,v)}\big|\mathcal{B}_{\vec h}(x,f)\big|^p\nu_d(\mathrm{d}x)+v^pz^{-\mathbf{u}}\Big].$$

Note that $\Lambda_p(\infty,f,\mathbf{u})\equiv 0$. Let K be a compactly supported function satisfying Assumption 2. The goal of Part I is to prove the following bound.

For any $n\geq 3,\ p>1,\mathbf{q}>1,R>1,D>0,0<\underline{\boldsymbol{v}}\leq\overline{\boldsymbol{v}}\leq\infty,\mathbf{u}\in(p/2,\infty],\mathbf{u}\geq\mathbf{q}$ and any $f\in\mathbb{F}_{g,\mathbf{u}}(R,D)\cap\mathbb{B}_{\mathbf{q},d}(D)$

$$\begin{aligned}\mathcal{R}_n^{(p)}[\widehat f_{\vec{\mathbf{h}}(\cdot)},f]\quad&\leq\quad C^{(1)}\Big[l_{\mathbb{H}}(\underline{\boldsymbol{v}})+\int_{\underline{\boldsymbol{v}}}^{\overline{\boldsymbol{v}}}v^{p-1}\{\Lambda(v,f)\wedge\Lambda(v,f,\mathbf{u})\}\mathrm{d}v\\&\qquad+\Lambda_p(\overline{\boldsymbol{v}},f,\mathbf{u})\Big]^{\frac1p}+\mathbf{C}_pn^{-\frac12}.\end{aligned}\tag{5.39}$$

Here $C^{(1)}$ is a constant independent of f and n. Its explicit expression can be found in the proof of the theorem. We remark also that only this constant depends on $\mathbf{q}$. Since the risk of our estimator is independent of $\underline{\boldsymbol{v}},\overline{\boldsymbol{v}}>0$ we can minimize the right hand side of (5.39) w.r.t. these parameters.

Auxiliary results. Let us prove several simple facts. First note that for any $n\geq 3$ for any $\vec h\in\mathcal{H}^d$

$$\lambda_n\big(\vec h\big)\leq c_3\Big[\ln(n)+\sum_{j=1}^d\big|\ln(h_j)\big|\Big]. \tag{5.40}$$

Next, it is easy to see that for any any $n \geq 3$ and $\vec{\eta}, \vec{h} \in (0,\infty)^d : \ \vec{\eta} \geq \vec{h}$

$$F_n\big(\vec{\eta}\big) \leq F_n\big(\vec{h}\big)\sqrt{l\Big(V_{\vec{\eta}}/V_{\vec{h}}\Big)},\ G_n\big(\vec{\eta}\big) \leq G_n\big(\vec{h}\big) l\Big(V_{\vec{\eta}}/V_{\vec{h}}\Big),$$

where $l(v) = v^{-1}(1+\ln v)$. Since $\vec{\eta} \geq \vec{h}$ implies $V_{\vec{\eta}} \geq V_{\vec{h}}$ and $l(v) \leq 1$ if $v \geq 1$, we have

$$F_n\big(\vec{\eta}\big) \leq F_n\big(\vec{h}\big), \quad G_n\big(\vec{\eta}\big) \leq G_n\big(\vec{h}\big), \quad \forall \vec{\eta}, \vec{h} \in (0,\infty)^d : \ \vec{\eta} \geq \vec{h}. \tag{5.41}$$

Then by (5.40) and the second inequality in (5.41), we have:

$$\sup_{\vec{\eta} \in \mathcal{H}^d:\ \vec{\eta} \geq \vec{h}} \frac{4M_\infty \lambda_n\big(\vec{\eta}\big)}{3n\prod_{j=1}^d \eta_j(\eta_j \wedge 1)^{\boldsymbol{\mu}_j(\alpha)}} \leq c_2 c_3 G_n\big(\vec{h}\big). \tag{5.42}$$

Now let us establish two bounds for $\|U_n^*\big(\cdot,\vec{h}\big)\|_\infty$.

1^0a. Let $\mathbf{u} = \infty$. We have in view of the second inequality in (5.10) for any $\vec{\eta} \in \mathcal{H}^d$

$$\sigma\big(x,\vec{\eta}\big) \leq \sqrt{D}\big\|M\big(\cdot,\vec{\eta}\big)\big\|_2 \leq M_2\sqrt{D}\prod_{j=1}^d \eta_j^{-\frac{1}{2}}(\eta_j \wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}, \quad \forall x \in \mathbb{R}^d.$$

It yields for any $x \in \mathbb{R}^d$ in view of the first inequality in (5.41)

$$\sup_{\vec{\eta} \in \mathcal{H}^d:\ \vec{\eta} \geq \vec{h}} \sqrt{2n^{-1}\lambda_n\big(\vec{\eta}\big)\sigma^2\big(x,\vec{\eta}\big)} \leq \sup_{\vec{\eta} \in \mathcal{H}^d:\ \vec{\eta} \geq \vec{h}} c_1\sqrt{c_3}F_n\big(\vec{\eta}\big) \leq c_1\sqrt{c_3}F_n\big(\vec{h}\big). \tag{5.43}$$

Then gathering (5.42), (5.43) and by definition of a, we have

$$\|U_n^*\big(\cdot,\vec{h}\big)\|_\infty \leq (196a)^{-1}\big[F_n\big(\vec{h}\big) + G_n\big(\vec{h}\big)\big]. \tag{5.44}$$

1^0b. Another bound for $\|U_n^*\big(\cdot,\vec{h}\big)\|_\infty$ is available regardless of the value of $\mathbf{u}$. Indeed for any $\vec{\eta} \in \mathcal{H}^d$ in view of the first inequality in (5.10)

$$\sigma\big(x,\vec{\eta}\big) \leq \big\|M\big(\cdot,\vec{\eta}\big)\big\|_\infty \leq M_\infty\prod_{j=1}^d \eta_j^{-1}(\eta_j \wedge 1)^{-\boldsymbol{\mu}_j(\alpha)}, \quad \forall x \in \mathbb{R}^d.$$

It yields for any $x \in \mathbb{R}^d$ and any $n \geq 3$

$$\begin{aligned}
&\sup_{\vec{\eta} \in \mathcal{H}^d:\ \vec{\eta} \geq \vec{h}} \sqrt{2n^{-1}\lambda_n\big(\vec{\eta}\big)\sigma^2\big(x,\vec{\eta}\big)} \\
&\leq \sup_{\vec{\eta} \in \mathcal{H}^d:\ \vec{\eta} \geq \vec{h}} \sqrt{2c_3}M_\infty\frac{\sqrt{\ln(n) + \sum_{j=1}^d \big|\ln(\eta_j)\big|}}{\sqrt{n}\prod_{j=1}^d \eta_j(\eta_j \wedge 1)^{\boldsymbol{\mu}_j(\alpha)}} \leq \sqrt{\frac{2c_3 n}{\ln(n)}}M_\infty G_n\big(\vec{h}\big).
\end{aligned}$$

Then gathering with (5.42) again, we have

$$\|U_n^*\big(\cdot,\vec{h}\big)\|_\infty \leq \sqrt{2c_3}M_\infty \vee (c_2c_3)\sqrt{n/\ln n}\, G_n\big(\vec{h}\big), \quad \forall \vec{h} \in \mathcal{H}^d.$$

Denoting $\vec{\mathbf{b}}[b]=(b,\dots,b)$, we obtain from (5.2.2)

$$\inf_{\vec{h}\in\mathbb{H}}\|U_n^*(\cdot,\vec{h})\|_\infty\leq\inf_{b\geq1}\big\|U_n^*(\cdot,\vec{\mathbf{b}}[b])\big\|_\infty=0. \tag{5.45}$$

2⁰. Let now $\mathbf{u}<\infty$. Let us prove that for any $\mathfrak{z}>0$, $\mathbf{s}\in\{1,\mathbf{u}\}$ and any $f\in\mathbb{F}_{g,\mathbf{u}}(R,D)$

$$\nu_d\bigg(x\in\mathbb{R}^d:\sup_{\vec{\eta}\in\mathcal{H}^d:\,\vec{\eta}\geq\vec{h}}\mathcal{U}_n\big(x,\vec{\eta},f\big)\geq\mathfrak{z}\bigg)\leq c_5\big[\widetilde{D}\mathfrak{z}^{-2}F_n^2\big(\vec{h}\big)\big]^{\mathbf{s}}, \tag{5.46}$$

where we have put $\mathcal{U}_n^2\big(\cdot,\vec{\eta},f\big)=2n^{-1}\lambda_n\big(\vec{\eta}\big)\sigma^2\big(\cdot,\vec{\eta}\big)$ and $\widetilde{D}=1$ if $\mathbf{s}=1$ and $\widetilde{D}=D$ if $\mathbf{s}=\mathbf{u}$. Indeed, if $\mathbf{s}=1$, applying the Markov inequality, we obtain in view of the second inequality in (5.10) for any $\vec{\eta}\in\mathcal{H}^d$

$$\begin{aligned}\nu_d\Big(x\in\mathbb{R}^d:\,\mathcal{U}_n\big(x,\vec{\eta},f\big)\geq\mathfrak{z}\Big)&\leq2(n\mathfrak{z}^2)^{-1}\lambda_n\big(\vec{\eta}\big)\int_{\mathbb{R}^d}\sigma^2\big(x,\vec{\eta}\big)\nu_d(\mathrm{d}x)\\&=2(n\mathfrak{z}^2)^{-1}\lambda_n\big(\vec{\eta}\big)\big\|M\big(\cdot,\vec{\eta}\big)\big\|_2^2\\&\leq2M_2^2(n\mathfrak{z}^2)^{-1}\frac{\lambda_n\big(\vec{\eta}\big)}{\prod_{j=1}^d\eta_j(\eta_j\wedge1)^{2\boldsymbol{\mu}_j(\alpha)}}\leq c_6\mathfrak{z}^{-2}F_n^2\big(\vec{\eta}\big).\end{aligned} \tag{5.47}$$

Here we have put $c_6=2M_2^2c_1^2c_3$ and to get the last inequality we have used (5.40). To get the similar result if $\mathbf{s}=\mathbf{u}$ we remark that $\sigma^2\big(\cdot,\vec{\eta}\big)=M^2\big(\cdot,\vec{\eta}\big)\star\mathfrak{p}(\cdot)$ and that $M^2\big(\cdot,\vec{\eta}\big)\in\mathbb{L}_1\big(\mathbb{R}^d\big)$ in view of the second inequality in (5.10). It remains to note that $f\in\mathbb{F}_{g,\mathbf{u}}(R,D)$ implies $\mathfrak{p}\in\mathbb{B}_{\mathbf{u},d}^{(\infty)}(D)$ and to apply the inequality (5.35).

It yields together with the second inequality in (5.10) for any $\vec{\eta}\in\mathcal{H}^d$

$$\nu_d\Big(x\in\mathbb{R}^d:\,\mathcal{U}_n\big(x,\vec{\eta},f\big)\geq\mathfrak{z}\Big)\leq\big[c_6C_{\mathbf{u}}D\mathfrak{z}^{-2}F_n^2\big(\vec{\eta}\big)\big]^{\mathbf{u}}. \tag{5.48}$$

Denoting $\widetilde{C}=1$ if $\mathbf{s}=1$ and $\widetilde{C}=C_{\mathbf{u}}$ if $\mathbf{s}=\mathbf{u}$, we get from (5.47), (5.48)

$$\nu_d\bigg(x\in\mathbb{R}^d:\sup_{\vec{\eta}\in\mathcal{H}^d:\,\vec{\eta}\geq\vec{h}}\mathcal{U}_n\big(x,\vec{\eta},f\big)\geq\mathfrak{z}\bigg)\leq\big[c_6\widetilde{C}\widetilde{D}\mathfrak{z}^{-2}\big]^{\mathbf{s}}\sum_{\vec{\eta}\in\mathcal{H}^d:\vec{\eta}\geq\vec{h}}F_n^{2\mathbf{s}}\big(\vec{\eta}\big).$$

It remains to note that since $\vec{\eta},\vec{h}\in\mathcal{H}^d$ and $\vec{\eta}\geq\vec{h}$ we can write $\eta_j=e^{m_j}h_j$ with $m_j\geq0$ for any $j=1,\dots,d$. Putting $\mathbf{m}=(m_1,\dots,m_d)\in\mathbb{N}^d$ the latter result yields together with the first inequality in (5.41)

$$\sum_{\vec{\eta}\in\mathcal{H}^d:\vec{\eta}\geq\vec{h}}F_n^{2\mathbf{s}}\big(\vec{\eta}\big)\leq F_n^{2\mathbf{s}}\big(\vec{h}\big)\sum_{\mathbf{m}\in\mathbb{N}^d}\Big(1+\sum_{j=1}^d m_j\Big)^{\mathbf{s}}e^{-\mathbf{s}\sum_{j=1}^d m_j}=:c_7F_n^{2\mathbf{s}}\big(\vec{h}\big).$$

Thus, (5.46) with $c_5=c_7[c_6\widetilde{C}]^{\mathbf{s}}$ is established.

3^0. Let $c_K \geq 1$ be s.t. $\mathrm{supp}(K) \subset [-c_K, c_K]^d$. We have $\forall \vec{h} \in (0,\infty)^d$

$$\big|S_{\vec{h}}(x,f)\big| = \Big|\int_{\mathbb{R}^d} K_{\vec{h}}(t-x)f(t)\nu_d(\mathrm{d}t)\Big| \leq (2c_K)^d \|K\|_\infty^d \mathfrak{M}[|f|](x).$$

If $\vec{h} = (h,\ldots,h), h \in (0,\infty)$, the latter inequality holds with $\mathfrak{m}[|f|]$ instead of $\mathfrak{M}[|f|]$. Thus,

$$\sup_{\vec{h}\in\mathbb{H}} \big|\mathcal{B}_{\vec{h}}(x,f)\big| \leq 3(2c_K)^d \|K\|_\infty \mathfrak{M}_{\mathbb{H}}[|f|](x) + |f(x)|, \quad \forall x \in \mathbb{R}^d, \tag{5.49}$$

where we have denoted $\mathfrak{M}_{\mathbb{H}} = \mathfrak{M}$ if $\mathbb{H} = \mathcal{H}^d$ and $\mathfrak{M}_{\mathbb{H}} = \mathfrak{m}$ if $\mathbb{H} = \mathcal{H}^d_{\mathrm{isotr}}$. Moreover, we deduce from (5.49) and (5.45) putting $T_{\vec{h}}(x,f) = \mathcal{B}_{\vec{h}}(x,f) + 49U_n^*\big(\cdot,\vec{h}\big)$ that

$$\inf_{\vec{h}\in\mathbb{H}} \big|T_{\vec{h}}(x,f)\big| \leq 3(2c_K)^d \|K\|_\infty \mathfrak{M}_{\mathbb{H}}[|f|](x) + |f(x)|. \tag{5.50}$$

Proof of 5.39. Put $\mathbf{T}(x,f) = \inf_{\vec{h}\in\mathbb{H}} |T_{\vec{h}}(x,f)|$ and introduce $\mathcal{C}_v = \big\{x \in \mathbb{R}^d : \mathbf{T}(x,f) \geq v\big\}, v > 0$. For any given $\overline{\boldsymbol{v}} > 0$ one has

$$\|\mathbf{T}(\cdot,f)\|_p^p \leq p\int_0^{\overline{\boldsymbol{v}}} v^{p-1}\nu_d\big(\mathcal{C}_v\big)\mathrm{d}v + \int_{\mathcal{C}_{\overline{\boldsymbol{v}}}} |\mathbf{T}(x,f)|^p \nu_d(\mathrm{d}x). \tag{5.51}$$

Note that the second term disappears if one chooses $\overline{\boldsymbol{v}} = \infty$. Denoting $\mathcal{W}_v = \{x \in \mathbb{R}^d :\ 49U_n^*\big(x,\vec{h}\big) \geq 2^{-1}v\}$ we have for any $\vec{h} \in \mathbb{H}$ and $v > 0$

$$\nu_d\big(\mathcal{C}_v\big) \leq \nu_d\big(\mathcal{A}(\vec{h},f,v)\big) + \nu_d\big(\mathcal{W}_v(\vec{h},f)\big); \tag{5.52}$$

$$|\mathbf{T}(x,f)|^p 1_{\mathcal{C}_v}(x) \leq 2^p\big|\mathcal{B}_{\vec{h}}(x,f)\big|^p 1_{\mathcal{A}(\vec{h},f,v)} + 98^p\big|U_n^*\big(x,\vec{h}\big)\big|^p 1_{\mathcal{W}_v}(x); \tag{5.53}$$

$$\nu_d\big(\mathcal{C}_v\big) \leq \nu_d\Big(x \in \mathbb{R}^d :\ 3(2c_K)^d\|K\|_\infty \mathfrak{M}_{\mathbb{H}}[|f|](x) + |f(x)| > v\Big). \tag{5.54}$$

The last inequality follows from (5.50).

1^0. Set $\mathcal{U}_n^*\big(x,\vec{h},f\big) = \sup\limits_{\vec{\eta}\in\mathcal{H}^d:\ \vec{\eta}\geq\vec{h}} \mathcal{U}_n\big(x,\vec{\eta},f\big)$. Noting that $U_n^*\big(x,\vec{h}\big) \leq \mathcal{U}_n^*\big(x,\vec{h},f\big) + (196a)^{-1}G_n\big(\vec{h}\big)$ in view of (5.42), we get for any $\vec{h} \in \mathfrak{H}(v)$

$$\mathcal{W}_v \subseteq \Big\{x \in \mathbb{R}^d :\ 49\mathcal{U}_n^*\big(x,\vec{h}\big) \geq 4^{-1}v\Big\} := \widetilde{\mathcal{W}}_v. \tag{5.55}$$

Applying (5.46) with $\mathbf{s} = 1$ we deduce from (5.52) that

$$\nu_d\big(\mathcal{C}_v\big) \leq \nu_d\big(\mathcal{A}(\vec{h},f,v)\big) + 196^2 c_5 v^{-2} F_n^2\big(\vec{h}\big), \quad \forall \vec{h} \in \mathfrak{H}(v).$$

Since the left hand side of the latter inequality is independent of $\vec{h}$ we get

$$\nu_d\big(\mathcal{C}_v\big) \leq \max[1, 196^2 c_5]\Lambda(v,f). \tag{5.56}$$

$\mathbf{2^0}$. Let us establish the following bounds, where c_9 is given in the paragraph $\mathbf{2^0b}$. below. For any $\mathbf{u}\in[1,\infty]$ and $v>0$

$$\nu_d\big(\mathcal{C}_v\big)\leq \max[1, c_5 196^2, c_5 196^{2\mathbf{u}}D^{\mathbf{u}}a^{2\mathbf{u}}]\{\Lambda(v,f)\wedge\Lambda(v,f,\mathbf{u})\}. \tag{5.57}$$

and for any $\mathbf{u}\in(p/2,\infty]$,

$$\int_{\mathcal{C}_v}|\mathbf{T}(x,f)|^p\nu_d(\mathrm{d}x)\leq \max[2^p, 98^p c_9]\Lambda_p\big(v,f,\mathbf{u}\big),\quad \forall v>0. \tag{5.58}$$

$\mathbf{2^0a}$. Let $\mathbf{u}=\infty$. Note that minimum over z in the definition of $\Lambda(\cdot,\cdot,\infty)$ and $\Lambda_p(\cdot,\cdot,\infty)$ is obviously attained for $z=2$. Also, we remark that $\mathcal{W}_v=\emptyset$ for any $\vec{h}\in\mathfrak{H}(v,2)$ in view of (5.44). Thus, we deduce from (5.52) and (5.53), since the left hand sides of both inequalities are independent of $\vec{h}$

$$\nu_d\big(\mathcal{C}_v\big)\leq\Lambda(v,f,\infty),\qquad \int_{\mathcal{C}_v}|\mathbf{T}(x,f)|^p\nu_d(\mathrm{d}x)\leq\Lambda_p(v,f,\infty). \tag{5.59}$$

This inequality and (5.56) ensure that (5.57) and (5.58) hold if $\mathbf{u}=\infty$.

$\mathbf{2^0b}$. Let $\mathbf{u}<\infty$. Applying (5.46) with $\mathbf{s}=\mathbf{u}$, we obtain in view of (5.55)

$$\nu_d\big(\mathcal{W}_v\big)\leq c_5 196^{2\mathbf{u}}D^{\mathbf{u}}v^{-2\mathbf{u}}F_n^{2\mathbf{u}}\big(\vec{h}\big)\leq c_5 196^{2\mathbf{u}}D^{\mathbf{u}}a^{2\mathbf{u}}z^{-\mathbf{u}},\ \ \forall\vec{h}\in\mathfrak{H}(v,z).$$

It yields together with (5.52)

$$\nu_d\big(\mathcal{C}_v\big)\leq\max[1, c_5 196^{2\mathbf{u}}D^{\mathbf{u}}a^{2\mathbf{u}}]\Lambda(v,f,\mathbf{u}). \tag{5.60}$$

This inequality and (5.56) ensure that (5.57) holds if $\mathbf{u}<\infty$.

What is more, we have in view of (5.42) and (5.55) for any $\vec{h}\in\mathfrak{H}(v)$

$$\big|U_n^*\big(x,\vec{h}\big)\big|^p 1_{\mathcal{W}_v}\leq 2^p\big|\mathcal{U}_n^*\big(x,\vec{h},f\big)\big|^p 1_{\widetilde{\mathcal{W}}_v}$$

Moreover, applying (5.46) with $\mathbf{s}=\mathbf{u}$, we have for any $y>0$ and $\vec{h}\in\mathfrak{H}(v,z)$

$$\nu_d\big(\widetilde{\mathcal{W}}_y\big)\leq c_5 196^{2\mathbf{u}}D^{\mathbf{u}}y^{-2\mathbf{u}}F_n^{2\mathbf{u}}\big(\vec{h}\big)\leq c_5 196^{2\mathbf{u}}D^{\mathbf{u}}y^{-2\mathbf{u}}(av)^{2\mathbf{u}}z^{-\mathbf{u}}.$$

Hence, if additionally $\mathbf{u}>p/2$, we have for any $\vec{h}\in\mathfrak{H}(v,z)$

$$\begin{aligned}\int_{\mathcal{W}_v}\big|U_n^*\big(x,\vec{h}\big)\big|^p\nu_d(\mathrm{d}x)&\leq 2^p p\int_v^\infty y^{p-1}\nu_d\big(\widetilde{\mathcal{W}}_y\big)\mathrm{d}y\\ &= c_5 196^{2\mathbf{u}}D^{\mathbf{u}}a^{2\mathbf{u}}2^p p v^{2\mathbf{u}}z^{-\mathbf{u}}\int_v^\infty y^{p-1-2\mathbf{u}}\mathrm{d}y=:c_9 v^p z^{-\mathbf{u}}.\end{aligned}$$

This yields together with (5.53)

$$\int_{\mathcal{C}_v}|\mathbf{T}(x,f)|^p\nu_d(\mathrm{d}x)\leq\max[2^p, 98^p c_9]\Lambda_p\big(v,f,\mathbf{u}\big). \tag{5.61}$$

This inequality ensures that (5.58) holds if $\mathbf{u} < \infty$.

$3^{\mathbf{0}}$. Recall that $f \in \mathbb{F}_g(R)$ implies that $f \in \mathbb{B}_{\mathbf{1},d}(R)$. Since additionally $f \in \mathbb{B}_{\mathbf{q},d}(D)$, $\mathbf{q} > 1$, Lemma 1 as well as (5.34) is applicable and we obtain in view of (5.54) $\nu_d(\mathcal{C}_v) \leq c_{10} v^{-1}(1+|\ln v|)^{t(\mathbb{H})}$ for any $v > 0$. It yields for any $\underline{\boldsymbol{v}} > 0$ and $p > 1$

$$p\int_0^{\underline{\boldsymbol{v}}} v^{p-1}\nu_d(\mathcal{C}_v)\mathrm{d}v \leq c_{11}\underline{\boldsymbol{v}}^{p-1}(1+|\ln \underline{\boldsymbol{v}}|)^{p-1+t(\mathbb{H})}. \tag{5.62}$$

In the case of $t(\mathbb{H}) = 0$ the last inequality is obvious and if $t(\mathbb{H}) = d-1$ it follows by integration by parts. The bound (5.39) follows now from (5.51), where the bound (5.62) is used when the integration is made over $[0, \underline{\boldsymbol{v}}]$, the estimate (5.57) for integration over $[\underline{\boldsymbol{v}}, \overline{\boldsymbol{v}}]$ and the bound (5.58) with $v = \overline{\boldsymbol{v}}$.

5.2.3. **Part II.** In the subsequent proof $c_1, c_2, \ldots,$ stand for constants depending only on $\vec{s}, \vec{q}, g, \mathcal{K}, d,\ R, D, \mathbf{u}$ and $\mathbf{q}$.

$1^{\mathbf{0}}$. We start with the following obvious observation. For any $\lambda : \mathbb{R}^d \to \mathbb{R}_+$, $\vec{u} \in \mathbb{R}^d$ and $J \in \mathfrak{J}$

$$[K_{\vec{u}} \circ \lambda]_J(x) \leq (2c_{\mathcal{K}}\|\mathcal{K}\|_\infty)^d \mathfrak{M}_J[\lambda](x), \quad \forall x \in \mathbb{R}^d. \tag{5.63}$$

Putting $C_1 = (2c_{\mathcal{K}}\|\mathcal{K}\|_\infty)^d$ we get for any $\vec{h}, \vec{\eta} \in \mathcal{H}^d$ in view of (5.63) and assertions of Lemma 2 that

$$B_{\vec{h},\vec{\eta}}(\cdot, f) \leq 2C_1 \sum_{j=1}^d \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](\cdot),\ B_{\vec{h}}(\cdot, f) \leq C_1 \sum_{j=1}^d \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](\cdot).$$

Thus noting that the right hand side of the first inequality above is independent of $\vec{\eta}$, we obtain

$$\mathcal{B}_{\vec{h}}(x, f) \leq 5C_1 \sum_{j=1}^d \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](x), \quad \forall x \in \mathbb{R}^d,\ \forall \vec{h} \in \mathcal{H}^d. \tag{5.64}$$

Applying (5.38) with $\mathbf{t} = \infty$, we have for any $v > 0$ in view of the definition of $J(\vec{h}, v)$

$$\begin{aligned}
\mathcal{B}_{\vec{h}}(x, f) &\leq 5C_1\Big[\sum_{j\in\bar{J}(\vec{h},v)} \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](x) + \sum_{j\in J(\vec{h},v)} \sup_{J\in\mathfrak{J}} \big\|\mathfrak{M}_J\big[b_{h_j,f,j}\big]\big\|_\infty\Big] \\
&\leq 5C_1 \sum_{j\in\bar{J}(\vec{h},v)} \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](x) + 5C_1 \sum_{j\in J(\vec{h},v)} \mathbf{B}_{j,\infty,\mathbb{F}}\big(h_j\big) \\
&\leq 5C_1 \sum_{j\in\bar{J}(\vec{h},v)} \sup_{J\in\mathfrak{J}} \mathfrak{M}_J\big[b_{h_j,f,j}\big](x) + 4^{-1}v, \quad \forall f \in \mathbb{F}. \qquad (5.65)
\end{aligned}$$

We obtain for any $f \in \mathbb{F}$, $v > 0$ and $\vec{s} = (s_1, \dots, s_d) \in (1, \infty)^d$, applying consecutively the Markov inequality and (5.38) with $\mathbf{t} = s_j$,

$$\nu_d \Big\{ \mathcal{A}\big(\vec{h}, f, v\big) \Big\} \leq \nu_d \Big(\cup_{J \in \mathfrak{J}} \cup_{j \in \bar{J}(\vec{h}, v)} \Big\{ x : 5C_1 \mathfrak{M}_J \big[b_{h_j, f, j} \big](x) \geq (4d)^{-1} v \Big\} \Big)$$

$$\leq c_1 \sum_{j \in \bar{J}(\vec{h}, v)} v^{-s_j} \big\| b_{h_j, f, j} \big\|_{s_j}^{s_j} \leq c_1 \sum_{j \in \bar{J}(\vec{h}, v)} v^{-s_j} \Big[\mathbf{B}_{j, s_j, \mathbb{F}} \big(h_j \big) \Big]^{s_j}. \tag{5.66}$$

Noting that the right hand side of the latter inequality is independent of f and the left hand side is independent of $\vec{s}$, we get for any $v > 0$ and $\vec{s} \in (1, \infty)^d$

$$c_1^{-1} \sup_{f \in \mathbb{F}} \{ \Lambda(v, f) \wedge \Lambda(v, f, \mathbf{u}) \} \leq \mathbf{\Lambda}_{\vec{s}}(v, \mathbb{F}, \mathbf{u}) \wedge \mathbf{\Lambda}_{\vec{s}}(v, \mathbb{F}). \tag{5.67}$$

$\mathbf{2^0}$. Note also that in view of (5.2.3), we have for any $v > 0$

$$\int_{\mathcal{A}(\vec{h}, f, v)} \big| \mathcal{B}_{\vec{h}}(x, f) \big|^p \nu_d(\mathrm{d}x) \tag{5.68}$$

$$\leq c_2 \int_{\mathcal{A}(\vec{h}, f, v)} \Big| \sum_{j \in \bar{J}(\vec{h}, v)} \sup_{J \in \mathfrak{J}} \mathfrak{M}_J \big[b_{h_j, f, j} \big](x) \Big|^p \nu_d(\mathrm{d}x) + c_3 v^p \nu_d \Big\{ \mathcal{A}\big(\vec{h}, f, v\big) \Big\}$$

$$\leq c_4 \Bigg[\sum_{j \in \bar{J}(\vec{h}, v)} \int_{\mathcal{A}(\vec{h}, f, v)} \Big| \sup_{J \in \mathfrak{J}} \mathfrak{M}_J \big[b_{h_j, f, j} \big](x) \Big|^p \nu_d(\mathrm{d}x) + v^p \nu_d \Big\{ \mathcal{A}\big(\vec{h}, f, v\big) \Big\} \Bigg].$$

For any $v > 0$ and $j = 1, \dots, d$, introduce $\mathcal{A}_j(v) = \mathcal{A}\big(\vec{h}, f, v\big) \cap \bar{\mathfrak{A}}_j(v)$, where

$$\mathfrak{A}_j(v) = \Big\{ x \in \mathbb{R}^d : \ \sup_{J \in \mathfrak{J}} \mathfrak{M}_J \big[b_{h_j, f, j} \big](x) \geq (40C_1)^{-1} v \Big\}.$$

Noting that in view of (5.2.3) for any $v > 0$ and any $j \in \bar{J}(\vec{h}, v)$

$$\mathcal{A}_j(v) \subseteq \Big\{ x \in \mathbb{R}^d : \ 5C_1 \sum_{k \in \bar{J}(\vec{h}, v),\ k \neq j} \sup_{J \in \mathfrak{J}} \mathfrak{M}_J \big[b_{h_j, f, k} \big](x) \geq v/8 \Big\}$$

$$\subseteq \Big\{ x \in \mathbb{R}^d : \ 5C_1 \sum_{k \in \bar{J}(\vec{h}, v)} \sup_{J \in \mathfrak{J}} \mathfrak{M}_J \big[b_{h_j, f, k} \big](x) \geq v/8 \Big\} =: \mathcal{A}^*\big(\vec{h}, f, v\big),$$

we deduce from (5.68) that for any $\vec{q}\in[p,\infty)^d$

$$\begin{aligned}\int_{\mathcal{A}\left(\vec{h},f,v\right)}\left|\mathcal{B}_{\vec{h}}(x,f)\right|^p\nu_d(\mathrm{d}x)&\leq c_4\sum_{j\in\bar{J}(\vec{h},v)}\int_{\mathfrak{A}_j(v)}\Big|\sup_{J\in\mathfrak{J}}\mathfrak{M}_J\big[b_{h_j,f,j}\big](x)\Big|^p\nu_d(\mathrm{d}x)\\&\quad+c_5v^p\Big[\nu_d\Big\{\mathcal{A}^*\big(\vec{h},f,v\big)\Big\}+\nu_d\Big\{\mathcal{A}\big(\vec{h},f,v\big)\Big\}\Big]\\&\leq c_6\sum_{j\in\bar{J}(\vec{h},v)}v^{p-q_j}\Big\|\sup_{J\in\mathfrak{J}}\mathfrak{M}_J\big[b_{h_j,f,j}\big]\Big\|_{q_j}^{q_j}\\&\quad+c_5v^p\Big[\nu_d\Big\{\mathcal{A}^*\big(\vec{h},f,v\big)\Big\}+\nu_d\Big\{\mathcal{A}\big(\vec{h},f,v\big)\Big\}\Big].\end{aligned}\tag{5.69}$$

It remains to note that similarly (5.2.3) for any $\vec{s}\in(1,\infty)^d$

$$\nu_d\Big\{\mathcal{A}^*\big(\vec{h},f,v\big)\Big\}\leq c_7\sum_{j\in\bar{J}}(\vec{h},v)v^{-s_j}\big\|b_{h_j,f,j}\big\|_{s_j}^{s_j}$$

and to apply (5.38) with $\mathbf{t}=q_j$ to the each term in the sum appeared in (5.69). All of this together with (5.2.3), applied with $\vec{s}=\vec{q}$ yields for any $v>0$ and $\vec{q}\in[p,\infty)^d$

$$\int_{\mathcal{A}\left(\vec{h},f,v\right)}\left|\mathcal{B}_{\vec{h}}(x,f)\right|^p\nu_d(\mathrm{d}x)\leq c_9\sum_{j\in\bar{J}(\vec{h},v)}v^{p-q_j}\Big[\mathbf{B}_{j,q_j,\mathbb{F}}\big(h_j\big)\Big]^{q_j}.$$

Noting that the right hand side of the latter inequality is independent of f and the left hand side is independent of $\vec{q}$, the we get

$$\sup_{f\in\mathbb{F}}\Lambda_p(v,f,\mathbf{u})\leq c_9v^p\mathbf{\Lambda}_{\vec{q}}(v,\mathbb{F},\mathbf{u}),\quad\forall v>0,\;\;\vec{q}\in[p,\infty)^d.\tag{5.70}$$

The first assertion follows from (5.67), (5.70) and (5.39).

$\mathbf{3^0}$. Remark that in view of (5.49) and (5.33) $f\in\mathbb{B}_{\mathbf{q},d}(D)$ implies

$$\big\|\mathcal{B}_{\vec{h}}(\cdot,f)\big\|_{\mathbf{q}}\leq\big[3(2c_{\mathcal{K}})^d\|\mathcal{K}\|_\infty^dC_{\mathbf{q}}+1\big]D,\quad\forall\vec{h}\in(0,\infty)^d,\tag{5.71}$$

where $C_{\mathbf{q}}$ is the constant which appeared in (5.33). Hence for any $v>0$ and $\mathbf{q}\in[p,\infty)$

$$\int_{\mathcal{A}\left(\vec{h},f,v\right)}\left|\mathcal{B}_{\vec{h}}(x,f)\right|^p\nu_d(\mathrm{d}x)\leq 2^{\mathbf{q}-p}v^{p-\mathbf{q}}\big\|\mathcal{B}_{\vec{h}}(\cdot,f)\big\|_{\mathbf{q}}^{\mathbf{q}}\leq c_{10}v^{p-\mathbf{q}}.\tag{5.72}$$

Remind that $\mathfrak{H}(v)\neq\emptyset$, $\mathfrak{H}(v,z)\neq\emptyset$ whatever $v>0$ and $z\geq 2$, see Remark 2. Hence, in view of (5.72) for any f

$$\Lambda_p(v,f,\mathbf{u})\leq\inf_{z\geq 2}\big[c_{10}v^{p-\mathbf{q}}+z^{-\mathbf{u}}\big]=c_{10}v^{p-\mathbf{q}}.$$

It remains to note that the right hand side of the obtained inequality is independent of f and the second assertion of the theorem follows from this inequality, (5.67) and (5.39).

$4^{\mathbf{0}}$. Since $C_\infty = 1$ we obtain in view of (5.71) for all $f \in \mathbb{B}_{\infty,d}(D)$

$$\big\|\mathcal{B}_{\vec{h}}(\cdot,f)\big\|_\infty \le \big[3(2c_{\mathcal{K}})^d\|\mathcal{K}\|_\infty^d + 1\big]D, \quad \forall \vec{h} \in (0,\infty)^d.$$

It yields for any $\vec{s} \in (1,\infty)$ in view of (5.2.3) if $\mathbf{q} = \infty$

$$\int_{\mathcal{A}(\vec{h},f,v)} \big|\mathcal{B}_{\vec{h}}(x,f)\big|^p \nu_d(\mathrm{d}x) \le c_{12} \sum_{j\in \bar{J}(\vec{h},v)} v^{-s_j}\Big[\mathbf{B}_{j,\mathbf{s},\mathbb{F}}\big(h_j\big)\Big]^{s_j}.$$

Since the left hand side of the obtained inequality is independent of f and the left hand side is independent of $\vec{s}$ we conclude that

$$\sup_{f\in\mathbb{F}} \Lambda_p(v,f,\mathbf{u}) \le c_{12}\mathbf{\Lambda}_{\vec{s}}(v,\mathbb{F},\mathbf{u}), \quad \forall v>0,\ \vec{s} \in (1,\infty)^d. \tag{5.73}$$

The third assertion of the theorem follows now from (5.67), (5.73) and (5.39).

$5^{\mathbf{0}}$. We have seen (Corollary 1), that $B^*_{\vec{h}}(\cdot,f) \le 2\sup_{\eta\in\mathcal{H}:\eta\le h} B_{\vec{\eta}}(\cdot,f)$ if $\vec{h} = (h,\dots,h) \in \mathcal{H}^d_{\text{isotr}}$. Therefore by definition of $\mathcal{B}_{\vec{h}}(\cdot,f)$:

$$\mathcal{B}_{\vec{h}}(\cdot,f) \le 5 \sup_{\eta\in\mathcal{H}:\eta\le h} B_{\vec{\eta}}(\cdot,f) \le 5 \sup_{\eta\in\mathcal{H}:\eta\le h} \sum_{j=1}^d \sup_{J\in\mathfrak{J}} \big[\big|K_{\vec{\eta}}\big| \circ b^*_{\eta,f,j}\big]_J(x). \tag{5.74}$$

where, remind $\vec{\eta} = (\eta,\dots,\eta) \in \mathcal{H}^d_{\text{isotr}}$. We remark that (5.74) is similar to (5.64) but the maximal operator is not involved in this bound. This, in its turn, allows to consider $\vec{s} \in [1,\infty)^d$. Indeed, similarly to (5.2.3) we have for any $v>0$, applying (5.36) with $\mathbf{t} = \infty$

$$\mathcal{B}_{\vec{h}}(x,f) \le 5 \sup_{\eta\in\mathcal{H}:\eta\le h} \sum_{j\in\bar{J}(\vec{h},v)} \sup_{J\in\mathfrak{J}} \big[\big|K_{\vec{\eta}}\big| \circ b^*_{\eta,f,j}\big]_J(x) + 4^{-1}v,\ \forall f \in \mathbb{F}. \tag{5.75}$$

We obtain for any $f \in \mathbb{F}$, $v>0$ and $\vec{s} = (s_1,\dots,s_d) \in [1,\infty)^d$ applying consecutively the Markov inequality and (5.36) with $\mathbf{t} = s_j$

$$\nu_d\Big(\mathcal{A}\big(\vec{h},f,v\big)\Big) \le c_{14} \textstyle\sum_{j\in\bar{J}(\boldsymbol{h},\boldsymbol{y})} v^{-s_j}\Big[\mathbf{B}^*_{j,s_j,\mathbb{F}}\big(h\big)\Big]^{s_j}.$$

We note that the obtained inequality coincides with (5.2.3) if one replaces $\mathbf{B}_{j,s_j,\mathbb{F}}(\cdot)$ by $\mathbf{B}^*_{j,s_j,\mathbb{F}}(\cdot)$. It remains to note that $\mathbf{B}_{j,s_j,\mathbb{F}}(\cdot) \le \mathbf{B}^*_{j,s_j,\mathbb{F}}(\cdot)$. Indeed,

$$b_{\mathbf{v},f,j}(x) = \lim_{k\to\infty} \sup_{h\in\mathcal{H}:\, e^{-k}\le h\le \mathbf{v}} b^*_{h,f,j}(x).$$

Therefore, by the monotone convergence theorem and the triangle inequality for any $s\in[1,\infty)$

$$
\begin{aligned}
\mathbf{B}_{j,s,\mathbb{F}}(\mathbf{h}) &:= \sup_{f\in\mathbb{F}}\|b_{\mathbf{v},f,j}\|_s = \sup_{f\in\mathbb{F}}\lim_{k\to\infty}\Big\|\sup_{h\in\mathcal{H}:\, e^{-k}\le h\le\mathbf{h}} b^*_{h,f,j}\Big\|_s\\
&\le \sup_{f\in\mathbb{F}}\lim_{k\to\infty}\sum_{h\in\mathcal{H}:\, e^{-k}\le h\le\mathbf{h}}\Big\|b^*_{h,f,j}\Big\|_s = \sup_{f\in\mathbb{F}}\sum_{h\in\mathcal{H}:\, h\le\mathbf{h}}\Big\|b^*_{h,f,j}\Big\|_s =: \mathbf{B}^*_{j,s,\mathbb{F}}(\mathbf{v}).
\end{aligned}
$$

The fourth statement of the theorem follows now from (5.67), (5.70), (5.72) and (5.39). ∎

6. Proof of Theorems 3 and 4.

In the subsequent proof $\mathbf{c},\mathbf{c}_1,C,C_1,..$ stand for constants that can depend on g,L_0,L_∞, Q,R, $\vec{\beta}$, $\vec{r}$, d and p, but are independent of $\vec{L}$ and n. These constants can be different on different appearances. The proofs are based on the application of Theorem 3 and on some auxiliary assertions presented below.

Let $\mathcal{J}_\infty=\{j=1,\dots,d:\ r_j=\infty\}$ and put $p_\pm=[\sup_{j\in\bar{\mathcal{J}}_\infty} r_j]\vee p$, where $\bar{\mathcal{J}}_\infty$ is complimentary to $\mathcal{J}_\infty$. Introduce

$$
q_j=\begin{cases} p_\pm, & j\in\bar{\mathcal{J}}_\infty,\\ \infty, & j\in\mathcal{J}_\infty,\end{cases}\qquad \gamma_j=\begin{cases}\frac{\beta_j\tau(p_\pm)}{\tau(r_j)}, & j\in\bar{\mathcal{J}}_\infty,\\ \beta_j, & j\in\mathcal{J}_\infty.\end{cases}\tag{6.1}
$$

Proposition 2. *Let $\ell\in\mathbb{N}^*$, $p>1$ and K satisfying Assumption 4 be fixed. Then for any $\vec{\beta}\in(0,\ell]^d$, $\vec{r}\in[1,\infty]^d$ and $\vec{L}\in(0,\infty)^d$ one can find $C_1>0$ independent of $\vec{L}$ such that $\forall\vec{h}\in\mathcal{H}^d$*

$$
\mathbf{B}_{j,r_j,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\big(h_j\big)\le C_1L_jh_j^{\beta_j},\quad j=1,\dots,d.\tag{6.2}
$$

If additionally $\tau(p^*)>0$ then

$$
\mathbf{B}_{j,q_j,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\big(h_j\big)\le C_1L_jh_j^{\gamma_j},\quad j=1,\dots,d,\tag{6.3}
$$

At last, (6.2) and (6.3) remain true if one replaces the quantity $\mathbf{B}$ by $\mathbf{B}^*$.

The proof of the proposition as well as the proofs of technical Lemmas 3, 4 and 5 are put into Appendix.

The bandwidth's construction presented below as well as auxiliary statements from the next section will be exploited not only for proving Theorems 3 and 4, but also in the consideration forming Part II of this work. By this reason we formulate them in a bit more general form than what is needed for our current purposes. Recall that $\mathbf{c}=\big(20d\big)^{-1}\big[\max(2c_{\mathcal{K}_\ell}\|\mathcal{K}_\ell\|_\infty,\|\mathcal{K}_\ell\|_1)\big]^{-d}$ and let $\boldsymbol{L}>0$ be any number satisfying

$$
\boldsymbol{L}\le 1\wedge(C_1^{-1}\mathbf{c})\wedge L_0.\tag{6.4}
$$

6.1. *Special set of bandwidths.* Set for any $r,\mathbf{s}\in[1,\infty]$

$$\varkappa_\alpha(r,\mathbf{s})=\tfrac{\mathbf{s}\omega(\alpha)(2+1/\beta(\alpha))}{(\mathbf{s}+\omega(\alpha))}-r,\quad \alpha\in[0,1].$$

Recall that $\delta_n=L(\alpha)n^{-1}\ln n$ and introduce for any $v>0$, $\mathbf{s}\in[1,\infty]$

$$\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{s})=\big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1}{\beta_j}}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{s}\omega(\alpha)}{(\mathbf{s}+\omega(\alpha))\beta_j r_j}}v^{\frac{1}{\beta_j}-\frac{\mathbf{s}\omega(\alpha)(2+1/\beta(\alpha))}{(\mathbf{s}+\omega(\alpha))\beta_j r_j}};\tag{6.5}$$

$$\widehat{\boldsymbol{\eta}}_j(v,\mathbf{s})=\big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1}{\gamma_j}}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{s}\upsilon(\alpha)}{(\mathbf{s}+\upsilon(\alpha))\gamma_j q_j}}v^{\frac{1}{\gamma_j}-\frac{\mathbf{s}\upsilon(\alpha)(2+1/\gamma(\alpha))}{(\mathbf{s}+\upsilon(\alpha))\gamma_j q_j}},\tag{6.6}$$

where $\frac{1}{\gamma(\alpha)}:=\sum_{j=1}^d\frac{2\boldsymbol{\mu}_j(\alpha)+1}{\gamma_j}$, $\quad\frac{1}{\upsilon(\alpha)}:=\sum_{j=1}^d\frac{2\boldsymbol{\mu}_j(\alpha)+1}{\gamma_j q_j}$.

The constant $\mathfrak{a}>0$ will be chosen differently in accordance with some special relationships between the parameters $\vec{\beta}$, $\vec{r}$, $\vec{\mu}$, α and p.

Determine $\boldsymbol{h}_j(\cdot,\mathbf{s})$ and $\mathfrak{h}_j(\cdot,\mathbf{s})$, $j=1,\dots,d$, from the relations

$$\boldsymbol{h}_j(v,\mathbf{s})=\max\big\{h\in\mathcal{H}:\ h\le\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{s})\big\},\quad v>0;\tag{6.7}$$

$$\mathfrak{h}_j(v,\mathbf{s})=\max\big\{h\in\mathcal{H}:\ h\le\widehat{\boldsymbol{\eta}}_j(v,\mathbf{s})\big\},\quad v>0,\tag{6.8}$$

and set $\vec{\boldsymbol{h}}(\cdot,\mathbf{s})=\big(\boldsymbol{h}_1(\cdot,\mathbf{s}),\dots,\boldsymbol{h}_d(\cdot,\mathbf{s})\big)$ and $\vec{\mathfrak{h}}(\cdot,\mathbf{s})=\big(\mathfrak{h}_1(\cdot,\mathbf{s}),\dots,\mathfrak{h}_d(\cdot,\mathbf{s})\big)$.

6.2. *Auxiliary statements.* All the results formulated below are proved in Section 7. Let

$$\mathfrak{z}(v)=2\big(\mathfrak{a}^{-2}\delta_n\big)^{-\frac{\omega(\alpha)}{\omega(\alpha)+\mathbf{u}}}v^{\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\omega(\alpha)+\mathbf{u}}},\quad \mathbf{u}\in[1,\infty],$$

and remark that $\mathfrak{z}(\cdot)\equiv 2$ if $\mathbf{u}=\infty$. Note also that

$$\mathfrak{z}(v)\ge 2,\quad \forall v\ge\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{1}{2+1/\beta(\alpha)}}=:\boldsymbol{v}.\tag{6.9}$$

Introduce the following notations: $\mu(\alpha)=\min_{j=1,\dots,d}\mu_j(\alpha)$,

$$X=\tfrac{1}{2\beta(\alpha)}-\tfrac{1}{2\beta(0)}=\textstyle\sum_{j=1}^d\frac{\boldsymbol{\mu}_j(\alpha)}{\beta_j},\quad Y=\tfrac{1}{2\omega(\alpha)}-\tfrac{1}{2\omega(0)}=\textstyle\sum_{j=1}^d\frac{\boldsymbol{\mu}_j(\alpha)}{\beta_j r_j}.$$

Recall that $z(\alpha)=\omega(\alpha)(2+1/\beta(\alpha))\beta(0)\tau(\infty)+1$ and define

$$\underline{\mathbf{v}}=(\mathfrak{a}^{-2}\delta_n)^{\frac{1}{1-1/\omega(\alpha)+1/\beta(\alpha)}},\quad \mathbf{v}=\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)\tau(\infty)\beta(0)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}}.\tag{6.10}$$

Set $\mathbf{u}^*=[-\tau(\infty)\beta(0)]^{-1}$ if $\tau(\infty)<0$ and let $\mathbf{u}^*=\infty$ if $\tau(\infty)\ge 0$. Put finally $\mathbf{y}=\mathbf{u}^*\vee p^*$ and $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)=Y-[X+1]\mathbf{y}^{-1}+1/\mathbf{u}$.

Proposition 3. *Let $\vec{\beta}$, $\vec{r}$, L_0,L_∞, $\vec{\mu}$, α and p be given. Assume that $\vec{L}\in[L_0,L_\infty]^d$. Then,*
1) there exists $\mathfrak{a}>0$ independent of $\vec{L}$ such that for all n large enough

$$\vec{\boldsymbol{h}}(v,\mathbf{1})\in\mathfrak{H}(v),\quad \forall v\in[\underline{\mathbf{v}},1],$$

2) there exists $\mathfrak{a}>0$ independent of $\vec{L}$ and $\mathbf{u}>1$ such that $\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\in\mathfrak{H}\big(\mathbf{v},\mathfrak{z}(\mathbf{v})\big)$ for all large n if either $\tau(\infty)\ge 0$ or $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)\ge 0$, $\tau(p^*)\ge 0$.

Recall that $\boldsymbol{v} \to 0, n \to \infty$, is defined in (6.9) and introduce

$$\mathbf{v_1} = \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{1}{1-\mathbf{u}/\omega(0)+1/\beta(0)}}, \quad \mathbf{v_3} = \big(\mathfrak{a}^{-2}\delta_n\big)^{-\frac{Y+1/\mathbf{u}}{\pi(\mathbf{u})\vee 0}}, \tag{6.11}$$

where $\pi(\mathbf{u}) = [1/\omega(0) - 1/\mathbf{u}][1+X] - 1/\beta(0)[Y+1/\mathbf{u}]$. Define also

$$\overline{\mathbf{v}} = \mathbf{v}1_{\{\tau(p^*)>0\}} + \mathbf{v_2}1_{\{\tau(p^*)\le 0\}}, \quad \mathbf{v_2} = \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\mathbf{u}\omega(1)}{\varkappa_1(p^*,\mathbf{u})(\omega(1)+\mathbf{u})}}. \tag{6.12}$$

Note that $\mathbf{v_1} \to \infty, n \to \infty$, if $\infty > \mathbf{u} \ge \mathbf{u}^* \vee p^*$ (it will be proved in Proposition 4 below). However $\mathbf{v_1} = 1$ if $\mathbf{u} = \infty$. As it is shown in the proof of Proposition 3, formulae (7.12), $\boldsymbol{v} < \mathbf{v}$ for all n large enough. Also $\mathbf{v_2} \to \infty, n \to \infty$, if $\varkappa_1(p^*, \mathbf{u}) < 0$. At last $\mathbf{v_3} \to \infty, n \to \infty$, since $\omega(0) > \omega(1)$. Moreover $\mathbf{v_3} = \infty$ if $\pi(\mathbf{u}) \le 0$. Introduce finally

$$\mathcal{I}_{\mathbf{u}}(\alpha) = \begin{cases} [\boldsymbol{v}, 1], & p^* = \infty \\ [\boldsymbol{v}, \mathbf{v_1}], & \alpha \ne 1,\ p^* < \infty; \\ [\boldsymbol{v}, \overline{\mathbf{v}}], & \alpha = 1,\ p^* < \infty,\ \ \mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) \ge 0; \\ [\boldsymbol{v}, \mathbf{v_3}], & \alpha = 1,\ p^* < \infty,\ \ \mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) < 0. \end{cases}$$

PROPOSITION 4. *Let $\vec{\beta}$, $\vec{r}$, L_0, L_∞, $\vec{\mu}$, α and p be given and let $\vec{L} \in [L_0, L_\infty]^d$, $\mathbf{u} \in [\mathbf{y}, \infty]$. Then, there exists $\mathfrak{a} > 0$ independent of $\vec{L}$ and $\mathbf{u}$ such that for all n large enough $\vec{\boldsymbol{h}}(v, \mathbf{u}) \in \mathfrak{H}\big(v, \mathfrak{z}(v)\big),\ v \in \mathcal{I}_{\mathbf{u}}(\alpha)$.*

In the current paper we will use the statements of Proposition 3 and 4 only with $\mathbf{u} = \infty$. In this context we remark that $\varkappa_\alpha(\cdot) \equiv \varkappa_\alpha(\cdot, \infty)$.

We finish this section with the following observations which will be useful in the sequel.

LEMMA 3. *For any $\mathbf{u} \in (1, \infty]$ and $\alpha \in [0, 1]$*

$$\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) \ge 0,\ \tau(p^*) \ge 0,\ \Rightarrow\ z(\alpha)/\omega(\alpha) - 1 + 2/\mathbf{u} \ge 0; \tag{6.13}$$

$$\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) \ge 0,\ \tau(p^*) \le 0,\ \Rightarrow\ \varkappa_\alpha(p^*, \mathbf{u}) < 0; \tag{6.14}$$

$$\varkappa_\alpha(p^*, \mathbf{u}) \le 0,\ \tau(p^*) > 0,\ \Rightarrow\ z(\alpha) + \omega(\alpha)/\mathbf{u} > 0; \tag{6.15}$$

$$\tau(\infty) \ge 0 \text{ or } \mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) \ge 0, \tau(p^*) \ge 0,\ \Rightarrow\ z(\alpha) + \omega(\alpha)/\mathbf{u} > 0; \tag{6.16}$$

$$\mu(\alpha) + 1/\mathbf{u} - 1/\mathbf{y} \ge 0 \quad \Rightarrow \quad \mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha) \ge 0. \tag{6.17}$$

Set $r(\alpha) = \frac{1-1/p}{1-1/\omega(\alpha)+1/\beta(\alpha)} \wedge \frac{\beta(\alpha)}{2\beta(\alpha)+1}$. If $\varkappa_\alpha(p^*) \ge 0$ one has

$$\varrho(\alpha) = r(\alpha), \quad \rho(\alpha) = r(\alpha) \wedge [\omega(\alpha)/p]. \tag{6.18}$$

If $\varkappa_\alpha(p^*) < 0$ one has

$$(6.19)\ \varrho(\alpha) = r(\alpha) \wedge \left[\frac{\tau(p)\omega(\alpha)\beta(0)}{z(\alpha)} 1_{\{\tau(p^*)>0\}} + \frac{\omega(\alpha)(1-p^*/p)}{\varkappa_\alpha(p^*)} 1_{\{\tau(p^*)\le 0\}}\right];$$

$$(6.20)\ \rho(\alpha) = r(\alpha) \wedge \left[\frac{\tau(p)\omega(\alpha)\beta(0)}{z(\alpha)} 1_{\{\tau(\infty)>0\}} + \frac{\omega(\alpha)}{p} 1_{\{\tau(\infty)\le 0\}}\right].$$

6.3. *Several bounds.* Let us collect some bounds for several terms appearing in Theorem 2 and used in the proofs of Theorems 3 and 4 simultaneously.

LEMMA 4. *For any* $v \in [\underline{\mathbf{v}}, \boldsymbol{v}] \cup \mathcal{I}_\infty(\alpha)$

$$\begin{aligned} \mathbf{\Lambda}_{\vec{r}}\big(v, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \infty\big) &\le \mathbf{C}_1 \delta_n^{\omega(\alpha)} v^{-\omega(\alpha)(2+1/\beta(\alpha))}, \quad \forall v \in \mathcal{I}_\infty(\alpha); \\ \mathbf{\Lambda}_{\vec{r}}\big(v, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})\big) &\le \mathbf{C}_1 \delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}} v^{-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\omega(\alpha)+1}}, \quad \forall v \in [\underline{\mathbf{v}}, \boldsymbol{v}]. \end{aligned}$$

If additionally $\tau(p^*) > 0$ *the following inequality with* $\vec{q}$ *defined in (6.1) holds*

$$\mathbf{v}^p \mathbf{\Lambda}_{\vec{q}}\big(\mathbf{v}, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \infty\big) \le \mathbf{C}_2 \delta_n^{\frac{\omega(\alpha)\tau(p)\beta(0)}{z(\alpha)}}.$$

$\mathbf{1^0}$. In view of the first and the second bounds from Lemma 4 and the definitions of $\underline{\mathbf{v}}$ and $\boldsymbol{v}$, choosing $\underline{\boldsymbol{v}} = \underline{\mathbf{v}}$, we get

$$(6.21)\quad \int_{\underline{\boldsymbol{v}}}^{\overline{\boldsymbol{v}}} v^{p-1}\big[\mathbf{\Lambda}_{\vec{r}}\big(v, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \infty\big) \wedge \mathbf{\Lambda}_{\vec{r}}\big(v, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})\big)\big]\mathrm{d}v$$
$$\le c_4\Big[\delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}} \underline{\mathbf{v}}^{p-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\omega(\alpha)+1}} 1_{\{\varkappa_\alpha(p)>p\omega(\alpha)\}}$$
$$+\delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}} \boldsymbol{v}^{p-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\omega(\alpha)+1}} 1_{\{\varkappa_\alpha(p)<p\omega(\alpha)\}} + \delta_n^{\omega(\alpha)} \boldsymbol{v}^{p-\omega(\alpha)(2+1/\beta(\alpha))} 1_{\{\varkappa_\alpha(p)>0\}}$$
$$+\delta_n^{\omega(\alpha)} \overline{\boldsymbol{v}}^{p-\omega(\alpha)(2+1/\beta(\alpha))} 1_{\{\varkappa_\alpha(p)<0\}} + \ln(n)\Big(\delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}} 1_{\{\varkappa_\alpha(p)=p\omega(\alpha)\}}$$
$$+\delta_n^{\omega(\alpha)} 1_{\{\varkappa_\alpha(p)=0\}}\Big)\Big] =: A_n + c_2 \delta_n^{\omega(\alpha)} \overline{\boldsymbol{v}}^{p-\omega(\alpha)(2+1/\beta(\alpha))} 1_{\{\varkappa_\alpha(p)<0\}}.$$

After elementary computations and taking into account (6.18), we obtain

$$(6.22)\qquad A_n \le c_5 \mathfrak{b}_n^p(\mathbb{H}) \delta_n^{p\rho(\alpha)}, \qquad A_n \le c_5 \mathfrak{b}_n^p(\mathbb{H}) \delta_n^{p\varrho(\alpha)}.$$

These bounds are not surprising because $\varrho(\alpha) = \rho(\alpha)$ if $\varkappa_\alpha(p) \ge 0$.

$\mathbf{2^0}$. Choosing $\underline{\boldsymbol{v}} = \underline{\mathbf{v}}$, we obtain $\ell_{\mathbb{H}}(\underline{\boldsymbol{v}}) \le c_6 \delta_n^{\frac{p-1}{1-1/\omega(\alpha)+1/\beta(\alpha)}} \big(\ln n\big)^{t(\mathbb{H})}$, which yields by (6.18), (6.19) and (6.20):

$$(6.23)\qquad \ell_{\mathbb{H}}(\underline{\boldsymbol{v}}) \le c_1 \mathfrak{b}_n^p(\mathbb{H}) \delta_n^{p\rho(\alpha)}, \quad \ell_{\mathbb{H}}(\underline{\boldsymbol{v}}) \le c_1 \mathfrak{b}_n^p(\mathbb{H}) \delta_n^{p\varrho(\alpha)}.$$

6.4. *Proof of Theorem 3.* Furthermore $\mathcal{F} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{F}_g(R) \cap \mathbb{B}_{\infty,d}(Q)$. Since $\mathbb{F}_g(R) \cap \mathbb{B}_{\infty,d}(Q) \subset \mathbb{F}_{g,\infty}(R, D)$ with $D = Q[1-\alpha+\alpha\|g\|_1]$, Theorem 2 with $\mathbf{u} = \infty$, $\mathbf{q} = \infty$, $D = Q[1-\alpha+\alpha\|g\|_1] \vee Q$ is applicable with $\mathbb{F} = \mathcal{F}$.

$\mathbf{1^0}$. Consider the case $\tau(\infty) \le 0$. Choose $\overline{\boldsymbol{v}} = 1$ and remark that the statements of Propositions 3 and 4 hold for any $v \in [\underline{\boldsymbol{v}}, \overline{\boldsymbol{v}}]$. Indeed, it suffices to note that $\mathcal{I}_\infty(\alpha) \supseteq [\underline{\boldsymbol{v}}, \overline{\boldsymbol{v}}] := [\underline{\mathbf{v}}, 1]$, because $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3} > 1$ and $\overline{\mathbf{v}} \ge 1$ if $\tau(\infty) \le 0, \tau(p^*) > 0, \mathbf{Z}_{\mathbf{y},\infty}(\alpha) \ge 0$ since in this case $\mathbf{v} \ge 1$ by (6.16). Then we can apply all the bounds obtained above, in particular we get from the first inequality of Lemma 4

$$\boldsymbol{\Lambda}_{\vec{r}}\big(1, \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \infty\big) \le C_1 \boldsymbol{\lambda}_1(1) \le c_6 \delta_n^{\omega(\alpha)} \le c_6 \mathfrak{b}_n^p(\mathbb{H}) \delta_n^{p\rho(\alpha)}, \tag{6.24}$$

since $\omega(\alpha) \ge p\rho(\alpha)$ in both considered cases in view of the second equality in (6.18) and of (6.20). Applying the third assertion of Theorem 2, we obtain from (6.21), (6.22), (6.24) and (6.23)

$$\sup_{f\in\mathcal{F}} \mathcal{R}^{(p)}_n[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}, f] \le C\Big[(c_1+c_4+c_5+c_6)\mathfrak{b}_n^p(\mathbb{H})\delta_n^{p\rho(\alpha)}\Big]^{\frac{1}{p}} \le c_7 \mathfrak{b}_n(\mathbb{H})\delta_n^{\rho(\alpha)},$$

and the assertion of Theorem 3 follows in both considered cases.

$\mathbf{2^0}$. Consider the case $\tau(\infty) > 0$.

Choose $\overline{\boldsymbol{v}} = \mathbf{v}$ and remark that the statements of Propositions 3 and 4 hold for any $v \in [\underline{\boldsymbol{v}}, \overline{\boldsymbol{v}}]$. Indeed, $\tau(\infty) > 0$ implies $\mathbf{v} < 1$ in view of (6.16) and, therefore, $[\underline{\mathbf{v}}, \mathbf{v}] \subseteq \mathcal{I}_\infty(\alpha)$. We deduce from (6.21), (6.22), third bound in Lemma 4 and (6.23), applying the first assertion of Theorem 2 that

$$\sup_{f\in\mathcal{F}} \mathcal{R}^{(p)}_n[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}, f] \le C\Big[c_8 \delta_n^{\frac{\omega(\alpha)\tau(p)\beta(0)}{z(\alpha)}} + (c_3+c_6)\mathfrak{b}_n^p(\mathbb{H})\delta_n^{p\rho(\alpha)}\Big]^{\frac{1}{p}} \tag{6.25}$$

This completes the proof of Theorem 3 in view of (6.20).

6.5. *Proof of Theorem 4.* In the following we assume $p^* < \infty$, since $p^* = \infty$ implies by definition of the anisotropic Nikol'skii class that $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \subset \mathbb{B}_{\infty,d}(L_\infty)$. Hence, the results in that case follow from Theorem 3 since $\varrho(\alpha) = \rho(\alpha)$ when $p^* = \infty$.

Moreover, we remark that the imposed condition $p > [\min_{j=1,\dots} \mu_j]^{-1}$ implies $\mathbf{Z}_{\mathbf{y},\infty}(\alpha) \ge 0$ in view of (6.17) proved in Lemma 3. This, first, makes the second assertion of Proposition 3 applicable in the case $\tau(p^*) > 0$. Next, it allows (recall that $p^* < \infty$ and $\alpha = 1$) to rewrite $\mathcal{I}_\infty(1)$ appeared in Proposition 4 as $\mathcal{I}_\infty(1) = [\boldsymbol{v}, \overline{\mathbf{v}}]$.

$\mathbf{1^0}$. Consider the case $\tau(p^*) > 0$.

Taking into account that $\vec{L} \in [L_0, L_\infty]$ we remark that in view of Nikol'skii (1977) [Theorem 6.9.1, Section 6.9] $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \subset \mathbb{B}_{p^*,d}(c_9 L_\infty)$, where c_9 is independent of $\vec{L}$. Thus, Theorem 2 is applicable with $\mathbf{u}=\infty$, $\mathbf{q}=p^*$ and $D=c_9L_\infty\vee Q$. Choose $\overline{\boldsymbol{v}}=\mathbf{v}$ and remark that the statements of Propositions 3 and 4 hold since $\overline{\mathbf{v}}=\mathbf{v}$. The assertion of the theorem is obtained from (6.21), (6.22), third bound in Lemma 4, (6.23), (6.19) and the first assertion of Theorem 2 by the same computations that led to (6.25).

$\mathbf{2^0}$. Consider the case $\tau(p^*)\le 0$. Since $\mathbf{Z}_{\mathbf{y},\infty}(\alpha)\ge 0$ we have $\varkappa_1(p^*)<0$ in view of (6.14) of Lemma 3. This in its turn implies that $p^*>p$ in this case because we consider only class parameters belonging to $\mathcal{P}_{p,\vec{\mu}}$. Since the definition of the anisotropic Nikol'skii class implies that $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\subset\mathbb{B}_{p^*,d}(L_\infty)$, we assert that the second assertion of Theorem 2 is applicable with $\mathbf{u}=\infty$, $\mathbf{q}=p^*$ and $D=L_\infty\vee Q$. Choose $\overline{\boldsymbol{v}}=\mathbf{v_2}$ and note that $\overline{\mathbf{v}}=\mathbf{v_2}$ in the considered case. Thus, we deduce from (6.21), (6.22), (6.23) and (6.19), denoting $\mathcal{F}=\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\cap\mathbb{F}_{g,\infty}(R,Q)$,

$$\sup_{f\in\mathcal{F}}\mathcal{R}_n^{(p)}\big[\widehat{f}_{\vec{\mathbf{h}}(\cdot)},f\big]\le C\Big[c_9\delta_n^{\omega(1)-\frac{\omega(1)\varkappa_1(p,\infty)}{\varkappa_1(p^*,\infty)}}+(c_3+c_6)\mathfrak{b}_n^p(\mathbb{H})\delta_n^{p\varrho(\alpha)}+\delta_n^{\frac{\omega(1)(p-p^*)}{\varkappa_1(p^*,\infty)}}\Big]^{\frac{1}{p}}$$

and the assertion of the theorem follows in this case. Theorem 4 is proved.

7. Proofs of Propositions 3 and 4. Without further mentioning we will assume that n is large enough to provide $\mathfrak{a}^{-2}\delta_n\le 1$.

LEMMA 5. *For any $\vec{\beta}$, $\vec{r}$, $\vec{\mu}$, $p\ge 1$ and $\alpha\in[0,1]$ the following is true.*

$$1/\gamma(\alpha)-1/\beta(\alpha)=\big[\tau(\infty)\beta(0)\big]^{-1}\big[1/\omega(\alpha)-1/\upsilon(\alpha)\big].$$

7.1. *Proof of Proposition 3.* We start the proof with several remarks which will be useful in the sequel. First, obviously there exists $0<\mathbf{T}:=T\big(\vec{\beta},\vec{r},\vec{\mu},p\big)<\infty$ independent of $\vec{L}$ such that

$$\lim_{n\to\infty}\sup_{\alpha\in\{0,1\}}\sup_{\mathbf{s}\in[1,\infty]}\sup_{v\in[\underline{\mathbf{v}},1\vee\mathbf{v}]}\sum_{j=1}^{d}\left\{\frac{\big|\ln\big(\boldsymbol{h}_j(v,\mathbf{1})\big)\big|+\big|\ln\big(\mathfrak{h}_j(v,\mathbf{s})\big)\big|}{\ln n}\right\}=\mathbf{T}. \tag{7.1}$$

Next, for any $\mathbf{s}\in[1,\infty]$ and any $v>0$

$$\begin{aligned}\frac{\ln n}{n}\prod_{j=1}^{d}\big(\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{s})\big)^{-1-2\boldsymbol{\mu}_j(\alpha)}&=\mathfrak{a}^2\boldsymbol{L}^{-\frac{1}{\beta(\alpha)}}\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)}{\omega(\alpha)+\mathbf{s}}}v^{\frac{2\mathbf{s}-\omega(\alpha)/\beta(\alpha)}{\mathbf{s}+\omega(\alpha)}}\\&=2\boldsymbol{L}^{-\frac{1}{\beta(\alpha)}}\mathfrak{a}^2v^2\mathfrak{z}^{-1}(v).\end{aligned}\tag{7.2}$$

1) Let us proceed to the proof of the first assertion. First we remark that for all $n\ge 3$

$$\vec{\boldsymbol{h}}(v,\mathbf{1}) \in (0,1]^d, \quad \forall v \in [\underline{\mathbf{v}},1]. \tag{7.3}$$

Indeed for any $v>0$ we have since $\boldsymbol{L} \le L_0$,

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{1}) \le \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)}{\mathbf{1}+\omega(\alpha)}} v^{r_j-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\mathbf{1}+\omega(\alpha)}}, \quad j \in \bar{\mathcal{J}}_\infty. \tag{7.4}$$

Therefore, for any $v \in [\underline{\mathbf{v}},1]$ one has in view of the definition of $\underline{\mathbf{v}}$

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{1}) \le \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)}{\mathbf{1}+\omega(\alpha)}} \underline{\mathbf{v}}^{1-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\mathbf{1}+\omega(\alpha)}} = 1, \; j \in \bar{\mathcal{J}}_\infty.$$

Note that for any $j \in \mathcal{J}_\infty$

$$\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{1}) = \big(\boldsymbol{L}L_j^{-1}v\big)^{\frac{1}{\beta_j}} \le v^{\frac{1}{\beta_j}} \le 1, \qquad \forall v \le 1.$$

and the proof of (7.3) is completed since $\boldsymbol{h}_j(\cdot,\mathbf{1}) \le \widetilde{\boldsymbol{\eta}}_j(\cdot,\mathbf{1})$ by construction.

Set $T_0 = \big[\mathbf{T}+2\big]\, e^{d+2\sum_{j=1}^d \mu_j(\alpha)} \boldsymbol{L}^{-\frac{1}{\beta(\alpha)}}$ and remark that in view of (7.1), (7.2) and (7.3) for all n large enough and any $v \in [\underline{\mathbf{v}},1]$

$$\begin{aligned} G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{1})\big) \;&\le\; \frac{(\mathbf{T}+2)\ln n}{n\prod_{j=1}^d \big(\boldsymbol{h}_j(v,\mathbf{1})\big)^{1+\boldsymbol{\mu}_j(\alpha)}} \le \frac{T_0 \boldsymbol{L}^{\frac{1}{\beta(\alpha)}} \ln n}{n\prod_{j=1}^d \big(\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{1})\big)^{1+\boldsymbol{\mu}_j(\alpha)}} \\ &\le\; \frac{T_0 \boldsymbol{L}^{\frac{1}{\beta(\alpha)}} \ln n}{n\prod_{j=1}^d \big(\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{1})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}} = T_0 \mathfrak{a}^{\frac{2}{1+\omega(\alpha)}} \delta_n^{\frac{\omega(\alpha)}{1+\omega(\alpha)}} v^{\frac{2-\omega(\alpha)/\beta(\alpha)}{\mathbf{1}+\omega(\alpha)}}. \end{aligned} \tag{7.5}$$

Here we have taken into account that $\boldsymbol{h}_j(v,\mathbf{s}) \ge e^{-1}\boldsymbol{\eta}_j(v,\mathbf{s})$. Since

$$T_0 \mathfrak{a}^{\frac{2}{1+\omega(\alpha)}} \delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}} v^{\frac{2-\omega(\alpha)/\beta(\alpha)}{1+\omega(\alpha)}} \le T_0\mathfrak{a}^2 v \quad \Leftrightarrow \quad v \ge \underline{\mathbf{v}},$$

denoting $\mathfrak{a} = \sqrt{a/T_0}$ we assert that

$$G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{1})\big) \le av, \quad \forall v \in [\underline{\mathbf{v}},1].$$

The first assertion is established.

2) Before proving the second assertion, let us make several remarks.

$\mathbf{1^0}$. For any $\mathbf{u} \in [1,\infty]$ the following is true.

$$\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u}) \;=\; \big(\boldsymbol{L}L_j^{-1}\mathbf{v}\big)^{\frac{1}{\beta_j}}, \;\; j \in \mathcal{J}_\infty; \tag{7.6}$$

$$\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u}) \;=\; \big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1}{\gamma_j}} \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)\tau(p_\pm)\beta(0)}{\gamma_j[z(\alpha)+\omega(\alpha)/\mathbf{u}]}}, \;\; j \in \bar{\mathcal{J}}_\infty. \tag{7.7}$$

The equality (7.6) follows directly from the definition of $\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u})$ since, remind $\gamma_j = \beta_j, q_j = \infty$ if $j \in \mathcal{J}_\infty$. Thus, let us prove the equality (7.7).

$$\widehat{\boldsymbol{\eta}}_j^{\gamma_j q_j}(\mathbf{v},\mathbf{u}) = \big(\boldsymbol{L}L_j^{-1}\big)^{p_\pm} \big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\mathbf{u}\upsilon(\alpha)}{\mathbf{u}+\upsilon(\alpha)}} \mathbf{v}^{p_\pm - \frac{\mathbf{u}\upsilon(\alpha)(2+1/\gamma(\alpha))}{\mathbf{u}+\upsilon(\alpha)}}, \quad \forall j \in \bar{\mathcal{J}}_\infty.$$

Here we used that $q_j = p_\pm$ for any $j \in \bar{\mathcal{J}}_\infty$. Using the definition of $\mathbf{v}$ we get

$$\widehat{\boldsymbol{\eta}}_j^{\gamma_j q_j}(\mathbf{v},\mathbf{u}) = \big(\boldsymbol{L}L_j^{-1}\big)^{p_\pm}\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\mathbf{u}\upsilon(\alpha)}{\mathbf{u}+\upsilon(\alpha)}+\frac{\omega(\alpha)\tau(\infty)\beta(0)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}\left[p_\pm-\frac{\mathbf{u}\upsilon(\alpha)(2+1/\gamma(\alpha))}{\mathbf{u}+\upsilon(\alpha)}\right]}$$

for any $j \in \bar{\mathcal{J}}_\infty$ Using the definition of $z(\alpha)$ we obtain

$$A := \tfrac{\mathbf{u}\upsilon(\alpha)}{\mathbf{u}+\upsilon(\alpha)} + \tfrac{\omega(\alpha)\tau(\infty)\beta(0)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}\left[p_\pm - \tfrac{\mathbf{u}\upsilon(\alpha)(2+1/\gamma(\alpha))}{\mathbf{u}+\upsilon(\alpha)}\right] = \tfrac{\omega(\alpha)\tau(\infty)\beta(0)p_\pm}{z(\alpha)+\omega(\alpha)/\mathbf{u}}$$
$$+\tfrac{\mathbf{u}\upsilon(\alpha)\left[1+\omega(\alpha)/\mathbf{u}-\omega(\alpha)\tau(\infty)\beta(0)\left\{1/\gamma(\alpha)-1/\beta(\alpha)\right\}\right]}{(\mathbf{u}+\upsilon(\alpha))(z(\alpha)+\omega(\alpha)/\mathbf{u})}.$$

We obtain applying Lemma 5

$$A = \tfrac{\omega(\alpha)\tau(\infty)\beta(0)p_\pm}{z(\alpha)+\omega(\alpha)/\mathbf{u}} + \tfrac{\mathbf{u}\upsilon(\alpha)\omega(\alpha)[1/s+1/\upsilon(\alpha)]}{(\mathbf{u}+\upsilon(\alpha))(z(\alpha)+\omega(\alpha)/\mathbf{u})} = \tfrac{\omega(\alpha)\tau(p_\pm)\beta(0)p_\pm}{z(\alpha)+\omega(\alpha)/\mathbf{u}}.$$

Thus, (7.7) is established.

$\mathbf{2^0}$. Next, let us prove that

$$\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u}) \in (0,1]^d, \quad \forall \mathbf{u} \in (1,\infty]. \tag{7.8}$$

If $\mathcal{J}_\infty \neq \emptyset$, which is equivalent to $p^* = \infty$, the definition of $\mathbf{v}$ implies that $\mathbf{v} \le 1$ for all n large enough, since $\tau(p^*) = \tau(\infty) \ge 0$ and in view of (6.16). We deduce from (7.6)

$$\mathfrak{h}_j(\mathbf{v},\mathbf{u}) \le \widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u}) = \big(\boldsymbol{L}L_j^{-1}\mathbf{v}\big)^{\frac{1}{\beta_j}} \le \mathbf{v}^{\frac{1}{\beta_j}} \le 1, \quad \forall j \in \mathcal{J}_\infty.$$

and (7.8) is proved for any $j \in \mathcal{J}_\infty$.

It remains to note that $\tau(p_\pm) \ge \tau(p^*)$ since $p^* \ge p_\pm$ and therefore, if $\tau(p^*) \ge 0$ we have $\mathfrak{h}_j(\mathbf{v},\mathbf{u}) \le \widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u}) \le 1$, for any $j \in \bar{\mathcal{J}}_\infty$ and all n large enough in view of (6.16), (7.7) and since $\boldsymbol{L}L_j^{-1} \le 1$. Thus, (7.8) is proved.

$\mathbf{3^0}$. For any $\mathbf{u} \in (1,\infty]$ one has

$$\mathfrak{a}^{-2}\delta_n \prod_{j=1}^d \widehat{\boldsymbol{\eta}}_j^{-1-2\boldsymbol{\mu}_j(\alpha)}(\mathbf{v},\mathbf{u}) \le T^{-1}(\alpha)\big(\mathfrak{a}^{-2}\delta_n\big)^{1-\frac{\omega(\alpha)\tau(\infty)\beta(0)/\beta(\alpha)+1}{z(\alpha)+\omega(\alpha)/\mathbf{u}}}; \tag{7.9}$$

$$\mathfrak{a}^{-2}\delta_n \prod_{j=1}^d \widehat{\boldsymbol{\eta}}_j^{-1}(\mathbf{v},\mathbf{u}) \le T^{-1}(0)\big(\mathfrak{a}^{-2}\delta_n\big)^{1-\frac{\omega(\alpha)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}} \tag{7.10}$$

where $T(\alpha) = \inf_{\vec{L}\in[L_0,L_\infty]^d} \prod_{j\in\mathcal{J}_\infty} \big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1+2\boldsymbol{\mu}_j(\alpha)}{\beta_j}} \prod_{j\in\bar{\mathcal{J}}_\infty} \big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1+2\boldsymbol{\mu}_j(\alpha)}{\gamma_j}}$.

Indeed, we have in view of (7.6), (7.7) and the definition of $\mathbf{v}$

$$\textstyle\prod_{j=1}^d \widehat{\boldsymbol{\eta}}_j^{1+2\boldsymbol{\mu}_j(\alpha)}(\mathbf{v},\mathbf{u}) \ge T^{-1}(\alpha)\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)\tau(p_\pm)\beta(0)}{\gamma_\pm(\alpha)[z(\alpha)+\omega(\alpha)/\mathbf{u}]}+\frac{\omega(\alpha)\tau(\infty)\beta(0)}{\beta_\infty(\alpha)[z(\alpha)+\omega(\alpha)/\mathbf{u}]}}$$

$$\textstyle\prod_{j=1}^d \widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u}) \ge T^{-1}(0)\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)\tau(p_\pm)\beta(0)}{\gamma_\pm(0)[z(\alpha)+\omega(\alpha)/\mathbf{u}]}+\frac{\omega(\alpha)\tau(\infty)\beta(0)}{\beta_\infty(0)[z(\alpha)+\omega(\alpha)/\mathbf{u}]}},$$

where we have put $\frac{1}{\beta_\infty(\alpha)} = \sum_{j\in\mathcal{J}_\infty}\frac{1+2\boldsymbol{\mu}(\alpha)}{\beta_j}$, $\frac{1}{\gamma_\pm(\alpha)} = \sum_{j\in\bar{\mathcal{J}}_\infty}\frac{1+2\boldsymbol{\mu}(\alpha)}{\gamma_j}$. Note that for any $\alpha \in [0,1]$

$$\tfrac{\tau(p_\pm)}{\gamma_\pm(\alpha)}+\tfrac{\tau(\infty)}{\beta_\infty(\alpha)}=\sum_{j\in\bar{\mathcal{J}}_\infty}\tfrac{(1+2\boldsymbol{\mu}(\alpha))\tau(r_j)}{\beta_j}+\tfrac{\tau(\infty)}{\beta_\infty(\alpha)}=\tfrac{\tau(\infty)}{\beta(\alpha)}+\tfrac{1}{\omega(\alpha)\beta(0)},$$

and (7.9) and (7.10) are established.

$4^{\mathbf{0}}$. Simple algebra shows that for any $\mathbf{u}\in[1,\infty]$

$$\big(\mathfrak{a}^{-2}\delta_n\big)^{1-\frac{\omega(\alpha)\tau(\infty)\beta(0)/\beta(\alpha)+1}{z(\alpha)+\omega(\alpha)/\mathbf{u}}}=2\mathbf{v}^2\mathfrak{z}^{-1}(\mathbf{v}),$$

and we deduce from (7.9) for any $\mathbf{u}\in(1,\infty]$ (recall that $\mathfrak{z}\equiv 2$ if $\mathbf{u}=\infty$)

$$\delta_n\prod_{j=1}^{d}\widehat{\boldsymbol{\eta}}_j^{\,-1-2\boldsymbol{\mu}_j(\alpha)}(\mathbf{v},\mathbf{u})\le 2T^{-1}(\alpha)\mathfrak{a}^2\mathbf{v}^2\mathfrak{z}^{-1}(\mathbf{v}). \tag{7.11}$$

Let us also prove that for any $\mathbf{u}\in[1,\infty]$ and all n large enough

$$\mathbf{v}>\boldsymbol{v}:=\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{1}{2+1/\beta(\alpha)}}\quad\Rightarrow\quad\mathfrak{z}(\mathbf{v})\ge 2. \tag{7.12}$$

The latter inclusion follows from (6.9). Indeed, if $\tau(\infty)\le 0$ then $\mathbf{v}\ge 1\ge\boldsymbol{v}$. If $\tau(\infty)>0$ then in view of (6.16)

$$\tfrac{\omega(\alpha)\tau(\infty)\beta(0)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}-\tfrac{1}{2+1/\beta(\alpha)}=-\tfrac{1+\omega(\alpha)/\mathbf{u}}{[z(\alpha)+\omega(\alpha)/\mathbf{u}][2+1/\beta(\alpha)]}<0,$$

so $\mathbf{v}>\boldsymbol{v}$. Note at last that for any $\mathbf{u}\in(1,\infty]$

$$\mathbf{v}\mathfrak{z}^{-1}(\mathbf{v})=2\big(\mathfrak{a}^{-2}\delta_n\big)^{\frac{\omega(\alpha)\tau(\mathbf{u})\beta(0)}{z(\alpha)+\omega(\alpha)/\mathbf{u}}}. \tag{7.13}$$

$5^{\mathbf{0}}$. Let us proceed to the proof of the second assertion. Choose $\mathfrak{a}^2<a^2T(\alpha)/(4T_0)<1$. We get from (7.1), (7.9) and (7.11) similarly to (7.5)

$$\begin{aligned}F_n^2\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\;&\le\;\frac{T_0\delta_n}{\prod_{j=1}^{d}\big(\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}}\le 2T_0T^{-1}(\alpha)\mathfrak{a}^2\mathbf{v}^2\mathfrak{z}^{-1}(\mathbf{v})\\ &\le\;2^{-1}a^2\mathbf{v}^2\mathfrak{z}^{-1}(\mathbf{v}).\end{aligned}\tag{7.14}$$

Thus to prove the assertion all we need to show is that $\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\in\mathfrak{H}(\mathbf{v})$, i.e. $G_n\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\le a\mathbf{v}$. Let us distinguish three cases.

$5^{\mathbf{0}}$a. Let $\tau(\infty)\ge 0$. We remark that the definition of $\mathbf{v}$ in this case yields $\mathbf{v}\le 1$ for all n large enough and we obtain from (7.11) and (7.12) that

$$\delta_n\prod_{j=1}^{d}\widehat{\boldsymbol{\eta}}_j^{\,-1-2\boldsymbol{\mu}_j(\alpha)}(\mathbf{v},\mathbf{u})\le T^{-1}(\alpha)\mathfrak{a}^2\mathbf{v}. \tag{7.15}$$

Then we have in view of (7.1), (7.8), (7.9) and (7.15) similarly to (7.5)

$$G_n\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\le\frac{T_0\delta_n}{\prod_{j=1}^{d}\big(\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}}\le T_0T^{-1}(\alpha)\mathfrak{a}^2\mathbf{v}\le a\mathbf{v}. \tag{7.16}$$

5^0b. Let $\tau(\infty)<0, \tau(p^*)\geq 0$ and $\alpha\neq 1$. Then by imposed assumption $\mathbf{u}\leq\mathbf{u}^*$, and, therefore, $\tau(\mathbf{u})\geq 0$. We get from (7.13), (6.16) and (7.14)

$$G_n\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)=F_n^2\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\leq a^2\mathbf{v}<a\mathbf{v}. \tag{7.17}$$

5^0c. Let $\tau(\infty)<0, \tau(p^*)\geq 0$, $\alpha=1$. We have as previously

$$\begin{aligned} G_n^2\big(\vec{h}(\mathbf{u})\big) &\leq \frac{(\mathbf{T}+2)\ln n}{n\prod_{j=1}^d\big(\mathfrak{h}_j(\mathbf{v},\mathbf{u})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}}\frac{(\mathbf{T}+2)\ln n}{n\prod_{j=1}^d\mathfrak{h}_j(\mathbf{v},\mathbf{u})}\\ &\leq 2T_0^2T^{-1}(1)T_1\mathfrak{a}^4\mathbf{v}^2\mathfrak{z}^{-1}(\mathbf{v})\bigg[\frac{T(0)\mathfrak{a}^{-2}\delta_n}{\prod_{j=1}^d\widehat{\boldsymbol{\eta}}_j(\mathbf{v},\mathbf{u})}\bigg]. \end{aligned} \tag{7.18}$$

Here we have used (7.11) and put $T_1=T^{-1}(0)\boldsymbol{L}^{-1/\beta(0)}$. Our goal now is to show that for any $\mathbf{u}\in[1,\infty]$ and all n large enough

$$T(0)\mathfrak{a}^{-2}\delta_n\mathfrak{z}^{-1}(\mathbf{v})\prod_{j=1}^d\widehat{\boldsymbol{\eta}}_j^{-1}(\mathbf{v},\mathbf{u})\leq 1. \tag{7.19}$$

In view of (7.10) and of the definition of $\mathfrak{z}(\cdot)$ in order to establish (7.19) it suffices to show that $z(1)/\omega(1)-1+2/\mathbf{u}\geq 0$. Since we assumed $\tau(\infty)<0$ and $\tau(p^*)\geq 0$, the required results follows from (6.13). Thus, (7.19) is proved. Then choosing $\mathfrak{a}$ such that $T_0(2T^{-1}(1)T_1)^{1/2}\mathfrak{a}^2\leq a$, we obtain from (7.18) and (7.19) that

$$G_n\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\leq T_0(2T^{-1}(\alpha)T_1)^{1/2}\mathfrak{a}^2\mathbf{v}\leq av.$$

for all all n large enough. The second assertion is proved. ∎

7.2. *Proof of Proposition 4.* We start the proof with several remarks which will be useful in the sequel.

1^0. Let us show that for all n large enough

$$\vec{\boldsymbol{h}}(v,\mathbf{u})\in(0,1]^d,\quad\forall v\in\mathcal{I}_{\mathbf{u}}(\alpha),\ \ \forall\mathbf{u}\geq\mathbf{y}. \tag{7.20}$$

In view of the definition of $\widetilde{\boldsymbol{\eta}}_j(\cdot,\mathbf{u}), j=1,\dots,d$,

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_jr_j}(v,\mathbf{u})=\big(\boldsymbol{L}L_j^{-1}\big)^{r_j}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}}v^{r_j-\frac{\mathbf{u}\omega(\alpha)(2+1/\beta(\alpha))}{\mathbf{u}+\omega(\alpha)}},\ j\in\bar{\mathcal{J}}_\infty. \tag{7.21}$$

Therefore, for any $v\in[\underline{\mathbf{v}},1]$ one has, taking into account that $\boldsymbol{L}\leq L_0$,

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_jr_j}(v,\mathbf{u})\leq\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}}\underline{\mathbf{v}}^{1-\frac{\mathbf{u}\omega(\alpha)(2+1/\beta(\alpha))}{\mathbf{u}+\omega(\alpha)}}=1,\quad j\in\bar{\mathcal{J}}_\infty.$$

It remains to note that $\boldsymbol{v}>\underline{\mathbf{v}}$ for all n large enough and, therefore,

$$\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{u})\leq 1,\ j\in\bar{\mathcal{J}}_\infty,\quad\forall v\in[\boldsymbol{v},1]\cap\mathcal{I}_{\mathbf{u}}(\alpha). \tag{7.22}$$

We also have in view of the definition of $\widetilde{\boldsymbol{\eta}}_j(\cdot,\mathbf{u}), j=1,\dots,d$,

$$\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{u}) = \big(\boldsymbol{L}L_j^{-1}v\big)^{\frac{1}{\beta_j}} \le 1, \quad j\in\mathcal{J}_\infty,$$

for any $v\le 1$. This together with (7.22) proves (7.20) in the cases when $\mathcal{I}_{\mathbf{u}}(\alpha)=[\boldsymbol{v},1]$. Noting that $p^*<\infty$ is equivalent to $\mathcal{J}_\infty=\emptyset$, we deduce from (7.21) for any $v\ge 1$ and any $j=1,\dots,d$

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{u}) \le \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}} v^{p^*-\frac{\mathbf{u}\omega(\alpha)(2+1/\beta(\alpha))}{\mathbf{u}+\omega(\alpha)}} \le \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}} v^{-\varkappa_\alpha(p^*,\mathbf{u})}.$$

Thus, for any $v\ge 1$, $j=1,\dots,d$ and for all n large enough

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{u}) \le 1_{\{\varkappa_\alpha(p^*,\mathbf{u})\ge 0\}} + \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}} \widetilde{\mathbf{v}}^{-\varkappa_\alpha(p^*,\mathbf{u})} 1_{\{\varkappa_\alpha(p^*,\mathbf{u})<0\}}, \tag{7.23}$$

where we denoted $\widetilde{\mathbf{v}}=\mathbf{v_1}$ if $\alpha\neq 1$ and $\widetilde{\mathbf{v}}=\overline{\mathbf{v}}$ if $\alpha=1$.

Let $\alpha=1, p^*<\infty, \tau(p^*)>0$. Then $\widetilde{\mathbf{v}}=\mathbf{v}$ and we have for any $j=1,\dots,d$ and $v\in[1,\mathbf{v}]$, using the definition of $\mathbf{v}$,

$$\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(1)}{\mathbf{u}+\omega(1)}-\frac{\varkappa_1(p^*,\mathbf{u})\omega(1)\tau(\infty)\beta(0)}{z(1)+\omega(1)/\mathbf{u}}} = \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{p^*\tau(p^*)\omega(1)}{z(1)+\omega(1)/\mathbf{u}}} \to 0,\ n\to\infty,$$

in view of (6.15). Hence, (7.20) follows from (7.23) in this case.

Let $\alpha=1, \tau(p^*)\le 0$. Then $\widetilde{\mathbf{v}}=\mathbf{v_2}$ and we have for any $v\in[1,\mathbf{v_2}]$ in view of the definition of $\mathbf{v_2}$

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{u}) \le \widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(\mathbf{v_2},\mathbf{u}) = \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(1)}{\mathbf{u}+\omega(1)}} \mathbf{v}_2^{-\varkappa_1(p^*,\mathbf{u})} = 1, \quad j=1,\dots,d.$$

and, therefore (7.20) follows from (7.23) in this case.

Let $\alpha\neq 1, \mathbf{u}<\infty$. First we note that $\tau(\infty)<0$ and $\mathbf{u}\ge\mathbf{y}$ imply

$$1-\mathbf{u}/\omega(0)+1/\beta(0) = 1-\mathbf{u}+\mathbf{u}\tau(\mathbf{u}) \le 1-\mathbf{u}+\mathbf{u}\tau\big(\mathbf{y}\big) \le 1-\mathbf{u}<0,$$

since either $\mathbf{y}=p^*$ that is equivalent to $\tau(p^*)\le 0$ or $\mathbf{y}=\mathbf{u}^*$ and then $\tau\big(\mathbf{y}\big)=0$. Thus $\mathbf{v_1}\to\infty, n\to\infty$ and, therefore, in view of (7.23) for any $v\in[1,\mathbf{v}_1]$ and $j=1,\dots,d$ one has

$$\widetilde{\boldsymbol{\eta}}_j^{\beta_j r_j}(v,\mathbf{u}) \le 1_{\{\varkappa_\alpha(p^*,\mathbf{u})\ge 0\}} + \big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(\alpha)}{\mathbf{u}+\omega(\alpha)}} \mathbf{v_1}^{-\varkappa_0(p^*,\mathbf{u})} 1_{\{\varkappa_\alpha(p^*,\mathbf{u})<0\}}$$

Note that $1-\mathbf{u}/\omega(0)+1/\beta(0) = \varkappa_0(p^*,\mathbf{u})\big[1/\mathbf{u}+1/\omega(0)\big]-(\mathbf{u}-p^*)\big[1/\mathbf{u}+1/\omega(0)\big]$ and, therefore

$$-\frac{\varkappa_0(p^*,\mathbf{u})}{1-\mathbf{u}/\omega(0)+1/\beta(0)} \ge -\frac{\mathbf{u}\omega(0)}{\mathbf{u}+\omega(0)} \quad\Rightarrow\quad \mathbf{v}_1^{-\varkappa_0(p^*,\mathbf{u})} \le \big\{\mathfrak{a}^{-2}\delta_n\big\}^{-\frac{\mathbf{u}\omega(0)}{\mathbf{u}+\omega(0)}}.$$

It remains to note that if $\tau(\infty)\ge 0$ then $\mathbf{u}^*=\infty$ and, therefore $\mathbf{u}=\infty$. It implies $\mathbf{v}_1=1$ and $\mathcal{I}_{\mathbf{u}}(\alpha)=[\boldsymbol{v},1]$ and this case has been already treated. This completes the proof of (7.20).

$\mathbf{2^0}$. Remark that there obviously exists $0<\mathbf{S}:=S\big(\vec{\beta},\vec{r},\vec{\mu},p\big)<\infty$ independent of $\vec{L}$ such that

$$\lim_{n\to\infty}(\ln n)^{-1}\sup_{\alpha\in\{0,1\}}\sup_{\mathbf{u}\in[1,\infty]}\sup_{v\in\mathcal{I}_{\mathbf{u}}(\alpha)}\sum_{j=1}^d\big|\ln\big(\boldsymbol{h}_j(v,\mathbf{u})\big)\big| = \mathbf{S}.$$

Hence, in view of (7.20) one has for all n large enough and $v\in\mathcal{I}_{\mathbf{u}}(\alpha)$

$$F_n\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big) \;\leq\; \sqrt{(\mathbf{S}+2)n^{-1}\ln n}\,\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{u})\big)^{-\frac12-\boldsymbol{\mu}_j(\alpha)}; \tag{7.24}$$

$$G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big) \;\leq\; (\mathbf{S}+2)n^{-1}\ln n\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{u})\big)^{-1-\boldsymbol{\mu}_j(\alpha)}. \tag{7.25}$$

Taking into account that $\boldsymbol{h}_j(v,\mathbf{u})\geq e^{-1}\widetilde{\boldsymbol{\eta}}_j(v,\mathbf{u})$ and setting $S_0=\big[\mathbf{S}+2\big]\,e^{d+2\sum_{j=1}^d\mu_j}\boldsymbol{L}^{-\frac{1}{\beta(1)}}$ we obtain from (7.2) for any $\alpha\in[0,1]$ and $v\in\mathcal{I}_{\mathbf{u}}(\alpha)$

$$(\mathbf{S}+2)n^{-1}\ln(n)\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{s})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}\leq 2S_0\mathfrak{a}^2v^2\mathfrak{z}^{-1}(v). \tag{7.26}$$

From now on we choose $\mathfrak{a}\leq a/(2S_0)<1$. It yields in view of (7.24), (7.26)

$$F_n^2\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big)\leq a^2v^2\mathfrak{z}^{-1}(v),\;\;\forall v\in\mathcal{I}_{\mathbf{u}}(\alpha). \tag{7.27}$$

$\mathbf{3^0}$. Since (7.27) holds, to finish the proof of Proposition (4) all we need to show is that $G_n\big(\vec{\mathfrak{h}}(\mathbf{v},\mathbf{u})\big)\leq a\mathbf{v},\forall v\in\mathcal{I}_{\mathbf{u}}(\alpha)$. Let us distinguish three cases.

$\mathbf{3^0a}$. Let $p^*=\infty$ or $\alpha\neq 1,\mathbf{u}=\infty$. First we note that in these cases $\mathcal{I}_{\mathbf{u}}(\alpha)=[\boldsymbol{v},1]$. Next in view of the second inequality in (7.24), (7.20), (7.26) and (7.27) we obtain for any $v\in\mathcal{I}_{\mathbf{u}}(\alpha)$

$$G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big)\leq\frac{(\mathbf{S}+2)\ln n}{n\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{u})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}}\leq a^2v^2\mathfrak{z}^{-1}(v)\leq av. \tag{7.28}$$

To get the last inequality we have used that $a<1$, $\mathfrak{z}(\cdot)\geq 2$ and $v\leq 1$.

$\mathbf{3^0b}$. Let $\alpha\neq 1,p^*<\infty,\mathbf{u}<\infty$. We have in view of (7.25) and (7.26)

$$G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big)\leq a^2v^2\mathfrak{z}^{-1}(v),\quad\forall v\in\mathcal{I}_{\mathbf{u}}(0).$$

Here we have also used that $\boldsymbol{\mu}_j(\alpha)=0$ for all j. Simple algebra shows that $v\mathfrak{z}^{-1}(v)=\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(0)}{\mathbf{u}+\omega(0)}}v^{\frac{\mathbf{u}-\omega(0)-\omega(0)/\beta(0)}{\mathbf{u}+\omega(0)}}$, $\mathbf{u}\neq\infty$, and since $\mathbf{u}-\omega(0)-\omega(0)/\beta(0)>0$ for any $\mathbf{u}\geq\mathbf{u}^*$, the result follows from

$$\sup_{v\in\mathcal{I}_{\mathbf{u}}(0)}v\mathfrak{z}^{-1}(v)=\mathbf{v_1}\mathfrak{z}^{-1}(\mathbf{v_1})=1.$$

$\mathbf{3^0c}$. Let $\alpha=1,p^*<\infty$. For any $v\in\mathcal{I}_{\mathbf{u}}(1)$ we have in view of the second inequality in (7.24) and (7.26), denoting $S_1=S_0\boldsymbol{L}^{-1/\beta(0)}$,

$$\begin{aligned}G_n^2\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big)\;&\leq\;\frac{(\mathbf{S}+2)\ln n}{n\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{u})\big)^{1+2\boldsymbol{\mu}_j(\alpha)}}\frac{(\mathbf{S}+2)\ln n}{n\prod_{j=1}^d\big(\boldsymbol{h}_j(v,\mathbf{u})\big)}\\ &\leq\; S_1\mathfrak{a}^2a^2v^2\mathfrak{z}^{-1}(v)\mathfrak{a}^{-2}\delta_n\prod_{j=1}^d\widetilde{\boldsymbol{\eta}}_j^{-1}(v,\mathbf{u}).\end{aligned} \tag{7.29}$$

Our goal now is to show that for all n large enough

$$\sup_{v\in\mathcal{I}_{\mathbf{u}}(1)} \mathfrak{a}^{-2}\delta_n\mathfrak{z}^{-1}(v)\prod_{j=1}^{d}\widetilde{\boldsymbol{\eta}}_j^{-1}(v,\mathbf{u}) \le 1. \tag{7.30}$$

Denoting $P(v)=\mathfrak{a}^{-2}\delta_n\mathfrak{z}^{-1}(v)\prod_{j=1}^{d}\widetilde{\boldsymbol{\eta}}_j^{-1}(v,\mathbf{u})$ we easily compute

$$P(v)=2^{-1}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{2\mathbf{u}\omega(1)(Y+1/\mathbf{u})}{\omega(1)+\mathbf{u}}}\, v^{\frac{2\mathbf{u}\omega(1)\pi(\mathbf{u})}{\mathbf{u}+\omega(1)}}, \quad v>0. \tag{7.31}$$

It yields obviously

$$\sup_{v\in\mathcal{I}_{\mathbf{u}}(1)} \mathfrak{a}^{-2}\delta_n\mathfrak{z}^{-1}(v)\prod_{j=1}^{d}\widetilde{\boldsymbol{\eta}}_j^{-1}(v,\mathbf{u}) \le \max\big[P(\boldsymbol{v}),P(\widetilde{\mathbf{v}})\big], \tag{7.32}$$

where $\widetilde{\mathbf{v}}\in\{\overline{\mathbf{v}},\mathbf{v_3}\}$. We deduce from (7.31) that for any $\mathbf{u}\in[1,\infty]$

$$P(\boldsymbol{v})=2^{-1}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{2+1/\beta(\alpha)-1/\beta(0}{2+1/\beta(\alpha)}}\to 0,\ n\to\infty. \tag{7.33}$$

$\mathbf{3^0c1}$. Consider the case $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)\ge 0$. Here $\widetilde{\mathbf{v}}=\overline{\mathbf{v}}$.

If $\tau(p^*)>0$ then $\overline{\mathbf{v}}=\mathbf{v}$. Moreover $\mathbf{y}=\mathbf{u}^*$ since $\mathbf{u}^*=\infty$ if $\tau(\infty)\ge 0$ and $\tau(\mathbf{u}^*)=0$ if $\tau(\infty)<0$. Hence $z(1)/\omega(1)-1+2/\mathbf{u}\ge 0$ in view of (6.13) and we have in view of the definition of $\mathbf{v}$

$$P(\mathbf{v})=2^{-1}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(1)(1/\omega(1)-1/\omega(0)+2/\mathbf{u})}{\omega(1)+\mathbf{u}}+\frac{\mathbf{u}\omega^2(1)\tau(\infty)\beta(0)\pi(\mathbf{u})}{[\mathbf{u}+\omega(1)][z(\alpha)+\omega(\alpha)/\mathbf{u}]}}. \tag{7.34}$$

Note that,

$$\begin{aligned}&\tfrac{\mathbf{u}\omega(1)(1/\omega(1)-1/\omega(0)+2/\mathbf{u})}{\omega(1)+\mathbf{u}}+\tfrac{\mathbf{u}\omega^2(1)\tau(\infty)\beta(0)\pi(\mathbf{u})}{[\mathbf{u}+\omega(1)][z(1)+\omega(1)/\mathbf{u}]}\\ &=1-\tfrac{\omega(1)[1/\omega(0)-1/\mathbf{u}]}{z(1)+\omega(1)/\mathbf{u}}-\tfrac{\omega(1)\tau(\infty)}{z(1)+\omega(1)/\mathbf{u}}=1-\tfrac{\omega(1)[1-1/\mathbf{u}]}{z(1)+\omega(1)/\mathbf{u}}>0.\end{aligned}$$

To get the last inequality we have used that

$$1-\tfrac{\omega(1)[1-1/\mathbf{u}]}{z(1)+\omega(1)/\mathbf{u}}>0 \quad\Leftrightarrow\quad z(1)/\omega(1)-1+2/\mathbf{u}>0.$$

Thus, we conclude that $P(\mathbf{v})\le 1$, for all large n, which together with (7.33) implies (7.30) in the considered case.

If $\tau(p^*)\le 0$ then $\overline{\mathbf{v}}=\mathbf{v_2}$ and moreover $\mathbf{y}=\mathbf{p}^*$. Also, $\varkappa_1(p^*,\mathbf{u})<0$ thanks to (6.14) of Lemma 3. We have in view of the definition of $\mathbf{v_2}$

$$P(\mathbf{v}_2)=2^{-1}\big\{\mathfrak{a}^{-2}\delta_n\big\}^{\frac{\mathbf{u}\omega(1)(1/\omega(1)-1/\omega(0)+2/\mathbf{u})}{\omega(1)+\mathbf{u}}+\frac{[\mathbf{u}\omega(1)]^2\pi(\mathbf{u})}{\varkappa_1(p^*,\mathbf{u})[\mathbf{u}+\omega(1)]^2}}. \tag{7.35}$$

After routine computations we come to the following equality

$$\tfrac{\mathbf{u}\omega(1)(1/\omega(1)-1/\omega(0)+2/\mathbf{u})}{\omega(1)+\mathbf{u}}+\tfrac{[\mathbf{u}\omega(1)]^2\pi(\mathbf{u})}{\varkappa_1(p^*,\mathbf{u})[\mathbf{u}+\omega(1)]^2}=-\tfrac{2\mathbf{u}\omega(1)p^*\big[Y-(X+1)(\mathbf{y})^{-1}+1/\mathbf{u}\big]}{\varkappa_1(p^*,\mathbf{u})[\mathbf{u}+\omega(1)]}\ge 0.$$

Hence, $P(\mathbf{v_2})\le 1$ for all n large enough, which together with (7.33) allows us to assert (7.30) in the considered case.

$\mathbf{3^0c3}$. Consider the case $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)<0$. Here $\widetilde{\mathbf{v}}=\mathbf{v_3}$. If $\pi(\mathbf{u})\le 0$ when $\mathbf{v_3}=\infty$ and obviously $P\big(\mathbf{v_3}\big)$=0. If $\pi(\mathbf{u})>0$ then $P\big(\mathbf{v_3}\big)=1$ in view of (7.31) and the definition of $\mathbf{v_3}$. This completes the proof (7.30).

Finally in the case $\mathbf{3^0c}$, choosing $\mathfrak{a}\le\sqrt{1/S_1}$, we deduce from (7.29) and (7.30) that for all n large enough $G_n\big(\vec{\boldsymbol{h}}(v,\mathbf{u})\big)\le\sqrt{S_1}\mathfrak{a}av\le av, v\in\mathcal{I}_{\mathbf{u}}(1)$.

8. Appendix. *Proof of Proposition 2.* In view of Lemma 5 in Lepski (2015), if $\tau(p^*)>0$ then

$$\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)\subseteq \mathbb{N}_{\vec{q},d}\big(\vec{\gamma},c_2\vec{L}\big), \tag{8.1}$$

where c_2 is independent on $\vec{L}$. Note also that $\gamma_j\leq\beta_j$ for any $j=1,\dots,d$.

$\mathbf{1^0}$. Let $\big(\vec{\pi},\vec{s}\big)$ be either $\big(\vec{\beta},\vec{r}\big)$ or $\big(\vec{\gamma},\vec{q}\big)$ and without further mentioning the couple $\big(\vec{\gamma},\vec{q}\big)$ is used below under the condition $\tau(p^*)>0$. We obviously have for any $\vec{\mathbf{h}}\in\mathcal{H}$

$$\begin{aligned} b_{\mathbf{h},f,j}(x) &:= \sup_{h\in\mathcal{H}:\,h\leq\mathbf{h}}\Big|\int_{\mathbb{R}}\mathcal{K}_\ell(u)\big[f\big(x+uh\mathbf{e}_j\big)-f(x)\big]\nu_1(\mathrm{d}u)\Big| \\ &= \sup_{h\in\mathcal{H}:\,h\leq\mathbf{h}}\Big|\int_{\mathbb{R}}\mathcal{K}_\ell(u)\big[\Delta_{uh,j}f(x)\big]\nu_1(\mathrm{d}u)\Big|. \end{aligned}$$

For $j=1,\dots,d$ we have

$$\begin{aligned} \int_{\mathbb{R}}\mathcal{K}_\ell(u)\Delta_{uh,j}f(x)\nu_1(\mathrm{d}u) &= \int_{\mathbb{R}}\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+1}\tfrac{1}{i}\mathcal{K}_\ell\Big(\frac{u}{i}\Big)\big[\Delta_{hu,j}f(x)\big]\nu_1(\mathrm{d}u) \\ &= (-1)^{\ell-1}\int_{\mathbb{R}}\mathcal{K}_\ell(z)\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+\ell}\big[\Delta_{izh,j}f(x)\big]\nu_1(\mathrm{d}z) \\ &= (-1)^{\ell-1}\int_{\mathbb{R}}\mathcal{K}_\ell(z)\big[\Delta^{\ell}_{zh,j}\,f(x)\big]\nu_1(\mathrm{d}z). \end{aligned}$$

The last equality follows from the definition of the ℓ-th order difference operator determined in Definition 1. Hence, for any $j\in\mathcal{J}_\infty$ we have in view of the definition of the Nikol'skii class (remind that $\gamma_j=\beta_j, j\in\mathcal{J}_\infty$)

$$\|b_{\mathbf{h},f,j}\|_\infty\leq L_j\sup_{h\in\mathcal{H}:\,h\leq\mathbf{h}}h_j^{\pi_j}\int_{\mathbb{R}}\big|\mathcal{K}_\ell(z)\big||z|^{\pi_j}\nu_1(\mathrm{d}z).$$

This yields for any $\mathbf{h}\in\mathcal{H}$

$$\mathbf{B}_{j,\infty,\mathbb{N}_{\vec{r},d}\left(\vec{\beta},\vec{L}\right)}(\mathbf{h})\leq c_1L_j\mathbf{h}^{\pi_j}, \tag{8.2}$$

and the first and the second assertions are proved for any $j\in\mathcal{J}_\infty$.

Let $j\in\bar{\mathcal{J}}_\infty$. Choosing $\mathbf{k}$ from the relation $e^{\mathbf{k}}=\mathbf{h}$ (recall that $\mathbf{h}\in\mathcal{H}$), we have for any $x\in\mathbb{R}^d$

$$\begin{aligned} b_{\mathbf{h},f,j}(x) &= \sup_{k\leq\mathbf{k}}\Big|\int_{\mathbb{R}}\mathcal{K}_\ell(z)\big[\Delta^{\ell}_{ze^k,j}\,f(x)\big]\nu_1(\mathrm{d}z)\Big| \\ &=: \lim_{l\to-\infty}\ \sup_{l\leq k\leq\mathbf{k}}\Big|\int_{\mathbb{R}}\mathcal{K}_\ell(z)\big[\Delta^{\ell}_{ze^k,j}\,f(x)\big]\nu_1(\mathrm{d}z)\Big|. \end{aligned}$$

Using monotone convergence theorem and the triangle inequality

$$\begin{aligned} \big\|b_{\mathbf{h},f,j}\big\|_{s_j} &= \lim_{l\to-\infty}\ \sup_{l\leq k\leq\mathbf{k}}\Big\|\int_{\mathbb{R}}\mathcal{K}_\ell(z)\big[\Delta^{\ell}_{ze^k,j}\,f(\cdot)\big]\nu_1(\mathrm{d}z)\Big\|_{s_j} \\ &\leq \sum_{k=-\infty}^{\mathbf{k}}\Big\|\int_{\mathbb{R}}\mathcal{K}_\ell(z)\big[\Delta^{\ell}_{ze^k,j}\,f(\cdot)\big]\nu_1(\mathrm{d}z)\Big\|_{s_j}. \end{aligned}$$

By the Minkowski inequality for integrals [see, e.g., (Folland 1999, Section 6.3)], we obtain

$$\big\|b_{\mathbf{v},f,j}\big\|_{s_j}\leq\sum_{k=-\infty}^{\mathbf{k}}\int_{\mathbb{R}}|\mathcal{K}_\ell(z)|\big\|\Delta^{\ell}_{ze^k,j}\,f\big\|_{s_j}\,\nu_1(\mathrm{d}z),\quad j=1,\dots,d.$$

Taking into account that $f \in \mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ and (8.1), we have for any $\mathbf{h} \in \mathcal{H}$ and $j = 1,\dots,d$,

$$\big\|b_{\mathbf{h},f,j}\big\|_{s_j} \le \bigg[\int_{\mathbb{R}} |\mathcal{K}_\ell(z)|\,|z|^{\beta_j}\nu_1(\mathrm{d}z)\bigg] L_j \sum_{k=-\infty}^{\mathbf{k}} e^{k\pi_j} \le c_1 L_j \mathbf{h}^{\pi_j}. \tag{8.3}$$

This proves the first and the second assertions for any $j \in \bar{\mathcal{J}}_\infty$.

$\mathbf{2^0}$. Set $\mathbb{F} = \mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ and recall that

$$\begin{aligned}\mathbf{B}^*_{j,s_j,\mathbb{F}}(\mathbf{h}) &:= \sup_{f\in\mathbb{F}} \textstyle\sum_{h\in\mathcal{H}:\, h\le\mathbf{h}} \big\| \int_{\mathbb{R}} \mathcal{K}_\ell(u)\big[f\big(x+uh\mathbf{e}_j\big)-f(x)\big]\nu_1(\mathrm{d}u)\big\|_{s_j}\\ &\le \sup_{f\in\mathbb{F}} \textstyle\sum_{h\in\mathcal{H}:\, h\le\mathbf{h}} \big\|b_{h,f,j}\big\|_{s_j}.\end{aligned}$$

Hence, the third assertion follows from (8.2) and (8.3). ∎

Proof of Lemma 3. Note that

$$\begin{aligned}z(\alpha)+\omega(\alpha)/\mathbf{u} &= \omega(\alpha)(2+1/\beta(\alpha))\beta(0)\tau(p^*)+1-\omega(\alpha)(2+1/\beta(\alpha))(p^*)^{-1}\\ &+\omega(\alpha)/\mathbf{u} = \omega(\alpha)(2+1/\beta(\alpha))\beta(0)\tau(p^*)-(p^*)^{-1}(1+\omega(\alpha)/\mathbf{u})\varkappa_\alpha(p^*,\mathbf{u}),\end{aligned}$$

and (6.15) follows. On the other hand we have

$$z(\alpha)/\omega(\alpha)-1+2/\mathbf{u} = \big(2+2X\big)\beta(0)\tau(\infty)+2Y+2/\mathbf{u}$$

and (6.13) is checked if $\tau(\infty)\ge 0$ since $X,Y\ge 0$. If $\tau(\infty)<0$ and $\tau(p^*)\ge 0$ then we note first that necessarily $\mathbf{u}^*\ge p^*$ since $\tau(\mathbf{u}^*)=0$ and $\tau(\cdot)$ is strictly decreasing. Hence $\mathbf{y}=\mathbf{u}^*$ and we have

$$z(\alpha)/\omega(\alpha)-1+2/\mathbf{u} = 2\big\{Y-(X+1)\mathbf{y}^{-1}+1/\mathbf{u}\big\} = 2\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)\ge 0$$

and (6.13) is established. Let us prove (6.14). We obviously have

$$\tfrac{\varkappa_\alpha(p^*,\mathbf{u})(\mathbf{u}+\omega(\alpha))}{p^*\mathbf{u}\omega(\alpha)} = -2\big[Y-(X+1)/p^*+1/\mathbf{u}\big]+\tau(p^*)-1+1/\mathbf{u}.$$

If $\tau(p^*)\le 0$ then necessarily $\mathbf{y}=p^*$ and, therefore, for any $\mathbf{u}>1$

$$\frac{\varkappa_\alpha(p^*,\mathbf{u})(\mathbf{u}+\omega(\alpha))}{p^*\mathbf{u}\omega(\alpha)} = -2\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)+\tau(p^*)-1+1/\mathbf{u}<0,$$

since we have supposed that $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)\ge 0$.

Let us prove (6.16). If $\tau(\infty)\ge 0$ then $z(\alpha)\ge 1$ and (6.16) follows. If $\mathbf{Z}_{\mathbf{y},\mathbf{u}}(\alpha)\ge 0$, $\tau(p^*)\ge 0$ then (6.16) follows from (6.13) since $\mathbf{u}>1$.

It remains to prove (6.17). If $\alpha\ne 1$ (6.17) is trivial because $\mu(\alpha)=Y=X=0$. If $\alpha=1$, noting that $r_j\le p^*\le\mathbf{y}$ for any $j=1,\dots,d$, we have

$$Y-[X+1]\mathbf{y}^{-1}+1/\mathbf{u}\ge\mu(1)[1-\tau(\mathbf{y})]-1/\mathbf{y}+1/\mathbf{u}\ge\mu(1)-1/\mathbf{y}+1/\mathbf{u}$$

and (6.17) follows. To get the last inequality we have used that $\tau(\mathbf{u}^*)=0$ and that $\tau(\cdot)$ is strictly decreasing, so $\tau(\mathbf{y})\le 0$. ∎

Proof of Lemma 4. First we remark that $\boldsymbol{h}_j(\cdot,\mathbf{1})\equiv\boldsymbol{h}_j(\cdot,\infty)\equiv\mathfrak{h}_j(\cdot,\infty)\leq\big(\boldsymbol{L}L_j^{-1}\big)^{\frac{1}{\beta_j}}$, $j\in\mathcal{J}_\infty$. Then, we get from (6.2) and (6.4) for any $j\in\mathcal{J}_\infty$

$$\begin{aligned}\mathbf{B}_{j,\infty,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\big(\boldsymbol{h}_j(v,\mathbf{1})\big) &= \mathbf{B}_{j,\infty,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\big(\boldsymbol{h}_j(v,\infty)\big)\\ &= \mathbf{B}_{j,\infty,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})}\big(\mathfrak{h}_j(v,\infty)\big)\leq\mathbf{c}v,\quad\forall v>0.\end{aligned}$$

It yields in particular that for any $v>0$

$$J\big(\vec{\boldsymbol{h}}(v,\mathbf{1}),v\big)\supseteq\mathcal{J}_\infty,\quad J\big(\vec{\boldsymbol{h}}(v,\infty),v\big)\supseteq\mathcal{J}_\infty,\quad J\big(\vec{\mathfrak{h}}(v,\infty),v\big)\supseteq\mathcal{J}_\infty, \tag{8.4}$$

Thus, putting

$$\boldsymbol{\lambda}_2(v)=\sum_{j\in\bar{\mathcal{J}}_\infty}v^{-r_j}L_j^{r_j}\big[\boldsymbol{h}_j(v,\mathbf{1})\big]^{r_j\beta_j}+v^{-2}(\ln n/n)\prod_{j=1}^{d}(\boldsymbol{h}_j(v,\mathbf{1}))^{-1-2\boldsymbol{\mu}_j(\alpha)},$$

we obtain in view of (6.2), Propositions 3, 4, (8.4) and the definition of $\vec{\boldsymbol{h}}(\cdot,\mathbf{s}),\mathbf{s}\in\{1,\infty\}$ that for any $v\in\mathcal{I}_\infty(\alpha)$ and $v\in[\underline{\mathbf{v}},1]$ respectively

$$\begin{aligned}\boldsymbol{\Lambda}_{\vec{r}}\big(v,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L}),\infty\big) &\leq C_1\sum_{j\in\bar{\mathcal{J}}_\infty}v^{-r_j}L_j^{r_j}\big[\boldsymbol{h}_j(v,\infty)\big]^{r_j\beta_j}\\ &\leq \mathbf{C}_3\delta_n^{\omega(\alpha)}v^{-\omega(\alpha)(2+1/\beta(\alpha))};\end{aligned}\tag{8.5}$$

$$\boldsymbol{\Lambda}_{\vec{r}}\big(v,\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})\big) \leq \mathbf{C}_4\boldsymbol{\lambda}_2(v)\leq\mathbf{C}_5\delta_n^{\frac{\omega(\alpha)}{\omega(\alpha)+1}}v^{-\frac{\omega(\alpha)(2+1/\beta(\alpha))}{\omega(\alpha)+1}}. \tag{8.6}$$

To get (8.6) we have used that for all n large enough and all $v\in[\underline{\mathbf{v}},1]$

$$F_n\big(\vec{\boldsymbol{h}}(v,\mathbf{1}))\big)\leq C_2(\ln n/n)\textstyle\prod_{j=1}^{d}(\boldsymbol{h}_j(v,\mathbf{1}))^{-1-2\boldsymbol{\mu}_j(\alpha)},$$

where C_2 is independent of $\vec{L}$. This follows from assertions (7.1) and (7.3) established in the proof of Proposition 3. The first and second assertions of the lemma follow now from (8.5) and (8.6) respectively.

Moreover, if $\tau(p^*)>0$ we get in view of (6.3), Propositions 3 and (8.4)

$$\mathbf{v}^p\boldsymbol{\Lambda}_{\vec{q}}\big(\mathbf{v},\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L}),\infty\big)\leq C_1\sum_{j\in\bar{\mathcal{J}}_\infty}\mathbf{v}^{-q_j}L_j^{q_j}\big[\mathfrak{h}_j(\mathbf{v},\infty)\big]^{q_j\gamma_j}\leq\mathbf{C}_2\delta_n^{\frac{\omega(\alpha)\tau(p)\beta(0)}{z(\alpha)}}.$$

and the third assertion of the lemma is established. ∎

Proof of Lemma 5. Note that

$$\begin{aligned}1/\gamma(\alpha)-1/\beta(\alpha)=1/\gamma_\pm(\alpha)-1/\beta_\pm(\alpha) &= \textstyle\sum_{j\in\mathcal{J}_\pm}\frac{1+2\boldsymbol{\mu}_j(\alpha)}{\beta_j}\big[\tau(r_j)/\tau(p_\pm)-1\big]\\ &= \big[\beta(0)\tau(p_\pm)\big]^{-1}\textstyle\sum_{j\in\mathcal{J}_\pm}\frac{1+2\boldsymbol{\mu}_j(\alpha)}{\beta_j}(1/r_j-1/p_\pm)\\ &= \big[\tau(p_\pm)\beta(0)\big]^{-1}\big[1/\omega(\alpha)-1/(\beta_\pm(\alpha)p_\pm)\big].\end{aligned}$$

Moreover, in view of the latter inequality

$$1/\omega(\alpha) - 1/\upsilon(\alpha) = 1/\omega(\alpha) - 1/(p_\pm\gamma_\pm(\alpha))$$
$$= 1/\omega(\alpha) - 1/(p_\pm\beta_\pm(\alpha)) - \big[\tau(p_\pm)\beta(0)p_\pm\big]^{-1}\big[1/\omega(\alpha) - 1/(\beta_\pm(\alpha)p_\pm)\big]$$
$$= \big\{1 - \big[\tau(p_\pm)\beta(0)p_\pm\big]^{-1}\big\}\big[1/\omega(\alpha) - 1/(\beta_\pm(\alpha)p_\pm)\big].$$

Note that $1 - \big[\tau(p_\pm)\beta(0)p_\pm\big]^{-1} = \tau(\infty)/\tau(p_\pm)$ and the lemma follows. ∎

References.

Akakpo, N. (2012). Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection. *Math. Methods Statist.* **21**, 1–28.

Butucea, C. and Tsybakov, A. (2008). Sharp optimality for density deconvolution with dominating bias. *Theory of Probability and Its Applications*, **52**, 1, pp 111-128.

Birgé, L. (2008). Model selection for density estimation with $\mathbb{L}_2$–loss. `arXiv:0808.1416v2, http://arxiv.org`

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, **34**, 3, 431–452.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 2, 569–609.

Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25**, 2626–2637.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539.

Efroimovich, S. Yu. (1986). Non-parametric estimation of the density with unknown smoothness. *Theory Probab. Appl.* **30**, 557–568.

Folland, G. B. (1999). *Real Analysis.* Second edition. Wiley, New York.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 3, 1257–1272.

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, **21**, 2, 600–610.

Fan, J. and Koo, J. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory*, **48**, 734–747.

Gach, D., Nickl, R., and Spokoiny, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. H.Poincaré (Probab-Stat)* **49**, 900–914.

Giné E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with application to adaptive estimation. *Probab. Theory and Related Fields* **143**, 569–596.

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

Goldenshluger, A. and Lepski, O.V. (2014). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory Related Fields*, 159, 479–543.

Golubev, G.K. (1992). Non-parametric estimation of smooth probability densities. *Probl. Inform. Transm.* **1**, 52–62.

Grafakos, L. (2008). *Classical Fourier Analysis.* Graduate texts in Mathematics, second edition, Vol.249. Springer scence+Business media.

de Guzman, M. (1975). *Differentiation of Integrals in R^n.* With appendices by Antonio Crdoba, and Robert Fefferman, and two by Roberto Moriyn. Lecture Notes in Mathematics, Vol. 481. Springer-Verlag, Berlin-New York.

Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.*, **35**, 4, 1535–1558.

Hasminskii, R. and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18**, 999–1010.

Hesse, C. H. (1995). Deconvolving a density from partially contaminated observations. *Journal of Multivariate Analysis* **55**, 246–260.

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

Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi–index denoising. *Probab. Theory Related Fields* **121**, 137–170.

Kerkyacharian, G., Thanh, M. and Picard, D. (2011). Localized spherical deconvilution. *Ann. Statist.* **39**, 2, 1042–1068.

Lepski, O.V. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Ann. Statist.* **41**, 2, 1005–1034.

Lepski, O.V. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43**, 3, 1178–1242.

Lepski, O.V., and Willer T. (2017). Lower bounds in the convolution structure density model. *Bernoulli*, **23**, 2, 884-926.

Lepski, O.V. (2017). A new approach to estimator selection. *Bernoulli, to appear.* `arXiv:1603.03934v1, http://arxiv.org`.

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

Masry, E. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic processes and their applications*, **47**, 1, 53–74.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics,* **193**, Lecture Notes in Statistics. Springer-Verlag, Berlin.

Nikol'skii, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya.* (in Russian). [Approximation of functions of several variables and imbedding theorems.] Second edition, revised and supplemented. Nauka, Moscow.

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

Rebelles, G. (2016) Structural adaptive deconvolution under L_p-losses. *Math. Methods Statist.* **25**, 1, 26-53.

Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2011). Adaptive density estimation: a curse of support? *J. Statist. Plann. Inference* **141**, 115-139.

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

Rigollet, P. and Tsybakov, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16**, 260–280.

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

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, **9**, 229–235.

Stefanski, L. A. and Carroll, R.J.(1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 1696-184.

Yuana, M. and Chenb, J. (2002). Deconvolving multidimensional density from partially contaminated observations. *Journal of Statistical Planning and Inference.* **104**, 147–160.

Institut de Mathématique de Marseille
Aix-Marseille Université
39, rue F. Joliot-Curie
13453 Marseille, France
E-mail: oleg.lepski@univ-amu.fr
E-mail: thomas.willer@univ-amu.fr