On adaptive minimax density estimation on R^d

A. Goldenshluger * **O. Lepski**

Department of Statistics
University of Haifa
Mount Carmel
Haifa 31905, Israel
e-mail: goldensh@stat.haifa.ac.il

Laboratoire d'Analyse, Topologie, Probabilités
Université Aix-Marseille
39, rue F. Joliot-Curie
13453 Marseille, France
e-mail: lepski@cmi.univ-mrs.fr

Abstract: We address the problem of adaptive minimax density estimation on $\mathbb{R}^d$ with $\mathbb{L}_p$–loss on the anisotropic Nikol'skii classes. We fully characterize behavior of the minimax risk for different relationships between regularity parameters and norm indexes in definitions of the functional class and of the risk. In particular, we show that there are four different regimes with respect to the behavior of the minimax risk. We develop a single estimator which is (nearly) optimal in order over the complete scale of the anisotropic Nikol'skii classes. Our estimation procedure is based on a data-driven selection of an estimator from a fixed family of kernel estimators.

AMS 2000 subject classifications: 62G05, 62G20.
Keywords and phrases: density estimation, oracle inequality, adaptive estimation, kernel estimators, $\mathbb{L}_p$–risk.

1. Introduction

Let $X_1, \dots, X_n$ be independent copies of random vector $X \in \mathbb{R}^d$ having density f with respect to the Lebesgue measure. We want to estimate f using observations $X^{(n)} = (X_1, \dots, X_n)$. By estimator we mean any $X^{(n)}$-measurable map $\hat{f} : \mathbb{R}^n \to \mathbb{L}_p\big(\mathbb{R}^d\big)$. Accuracy of an estimator $\hat{f}$ is measured by the $\mathbb{L}_p$–risk

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

where $\mathbb{E}_f$ denotes expectation with respect to the probability measure $\mathbb{P}_f$ of the observations $X^{(n)} = (X_1, \dots, X_n)$, and $\|\cdot\|_p$, $p \in [1, \infty)$, is the $\mathbb{L}_p$-norm on $\mathbb{R}^d$. The objective is to construct an estimator of f with small $\mathbb{L}_p$–risk.

In the framework of the minimax approach density f is assumed to belong to a functional class Σ, which is specified on the basis of prior information on f. Given a functional class Σ, a natural accuracy measure of an estimator $\hat{f}$ is its maximal $\mathbb{L}_p$–risk over Σ,

$$\mathcal{R}_p^{(n)}[\hat{f}; \Sigma] = \sup_{f \in \Sigma} \mathcal{R}_p^{(n)}[\hat{f}, f].$$

The main question is:

*Supported by the ISF grant No. 104/11.

(i) how to construct a *rate–optimal*, or *optimal in order*, estimator $\hat{f}_*$ such that

$$\mathcal{R}_p^{(n)}[\hat{f}_*;\Sigma] \asymp \phi_n(\Sigma) := \inf_{\hat{f}} \mathcal{R}_p^{(n)}[\hat{f};\Sigma], \quad n\to\infty?$$

Here the infimum is taken over all possible estimators. We refer to the outlined problem as the *problem of minimax density estimation with $\mathbb{L}_p$–loss on the class Σ.*

Although the minimax approach provides a fair and convenient criterion for comparison between different estimators, it lacks some flexibility. Typically Σ is a class of functions that is determined by some *hyper-parameter*, say, α. (We write $\Sigma = \Sigma_\alpha$ in order to indicate explicitly dependence of the class Σ on the corresponding hyper-parameter α.) In general, it turns out that an estimator which is optimal in order on the class Σ_α is not optimal on the class $\Sigma_{\alpha'}$. This fact motivates the following question:

(ii) is it possible to construct an estimator $\hat{f}_*$ that is optimal in order on some scale of functional classes $\{\Sigma_\alpha, \alpha\in A\}$ and not only on one class Σ_α? In other words, is it possible to construct an estimator $\hat{f}_*$ such that for any $\alpha\in A$ one has

$$\mathcal{R}^{(n)}[\hat{f}_*;\Sigma_\alpha] \asymp \phi_n(\Sigma_\alpha), \quad n\to\infty?$$

We refer to this question as the *problem of adaptive minimax density estimation on the scale of classes $\{\Sigma_\alpha, \alpha\in A\}$.*

The minimax and adaptive minimax density estimation with $\mathbb{L}_p$–loss is a subject of the vast literature, see for example Bretagnolle and Huber (1979), Ibragimov and Khasminskii (1980, 1981), Devroye and Györfi (1985), Devroye and Lugosi (1996, 1997, 2001), Efroimovich (1986, 2008), Hasminskii and Ibragimov (1990), Donoho et al. (1996), Golubev (1992), Kerkyacharian, Picard and Tribouley (1996), Rigollet (2006), Massart (2007)[Chapter 7], Samarov and Tsybakov (2007), Rigollet and Tsybakov (2007) and Birgé (2008). It is not our aim here to provide a complete review of the literature on density estimation with $\mathbb{L}_p$-loss. Below we will only discuss results that are directly related to our study. First we review papers dealing with the one–dimensional setting; then we proceed with the multivariate case.

The problem of minimax density estimation on $\mathbb{R}^1$ with $\mathbb{L}_p$–loss, $p\in[2,\infty)$, was studied by Bretagnolle and Huber (1979). In this paper the functional class Σ is the class of all densities such that $\left[\|f^{(\beta)}\|_p \|f\|_{p/2}^{\beta}\right]^{1/(2\beta+1)} \le L < \infty$, where $f^{(\beta)}$ is the generalized derivative of order β. It was shown there that

$$\phi_n(\Sigma) \asymp n^{-\frac{1}{2+1/\beta}}, \ \ \forall p\in[2,\infty).$$

Note that the same parameter p appears in the definitions of the risk and of the functional class.

The problem of adaptive minimax density estimation on a compact interval of $\mathbb{R}^1$ with $\mathbb{L}_p$–loss was addressed in Donoho et al. (1996). In this paper class Σ is the Besov functional class $\mathbb{B}_{r\theta}^{\beta}(L)$, where parameter β stands for the regularity index, and r is the index of the norm in which the regularity is measured. It is shown there that there is an elbow in the rates of convergence for the minimax risk according to whether $p \le r(2\beta+1)$ (called in the literature *the dense zone*) or $p \ge r(2\beta+1)$ (*the sparse zone*). In particular,

$$\phi_n\big(\mathbb{B}_{r\theta}^{\beta}(L)\big) \ge \begin{cases} n^{-\frac{1}{2+1/\beta}}, & p \le r(2\beta+1), \\ (\ln n/n)^{\frac{1-1/(\beta r)+1/(\beta p)}{1-1/(\beta r)+1/(2\beta)}}, & p \ge r(2\beta+1). \end{cases} \tag{1.1}$$

Donoho et al. (1996) develop a wavelet–based hard–thresholding estimator that achieves the indicated rates (up to a $\ln n$–factor in the dense zone) for a scale of the Besov classes $\mathbb{B}_{r,\theta}^{\beta}(L)$ under additional assumption $\beta r > 1$.

It is quite remarkable that if the assumption that the underlying density has compact support is dropped, then the minimax risk behavior becomes completely different. Specifically, Juditsky and Lambert–Lacroix (2004) studied the problem of adaptive minimax density estimation on $\mathbb{R}^1$ with Σ being the Hölder class $\mathbb{N}_{\infty,1}(\beta, L)$. Their results are in striking contrast with those of Donoho et al. (1996): it is shown that

$$\phi_n\big(\mathbb{N}_{\infty,1}(\beta, L)\big) \geq \begin{cases} n^{-\frac{1}{2+1/\beta}}, & p > 2 + 1/\beta, \\ n^{-\frac{1-1/p}{1+1/\beta}}, & 1 \leq p \leq 2 + 1/\beta. \end{cases}$$

Juditsky and Lambert–Lacroix (2004) develop a wavelet–based estimator that achieves the indicated rates up to a logarithmic factor on a scale of the Hölder classes. Note that if the aforementioned results of Donoho et al. (1996) for densities with compact support are applied to the Hölder class, $r = \infty$, then the rate is $n^{-1/(2+1/\beta)}$ for any $p \geq 1$. Thus, the rate corresponding to the zone $1 \leq p \leq 2 + 1/\beta$, does not appear in the case of compactly supported densities.

In a recent paper, Reynaud–Bouret et al. (2011) consider the problem of adaptive density estimation on $\mathbb{R}^1$ with $\mathbb{L}_2$–losses on the Besov classes $\mathbb{B}^{\beta}_{r\theta}(L)$. It is shown there that

$$\phi_n\big((\mathbb{B}^{\beta}_{r\theta}(L)\big) \geq \begin{cases} n^{-\frac{1}{2+1/\beta}}, & 2/(2\beta+1) < r \leq 2, \\ n^{-\frac{1}{1-1/(\beta r)+1/\beta}}, & r > 2. \end{cases}$$

They also proposed a wavelet–based estimator that achieves the indicated rates up to a logarithmic factor for a scale of Besov classes under additional assumption $2\beta r > 2 - r$. It follows from Donoho et al. (1996) that if $p = 2$ and the density is compactly supported then the corresponding rates are $\phi_n(\Sigma) \asymp n^{-1/(2+1/\beta)}$ for all $r \geq 2/(2\beta + 1)$. Hence the rate corresponding to the zone $r > 2$, $p = 2$, does not appear in the case of the compactly supported densities.

As for the multivariate setting, Ibragimov and Khasminskii in a series of papers [Ibragimov and Khasminskii (1980, 1981), and Hasminskii and Ibragimov (1990)] studied the problem of minimax density estimation with $\mathbb{L}_p$–loss on $\mathbb{R}^d$. Together with some classes of infinitely differentiable densities, they considered the anisotropic Nikolskii's classes $\Sigma = \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$, where $\vec{\beta} = (\beta_1, \ldots, \beta_d)$, $\vec{r} = (r_1, \ldots, r_d)$ and $\vec{L} = (L_1, \ldots, L_d)$ (for the precise definition see Section 3.1). It was shown that if $r_i = p$ for all $i = 1, \ldots, d$ then

$$\phi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})\big) \asymp \begin{cases} n^{-\frac{1-1/p}{1-1/(\beta p)+1/\beta}}, & p \in [1, 2), \\ n^{-\frac{1}{2+1/\beta}}, & p \in [2, \infty). \end{cases} \tag{1.2}$$

Here β is the parameter defined by the relation $1/\beta = \sum_{j=1}^{d} 1/\beta_j$. It should be stressed that in the cited papers the same norm index p is used in the definitions of the risk and of the functional class. We also refer to the recent paper by Mason (2009), where further discussion of these results can be found.

Delyon and Juditsky (1996) generalized the results of Donoho et al. (1996) to the minimax density estimation on a bounded interval of $\mathbb{R}^d$, $d \geq 1$ over a collection of the isotropic Besov classes. In particular, they showed that the minimax rates of convergence given by (1.1) hold with $1/(\beta r)$ and $1/\beta$ replaced by $d/(\beta r)$ and d/β respectively. Comparing rates in (1.2) with the asymptotics of minimax risk found in Delyon and Juditsky (1996) with $r = p$ we conclude that the rate in (1.2) in the zone $p \in [1, 2)$ does not appear for compactly supported densities.

Recently Goldenshluger and Lepski (2011b) developed an adaptive minimax estimator over a scale of classes $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$; in particular, if $r_i = p$ for all $i = 1, \ldots, d$ then their estimator attains

the minimax rates indicated in (1.2). Note that in the considered setting the norm indexes in the definitions of the risk and the functional class coincide.

The results discussed above show that there is an essential difference between the problems of density estimation on the whole space and on a compact interval. The literature on density estimation on the whole space is quite fragmented, and relationships between aforementioned results are yet to be understood. These relationships become even more complex and interesting in the multivariate setting where the density to be estimated belongs to a functional class with anisotropic and inhomogeneous smoothness. The problem of minimax estimation under $\mathbb{L}_p$–loss over homogeneous Sobolev $\mathbb{L}_q$–balls ($q \neq p$) was initiated in Nemirovski (1985) in the regression model on the unit cube of $\mathbb{R}^d$. For the first time, functional classes with anisotropic and inhomogeneous smoothness were considered in Kerkyacharian et al. (2001, 2008) for the Gaussian white noise model on a compact subset of $\mathbb{R}^d$. In the density estimation model Akakpo (2012) studied the case $p = 2$ and considered compactly supported densities on $[0, 1]^d$.

To the best of our knowledge, the problem of estimating a multivariate density from anisotropic and inhomogeneous functional classes on $\mathbb{R}^d$ was not considered in the literature. This problem is a subject of the current paper. Our results cover the existing ones and generalize them in the following directions.

1. We fully characterize behavior of the minimax risk for all possible relationships between regularity parameters and norm indexes in the definition of the functional classes and of the risk. In particular, we discover that there are four different regimes with respect to the minimax rates of convergence: *tail, dense and sparse zones*, and the last zone, in its turn, is subdivided in two regions. Existence of these regimes is not a consequence of the multivariate nature of the problem or the considered functional classes; in fact, these regimes appear already in the dimension one. Thus our results reveal all possible zones with respect to the rates of convergence in the problem of density estimation on $\mathbb{R}^d$ and explain different results on rates of convergence in the existing literature. In particular, results in Juditsky and Lambert–Lacroix (2004) and Reynaud–Bouret et al. (2011) pertain to the rates of convergence in the tail and dense zones, while those in Donoho et al. (1996) and Delyon and Juditsky (1996) correspond to the dense zone and to a subregion of the sparse zone.

2. We propose an estimator that is based upon a data–driven selection from a family of kernel estimators, and establish for it a point–wise oracle inequality. Then we use this inequality for derivation of bounds on the $\mathbb{L}_p$–risk over a collection of the Nikol'skii functional classes. Since the construction of our estimator does not use any prior information on the class parameters, it is adaptive minimax over a scale of these classes. Moreover, we believe that the method of deriving $\mathbb{L}_p$–risk bounds from point–wise oracle inequalities employed in the proof of Theorem 2 is of interest in its own right. It is quite general and can be applied to other nonparametric estimation problems.

3. Another issue studied in the present paper is related to the existence of the tail zone. This zone does not exist in the problem of estimating compactly supported densities. Then a natural question arises: what is a general condition on f which ensures the same asymptotics of the minimax risk on $\mathbb{R}^d$ as in the case of compactly supported densities? We propose a *tail dominance condition* and show that, in a sense, it is the weakest possible condition under which the tail zone disappears. We also show that this condition guarantees existence of a consistent estimator under $\mathbb{L}_1$-loss. Recall that smoothness alone is not sufficient in order to guarantee consistency of density estimators in $\mathbb{L}_1(\mathbb{R}^d)$ [see Ibragimov and Khasminskii (1981)].

The paper is structured as follows. In Section 2 we define our estimation procedure and derive the corresponding point–wise oracle inequality. Section 3 presents upper and lower bounds on

the minimax risk. We also discuss the obtained results and relate them to the existing results in the literature. The same estimation problem under the tail dominance condition is studied in Section 4. Sections 5–7 contain proofs of Theorems 1–4; proofs of auxiliary results are relegated to Appendices A and B.

The following notation and conventions are used throughout the paper. For vectors $u, v \in \mathbb{R}^d$ the operations u/v, $u \vee v$, $u \wedge v$ and inequalities such as $u \le v$ are all understood in the coordinate–wise sense. For instance, $u \vee v = (u_1 \vee v_1, \dots, u_d \vee v_d)$. All integrals are taken over $\mathbb{R}^d$ unless the domain of integration is specified explicitly. For a Borel set $\mathcal{A} \subset \mathbb{R}^d$ symbol $|\mathcal{A}|$ stands for the Lebesgue measure of $\mathcal{A}$; if $\mathcal{A}$ is a finite set, $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$.

2. Estimation procedure and point–wise oracle inequality

In this section we define our estimation procedure and derive an upper bound on its point–wise risk.

2.1. Estimation procedure

Our estimation procedure is based on data-driven selection from a family of kernel estimators. The family of estimators is defined as follows.

2.1.1. Family of kernel estimators

Let $K : [-1/2, 1/2]^d \to \mathbb{R}^1$ be a fixed kernel such that $K \in \mathbb{C}(\mathbb{R}^d)$, $\int K(x)\mathrm{d}x = 1$, and $\|K\|_\infty < \infty$. Let

$$\mathcal{H} = \Big\{ h = (h_1, \dots, h_d) \in (0,1]^d : \ h_j = 2^{-k_j}, k_j = 0, \dots, \log_2 n, \ j = 1, \dots, d \Big\};$$

without loss of generality we assume that $\log_2 n$ is integer.

Given a *bandwidth* $h \in \mathcal{H}$, define the corresponding kernel estimator of f by the formula

$$\hat{f}_h(x) := \frac{1}{nV_h} \sum_{i=1}^n K\left(\frac{X_i - x}{h} \right) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - x), \tag{2.1}$$

where $V_h := \prod_{j=1}^d h_j$, $K_h(\cdot) := (1/V_h)K(\cdot/h)$. Consider the family of kernel estimators

$$\mathcal{F}(\mathcal{H}) := \{\hat{f}_h, h \in \mathcal{H}\}.$$

The proposed estimation procedure is based on data–driven selection of an estimator from $\mathcal{F}(\mathcal{H})$.

2.1.2. Auxiliary estimators

Our selection rule uses auxiliary estimators that are constructed as follows. For any pair $h, \eta \in \mathcal{H}$ define the kernel $K_h * K_\eta$ by the formula $[K_h * K_\eta](t) = \int K_h(t-y)K_\eta(y)\mathrm{d}y$. Let $\hat{f}_{h,\eta}(x)$ denote the estimator associated with this kernel:

$$\hat{f}_{h,\eta}(x) = \frac{1}{n} \sum_{i=1}^n K_{h,\eta}(X_i - x), \quad K_{h,\eta} = K_h * K_\eta.$$

The following representation of kernels $K_{h,\eta}$ will be useful: for any $h, \eta \in \mathcal{H}$

$$[K_h * K_\eta](t) \;=\; \frac{1}{V_{h\vee\eta}} Q_{h,\eta}\Big(\frac{t}{h\vee\eta}\Big), \tag{2.2}$$

where function $Q_{h,\eta}$ is given by the formula

$$Q_{h,\eta}(t) = \int K\big(v(y, t-\nu y)\big) K\big(v(t-\nu y, y)\big)\mathrm{d}y, \quad \nu := \frac{h\wedge\eta}{h\vee\eta}. \tag{2.3}$$

Here function $v : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is defined by

$$v_j(y,z) = \begin{cases} y_j, & h_j \le \eta_j, \\ z_j, & h_j > \eta_j, \end{cases}, \qquad j = 1, \ldots, d.$$

The representation (2.2)–(2.3) is obtained by a straightforward change of variables in the convolution integral [see the proof of Lemma 12 in Goldenshluger and Lepski (2011a)]. We also note that $\mathrm{supp}(Q_{h,\eta}) \subseteq [-1,1]^d$, and $\|Q_{h,\eta}\|_\infty \le \|K\|_\infty^2$ for all h, η. In the special case where $K(t) = \prod_{i=1}^d k(t_i)$ for some univariate kernel $k : [-1/2, 1/2] \to \mathbb{R}^1$ we have

$$Q_{h,\eta}(t) = \prod_{i=1}^d \int k(t_i - \nu_i u_i) k(u_i)\mathrm{d}u_i, \quad \nu_i = (h_i \wedge \eta_i)/(h_i \vee \eta).$$

We also define

$$Q(t) = \sup_{h,\eta\in\mathcal{H}} \Big| \int K\big(v(y, t-\nu y)\big) K\big(v(t-\nu y, y)\big)\mathrm{d}y\Big|,$$

and note that $\mathrm{supp}(Q) \subseteq [-1,1]^d$, and $\|Q\|_\infty \le \|K\|_\infty^2$.

2.1.3. Stochastic errors of kernel estimators and their majorants

Uniform moment bounds on stochastic errors of kernel estimators $\hat{f}_h(x)$ and $\hat{f}_{h,\eta}(x)$ will play an important role in the construction of our selection rule. Let

$$\begin{aligned} \xi_h(x) \;&=\; \frac{1}{n}\sum_{i=1}^n K_h(X_i - x) - \int K_h(t-x) f(t)\mathrm{d}t, \\ \xi_{h,\eta}(x) \;&=\; \frac{1}{n}\sum_{i=1}^n K_{h,\eta}(X_i - x) - \int K_{h,\eta}(t-x) f(t)\mathrm{d}t \end{aligned} \tag{2.4}$$

denote the stochastic errors of $\hat{f}_h$ and $\hat{f}_{h,\eta}$ respectively. In order to construct our selection rule we need to find uniform upper bounds (*majorants*) on ξ_h and $\xi_{h,\eta}$, i.e. we need to find functions M_h and $M_{h,\eta}$ such that moments of random variables

$$\sup_{h\in\mathcal{H}} \big[|\xi_h(x)| - M_h(x)\big]_+, \quad \sup_{h,\eta\in\mathcal{H}} \big[|\xi_{h,\eta}(x)| - M_{h,\eta}(x)\big]_+ \tag{2.5}$$

are "small" for each $x \in \mathbb{R}^d$. We will be also interested in the integrability properties of these moments.

It turns out that the majorants $M_h(x)$ and $M_{h,\eta}(x)$ can be defined in the following way. For a function $g : \mathbb{R}^d \to \mathbb{R}^1$ let

$$A_h(g,x) = \int |g_h(t-x)|f(t)\mathrm{d}t, \quad g_h(\cdot) = V_h^{-1} g\big(\,\cdot\,/h\big), \quad h \in \mathcal{H}. \tag{2.6}$$

Now define

$$M_h(g,x) = \sqrt{\frac{\varkappa A_h(g,x)\ln n}{nV_h}} + \frac{\varkappa \ln n}{nV_h}, \tag{2.7}$$

where $\varkappa$ is a positive constant to be specified. In Lemma 2 in Section 5 we show that under appropriate choice of parameter $\varkappa$ functions

$$M_h(x) := M_h(K,x), \quad M_{h,\eta}(x) := M_{h\vee\eta}(Q,x) \tag{2.8}$$

uniformly majorate ξ_h and $\xi_{h,\eta}$ in the sense that the moments of random variables in (2.5) are "small".

It should be noted, however, that functions $M_h(x)$ and $M_{h,\eta}(x)$ given by (2.8) cannot be directly used in construction of the selection rule because they depend on unknown density f to be estimated. We will use empirical counterparts of $M_h(x)$ and $M_{h,\eta}(x)$ instead.

For $g : \mathbb{R}^d \to \mathbb{R}^1$ we let

$$\hat{A}_h(g,x) = \frac{1}{n}\sum_{i=1}^{n} |g_h(X_i - x)|,$$

and define

$$\hat{M}_h(g,x) = 4\sqrt{\frac{\varkappa \hat{A}_h(g,x)\ln n}{nV_h}} + \frac{4\varkappa \ln n}{nV_h}. \tag{2.9}$$

2.1.4. Selection rule and final estimator

Now we are in a position to define our selection rule. For every $x \in \mathbb{R}^d$ let

$$\begin{aligned}\hat{R}_h(x) \;=\; \sup_{\eta\in\mathcal{H}} \Big[|\hat{f}_{h,\eta}(x) - \hat{f}_\eta(x)| - \hat{M}_{h\vee\eta}(Q,x) - \hat{M}_\eta(K,x)\Big]_+ \\ +\; \sup_{\eta\ge h} \hat{M}_\eta(Q,x) + \hat{M}_h(K,x), \qquad h \in \mathcal{H}.\end{aligned} \tag{2.10}$$

The selected bandwidth $\hat{h}(x)$ and the corresponding estimator are defined by

$$\hat{h}(x) = \arg\inf_{h\in\mathcal{H}} \hat{R}_h(x), \quad \hat{f}(x) = \hat{f}_{\hat{h}(x)}(x), \quad x \in \mathbb{R}^d. \tag{2.11}$$

Note that the estimation procedure is completely determined by the family of kernel estimators $\mathcal{F}(\mathcal{H})$ and by the constant $\varkappa$ appearing in the definition of $\hat{M}_h$.

We have to ensure that the map $x \mapsto \hat{f}_{\hat{h}(x)}(x)$ is an $X^{(n)}$-measurable Borel function. This follows from continuity of K and the fact that $\mathcal{H}$ is a discrete set; for details see Appendix A, Section A.1.

The main idea behind the construction of the selection procedure (2.10)–(2.11) is the following. The expression $\hat{M}_{h\vee\eta}(Q,x) + \hat{M}_\eta(K,x)$ appearing in the square brackets in (2.10) dominates with large probability the stochastic part of the difference $|\hat{f}_{h,\eta}(x) - \hat{f}_\eta(x)|$. Consequently, the first term on the right hand side of (2.10) serves as a proxy for the deterministic part of $|\hat{f}_{h,\eta}(x) - \hat{f}_\eta(x)|$ which is the absolute value of the difference of biases of kernel estimates $\hat{f}_{h,\eta}(x)$ and $\hat{f}_\eta(x)$. The latter, in its own turn, is closely related to the bias of the estimator $\hat{f}_h(x)$. Thus, the first term on the right hand side of (2.10) is a proxy for the bias of $\hat{f}_h(x)$, while the second term is an upper bound on the standard deviation of $\hat{f}_h(x)$.

2.2. Point–wise oracle inequality

Let $B_h(f,t)$ be the bias of the kernel estimator $\hat{f}_h(t)$,

$$B_h(f,t) = \int K_\eta(y-t)f(y)\mathrm{d}y - f(t), \tag{2.12}$$

and define

$$\bar{B}_h(f,x) = |B_h(f,x)| \vee \sup_{\eta\in\mathcal{H}} \Big| \int K_\eta(t-x)B_h(f,t)\mathrm{d}t\Big|. \tag{2.13}$$

Theorem 1. *For any $x \in \mathbb{R}^d$ one has*

$$|\hat{f}(x) - f(x)| \le \inf_{h\in\mathcal{H}} \big\{4\bar{B}_h(f,x) + 60\sup_{\eta\ge h} M_\eta(Q,x) + 61M_h(K,x)\big\} + 7\zeta(x) \ + \ 18\chi(x), \tag{2.14}$$

where

$$\zeta(x) \ := \ \sup_{h\in\mathcal{H}}[|\xi_h(x)| - M_h(K,x)]_+ \ \vee \ \sup_{h,\eta\in\mathcal{H}}[|\xi_{h,\eta}(x)| - M_{h\vee\eta}(Q,x)]_+, \tag{2.15}$$

$$\chi(x) \ := \ \max_{g\in\{K,Q\}}\sup_{h\in\mathcal{H}}\big[|\hat{A}_h(g,x) - A_h(g,x)| - M_h(g,x)\big]_+. \tag{2.16}$$

Furthermore, for any $q \ge 1$ if $\varkappa \ge [\|K\|_\infty \vee 1]^2[(4d+2)q + 4(d+1)]$ then

$$\int \mathbb{E}_f\big\{[\zeta(x)]^q + [\chi(x)]^q\big\}\mathrm{d}x \ \le \ Cn^{-q/2}, \ \ \forall n \ge 3, \tag{2.17}$$

where C is the constant depending on d, q and $\|K\|_\infty$ only.

We remark that Theorem 1 does not require any conditions on the estimated density f.

3. Adaptive estimation over anisotropic Nikol'skii classes

In this section we study properties of the estimator defined in (2.10)–(2.11). The point–wise oracle inequality of Theorem 1 is the key technical tool for bounding $\mathbb{L}_p$-risk of this estimator on the anisotropic Nikol'skii classes.

3.1. Anisotropic Nikol'skii classes

Let $(e_1,\dots,e_d)$ denote the canonical basis of $\mathbb{R}^d$. For 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

$$\Delta_{u,j}g(x) = g(x+ue_j) - g(x), \quad j = 1,\dots,d.$$

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

$$\Delta^k_{u,j}g(x) = \Delta_{u,j}\Delta^{k-1}_{u,j}g(x) = \sum_{l=1}^{k}(-1)^{l+k}\binom{k}{l}\Delta_{ul,j}g(x). \tag{3.1}$$

Definition 1. *For given real numbers* $\vec{r} = (r_1, \ldots, r_d)$, $r_j \in [1, \infty]$, $\vec{\beta} = (\beta_1, \ldots, \beta_d)$, $\beta_j > 0$, *and* $\vec{L} = (L_1, \ldots, L_d)$, $L_j > 0$, $j = 1, \ldots, d$, *we say that function* $g : \mathbb{R}^d \to \mathbb{R}^1$ *belongs to the anisotropic Nikol'skii class* $\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$ *if*

(i) $\|g\|_{r_j} \le L_j$ *for all* $j = 1, \ldots, d$;
(ii) *for every* $j = 1, \ldots, d$ *there exists natural number* $k_j > \beta_j$ *such that*

$$\left\| \Delta_{u,j}^{k_j} g \right\|_{r_j} \le L_j |u|^{\beta_j}, \quad \forall u \in \mathbb{R}^d, \quad \forall j = 1, \ldots, d. \tag{3.2}$$

The anisotropic Nikol'skii class is a specific case of the anisotropic Besov class, often encountered in the nonparametric estimation literature. In particular, $\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \cdot\big) = \mathbb{B}^{\beta_1,\ldots,\beta_d}_{r_1,\ldots,r_d;\infty,\ldots,\infty}(\cdot)$, see (Nikol'skii 1977, Section 4.3.4).

3.2. Construction of kernel K

We will use the following specific kernel K in the definition of the family $\mathcal{F}(\mathcal{H})$ [see, e.g., Kerkyacharian et al. (2001) or Goldenshluger and Lepski (2011b)].

Let ℓ be an integer number, and let $w : [-1/(2\ell), 1/(2\ell)] \to \mathbb{R}^1$ be a function satisfying $\int w(y)\mathrm{d}y = 1$, and $w \in \mathbb{C}(\mathbb{R}^1)$. Put

$$w_\ell(y) = \sum_{i=1}^{\ell} \binom{\ell}{i} (-1)^{i+1} \frac{1}{i} w\Big(\frac{y}{i}\Big), \qquad K(t) = \prod_{j=1}^{d} w_\ell(t_j), \quad t = (t_1, \ldots, t_d). \tag{3.3}$$

The kernel K constructed in this way is bounded, supported on $[-1/2, 1/2]^d$, belongs to $\mathbb{C}(\mathbb{R}^d)$ and satisfies

$$\int K(t)\mathrm{d}t = 1, \quad \int K(t) t^k \mathrm{d}t = 0, \ \ \forall |k| = 1, \ldots, \ell - 1,$$

where $k = (k_1, \ldots, k_d)$ is the multi–index, $k_i \ge 0$, $|k| = k_1 + \cdots + k_d$, and $t^k = t_1^{k_1} \cdots t_d^{k_d}$ for $t = (t_1, \ldots, t_d)$.

3.3. Main results

Let $\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$ be the anisotropic Nikol'skii functional class. Put

$$\frac{1}{\beta} := \sum_{j=1}^{d} \frac{1}{\beta_j}, \qquad \frac{1}{s} := \sum_{j=1}^{d} \frac{1}{\beta_j r_j}, \qquad L_\beta := \prod_{j=1}^{d} L_j^{1/\beta_j},$$

and define

$$\nu = \begin{cases} \frac{1-1/p}{1-1/s+1/\beta}, & p < \frac{2+1/\beta}{1+1/s}, \\ \frac{\beta}{2\beta+1}, & \frac{2+1/\beta}{1+1/s} \le p \le s(2+1/\beta), \\ s/p, & p > s(2+1/\beta), \ s < 1, \\ \frac{1-1/s+1/(p\beta)}{2-2/s+1/\beta}, & p > s(2+1/\beta), \ s \ge 1, \end{cases} \tag{3.4}$$

$$\mu_n = \begin{cases} \ln^{\frac{d}{p}}(n), & p \le \frac{2+1/\beta}{1+1/s}; \\ (\ln n)^{1/p}, & p = s(2+1/\beta), \\ 1, & \text{otherwise.} \end{cases}$$

In contrast to Theorem 1 proved over the set of all probability densities, the adaptive results presented below require the additional assumption: the estimated density should be uniformly bounded. For this purpose we define for $M > 0$

$$\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L},M\big) \;:=\; \mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big) \;\cap\; \{f : \|f\|_\infty \le M\}\,.$$

Note, however, that if $J := \{j = 1,\dots,d : \;\; r_j = \infty\}$ then $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L},M\big) = \mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ with $M = \inf_J L_j$. Moreover, in view of the embedding theorem for the anisotropic Nikol'skii classes [see Section 6.1 below], condition $s > 1$ implies that the density to be estimated belongs to a class of uniformly bounded and continuous functions. Thus, if $s > 1$ one has $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L},M\big) = \mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ with some M completely determined by $\vec{L}$.

The asymptotic behavior of the $\mathbb{L}_p$-risk on class $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},M)$ is characterized in the next two theorems.

Let family $\mathcal{F}(\mathcal{H})$ be associated with kernel (3.3). Let $\hat{f}$ denote the estimator given by the selection rule (2.10)–(2.11) with $\varkappa = (\|K\|_\infty \vee 1)^2[(4d+2)p + 4(d+1)]$ that is applied to the family $\mathcal{F}(\mathcal{H})$.

Theorem 2. *For any $M > 0$, $L_0 > 0$, $\ell \in \mathbb{N}^*$, any $\vec{\beta} \in (0,\ell]^d$, $\vec{r} \in (1,\infty]^d$, any $\vec{L}$ satisfying $\min_{j=1,\dots,d} L_j \ge L_0$, and any $p \in (1,\infty)$ one has*

$$\limsup_{n\to\infty}\left\{\mu_n\Big(\frac{L_\beta \ln n}{n}\Big)^{-\nu} \mathcal{R}_p^{(n)}\big[\hat{f}\,;\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L},M\big)\big]\right\} \le C < \infty.$$

Here constant C does not depend on $\vec{L}$ in the cases $p \le s(2+1/\beta)$ and $p \ge s(2+1/\beta)$, $s < 1$.

Remark 1.

1. Condition $\min_{j=1,\dots,d} L_j \ge L_0$ ensures independence of the constant C on $\vec{L}$ in the cases $p \le s(2+1/\beta)$ and $p \ge s(2+1/\beta)$, $s < 1$. If $p \ge s(2+1/\beta)$, $s \ge 1$ then C depends on $\vec{L}$, and the corresponding expressions can be easily extracted from the proof of the theorem. We note that in this case the map $\vec{L} \mapsto C(\vec{L})$ is bounded on every closed cube of $(0,\infty)^d$.
2. We consider the case $1 < p < \infty$ only, not including $p = 1$ and $p = \infty$. It is well–known, Ibragimov and Khasminskii (1981), that smoothness alone is not sufficient in order to guarantee consistency of density estimators in $\mathbb{L}_1(\mathbb{R}^d)$; see also Theorem 3 for a lower bound. The case $p = \infty$ was considered recently in Lepski (2012).
3. As it was discussed above, Theorem 2 requires uniform boundedness of the estimated density, i.e. $\|f\|_\infty \le M < \infty$. We note however that our estimator $\hat{f}$ is fully adaptive, i.e., its construction does not use any information on the parameters $\vec{\beta},\vec{r},\vec{L}$ and M.

Now we present lower bounds on the minimax risk. Define

$$\alpha_n = \begin{cases} \ln n, & p > s(2+1/\beta),\; s \ge 1, \\ 1, & \text{otherwise.} \end{cases}$$

Theorem 3. *Let $\vec{\beta} \in (0,\infty)^d$, $\vec{r} \in [1,\infty]^d$, $\vec{L} \in (0,\infty)^d$ and $M > 0$ be fixed.*
(i) *There exists $c > 0$ such that*

$$\liminf_{n\to\infty}\left\{\Big(\frac{L_\beta \alpha_n}{n}\Big)^{-\nu} \inf_{\widetilde{f}} \mathcal{R}_p^{(n)}\big[\widetilde{f};\; \mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},M)\big]\right\} \ge c, \quad \forall p \in [1,\infty),$$

where the infimum is taken over all possible estimators $\widetilde{f}$. If $\min_{j=1,\dots,d} L_j \ge L_0 > 0$ then in the cases $p \le s(2+1/\beta)$ or $p \ge s(2+1/\beta)$ and $s < 1$ the constant c is independent of $\vec{L}$.

(ii) *Let $p = \infty$ and $s \le 1$; then there is no consistent estimator, i.e., for some $c > 0$*

$$\liminf_{n\to\infty} \inf_{\tilde f} \sup_{f\in \mathbb{N}_{\vec r,d}(\vec\beta,\vec L,M)} \mathbb{E}_f \big\|\tilde f - f\big\|_\infty \ > \ c.$$

Remark 2.

1. Inspection of the proof shows that if $\max_{j=1,\dots,d} L_j \le L_\infty < \infty$ then the statement (i) is valid with constant c depending on $\vec\beta, \vec r$, L_0, L_∞, d and M only.
2. As it was mentioned above, adaptive minimax density estimation on $\mathbb{R}^d$ under $\mathbb{L}_\infty$–loss was a subject of the recent paper Lepski (2012). A minimax adaptive estimator is constructed in this paper under assumption $s > 1$. Thus, statement (ii) of Theorem 3 finalizes the research on adaptive density estimation in the supremum norm. It is interesting to note that the minimax rates in the case $p = \infty$ coincide with those of Theorem 2 if we put formally $p = \infty$.

3.4. Discussion

The results of Theorem 2 together with the matching lower bounds of Theorem 3 provide complete classification of minimax rates of convergence in the problem of density estimation on $\mathbb{R}^d$. In particular, we discover four different zones with respect to the minimax rates of convergence.

- *Tail zone* corresponds to "small" p, $1 < p \le \frac{2+1/\beta}{1+1/s}$. This zone does not appear if density f is assumed to be compactly supported, or some tail dominance condition is imposed, see Section 4.
- *Dense zone* is characterized by the "intermediate" range of p, $\frac{2+1/\beta}{1+1/s} \le p \le s(2+1/\beta)$. Here the "usual" rate of convergence $n^{-\beta/(2\beta+1)}$ holds.
- *Sparse zone* corresponds to "large" p, $p \ge s(2+1/\beta)$. As Theorems 2 and 3 show, this zone, in its turn, is subdivided into two regions with $s \ge 1$ and $s < 1$. This phenomenon was not observed in the existing literature even for settings with compactly supported densities. For other statistical models (regression, white Gaussian noise etc) this result is also new.

It is important to emphasize that existence of these zones is not related to the multivariate nature of the problem or to the anistropic smoothness of the estimated density. In fact, these results hold already for the one–dimensional case, and this, to a limited degree, was observed in the previous works. In the subsequent remarks we discuss relationships between our results and the existing results in the literature, and comment on some open problems.

1. In Donoho et al. (1996), Delyon and Juditsky (1996) and Kerkyacharian et al. (2008) the sparse zone is defined as $p > 2(1+1/\beta)$, $s > 1$. Recall that condition $s > 1$ implies that the density to be estimated belongs to a class of uniformly bounded and continuous functions. In the sparse zone we consider also the case $s \le 1$, but density f is assumed to be uniformly bounded. It turns out that in this zone the rate corresponding to the index $\nu = s/p$ emerges.

2. 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 $s = \infty$, while Reynaud–Bouret et al. (2011) deal with the case of $p = 2$ and $\beta > 1/r - 1/2$. Both settings rule out the sparse zone. The rates of convergence in the dense zone obtained in the aforementioned papers are easily recovered from our results. However, in the tail zone our bound contains additional $\ln(n)$-factor.

3. In the previous papers on adaptive estimation of densities with unbounded support [cf. Juditsky and Lambert–Lacroix (2004) and Reynaud–Bouret et al. (2011)] the developed estimators are explicitly shrunk to zero. This shrinkage is used in bounding the minimax risk on the whole space. We do not employ shrinkage in our estimator construction. We derive bounds on the $\mathbb{L}_p$–risk by integration of the point–wise oracle inequality (2.14). The key elements of this derivation are inequality (2.17) and statement (i) of Proposition 2. The inequality (2.17) is based on the following fact: the errors $\zeta(x)$ and $\chi(x)$ are integrable by the accurate choice of the majorant. Indeed, Section 5.3 shows that these errors are not equal to zero with probability which is integrable and "negligible" in the regions where the density is "small". This leads to integrability of the remainders in (2.14). As for Proposition 2, it is related to the integrability of the main term in (2.14). The main problem here is that the majorant $M_h(\cdot, x)$ itself is not integrable. To overcome this difficulty we use the integrability of the estimator $\hat{f}$, approximation properties of the density f, and (2.17).

4. In the context of the Gaussian white noise model on a compact interval Kerkyacharian et al. (2001) developed an adaptive estimator that achieves the rate of convergence $(\ln n/n)^{\beta/(2\beta+1)}$ on the anisotropic Nikol'skii classes under condition $\sum_{i=1}^d [\frac{1}{\beta_i}(\frac{p}{r_i}-1)]_+ < 2$. This restriction determines a part of the dense zone, and our Theorem 2 improves on this result. In fact, our estimator achieves the rate $(\ln n/n)^{\beta/(2\beta+1)}$ in the zone $\sum_{i=1}^d \frac{1}{\beta_i}(\frac{p}{r_i}-1) \le 2$ which is equivalent to $p \le s(2+1/\beta)$.

5. It follows from Theorem 3 that the upper bound of Theorem 2 is sharp in the zone $p > s(2+1/\beta)$, $s > 1$, and it is nearly sharp up to a logarithmic factor in all other zones. This extra logarithmic factor is a consequence of the fact that we use the point–wise selection procedure (2.10)–(2.11). We also have extra $\ln n$–term on the boundaries $p = \frac{2+1/\beta}{1+1/s}$, $p = s(2+1/\beta)$.

Conjecture 1. *The rates found in Theorem 3 are optimal.*

Thus, if our conjecture is true, the construction of an estimator achieving the rates of Theorem 3 in the tail and dense zones remains an open problem.

6. Theorem 2 is proved under assumption $\vec{r} \in (1,\infty]^d$, i.e., we do not include the case where $r_j = 1$ for some $j = 1, \dots, d$. This is related to the construction of our selection rule, and to the necessity to bound $\mathbb{L}_{r_j}$–norm, $j = 1, \dots, d$ of the term $\bar{B}_h(f,x)$; see (2.13) and (2.14). In our derivations for this purpose we use properties of the strong maximal operator [for details see Section 6.1], and it is well–known that this operator is not of the weak $(1,1)$–type in dimensions $d \ge 2$. Nevertheless, using inequality (6.5) we were able to obtain the following result.

Corollary 1. *Let $\vec{r}$ be such that $r_j = 1$ for some $j = 1, \dots, d$. Then the result of Theorem 2 remains valid if the normalizing factor $\left(n^{-1}\ln n\right)^{\nu}$ is replaced by $\left(n^{-1}[\ln n]^d\right)^{\nu}$.*

The proof of Corollary 1 coincides with the proof of Theorem 2 with the only difference that bounds in the proof of Proposition 1 should use (6.5) instead of the Chebyshev inequality. This will result in an extra $(\ln n)^{d-1}$-factor. We note that the results of Theorem 2 and Corollary 1 coincide if $d = 1$. It is not surprising because in the dimension $d = 1$ the strong maximal operator is the Hardy-Littlewood maximal function which is of the weak (1,1)–type.

4. Tail dominance condition

Let $g : \mathbb{R}^d \to \mathbb{R}^1$ be a locally integrable function. Define the map $g \mapsto g^*$ by the formula

$$g^*(x) := \sup_{h \in (0,2]^d} \frac{1}{V_h} \int_{\Pi_h(x)} g(t)\mathrm{d}t, \quad x \in \mathbb{R}^d, \tag{4.1}$$

where $\Pi_h(x) = [x_1 - h_1/2, x_1 + h_1/2] \times \cdots \times [x_d - h_d/2, x_d + h_d/2]$. In fact, formula (4.1) defines the maximal operator associated with the differential basis $\cup_{x\in\mathbb{R}^d}\{\Pi_h(x), h \in (0, 2]\}$, see Guzman (1975).

Consider the following set of functions: for any $\theta \in (0, 1]$ and $R \in (0, \infty)$ let

$$\mathbb{G}_\theta(R) = \big\{g : \mathbb{R}^d \to \mathbb{R} : \ \ \|g^*\|_\theta \le R\big\}. \tag{4.2}$$

Note that, although we keep the previous notation $\|g\|_\theta = (\int |g(x)|^\theta \mathrm{d}x)^{1/\theta}$, $\|\cdot\|_\theta$ is not longer a norm if $\theta \in (0, 1)$.

The assumption that $f \in \mathbb{G}_\theta(R)$ for some $\theta \in (0, 1]$ and $R > 0$ imposes restrictions on the tail of the density f. In particular, the set of densities, uniformly bounded and compactly supported on a cube of $\mathbb{R}^d$, is embedded in the set $\mathbb{G}_\theta(\cdot)$ for any $\theta \in (0, 1]$ (for details, see Section 7.4). We will refer to the assumption $f \in \mathbb{G}_\theta(R)$ as *the tail dominance condition.*

In this section we study the problem of adaptive density estimation under the tail dominance condition. We show that under this condition the minimax rate of convergence can be essentially improved in the tail zone. In particular, if $\theta \le \theta^*$ for some $\theta^* < 1$ given below then the tail zone disappears.

For any $\theta \in (0, 1]$ let

$$\nu^*(\theta) = \max\left\{\frac{1 - \theta/p}{1 - \theta/s + 1/\beta}, \ \frac{1}{2 + 1/\beta}\right\},$$

and define

$$\nu(\theta) = \begin{cases} \nu^*(\theta), & p \le s(2 + 1/\beta), \\ \nu, & p > s(2 + 1/\beta), \end{cases} \qquad \mu_n(\theta) = \begin{cases} (\ln n)^{1/p}, & p \in \{\frac{2+1/\beta}{1/\theta+1/s}, \ s(2 + 1/\beta)\}, \\ 1, & \text{otherwise}, \end{cases} \tag{4.3}$$

where ν is defined in (3.4).

Theorem 4. *The following statements hold.*

(i) *For any $\theta \in (0, 1]$ and $R > 0$, Theorem 2 remains valid if one replaces $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ by $\mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$, ν by $\nu(\theta)$ and μ_n by $\mu_n(\theta)$. The constant C may depend on θ and R.*

(ii) *For any $\theta \in (0, 1]$, $\vec{\beta}, \vec{L} \in (0, \infty)^d$, $\vec{r} \in [1, \infty]^d$ and $M > 0$ one can find $R > 0$ such that Theorem 3 remains valid if one replaces $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ by $\mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$, ν by $\nu(\theta)$, and μ_n by $\mu_n(\theta)$.*

Remark 3.

1. The leads to improvement of the rates of convergence in the whole tail zone. In particular, if $f \in G_1(R)$ then the additional $\ln^{\frac{d}{p}}(n)$-factor disappears, cf. μ_n and $\mu_n(1)$. Moreover, if $\theta < 1$ under tail dominance condition condition the faster convergence rate of the dense zone is achieved over a wider range of values of p, $\frac{2+1/\beta}{1/\theta+1/s} \le p \le s(2 + 1/\beta)$. Additionally, if

$$\theta < \theta^* := \frac{ps}{s(2 + 1/\beta) - p},$$

then the tail zone disappears. Note that $\theta^* \in (0, 1)$ whenever $p \le \frac{2+1/\beta}{1+1/s}$.
As it was mentioned above, the set of uniformly bounded and compactly supported on a cube of $\mathbb{R}^d$ densities is embedded in the set $\mathbb{G}_\theta(\cdot)$ for any $\theta \in (0, 1]$. This fact explains why the tail zone does not appear in problems of estimating compactly supported densities.

2. We would like to emphasize that the couple (θ, R) is not used in the construction of the estimation procedure; thus, our estimator is adaptive with respect to (θ, R) as well. In particular, if the tail dominance condition does not hold, our estimator achieves the rate of Theorem 2. On the other hand, if this assumption holds, the rate of convergence is improved automatically in the tail zone.
3. The second statement of the theorem is proved under assumption that R is large enough. The fact that R cannot be chosen arbitrary small is not technical; the parameters $\vec{\beta}$, $\vec{L}$, $\vec{r}$, M,θ and R are related to each other. In particular, one can easily provide lower bounds on R in terms of the other parameters of the class. For instance, by the Lebesgue differentiation theorem, $f(x) \le f^*(x)$ almost everywhere; therefore for any density $f \in \mathbb{G}_\theta(R)$ such that $\|f\|_\infty \le M$ one has

$$1 = \int f \le M^{1-\theta}\|f^*\|_\theta^\theta \le M^{1-\theta}R^\theta \;\;\Rightarrow\;\; R \ge M^{1-1/\theta}.$$

Another lower bound on R in terms of $\vec{L}$, $\vec{r}$ and θ can be established using the Littlewood interpolation inequality [see, e.g., (Garling 2007, Section 5.5)]. Let $0 < q_0 < q_1$ and $\alpha \in (0,1)$ be arbitrary numbers; then the Littlewood inequality states that $\|g\|_q \le \|g\|_{q_0}^{1-\alpha}\|g\|_{q_1}^{\alpha}$, where q is defined by relation $\frac{1}{q} = \frac{1-\alpha}{q_0} + \frac{\alpha}{q_1}$. Now, suppose that $f \in \mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L})$, and choose $q_0 = \theta$, $q_1 = r_i$ and $\alpha = \frac{1-\theta}{1-\theta/r_i}$; then $q = 1$ and by the Littlewood inequality we have

$$1 = \|f\|_1 \le \|f\|_\theta^{1-\alpha}\|f\|_{r_i}^{\alpha} \le R^{\frac{r_i\theta-\theta}{r_i-\theta}} L_i^{\frac{r_i-r_i\theta}{r_i-\theta}}, \quad i = 1,\dots,d \;\;\Rightarrow\;\; R \ge \max_{i=1,\dots,d} L_i^{\frac{r_i\theta-r_i}{r_i-\theta}}.$$

4. Another interesting observation is related to the specific case $p = 1$. Recall that the condition $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ alone is not sufficient for existence of consistent estimators. However, for any $\theta \in (0,1)$ we can show

$$\inf_{\tilde{f}} \mathcal{R}_1^{(n)}\big[\tilde{f};\ \mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)\big] \;\le\; C\left[\frac{L_\beta(\ln n)^d}{n}\right]^{\frac{1-\theta}{1-\theta/s+1/\beta}} \;\to\; 0, \;\; n \to \infty.$$

This result follows from the proof of Theorem 4 and (6.5).

Now we argue that condition $f \in \mathbb{G}_{\theta^*}(R)$ is, in a sense, the weakest possible ensuring the "usual" rate of convergence, corresponding to the index $\nu = \beta/(2\beta+1)$, in the whole zone $p \le s(2+1/\beta)$. Indeed, in view of Theorem 4, the minimax rate of convergence on the class $\mathbb{G}_{\theta^*}(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$, say $\overline{\psi}_n(\theta^*)$, satisfies

$$c\big(L_\beta/n\big)^{\frac{\beta}{2\beta+1}} \;\le\; \overline{\psi}_n(\theta^*) \;\le\; C(\ln n)^{1/p}\big(L_\beta \ln n/n\big)^{\frac{\beta}{2\beta+1}},$$

where the constants c and C may depend on R. On the other hand, if $\underline{\psi}_n(\theta^*)$ denotes the minimax rate of convergence on the class $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M) \setminus \mathbb{G}_{\theta^*}(R)$ then

$$c\big(L_\beta/n\big)^{\frac{1-1/p}{1-1/s+1/\beta}} \;\le\; \underline{\psi}_n(\theta^*) \;\le\; C\big(L_\beta \ln n/n\big)^{\frac{1-1/p}{1-1/s+1/\beta}}, \tag{4.4}$$

provided that $p \le \frac{2+1/\beta}{1+1/s}$. The upper bound in (4.4) is one of the statements of Theorem 2, while the lower bound is proved in Section 7.5.

Thus we conclude that there is no tail zone in estimation over the class $G_{\theta^*}(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$, and this zone appears when considering the class $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M) \setminus \mathbb{G}_{\theta^*}(R)$. In this sense $f \in \mathbb{G}_{\theta^*}(R) \cap \mathbb{N}_{\vec{r}}(\vec{\beta}, \vec{L}, M)$ is the necessary and sufficient condition eliminating the tail zone.

5. Proof of Theorem 1

First we state two auxiliary results, Lemmas 1 and 2, and then turn to the proof of the theorem. Proof of measurability of our estimator and proofs of Lemmas 1 and 2 are given in Appendix A.

5.1. Auxiliary lemmas

For any $g : \mathbb{R}^d \to \mathbb{R}^1$ denote

$$\check{M}_h(g,x) = \sqrt{\frac{\varkappa \hat{A}_h(g,x)\ln n}{nV_h}} + \frac{\varkappa \ln n}{nV_h}\ .$$

Lemma 1. *Let* $\chi_h(g,x) = \big[|\hat{A}_h(g,x) - A_h(g,x)| - M_h(g,x)\big]_+$, $h \in \mathcal{H}$; *then*

$$[\check{M}_h(g,x) - 5M_h(g,x)]_+ \le \frac{1}{2}\chi_h(g,x), \quad [M_h(g,x) - 4\check{M}_h(g,x)]_+ \le 2\chi_h(g,x).$$

The next lemma establishes moment bounds on the following four random variables:

$$\begin{aligned}
\zeta_1(x) &= \sup_{h\in\mathcal{H}}[|\xi_h(x)| - M_h(K,x)]_+;\\
\zeta_2(x) &= \sup_{h,\eta\in\mathcal{H}}[|\xi_{h,\eta}(x)| - M_{h\vee\eta}(Q,x)]_+;\\
\zeta_3(x) &:= \sup_{h\in\mathcal{H}}[|A_h(K,x) - \hat{A}_h(K,x)| - M_h(K,x)]_+;\\
\zeta_4(x) &:= \sup_{h\in\mathcal{H}}[|A_h(Q,x) - \hat{A}_h(Q,x)| - M_h(Q,x)]_+.
\end{aligned} \tag{5.1}$$

Denote $\mathrm{k}_\infty = \|K\|_\infty \vee 1$ and

$$F(x) = \int \mathbf{1}_{[-1,1]^d}(t-x)f(t)\mathrm{d}t.$$

Lemma 2. *Let* $q \ge 1$, $l \ge 1$ *be arbitrary numbers. If* $\varkappa \ge \mathrm{k}_\infty^2[(2q+4)d + 2l]$ *then for all* $x \in \mathbb{R}^d$

$$\mathbb{E}_f[\zeta_j(x)]^q \ \ \le \ \ C_0 n^{-q/2}\big\{F(x) \vee n^{-l}\big\}, \quad j = 1,2,3,4, \tag{5.2}$$

where constant C_0 *depends on* d, q, *and* k_∞ *only.*

5.2. Proof of oracle inequality (2.14)

We recall the standard error decomposition of the kernel estimator: for any $h \in \mathcal{H}$ one has

$$|\hat{f}_h(x) - f(x)| \le |B_h(f,x)| + |\xi_h(x)|,$$

where $B_h(f,x)$ and $\xi_h(x)$ are given in (2.12) and (2.4) respectively. Similar error decomposition holds for auxiliary estimators $\hat{f}_{h,\eta}(x)$; the corresponding bias and stochastic error are denoted by $B_{h,\eta}(f,x)$ and $\xi_{h,\eta}(x)$.

1^0. The following relation for the bias $B_{h,\eta}(f,x)$ of $\hat{f}_{h,\eta}(x)$ holds:

$$B_{h,\eta}(f,x) - B_\eta(f,x) = \int K_\eta(t-x)B_h(f,t)\mathrm{d}t, \quad \forall h,\eta \in \mathcal{H}. \tag{5.3}$$

Indeed, using the Fubini theorem and the fact that $\int K_h(x)\mathrm{d}x = 1$ for all $h \in \mathcal{H}$ we have

$$\begin{aligned}\int [K_h * K_\eta](t-x)f(t)\mathrm{d}t &= \int\Big[\int K_h(t-y)K_\eta(y-x)\mathrm{d}y\Big]f(t)\mathrm{d}t \\ &= \int K_\eta(y-x)f(y)\mathrm{d}y \\ &\quad + \int K_\eta(y-x)\Big[\int K_h(t-y)[f(t)-f(y)]\mathrm{d}t\Big]\mathrm{d}y.\end{aligned}$$

It remains to note that $\int K_h(t-y)[f(t)-f(y)]\mathrm{d}t = B_h(f,y)$ and to subtract $f(x)$ from the both sides of the above equality. Thus, (5.3) is proved.

2^0. By the triangle inequality we have for any $h \in \mathcal{H}$

$$|\hat{f}_{\hat{h}}(x) - f(x)| \;\le\; |\hat{f}_{\hat{h}}(x) - \hat{f}_{\hat{h},h}(x)| + |\hat{f}_{\hat{h},h}(x) - \hat{f}_h(x)| + |\hat{f}_h(x) - f(x)|. \tag{5.4}$$

We bound each term on the right hand side separately.

First we note that, by (5.3) and (2.13), for any $h \in \mathcal{H}$

$$\begin{aligned}\hat{R}_h(x) - \sup_{\eta\ge h}\hat{M}_\eta(Q,x) - \hat{M}_h(K,x) &= \sup_{\eta\in\mathcal{H}}\Big[|\hat{f}_{h,\eta}(x) - \hat{f}_\eta(x)| - \hat{M}_{h\vee\eta}(Q,x) - \hat{M}_\eta(K,x)\Big]_+ \\ &\le \bar{B}_h(f,x) + \sup_{\eta\in\mathcal{H}}\Big[|\xi_{h,\eta}(x) - \xi_\eta(x)| - \hat{M}_{h\vee\eta}(Q,x) - \hat{M}_\eta(K,x)\Big]_+.\end{aligned}$$

Thus, for any $h \in \mathcal{H}$

$$\hat{R}_h(x) \;\le\; \bar{B}_h(f,x) + 2\hat{\zeta}(x) + \hat{M}_h(K,x) + \sup_{\eta\ge h}\hat{M}_\eta(Q,x), \tag{5.5}$$

where we put

$$\hat{\zeta}(x) := \sup_{h,\eta\in\mathcal{H}}\big[|\xi_{h,\eta}(x)| - \hat{M}_{h\vee\eta}(Q,x)\big]_+ \;\vee\; \sup_{h\in\mathcal{H}}\big[|\xi_h(x)| - \hat{M}_h(K,x)\big]_+.$$

Second, by (5.3) and $\hat{f}_{h,\eta} \equiv \hat{f}_{\eta,h}$ for any $h, \eta \in \mathcal{H}$ we have

$$\begin{aligned}|\hat{f}_{h,\eta}(x) - \hat{f}_h(x)| &\le |B_{\eta,h}(f,x) - B_h(f,x)| + |\xi_{h,\eta}(x) - \xi_h(x)| \\ &\le B_\eta(f,x) + \big[|\xi_{h,\eta}(x) - \xi_h(x)| - \hat{M}_{h\vee\eta}(Q,x) - \hat{M}_h(K,x)\big] + \sup_{\eta\ge h}\hat{M}_\eta(Q,x) + \hat{M}_h(K,x) \\ &\le \bar{B}_\eta(f,x) + 2\hat{\zeta}(x) + \hat{R}_h(x),\end{aligned}$$

where the last inequality holds by definition of $\hat{R}_h(x)$ [see (2.10)]. There inequalities imply the following upper bound on the first term on the right hand side of (5.4): for any $h \in \mathcal{H}$

$$\begin{aligned}|\hat{f}_{\hat{h},h}(x) - \hat{f}_{\hat{h}}(x)| &\le \bar{B}_h(f,x) + \hat{R}_{\hat{h}}(x) + 2\hat{\zeta}(x) \\ &\le \bar{B}_h(f,x) + \hat{R}_h(x) + 2\hat{\zeta}(x) \\ &\le 2\bar{B}_h(f,x) + 4\hat{\zeta}(x) + \sup_{\eta\ge h}\hat{M}_\eta(Q,x) + \hat{M}_h(K,x);\end{aligned} \tag{5.6}$$

where we have used the fact that $\hat{R}_{\hat{h}}(x) \le \hat{R}_h(x)$ for all $h \in \mathcal{H}$, and inequality (5.5).

Now we turn to bounding the second term on the right hand side of (5.4). We get for any $h \in \mathcal{H}$

$$\begin{aligned}|\hat{f}_{\hat{h},h}(x) - \hat{f}_h(x)| \;&=\; |\hat{f}_{\hat{h},h}(x) - \hat{f}_h(x)| \pm \big[\hat{M}_{\hat{h}\vee h}(Q,x) + \hat{M}_h(K,x)\big] \\ &\le\; \hat{R}_{\hat{h}}(x) + \sup_{\eta\ge h}\hat{M}_\eta(Q,x) + \hat{M}_h(K,x) \\ &\le\; \bar{B}_h(f,x) + 2\hat{\zeta}(x) + 2\sup_{\eta\ge h}\hat{M}_\eta(Q,x) + 2\hat{M}_h(K,x), \end{aligned} \tag{5.7}$$

where we again used (5.5) and the fact that $\hat{R}_{\hat{h}}(x) \le \hat{R}_h(x)$ for all $h \in \mathcal{H}$.

Finally for any $h \in \mathcal{H}$

$$|\hat{f}_h(x) - f(x)| \le |B_h(f,x)| + |\xi_h(x)| \le \bar{B}_h(f,x) + M_h(K,x) + \zeta(x).$$

Thus, combining (5.6), (5.7) and (5.4) we obtain

$$|\hat{f}_{\hat{h}}(x) - f(x)| \;\le\; \inf_{h\in\mathcal{H}}\Big\{4\bar{B}_h(f,x) + 3\sup_{\eta\ge h}\hat{M}_\eta(Q,x) + 3\hat{M}_h(K,x) + M_h(K,x)\Big\} + 6\hat{\zeta}(x) + \zeta(x). \tag{5.8}$$

3^0. In order to complete the proof we note that by the first inequality of Lemma 1 for any $g : \mathbb{R}^d \to \mathbb{R}^1$

$$\hat{M}_h(g,x) \le 20M_h(g,x) + 2\chi_h(g,x).$$

In addition, by the second inequality in Lemma 1

$$\begin{aligned}&|\xi_h(x)| - \hat{M}_h(K,x) = |\xi_h(x)| - M_h(K,x) + M_h(K,x) - \hat{M}_h(K,x) \le \zeta(x) + 2\chi(x), \\ &|\xi_{h,\eta}(x)| - \hat{M}_{h\vee\eta}(Q,x) = |\xi_{h,\eta}(x)| - M_{h\vee\eta}(Q,x) + M_{h\vee\eta}(Q,x) - \hat{M}_{h\vee\eta}(Q,x) \le \zeta(x) + 2\chi(x),\end{aligned}$$

so that $\hat{\zeta}(x) \le \zeta(x) + 2\chi(x)$. Substituting these bounds in (5.8) we obtain

$$|\hat{f}(x) - f(x)| \le \inf_{h\in\mathcal{H}}\Big\{4\bar{B}_h(f,x) + 60\sup_{\eta\ge h}M_\eta(Q,x) + 61M_h(K,x)\Big\} + 7\zeta(x) + 18\chi(x),$$

as claimed. ∎

5.3. Proof of moment bounds (2.17)

Let $\zeta_j(x)$, $j = 1,\dots,4$ be defined by (5.1). Then

$$\mathbb{E}_f[\zeta_j(x)]^q \;\;\le\;\; C_0 n^{-q/2}\big\{F(x) \vee n^{-l}\big\},$$

as claimed in Lemma 2.

Let $T_1 = \{x \in \mathbb{R}^d : F(x) \ge n^{-l}\}$ and $T_2 = \mathbb{R}^d \setminus T_1$. Therefore

$$\int_{T_1}\mathbb{E}_f[\zeta_j(x)]^q \mathrm{d}x \le C_0 n^{-q/2}\int_{T_1}F(x)\mathrm{d}x \le C_0 n^{-q/2}\int F(x)\mathrm{d}x = 2^d C_0 n^{-q/2}. \tag{5.9}$$

Now we analyze integrability on the set T_2. We consider only the case $j = 1, 2$ since computations for $j = 3, 4$ are the same as for $j = 1$.

Let $U_{\max}(x) = [x-1, x+1]^d$ and define the event $D(x) = \big\{ \sum_{i=1}^n \mathbf{1}[X_i \in U_{\max}(x)] < 2\big\}$, and let $\bar{D}(x)$ denote the complementary event. First we argue that for $j = 1, 2$

$$\zeta_j(x)\mathbf{1}\{D(x)\} = 0, \quad \forall x \in T_2. \tag{5.10}$$

Indeed, if $x \in T_2$ then for any $h \in \mathcal{H}$

$$|\mathbb{E}_f K_h(X_i - x)| \le n^d \mathrm{k}_\infty F(x) \le \mathrm{k}_\infty n^{d-l}, \quad |\mathbb{E}_f Q_h(X_i - x)| \le n^d \mathrm{k}_\infty^2 F(x) \le \mathrm{k}_\infty^2 n^{d-l}.$$

Here we have used that $\mathcal{H} = \big[1/n, 1\big]^d$ and that $\mathrm{supp}(K) = [-1/2, 1/2]^d$, $\mathrm{supp}(Q) = [-1, 1]^d$.

Hence, by definition of $\xi_h(x)$, for any $h \in \mathcal{H}$ one has for any $l \ge d+1$

$$\begin{aligned} |\xi_h(x)|\mathbf{1}\{D(x)\} \;&\le\; \Big|\frac{1}{n}\sum_{i=1}^n K_h(X_i - x)\Big|\mathbf{1}\{D(x)\} + \mathrm{k}_\infty n^{d-l} \\ &\le\; \frac{2\mathrm{k}_\infty}{nV_h} + \mathrm{k}_\infty n^{d-l} \;\le\; \frac{4\mathrm{k}_\infty}{nV_h} \le M_h(K, x), \end{aligned}$$

where we have used that $n^{d-l} \le (nV_h)^{-1}$ for $l \ge d+1$, $\varkappa \ln n \ge 4\mathrm{k}_\infty$ by the condition on $\varkappa$ [see also definition of $M_h(K, x)$], and $n \ge 3$. Therefore $\zeta_1(x)\mathbf{1}\{D(x)\} = 0$ for $x \in T_2$. By the same reasoning for $\zeta_2(x)$ we obtain that $\zeta_2(x)\mathbf{1}\{D(x)\} = 0$, $\forall x \in T_2$ because $\varkappa \ln n \ge 4\mathrm{k}_\infty^2$. Thus (5.10) is proved. Using (5.10) we can write

$$\begin{aligned} \int_{T_2} \mathbb{E}_f[\zeta_j(x)\big]^q \mathbf{1}\{\bar{D}(x)\}\mathrm{d}x \;&\le\; \int_{T_2} \mathbb{E}_f\Big(\big[\sup_{h\in\mathcal{H}} |\xi_h(x)|^q \vee \sup_{h,\eta\in\mathcal{H}} |\xi_{h,\eta}(x)|^q\big]\mathbf{1}\{\bar{D}(x)\}\Big)\mathrm{d}x \\ &\le\; \big(2\mathrm{k}_\infty^2 n^d\big)^q \int_{T_2} \mathbb{P}_f\{\bar{D}(x)\}\mathrm{d}x. \end{aligned} \tag{5.11}$$

Now we bound from above the integral on the right hand side of the last display formula. For any $z > 0$ we have in view of the exponential Markov inequality

$$\begin{aligned} \mathbb{P}_f\{\bar{D}(x)\} \;&=\; \mathbb{P}_f\Big\{\sum_{i=1}^n \mathbf{1}[X_i \in U_{\max}(x)] \ge 2\Big\} \le e^{-2z}\big[e^z F(x) + 1 - F(x)\big]^n \\ &=\; e^{-2z}\big[(e^z - 1)F(x) + 1\big]^n \le \exp\{-2z + n(e^z - 1)F(x)\}. \end{aligned}$$

Minimizing the right hand side w.r.t. z we find $z = \ln 2 - \ln\{nF(x)\}$ and, therefore,

$$\mathbb{P}_f\big\{\bar{D}(x)\big\} \le 4^{-1} n^2 F^2(x) \exp\{2 - nF(x)\} \le (e^2/4) n^2 F^2(x).$$

Since $F(x) \le n^{-l}$ for any $x \in T_2$ we obtain

$$\int_{T_2} \mathbb{P}_f\{\bar{D}(x)\}\mathrm{d}x \le (e^2/4) n^{2-l} \int F(x)\mathrm{d}x = 2^d\big(e^2/4\big) n^{2-l}.$$

Combining this inequality with (5.11) we obtain

$$\int_{T_2} \mathbb{E}_f[\zeta_j(x)\big]^q \mathbf{1}\{\bar{D}(x)\}\mathrm{d}x \le 2^d (2\mathrm{k}_\infty^2)^q (e^2/4) n^{2+dq-l}. \tag{5.12}$$

Choosing $l = (d+1)q + 2$ we come to the assertion of the theorem in view of (5.9) and (5.12). ∎

6. Proofs of Theorem 2 and statement (i) of Theorem 4

The proofs of Theorem 2 and of statement (i) of Theorem 4 go along similar lines. That is why we state our auxiliary results (Propositions 1 and 2) in the form that is suitable for the use in the proof of Theorem 4.

This section is organized as follows. First, in Subsection 6.1 we present and discuss some facts from functional analysis. Then in Lemma 3 of Subsection 6.2 we state an auxiliary result on approximation properties of the kernel K defined in (3.3). Proof outline and notation are discussed in Subsection 6.3. Subsection 6.4 presents two auxiliary propositions, and the proofs of Theorem 2 and statement (i) of Theorem 4 are completed in Subsections 6.5 and 6.6. Proofs of the auxiliary results, Lemma 3 and Propositions 1 and 2 are given in Appendix B.

In the subsequent proof $c_i, C_i, \bar{c}_i, \bar{C}_i, \hat{c}_i, \hat{C}_i, \tilde{c}_i, \tilde{C}_i, \ldots$, stand for constants that can depend on L_0, M $\vec{\beta}$, $\vec{r}$, d and p, but are independent of $\vec{L}$ and n. These constants can be different on different appearances. In the case when the assumption $f \in \mathbb{G}_\theta(R)$ with $\theta \in (0,1)$ is imposed, they may also depend on θ and R.

6.1. Preliminaries

We present an embedding theorem for the anisotropic Nikol'skii classes and discuss some properties of the strong maximal operator.

6.1.1. Embedding theorem

The statement given below in (6.2) is a particular case of the embedding theorem for anisotropic Nikol'skii classes $\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$; see (Nikol'skii 1977, Section 6.9.1.).

For the fixed class parameters $\vec{\beta}$ and $\vec{r}$ define

$$\tau(p) = 1 - \sum_{j=1}^{d} \frac{1}{\beta_j}\left(\frac{1}{r_j} - \frac{1}{p}\right), \quad \tau_i = 1 - \sum_{j=1}^{d} \frac{1}{\beta_j}\left(\frac{1}{r_j} - \frac{1}{r_i}\right), \ \ i = 1, \ldots, d,$$

and put

$$q_i = r_i \vee p, \qquad \gamma_i = \begin{cases} \frac{\beta_i \tau(p)}{\tau_i}, & r_i < p, \\ \beta_i, & r_i \geq p. \end{cases} \tag{6.1}$$

Let $\tau(p) > 0$ and $\tau_i > 0$ for all $i = 1, \ldots, d$; then for any $p \geq 1$ one has

$$\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big) \subseteq \mathbb{N}_{\vec{q},d}\big(\vec{\gamma}, c\vec{L}\big), \tag{6.2}$$

where constant $c > 0$ is independent of $\vec{L}$ and p.

6.1.2. Strong maximal function

Let $g : \mathbb{R}^d \to \mathbb{R}$ be a locally integrable function. We define the strong maximal function $g^\star$ of g by formula

$$g^\star(x) := \sup_H \frac{1}{|H|} \int_H g(t)\mathrm{d}t, \quad x \in \mathbb{R}^d, \tag{6.3}$$

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 worth noting that the *Hardy–Littlewood maximal function* is defined by (6.3) with the supremum taken over all cubes with sides parallel to the coordinate axes, centered at x.

It is well known that the strong maximal operator $g \mapsto g^\star$ is of the strong (p,p)–type for all $1 < p \le \infty$, i.e., if $g \in \mathbb{L}_p(\mathbb{R}^d)$ then $g^\star \in \mathbb{L}_p(\mathbb{R}^d)$ and there exists a constant $\bar{C}$ depending on p only such that

$$\|g^\star\|_p \le \bar{C}\|g\|_p, \quad p \in (1,\infty].$$

Let g^* be defined defined in (4.1). Since obviously $g^*(x) \le g^\star(x)$ for all $x \in \mathbb{R}^d$ we have

$$\|g^*\|_p \le \bar{C}\|g\|_p, \quad p \in (1,\infty]. \tag{6.4}$$

In distinction to the Hardy–Littlewood maximal function, the strong maximal operator is not of the weak (1,1)–type. In fact, the following statement holds: there exists constant C depending on d only such that

$$\big|\{x : g^\star(x) \ge \alpha\}\big| \le C \int \frac{|g(x)|}{\alpha}\left\{1 + \left(\ln_+ \frac{|g(x)|}{\alpha}\right)^{d-1}\right\}\mathrm{d}x, \quad \forall \alpha > 0. \tag{6.5}$$

We refer to Guzman (1975) for more details.

6.2. Approximation properties of kernel K

The next lemma establishes an upper bound on norm of the bias $B_h(f,\cdot)$ of kernel estimator $\hat{f}_h$ when f belongs to the anisotropic Nikol'skii class.

Lemma 3. *Let $f \in \mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L})$. Let $\hat{f}_h$ be the estimator (2.1) associated with kernel (3.3) with $\ell > \max_{j=1,\dots,d}\beta_j$. Then $B_h(f,x)$ can represented as the sum $B_h(f,x) = \sum_{j=1}^d B_{h,j}(f,x)$ with functions $B_{h,j}(f,x)$ satisfying the following inequalities:*

$$\big\|B_{h,j}(f,\cdot)\big\|_{r_j} \le C_1 L_j h_j^{\beta_j}, \quad \forall j = 1,\dots,d. \tag{6.6}$$

Moreover, if $s \ge 1$, then for any $p \ge 1$

$$\big\|B_{h,j}(f,\cdot)\big\|_{q_j} \le C_2 L_j h_j^{\gamma_j}, \quad \forall j = 1,\dots,d, \tag{6.7}$$

where $\vec{\gamma} = \vec{\gamma}(p)$ and $\vec{q} = \vec{q}(p)$ are defined in (6.1). Here C_1 and C_2 are constants independent of $\vec{L}$ and p.

6.3. Proof outline and notation

The starting point of our proof is the pointwise oracle inequality (2.14) together with the moment bound (2.17). Denote

$$\bar{U}_f(x) = \inf_{h\in\mathcal{H}}\Big\{\bar{B}_h(f,x) + \sup_{\eta\ge h} M_\eta(K\vee Q,x)\Big\}; \tag{6.8}$$

then, taking into account that $M_\eta(K\vee Q,x)$ is greater than $M_\eta(K,x)$ and $M_\eta(Q,x)$ for any x and η [see (2.6) and (2.9)], and using (2.14), we have

$$|\hat{f}(x) - f(x)| \le c_0[\bar{U}_f(x) + \omega(x)],$$

where c_0 is an absolute constant, and $\omega(x) := \zeta(x) + \chi(x)$ with $\zeta(x)$ and $\chi(x)$ defined in (2.15) and (2.16). Therefore, by (2.17) applied with $q = p$ and by the Fubini theorem, there exists constant $\bar{c}_0 > 0$ such that for any probability density f and any Borel set $\mathcal{A} \subseteq \mathbb{R}^d$ one has

$$\mathbb{E}_f \int_{\mathcal{A}} |\hat{f}(x) - f(x)|^p \mathrm{d}x \ \le \ \bar{c}_0 \Big[\int_{\mathcal{A}} \bar{U}_f^p(x) \mathrm{d}x + n^{-p/2} \Big]. \tag{6.9}$$

Recall that $\mathrm{k}_\infty = \|K\|_\infty \vee 1$; by definition of $\bar{B}_h(f,x)$ [see (2.13)] and by Lemma 3 one has

$$\bar{B}_h(f,x) \le \mathrm{k}_\infty \sum_{j=1}^d B_{h,j}^*(f,x),$$

where $B_{h,j}^*(f,x)$ is the strong maximal function of $|B_{h,j}(f,x)|$, $j = 1, \dots, d$. Therefore if we let

$$U_f(x) := \inf_{h \in \mathcal{H}} \Big\{ \max_{j=1,\dots,d} B_{h,j}^*(f,x) + \sup_{\eta \ge h} M_\eta(K \vee Q, x) \Big\}, \tag{6.10}$$

then

$$\bar{U}_f(x) \le \mathrm{k}_\infty U_f(x), \quad \forall x \in \mathbb{R}^d. \tag{6.11}$$

The key element of the proof is derivation of upper bounds on the integral

$$J := \int_{\mathbb{R}^d} U_f^p(x) \mathrm{d}x.$$

These bounds will be established by division of $\mathbb{R}^d$ in "slices", and appropriate choice of bandwidth $h \in \mathcal{H}$ on every "slice". For this purpose the following bounds on norms of $B_{h,j}^*(f,\cdot)$ will be used. Inequality (6.4) and the first assertion of Lemma 3 imply that for any $p > 1$, $\vec{r} \in (1,\infty]^d$ and any $f \in \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$ one has

$$\big\| B_{h,j}^*(f,\cdot) \big\|_{r_j} \le \bar{c}_1 L_j h_j^{\beta_j}, \quad \forall j = 1, \dots, d, \tag{6.12}$$

Moreover, if $s \ge 1$ then, by the second assertion of Lemma 3, for any $p > 1$, $\vec{r} \in (1,\infty]^d$ and $f \in \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big)$

$$\big\| B_{h,j}^*(f,\cdot) \big\|_{q_j} \le \bar{c}_2 L_j h_j^{\gamma_j}, \quad \forall j = 1, \dots, d. \tag{6.13}$$

Let $\delta := \ln n / n$, $\varphi := (L_\beta \delta)^{\beta/(2\beta+1)}$. Let $m_0(\theta)$, $\theta \in (0,1]$, be an integer number to be specified later; see (6.19) below. For $m \in \mathbb{Z}$, $m \ge m_0(\theta)$ define "slices"

$$\mathcal{X}_m := \big\{ x \in \mathbb{R}^d : 2^m \varphi < U_f(x) \le 2^{m+1} \varphi \big\}, \ \ \mathcal{X}_{m_0(\theta)}^- := \big\{ x \in \mathbb{R}^d : U_f(x) \le 2^{m_0(\theta)} \varphi \big\},$$

and consider the corresponding integrals

$$J_m := \int_{\mathcal{X}_m} U_f^p(x) \mathrm{d}x, \quad J_{m_0}^- := \int_{\mathcal{X}_{m_0(\theta)}^-} U_f^p(x) \mathrm{d}x.$$

With this notation, using (6.9) and (6.11) we can write

$$\begin{aligned}\mathbb{E}_f\|\hat{f}-f\|_p^p &\le \mathbb{E}_f\int_{\mathcal{X}^-_{m_0(\theta)}}|\hat{f}(x)-f(x)|^p\mathrm{d}x \;+\; \tilde{c}_1\sum_{m=m_0(\theta)}^{\infty}\int_{\mathcal{X}_m}U_f^p(x)\mathrm{d}x \;+\; \tilde{c}_2 n^{-p/2}\\ &=: J^-_{m_0(\theta)} + \tilde{c}_1\sum_{m=m_0(\theta)}^{\infty}J_m \;+\; \tilde{c}_2 n^{-p/2}. \qquad (6.14)\end{aligned}$$

The rest of the proof consists of bounding the integrals $J^-_{m_0(\theta)}$ and J_m on the right hand side of (6.14) and combining these bounds in different zones.

The following notation will be used in the subsequent proof. For the sake of brevity we will write

$$M_\eta(x) := M_\eta(K\vee Q, x), \quad A_\eta(x) := A_\eta(K\vee Q, x), \quad \forall \eta\in\mathcal{H}.$$

We let $I := \{1,\dots,d\}$, and

$$I_+ := \{j\in I : p\le r_j<\infty\},\;\; I_- := \{j\in I: 1<r_j<p\},\;\; I_\infty := \{j\in I: r_j=\infty\}.$$

With $\vec{\gamma}=(\gamma_1,\dots,\gamma_d)$ and $\vec{q}=(q_1,\dots,q_d)$ given by (6.1) we define quantities γ, υ and L_γ by the formulas

$$\frac{1}{\gamma} := \sum_{j=1}^d\frac{1}{\gamma_j}, \qquad \frac{1}{\upsilon} := \sum_{j=1}^d\frac{1}{\gamma_j q_j}, \qquad L_\gamma := \prod_{j=1}^d L_j^{1/\gamma_j}. \qquad (6.15)$$

Note some useful inequalities between the quantities defined above. First, $\gamma_j<\beta_j$ for all $j\in I_-$ which is a consequence of the fact that $\tau(p)<\tau_j$ for $j\in I_-$. This implies

$$\frac{1}{\gamma}-\frac{1}{\beta} = \sum_{j\in I_-}\Big(\frac{1}{\gamma_j}-\frac{1}{\beta_j}\Big) > 0. \qquad (6.16)$$

Next, if $s\ge 1$ then

$$\frac{1}{s} > \frac{1}{\upsilon}. \qquad (6.17)$$

We have

$$\frac{1}{s}-\frac{1}{\upsilon} = \sum_{j\in I_-}\Big(\frac{1}{\beta_j r_j}-\frac{1}{\gamma_j p}\Big) = \sum_{j\in I_-}\frac{1}{\beta_j}\Big(\frac{1}{r_j}-\frac{\tau_j}{\tau(p)p}\Big).$$

Hence (6.17) will be proved if we show that $r_j^{-1}\tau(p)p\ge\tau_j$ for all $j\in I_-$. Indeed,

$$\frac{\tau(p)p}{r_j} = \frac{p(1-1/s)+1/\beta}{r_j} \ge 1-\frac{1}{s}+\frac{1}{\beta r_j} := \tau_j,$$

where to get the second inequality we have used that $r_j\le p$ for any $j\in I_-$ and that $s\ge 1$.

Finally, remark also that

$$p-\upsilon(2+1/\gamma)<0. \qquad (6.18)$$

Indeed, since $r_j\ge p$ for any $j\in I_+\cup I_\infty$,

$$\frac{p}{\upsilon} = \sum_{j\in I_+}\frac{p}{\beta_j r_j}+\sum_{j\in I_-}\frac{1}{\gamma_j} \le \sum_{j\in I_+}\frac{1}{\beta_j}+\sum_{j\in I_-}\frac{1}{\gamma_j} = \frac{1}{\gamma}.$$

This yields $p\le\upsilon/\gamma$, and (6.18) follows.

6.4. Auxiliary results

For $\theta \in (0,1]$ and for some constant $\hat{c}_1 > 0$ define

$$m_0(\theta) := \min\left\{m \in \mathbb{Z} : 2^{m_0(\theta)\left(\frac{1-\theta/s+1/\beta}{1+\theta/s}\right)} > \hat{c}_1 \varkappa\varphi\right\}. \tag{6.19}$$

Note that $1 - \theta/s + 1/\beta > 0$ for any $\theta \in (0,1]$, since $s \geq \beta$ by $r_j \geq 1$, $j = 1, \ldots, d$. Therefore $m_0(\theta) < 0$ for large enough n.

It will be convenient to introduce the following notation

$$m_1 := \min\left\{m \in \mathbb{Z} : 2^{m[\upsilon(2+1/\gamma)-s(2+1/\beta)]} \geq \left(L_\gamma/L_\beta\right)^\upsilon \varphi^{\upsilon(1/\beta-1/\gamma)}\right\}. \tag{6.20}$$

It follows from this definition that

$$\left[\left(L_\gamma/L_\beta\right)\varphi^{1/\beta-1/\gamma}\right]^{\frac{\upsilon}{\upsilon(2+1/\gamma)-s(2+1/\beta)}} \leq 2^{m_1} \leq 2\left[\left(L_\gamma/L_\beta\right)\varphi^{1/\beta-1/\gamma}\right]^{\frac{\upsilon}{\upsilon(2+1/\gamma)-s(2+1/\beta)}}. \tag{6.21}$$

In view of (6.16) and (6.17)

$$\upsilon\Big(2+\frac{1}{\gamma}\Big) - s\Big(2+\frac{1}{\beta}\Big) = s\upsilon\left[\Big(2+\frac{1}{\beta}\Big)\Big(\frac{1}{s}-\frac{1}{\upsilon}\Big) + \frac{1}{s}\Big(\frac{1}{\gamma}-\frac{1}{\beta}\Big)\right] > 0; \tag{6.22}$$

hence $m_1 > 1$ for large n.

The bounds on $J^-_{m_0(\theta)}$ and J_m are given in the next two propositions.

Proposition 1. *There exist constants $\hat{c}_1, \hat{c}_2 > 0$ and $\hat{C}_1, \hat{C}_2 > 0$ such that any n large enough the following statements hold.*

(i) *For any probability density f and any $m_0(1) \leq m \leq 0$*

$$J_m \leq \hat{C}_1\, 2^{m\left(p-\frac{2+1/\beta}{1+1/s}\right)}\varphi^p. \tag{6.23}$$

(ii) *Let $f \in \mathbb{G}_\theta(R)$, $\theta \in (0,1)$; then for any $m_0(\theta) \leq m \leq 0$ one has*

$$J_m \leq \hat{C}_1\, 2^{m\left(p-\frac{2+1/\beta}{1/\theta+1/s}\right)}\varphi^p. \tag{6.24}$$

(iii) *For any $m \in \mathbb{Z}$ satisfying $1 \leq 2^m \leq \hat{c}_2\varphi^{-1}$ and any probability density f one has*

$$J_m \;\leq\; \hat{C}_2 2^{m[p-s(2+1/\beta)]}\varphi^p. \tag{6.25}$$

(iv) *Let $s \geq 1$; then for any $m \in \mathbb{Z}$ such that $m \geq m_1$, $\;2^m \leq \hat{c}_2\varphi^{-1}$ and any probability density f one has*

$$J_m \;\leq\; \hat{C}_2\varphi^p\left[\frac{L_\gamma\varphi^{1/\beta}}{L_\beta\varphi^{1/\gamma}}\right]^\upsilon 2^{m[p-\upsilon(2+1/\gamma)]}\;. \tag{6.26}$$

Proposition 2. *There exist constants $\hat{C}_3, \hat{C}_4 > 0$ such that the following statements hold.*

(i) *Let ν is defined in (3.4). Then for all large enough n and for any density f one has*

$$J^-_{m_0(1)} \;=\; \mathbb{E}_f \int_{\mathcal{X}^-_{m_0(1)}} \left|\hat{f}(x) - f(x)\right|^p \mathrm{d}x \;\leq\; \hat{C}_3 \ln^d(n)(L_\beta\delta)^{p\nu}. \tag{6.27}$$

(ii) *Let $\nu(\theta)$ is defined in (4.3). Then for any $\theta \in (0,1]$ and for all n large enough*

$$\sup_{f\in\mathbb{G}_\theta(R)} \mathbb{E}_f \int_{\mathcal{X}^-_{m_0(\theta)}} \left|\hat{f}(x) - f(x)\right|^p \mathrm{d}x \leq \hat{C}_4(L_\beta\delta)^{p\nu(\theta)}. \tag{6.28}$$

6.5. Proof of Theorem 2

Using (6.14) and inequality (6.27) of Proposition 2 we obtain

$$\mathbb{E}_f\|\hat{f}-f\|_p^p \;\le\; c_1\pi_n\big(L_\beta\delta\big)^{p\nu} + c_2\sum_{m=m_0(1)}^{\infty} J_m, \tag{6.29}$$

where $\pi_n = \ln^d(n)$ if $p \le \frac{2+1/\beta}{1+1/s}$ and $\pi_n = 1$ otherwise.

We proceed with bounding the second term on the right hand side of the last display formula. First, because $\|f\|_\infty \le M$,

$$\max_{j=1,\dots,d}\|B^*_{h,j}(f,\cdot)\|_\infty \le 2^d M\mathrm{k}_\infty^2, \qquad \sup_{\eta>0}\|A_\eta\|_\infty \le 2^d M\mathrm{k}_\infty^2.$$

This implies that there exists constant $c_3 > 0$ with the following property:

$$m_2 := \min\{m\in\mathbb{Z} : 2^m \ge c_3\varphi^{-1}\} \;\;\Rightarrow\;\; J_m = 0,\ \forall m \ge m_2.$$

Thus the sum on right hand side of (6.29) extends from $m_0(1)$ to m_2.

1^0. *Tail zone:* $p < \frac{2+1/\beta}{1+1/s}$. Using bounds (6.23) and (6.25) of Proposition 1, we obtain

$$\sum_{m=m_0(1)}^{\infty} J_m \;\le\; c_4\varphi^p\Big[\sum_{m=m_0(1)}^{0} 2^{m(p-\frac{2+1/\beta}{1+1/s})} + \sum_{m=1}^{m_2} 2^{m[p-s(2+1/\beta)]}\Big] \;\le\; c_5\,\varphi^p 2^{m_0(1)(p-\frac{2+1/\beta}{1+1/s})},$$

where the last inequality follows from the fact that $m_0(1) < 0$ and $p < \frac{2+1/\beta}{1+1/s} < s(2+1/\beta)$. Using (6.19), after straightforward algebra we obtain that

$$\sum_{m=m_0(1)}^{\infty} J_m \;\le\; c_6\big(L_\beta\delta\big)^{\frac{p-1}{1+1/\beta-1/s}} \le c_6\,\big(L_\beta\delta\big)^{p\nu}.$$

2^0. *Dense zone:* $\frac{2+1/\beta}{1+1/s} < p < s(2+\frac{1}{\beta})$. Because $p > \frac{2+1/\beta}{1+1/s}$, by Proposition 1, inequality (6.23) with $\theta = 1$,

$$\sum_{m=m_0(1)}^{0} J_m \;\le\; c_7\varphi^p\sum_{m=m_0(1)}^{0} 2^{m(p-\frac{2+1/\beta}{1+1/s})} \;\le\; c_8\,\varphi^p = c_8(L_\beta\delta)^{\frac{p\beta}{2\beta+1}}. \tag{6.30}$$

Furthermore, because $p < s(2+\frac{1}{\beta})$ we have by Proposition 1, inequality (6.25), that

$$\sum_{m=1}^{m_2} J_m \;\le\; c_9\varphi^p\sum_{m=1}^{m_2} 2^{m(p-s(2+\frac{1}{\beta}))} = c_{10}(L_\beta\delta)^{\frac{p\beta}{2\beta+1}}.$$

Thus, in the dense zone

$$\sum_{m=m_0(1)}^{m_2} J_m \le c_{11}(L_\beta\delta)^{\frac{p\beta}{2\beta+1}} \le c_{11}(L_\beta\delta)^{p\nu}.$$

3^0. *Sparse zone:* $p > s(2+1/\beta)$, $s < 1$. First we note that the bound in (6.30) remains true since $p > s(2+1/\beta)$. By the same reason in view of Proposition 1, inequality (6.25),

$$\sum_{m=1}^{m_2} J_m \le c_{12}\varphi^p 2^{m_2(p-s(2+\frac{1}{\beta}))} \le c_{13}\varphi^{s(2+\frac{1}{\beta})} = c_{13}\big(L_\beta\delta\big)^s \le c_{13}(L_\beta\delta)^{p\nu}. \tag{6.31}$$

Here we have used the definition of m_2. It remains to note that conditions $p > s(2+1/\beta)$, $s < 1$ imply that $\varphi^p\delta^{-s} \to 0$ as $n \to 0$. Therefore the statement of the theorem follows from (6.30) and (6.31).

4^0. *Sparse zone:* $p > s(2+1/\beta)$, $s \ge 1$. We need to bound only $\sum_{m=1}^{m_2} J_m$, because (6.30) remains true. By inequality (6.25) of Proposition 1 and because $p > s(2+1/\beta)$

$$\sum_{m=1}^{m_1} J_m \le c_{14}\varphi^p 2^{m_1(p-s(2+\frac{1}{\beta}))}.$$

Next, we have in view of the inequality (6.26) of Proposition 1

$$\sum_{m=m_1+1}^{m_2} J_m \le c_{15}\varphi^p\left[\frac{L_\gamma\varphi^{1/\beta}}{L_\beta\varphi^{1/\gamma}}\right]^\upsilon \sum_{m=m_1+1}^{m_2} 2^{m[p-\upsilon(2+1/\gamma)]}.$$

Since $p - \upsilon(2+1/\gamma) < 0$ [see (6.18)],

$$\sum_{m=m_1+1}^{m_2} J_m \le c_{16}\varphi^p\left[\frac{L_\gamma\varphi^{1/\beta}}{L_\beta\varphi^{1/\gamma}}\right]^\upsilon 2^{m_1(p-\upsilon[2+1/\gamma])} \le c_{16}\varphi^p 2^{m_1(p-s[2+1/\beta])}.$$

In order to obtain the second inequality we have used (6.21). Thus,

$$\sum_{m=1}^{m_2} J_m \le c_{17}\varphi^p 2^{m_1[p-s(2+1/\beta)]}.$$

Using equality (6.22) and (6.21) we obtain

$$\sum_{m=1}^{m_1} J_m \le c_{20}\Big(L_\gamma/L_\beta\Big)^{\frac{p-s(2+1/\beta)}{s(2+1/\beta)(1/s-1/\upsilon)+(1/\gamma-1/\beta)}}(L_\beta\delta)^{\frac{p(1/s-1/\upsilon)+1/\gamma-1/\beta}{(2+1/\beta)(1/s-1/\upsilon)+(1/\gamma-1/\beta)s^{-1}}}.$$

The statement of the theorem is now obtained by the following routine computations. Denote

$$A = \frac{1}{s_-} - \frac{1}{p\beta_-}, \qquad \frac{1}{s_-} = \sum_{j\in I_-}\frac{1}{\beta_j r_j}, \qquad \frac{1}{\beta_-} = \sum_{j\in I_-}\frac{1}{\beta_j}, \qquad \frac{1}{\gamma_-} = \sum_{j\in I_-}\frac{1}{\gamma_j}.$$

First, we remark that

$$p\Big(\frac{1}{s} - \frac{1}{\upsilon}\Big) + \frac{1}{\gamma} - \frac{1}{\beta} = \frac{p}{s_-} - \frac{1}{\gamma_-} + \frac{1}{\gamma_-} - \frac{1}{\beta_-} = \frac{p}{s_-} - \frac{1}{\beta_-} = Ap. \tag{6.32}$$

Next,

$$\begin{aligned}
\frac{1}{\gamma_-} = \sum_{j\in I_-}\frac{\tau_j}{\tau(p)\beta_j} &= \frac{1}{\tau(p)}\sum_{j\in I_-}\frac{1}{\beta_j}[1-1/s+1/(r_j\beta)] = \frac{1-1/s}{\tau(p)\beta_-} + \frac{1}{\tau(p)\beta s_-} \\
&= \frac{1-1/s}{\tau(p)\beta_-} + \frac{1}{\tau(p)\beta}\left(\frac{1}{s_-} - \frac{1}{p\beta_-}\right) + \frac{1}{\tau(p)\beta p\beta_-} \\
&= \frac{1}{\tau(p)\beta_-}\left(1 - \frac{1}{s} + \frac{1}{p\beta}\right) + \frac{A}{\tau(p)\beta} = \frac{1}{\beta_-} + \frac{A}{\tau(p)\beta}.
\end{aligned}$$

Hence, $1/\gamma - 1/\beta = 1/\gamma_- - 1/\beta_- = A/(\tau(p)\beta)$, which implies that

$$\frac{1}{s} - \frac{1}{\upsilon} = \frac{1}{s_-} - \frac{1}{p\gamma_-} = A + \frac{1}{p}\Big(\frac{1}{\beta_-} - \frac{1}{\gamma_-}\Big) = A\Big(1 - \frac{1}{p\tau(p)\beta}\Big).$$

Two last equalities yield

$$\Big(2 + \frac{1}{\beta}\Big)\Big(\frac{1}{s} - \frac{1}{\upsilon}\Big) + \Big(\frac{1}{\gamma} - \frac{1}{\beta}\Big)\frac{1}{s} = \frac{A}{\tau(p)}\Big[\Big(2 + \frac{1}{\beta}\Big)\Big(\tau(p) - \frac{1}{p\beta}\Big) + \frac{1}{s\beta}\Big] = \frac{A}{\tau(p)}\Big(2 + \frac{1}{\beta} - \frac{2}{s\beta}\Big),$$

where the last equality follows from the fact that $\tau(p) - 1/(p\beta) = 1 - 1/s$. This together with (6.32). leads to the statement of the theorem in the sparse zone.

5^0. *Boundary zones:* $p = s(2 + \frac{1}{\beta})$, $p = \frac{2+1/\beta}{1+1/s}$. Here the proof coincides with the proof for the dense zone with the only difference that the corresponding sums equal $|m_1|$ and m_2 respectively. ∎

6.6. Proof of statement (i) of Theorem 4

In view of (6.14) and by bound (6.28) of Proposition 2,

$$\mathbb{E}_f\|\hat{f} - f\|_p^p \;\le\; c_1\big(L_\beta\delta\big)^{p\nu(\theta)} + c_2 \sum_{m=m_0(\theta)}^{\infty} J_m.$$

If $p < \frac{2+1/\beta}{1/\theta+1/s}$ then, using bounds (6.24) and (6.25) of Proposition 1, we have

$$\sum_{m=m_0(\theta)}^{\infty} J_m \;\le\; c_3\varphi^p \sum_{m=m_0(\theta)}^{m_2} 2^{m\left(p - \frac{2+1/\beta}{1/\theta+1/s}\right)} \;\le\; c_4\,\varphi^p 2^{m_0(\theta)\left(p - \frac{2+1/\beta}{1/\theta+1/s}\right)} = c_5(L_\beta\delta)^{\frac{p-\theta}{1-\theta/s+1/\beta}},$$

and the assertion of the theorem follows. If $s(2 + 1/\beta) \ge p \ge \frac{2+1/\beta}{1/\theta+1/s}$ then

$$\sum_{m=m_0(\theta)}^{\infty} J_m \;\le\; c_6\mu_n^p(\theta)\varphi^p \le c_7\mu_n^p(\theta)\big(L_\beta\delta\big)^{p\nu(\theta)}.$$

7. Proofs of Theorem 3, statement (ii) of Theorem 4 and the lower bound in (4.4)

The proof is organized as follows. First, we formulate two auxiliary statements, Lemmas 4 and 5. Second, we present a general construction of a finite set of functions employed in the proof of lower bounds. Then we specialize the constructed set of functions in different regimes and derive the announced lower bounds.

7.1. Auxiliary lemmas

The first statement given in Lemma 4 is a simple consequence of Theorem 2.4 from Tsybakov (2009). Let $\mathbb{F}$ be a given set of probability densities.

Lemma 4. *Assume that for any sufficiently large integer n one can find a positive real number ρ_n and a finite subset of functions $\{f^{(0)}, f^{(j)},\ j \in \mathcal{J}_n\} \subset \mathbb{F}$ such that*

$$\left\|f^{(i)} - f^{(j)}\right\|_p \geq 2\rho_n, \qquad \forall i, j \in \mathcal{J}_n \cup \{0\} :\ i \neq j; \tag{7.1}$$

$$\limsup_{n\to\infty} \frac{1}{|\mathcal{J}_n|^2} \sum_{j\in\mathcal{J}_n} \mathbb{E}_{f^{(0)}} \left\{ \frac{\mathrm{d}\mathbb{P}_{f^{(j)}}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}(X^{(n)}) \right\}^2 =: C < \infty. \tag{7.2}$$

Then for any $q \geq 1$

$$\liminf_{n\to\infty} \inf_{\tilde{f}} \sup_{f\in\mathbb{F}} \rho_n^{-1} \left(\mathbb{E}_f \left\| \tilde{f} - f \right\|_p^q \right)^{1/q} \geq \left(\sqrt{C} + \sqrt{C+1} \right)^{-2/q},$$

where infimum on the left hand side is taken over all possible estimators.

We will apply Lemma 4 with $\mathbb{F} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ in the proof of Theorem 3 and with $\mathbb{F} = \mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ in the proof of statement (ii) of Theorem 4.

Next we quote the Varshamov–Gilbert lemma [see, e.g., Lemma 2.9 in Tsybakov (2009)].

Lemma 5 (Varshamov–Gilbert)**.** *Let ϱ_m be the Hamming distance on $\{0,1\}^m$, $m \in \mathbb{N}^*$, i.e.*

$$\varrho_m(a,b) = \sum_{j=1}^m \mathbf{1}\left\{a_j \neq b_j\right\}, \quad a, b \in \{0,1\}^m.$$

For any $m \geq 8$ there exists a subset $\mathcal{P}_m$ of $\{0,1\}^m$ such that $\left|\mathcal{P}_m\right| \geq 2^{m/8}$, and

$$\varrho_m\big(a, a'\big) \geq \frac{m}{8}, \qquad \forall a, a' \in \mathcal{P}_m.$$

7.2. Proof of Theorem 3. General construction of a finite set of functions

1^0. For any $t \in \mathbb{R}$ set

$$\Lambda(t) = \left(\int_{-1}^1 e^{-1/(1-u^2)} \mathrm{d}u \right)^{-1} e^{-1/(1-t^2)}\ \mathbf{1}_{[-1,1]}(t).$$

Note that Λ is a probability density compactly supported on $[-1,1]$ and infinitely differentiable on the real line, $\Lambda \in \mathbb{C}^\infty(\mathbb{R}^1)$. Obviously, for any $\alpha > 0$ and $r \geq 1$ there exists constant $c_1 = c_1(\alpha, r) < \infty$ such that

$$\Lambda \in \mathbb{N}_{r,1}(\alpha, c_1). \tag{7.3}$$

Define

$$\bar{f}^{(0)}(x) = \prod_{l=1}^d \left[\frac{1}{N} \int_{\mathbb{R}^1} \Lambda(y - x_l) \mathbf{1}_{[-\frac{N}{2}, \frac{N}{2}]}(y) \mathrm{d}y \right], \qquad x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

where parameter $N = N(n) > 8$ will be chosen later. By construction, $\bar{f}^{(0)}$ is a probability density for any choice of N, $\mathrm{supp}(\bar{f}^{(0)}) = [-N/2 - 1, N/2 + 1]^d$, and

$$\bar{f}^{(0)}(x) = N^{-d}, \quad \forall x \in \big[-N/2 + 1, N/2 - 1 \big]^d. \tag{7.4}$$

Moreover, in view of (7.3) and by the Young inequality, there exist constants $\vec{C} = \big(\tilde{C}_1, \dots, \tilde{C}_d\big)$ depending on $\vec{\beta}$ and $\vec{r}$ only such that

$$\bar{f}^{(0)} \in \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{C}\big). \tag{7.5}$$

Note that $\vec{C}$ do not depend on N.

Let $L_0 > 0$ be fixed, and let $f^{(0)}(x) = \varkappa^d \bar{f}^{(0)}\big(x\varkappa\big)$, where $\varkappa > 0$ is chosen in such a way that $f^{(0)}$ belongs to the class $\mathbb{N}_{\vec{r},d}(\vec{\beta}, 2^{-1}\vec{L}_0)$, where $\vec{L}_0 = (L_0, \dots, L_0)$. The existence of such $\varkappa$ independent of N and determined by $\vec{\beta}$, $\vec{r}$ and L_0 is guaranteed by (7.5). Note also that $f^{(0)}$ is a probability density. Moreover, we remark that $\|\bar{f}^{(0)}\|_\infty \le N^{-d}$ since $\int |\Lambda| = 1$. Thus,

$$f^{(0)} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}_0/2, M/2), \tag{7.6}$$

provided that $N > (2M^{-1})^{1/d}\varkappa$. This condition is assumed to be fulfilled.

2^0. Put for any $t \in \mathbb{R}^1$

$$g(t) = \int_{\mathbb{R}^1} \Lambda(y - t)\Big[\mathbf{1}_{[0,1]}(y) - \mathbf{1}_{[-1,0]}(y)\Big]\mathrm{d}y.$$

We obviously have $g \in \mathbb{C}^\infty(\mathbb{R}^1)$, and

$$\text{(i)} \quad \int_{\mathbb{R}^1} g(y)\mathrm{d}y = 0, \qquad \text{(ii)} \ \ \mathrm{supp}(g) \subseteq [-2, 2], \qquad \text{(iii)} \ \ \|g\|_\infty \le 1. \tag{7.7}$$

For any $l = 1, \dots, d$ let $(20\varkappa)^{-1} > \sigma_l = \sigma_l(n) \to 0$, $n \to \infty$, be the sequences to be specified later. Let $M_l = (20\varkappa\sigma_l)^{-1}N$, and without loss of generality assume that M_l, $l = 1, \dots, d$ are integer numbers. Define also

$$x_{j,l} = -\frac{N-4}{4\varkappa} + 8j\sigma_l, \quad j = 1, \dots, M_l, \ \ l = 1, \dots, d,$$

and let $\mathcal{M} = \{1, \dots, M_1\} \times \cdots \times \{1, \dots, M_d\}$. For any $m = (m_1, \dots, m_d) \in \mathcal{M}$ define

$$G_m(x) = \prod_{l=1}^{d} g\left(\frac{x_l - x_{m_l,l}}{\sigma_l}\right), \quad x \in \mathbb{R}^d,$$

$$\Pi_m = \big[x_{m_1,1} - 3\sigma_1, x_{m_1,1} + 3\sigma_1\big] \times \cdots \times \big[x_{m_d,d} - 3\sigma_d, x_{m_d,d} - 3\sigma_d\big] \subset \mathbb{R}^d.$$

Several remarks on these definitions are in order. First, in view of (7.7)(ii)

$$\mathrm{supp}\big(G_m\big) \subset \Pi_m, \quad \forall m \in \mathcal{M}, \tag{7.8}$$

$$\Pi_m \cap \Pi_j = \emptyset, \quad \forall m, j \in \mathcal{M} : \ \ m \ne j. \tag{7.9}$$

Second, since $g \in \mathbb{C}^\infty(\mathbb{R}^1)$, we have that $G_m \in \mathbb{C}^\infty(\mathbb{R}^d)$ for any $m \in \mathcal{M}$. Moreover, for any $l = 1, \dots, d$, any $|h| \le \sigma_l$ and any integer k

$$\mathrm{supp}\Big\{\Delta_{h,l}\left(D_l^k G_m\right)\Big\} \subseteq \Pi_m, \quad \forall m \in \mathcal{M}, \tag{7.10}$$

where $D_l^k G$ stands for the kth order derivative of a function G with respect to the variable x_l, and $\Delta_{h,l}$ is the first order difference operator with step size h in direction of the variable x_l.

For $m \in \mathcal{M}$ define

$$\pi(m) = \sum_{j=1}^{d-1}(m_j - 1)\Big(\prod_{l=j+1}^{d} M_l\Big) + m_d.$$

It is easily checked that π defines enumeration of the set $\mathcal{M}$, and $\pi : \mathcal{M} \to \{1, 2 \dots, |\mathcal{M}|\}$ is a bijection. Let W be a subset of $\{0,1\}^{|\mathcal{M}|}$. Define a family of functions $\{F_w, w \in W\}$ by

$$F_w(x) = A \sum_{m \in \mathcal{M}} w_{\pi(m)} G_m(x), \quad x \in \mathbb{R}^d,$$

where w_j, $j = 1, \dots, |\mathcal{M}|$ are the coordinates of w, and A is a parameter to be specified. It follows from (7.7)(iii), (7.8) and (7.9) that

$$\big\|F_w\big\|_\infty \le A, \quad \forall w \in W, \tag{7.11}$$

and (7.7)(i) implies that

$$\int_{\mathbb{R}^d} F_w(x)\mathrm{d}x = 0, \quad \forall w \in W. \tag{7.12}$$

3^0. Now we find conditions which guarantee that $F_w \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, 2^{-1}\vec{L})$ for any $w \in W$.

Fix $l = 1, \dots, d$, and let $k_l = \lfloor \beta_l \rfloor + 1$ if $\beta_l \notin \mathbb{N}^*$, and $k_l = \lfloor \beta_l \rfloor + 2$ if $\beta_l \in \mathbb{N}^*$ (here $\lfloor x \rfloor$ stands for the maximal integer number strictly less than x).

First, for any $w \in W$ and $h \in \mathbb{R}$

$$\Big\|\Delta_{h,l}^{k_l} F_w\Big\|_{r_l} = \Big\|\Delta_{h,l}^{k_l-1}(\Delta_{h,l} F_w)\Big\|_{r_l} \le |h|^{k_l-1}\Big\|\Delta_{h,l}(D_l^{k_l-1} F_w)\Big\|_{r_l}, \tag{7.13}$$

where the last inequality is found in (Nikol'skii 1977, Section 4.4.4). Next, in view of (7.9) and (7.10) we obtain for any $w \in W$ and any $r_l \ne \infty$

$$\begin{aligned}
\Big\|\Delta_{h,l}(D_l^{k_l-1} F_w)\Big\|_{r_l}^{r_l} &= \sum_{j \in \mathcal{M}} \int_{\Pi_j} \Big|\Delta_{h,l}(D_l^{k_l-1} F_w)(x)\Big|^{r_l}\mathrm{d}x \\
&= A^{r_l} \sum_{j \in \mathcal{M}} w_{\pi(j)} \int_{\Pi_j} \Big|\Delta_{h,l}(D_l^{k_l-1} G_j)(x)\Big|^{r_l}\mathrm{d}x \\
&\le A^{r_l} S_W \big\|g\big\|_{r_l}^{(d-1)r_l} \sigma_l^{-(k_l-1)r_l}\Big(\prod_{j=1}^{d}\sigma_j\Big)\Big\|g^{(k_l-1)}\Big(\cdot - \frac{h}{\sigma_l}\Big) - g^{(k_l-1)}(\cdot)\Big\|_{r_l}^{r_l},
\end{aligned}$$

where we have put $S_W := \sup_{w \in W} |\{j : \ w_j \ne 0\}|$. Thus, for any $r_l \ne \infty$ we have

$$\Big\|\Delta_{h,l}(D_l^{k_l-1} F_w)\Big\|_{r_l} \le A\big\|g\big\|_{r_l}^{(d-1)r_l} \sigma_l^{-(k_l-1)}\Big(S_W \prod_{j=1}^{d}\sigma_j\Big)^{\frac{1}{r_l}}\Big\|g^{(k_l-1)}\Big(\cdot - \frac{h}{\sigma_l}\Big) - g^{(k_l-1)}(\cdot)\Big\|_{r_l}. \tag{7.14}$$

Similarly, we get for any $w \in W$

$$\begin{aligned}
\Big\|\Delta_{h,l}(D_l^{k_l-1} F_w)\Big\|_{\infty} &= \sup_{j \in \mathcal{M}} \sup_{x \in \Pi_j} \Big|\Delta_{h,l}(D_l^{k_l-1} F_w)(x)\Big| \\
&= A \sup_{j \in \mathcal{M}} w_{\pi(j)} \sup_{x \in \Pi_j} \Big|\Delta_{h,l}(D_l^{k_l-1} G_j)(x)\Big| \\
&\le A\big\|g\big\|_{\infty}^{(d-1)} \sigma_l^{-(k_l-1)}\Big\|g^{(k_l-1)}\Big(\cdot - \frac{h}{\sigma_l}\Big) - g^{(k_l-1)}(\cdot)\Big\|_{\infty}.
\end{aligned} \tag{7.15}$$

In view of (7.7)(ii) and $|h| \le \sigma_l$, function $g^{(k_l-1)}\big(\,\cdot - [h/\sigma_l]\big) - g^{(k_l-1)}(\cdot)$ is supported on $[-3,3]$. Therefore the fact that $g \in \mathbb{C}^\infty(\mathbb{R}^1)$ implies for any $r_l \in [1,\infty]$

$$\Big\| g^{(k_l-1)}\big(\,\cdot - h/\sigma_l\big) - g^{(k_l-1)}(\cdot)\Big\|_{r_l} \le 6^{1/r_l}\big\|g^{(k_l)}\big\|_\infty (h/\sigma_l) \le 6^{1/r_l}\big\|g^{(k_l)}\big\|_\infty |h/\sigma_l|^{\beta_l - k_l + 1}.$$

In the last inequality we have used that $0 \le \beta_l - k_l + 1 \le 1$ by definition of k_l. Combining this with (7.13), (7.14) and (7.15) we have for any $|h| \le \sigma_l$ and any $r_l \in [1,\infty]$

$$\Big\|\Delta_{h,l}^{k_l} F_w\Big\|_{r_l} \le A|h|^{\beta_l} 6^{1/r_l} \big\|g\big\|_{r_l}^{d-1} \big\|g^{(k_l)}\big\|_\infty \sigma_l^{-\beta_l}\bigg(S_W \prod_{j=1}^d \sigma_j\bigg)^{1/r_l}. \tag{7.16}$$

If $|h| \ge \sigma_l$ then we note that $\Delta_{h,l}(D_l^{k_l-1}F_w)(\cdot) = (D_l^{k_l-1}F_w)(\cdot - he_l) - (D_l^{k_l-1}F_w)(\cdot)$, and by the triangle inequality

$$\Big\|\Delta_{h,l}(D_l^{k_l-1}F_w)\Big\|_{r_l} \le 2\Big\|D_l^{k_l-1}F_w\Big\|_{r_l} \le 2\Big\|D_l^{k_l-1}F_w\Big\|_{r_l} |h/\sigma_l|^{\beta_l - k_l + 1}.$$

In view of (7.8) and (7.9) we get for any $w \in W$ and any $r_l \ne \infty$

$$\begin{aligned}\Big\|D_l^{k_l-1}F_w\Big\|_{r_l}^{r_l} &= \sum_{j\in\mathcal{M}} \int_{\Pi_j} \Big|D_l^{k_l-1}F_w(x)\Big|^{r_l} \mathrm{d}x = A^{r_l} \sum_{j\in\mathcal{M}} w_{\pi(j)} \int_{\Pi_j} \Big|D_l^{k_l-1}G_j(x)\Big|^{r_l}\mathrm{d}x \\ &\le A^{r_l} S_W \big\|g\big\|_{r_l}^{(d-1)r_l} \Big\|g^{(k_l-1)}\Big\|_{r_l}^{r_l} \sigma_l^{(1-k_l)r_l}\bigg(\prod_{j=1}^d \sigma_j\bigg).\end{aligned}$$

Moreover,

$$\begin{aligned}\Big\|D_l^{k_l-1}F_w\Big\|_\infty &= \sup_{j\in\mathcal{M}} \sup_{x\in\Pi_j} \Big|D_l^{k_l-1}F_w(x)\Big| = A \sup_{j\in\mathcal{M}} w_{\pi(j)} \sup_{x\in\Pi_j} \Big|D_l^{k_l-1}G_j(x)\Big| \\ &\le A\big\|g\big\|_\infty^{(d-1)} \Big\|g^{(k_l-1)}\Big\|_\infty \sigma_l^{(1-k_l)}.\end{aligned}$$

We obtain finally from (7.13) that for any $|h| \ge \sigma_l$ and any $r_l \in [1,\infty]$

$$\Big\|\Delta_{h,l}^{k_l} F_w\Big\|_{r_l} \le A|h|^{\beta_l} 2 \big\|g\big\|_{r_l}^{d-1} \big\|g^{(k_l-1)}\big\|_{r_l} \sigma_l^{-\beta_l}\bigg(S_W \prod_{j=1}^d \sigma_j\bigg)^{1/r_l}. \tag{7.17}$$

Combining (7.16) and (7.17) we conclude that for any $w \in W$ and $r_l \in [1,\infty]$

$$\Big\|\Delta_{h,l}^{k_l} F_w\Big\|_{r_l} \le C_1 A|h|^{\beta_l} \sigma_l^{-\beta_l}\bigg(S_W \prod_{j=1}^d \sigma_j\bigg)^{1/r_l}, \quad \forall h \in \mathbb{R}^1,$$

where $C_1 = \max_l(\|g\|_{r_l}^{d-1} \max\{6^{1/r_l}\|g^{(k_l)}\|_\infty, 2\|g^{(k_l-1)}\|_{r_l}\})$. Thus, if

$$A\sigma_l^{-\beta_l}\bigg(S_W \prod_{j=1}^d \sigma_j\bigg)^{1/r_l} \le (2C_1)^{-1} L_l, \quad \forall l = 1,\dots,d \tag{7.18}$$

then $F_w \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, 2^{-1}\vec{L})$ for any $w \in W$.

4^0. Define for any $w \in W$

$$f_w(x) = f^{(0)}(x) + F_w(x), \qquad x \in \mathbb{R}^d.$$

Remind that $f^{(0)}$ is the probability density belonging to $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}_0/2, M/2)$. Therefore, in view of (7.12) and under condition (7.18), for any $w \in W$

$$\int_{\mathbb{R}^d} f_w(x)\mathrm{d}x = 1, \quad f_w \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \tag{7.19}$$

where the latter inclusion holds because $\min_{j=1,\dots,d} L_j \ge L_0$.

By construction of F_w, for any $w \in W$

$$F_w(x) = 0, \qquad \forall x \notin \Big[-\frac{1}{4\varkappa}(N-4),\ \frac{1}{4\varkappa}(N+4)\Big]^d. \tag{7.20}$$

This yields

$$f_w(x) = f^{(0)}(x) \ge 0, \quad \forall x \notin \Big[-\frac{1}{4\varkappa}(N-4),\ \frac{1}{4\varkappa}(N+4)\Big]^d. \tag{7.21}$$

On the other hand, by (7.4)

$$f^{(0)}(x) = \varkappa^d N^{-d}, \qquad \forall x \in \Big[-\frac{1}{4\varkappa}(N-4),\ \frac{1}{4\varkappa}(N+4)\Big]^d. \tag{7.22}$$

Therefore, if we require

$$A \le \varkappa^d N^{-d}, \tag{7.23}$$

this together with (7.13) implies

$$f_w(x) \ge 0, \qquad \forall x \in \Big[-\frac{1}{4\varkappa}(N-4),\ \frac{1}{4\varkappa}(N+4)\Big]^d.$$

We conclude that $f_w \ge 0$ for any $w \in W$. Moreover, we get from (7.6), (7.11) and (7.23) that $\|f_w\|_\infty \le M$ for any $w \in W$.

All this, together with (7.19), shows that $\{f^{(0)}, f_w, w \in W\}$ is a finite set of probability densities from $\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$. Thus Lemma 4 is applicable with $\mathcal{J}_n = W$ and $\mathbb{F} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$.

5^0. Suppose now that the set W is chosen so that

$$\varrho_{|\mathcal{M}|}\big(w, w'\big) \ge B, \quad \forall w, w' \in W, \tag{7.24}$$

where, we remind, $\varrho_{|\mathcal{M}|}$ is the Hamming distance on $\{0,1\}^{|\mathcal{M}|}$. Here $B = B(n) \ge 1$ is a parameter to be specified. Then we deduce from (7.19), (7.8) and (7.9), that for all $w, w' \in W$

$$\begin{aligned}
\|f_w - f_{w'}\|_p^p &= \|F_w - F_{w'}\|_p^p = A^p \sum_{j\in\mathcal{M}} \Big|w_{\pi(j)} - w'_{\pi(j)}\Big| \int_{\Pi_j} \Big|G_j(x)\Big|^p \mathrm{d}x \\
&= A^p \big\|g\big\|_p^{dp} \Big(\prod_{j=1}^d \sigma_j\Big) \sum_{j\in\mathcal{M}} \Big|w_{\pi(j)} - w'_{\pi(j)}\Big| = A^p \big\|g\big\|_p^{dp} \Big(\prod_{j=1}^d \sigma_j\Big) \varrho_{|\mathcal{M}|}\big(w, w'\big) \\
&\ge \big\|g\big\|_p^{dp} A^p B \Big(\prod_{j=1}^d \sigma_j\Big).
\end{aligned} \tag{7.25}$$

Here we have used that the map π is a bijection. Putting $C_2 = \frac{1}{2}\|g\|_p^d$, we conclude that condition (7.1) of Lemma 4 is fulfilled with

$$\rho_n = C_2 A\Big(B\prod_{j=1}^{d}\sigma_j\Big)^{1/p}. \tag{7.26}$$

Let us remark that (7.26) remains true if we formally put $p=\infty$. Indeed, similarly to (7.25),

$$\|f_w - f_{w'}\|_\infty = \|F_w - F_{w'}\|_\infty = A\sup_{j\in\mathcal{M}}|w_j - w'_j|\|g\|_\infty^d \geq A\|g\|_\infty^d. \tag{7.27}$$

Here we have used (7.9), the fact that the map π is a bijection and, that $w\neq w'$ for all $w, w'\in W$ in view of (7.24).

Now we verify condition (7.2) of Lemma 4. First observe that

$$\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(n)}\big) = \prod_{k=1}^{n}\frac{f_w(X_k)}{f^{(0)}(X_k)}.$$

Since X_k, $k=1,\dots,n$ are i.i.d. random vectors, we have for any $w\in W$

$$\begin{aligned}\mathbb{E}_{f^{(0)}}\left\{\prod_{k=1}^{n}\frac{f_w(X_k)}{f^{(0)}(X_k)}\right\}^2 &= \left\{\int_{\mathbb{R}^d}\frac{f^{(0)}(x)+2F_w(x)+F_w^2(x)}{f^{(0)}(x)}\mathrm{d}x\right\}^n\\ &= \left\{1+\int_{\mathbb{R}^d}\frac{F_w^2(x)}{f^{(0)}(x)}\mathrm{d}x\right\}^n.\end{aligned}$$

The last equality follows from (7.12). By (7.20) and (7.22),

$$\int_{\mathbb{R}^d}\frac{F_w^2(x)}{f^{(0)}(x)}\mathrm{d}x = \varkappa^{-d}N^d\|F_w\|_2^2;$$

hence for any $w\in W$

$$\mathbb{E}_{f^{(0)}}\left\{\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(n)}\big)\right\}^2 = \Big\{1+\varkappa^{-d}N^d\|F_w\|_2^2\Big\}^n \leq \exp\Big\{n\varkappa^{-d}N^d\|F_w\|_2^2\Big\}.$$

Repeating computations that led to (7.17) we have

$$\|F_w\|_2^2 \leq A^2\|g\|_2^{2d}S_W\prod_{j=1}^{d}\sigma_j.$$

The right hand side of the latter inequality does not depend on w; hence we

$$\frac{1}{|W|^2}\sum_{w\in W}\mathbb{E}_{f^{(0)}}\left\{\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\big(X^{(n)}\big)\right\}^2 \leq \exp\Big\{C_3nA^2S_WN^d\Big(\prod_{j=1}^{d}\sigma_j\Big) - \ln\big(|W|\big)\Big\},$$

where we have put $C_3 = \varkappa^{-d}\|g\|_2^{2d}$. Therefore, if

$$C_4nA^2S_WN^d\prod_{j=1}^{d}\sigma_j \leq \ln\big(|W|\big) \tag{7.28}$$

then condition (7.2) of Lemma 4 is fulfilled with $C = 1$.

In order to apply Lemma 4 it remains to specify the set W and the parameters A, N, σ_j, $j = 1, \dots, d$ so that the relationships (7.18), (7.23), (7.24), and (7.28) are simultaneously fulfilled. According to (4), under these conditions the lower bound is given by ρ_n in (7.26).

7.3. Proof of Theorem 3. Derivation of lower bounds in different zones

We begin with the construction of the set W. Let $m \geq 8$ be an integer number whose choice will be made later, and, without loss of generality, assume that $|\mathcal{M}|/m$ is integer. Let $\mathcal{P}_m$ be a subset of $\{0,1\}^m$ such that

$$|\mathcal{P}_m| \geq 2^{m/8}, \quad \varrho_m(z, z') \geq m/8, \ \ \forall z, z' \in \mathcal{P}_m. \tag{7.29}$$

Existence of such set $\mathcal{P}_m$ is guaranteed by Lemma 5. Let $\mathcal{J} := \{1 + \frac{j}{m}|\mathcal{M}|, \ j = 0, \dots, m-1\}$, and note that $\mathcal{J} \subseteq \{1, \dots, |\mathcal{M}|\}$ with the equality in the case $m = |\mathcal{M}|$. Define the map $\Upsilon : \mathcal{P}_m \to \{0,1\}^{|\mathcal{M}|}$ by

$$\Upsilon_j[a] = \begin{cases} a_j, & j \in \mathcal{J}, \\ 0, & j \in \{1, \dots, |\mathcal{M}|\} \setminus \mathcal{J}, \end{cases}$$

and let $W = \Upsilon(\mathcal{P}_m)$. Obviously, $\varrho_{|\mathcal{M}|}(w, w') = \varrho_{|\mathcal{M}|}(\Upsilon[a], \Upsilon[a']) = \varrho_m(a, a')$ for all $w, w' \in W$; therefore (7.29) implies that

$$|W| \geq 2^{m/8}, \quad \varrho_{|\mathcal{M}|}(w, w') \geq 8^{-1}m, \ \ \forall w, w' \in W. \tag{7.30}$$

With such a set W, $S_W \leq m$; moreover, since $\ln(|W|) \geq m \ln 2/8$, condition (7.28) holds true if

$$A^2 n N^d \prod_{j=1}^{d} \sigma_j \leq (8C_4)^{-1} \ln 2. \tag{7.31}$$

We also note that condition (7.18) is fulfilled if we require

$$A\sigma_l^{-\beta_l} \left(m \prod_{j=1}^{d} \sigma_j \right)^{1/r_l} \leq (2C_1)^{-1} L_l, \quad \forall l = 1, \dots, d. \tag{7.32}$$

In addition, (7.24) holds with $B = m/8$.

7.3.1. Tail zone: $p \leq \frac{2+1/\beta}{1+1/s}$

Let $m = |\mathcal{M}|$. By construction, $|\mathcal{M}| = \prod_{l=1}^{d} M_l = (20\varkappa)^{-d} N^d \prod_{l=1}^{d} \sigma_l^{-1}$ and, therefore (7.32) is reduced to

$$A\sigma_l^{-\beta_l} N^{d/r_l} \leq C_5 L_l. \tag{7.33}$$

Thus, choosing

$$\sigma_l = C_6 A^{1/\beta_l} L_l^{-1/\beta_l} N^{\frac{d}{\beta_l r_l}}, \tag{7.34}$$

we guarantee the fulfillment of (7.33) provided that $C_6 \geq \max_{l=1,\dots,d} C_5^{-1/\beta_l}$. Moreover, with this choice (7.31) is reduced to

$$A^{2+1/\beta} N^{d(1+1/s)} \leq C_7 L_\beta n^{-1}, \tag{7.35}$$

where, as before, $L_\beta = \prod_{l=1}^d L_l^{1/\beta_l}$. Moreover, we have from (7.26)

$$\rho_n = C_8 A N^{d/p}, \quad C_8 = C_3 (160\varkappa)^{-1/p}. \tag{7.36}$$

Let $N^d = C_9 A^{-1}$, where constant $C_9 \leq \varkappa^d$ will be specified below; then (7.23) holds. Next, in view of (7.35) and (7.36)

$$A = C_{10}(L_\beta/n)^{\frac{1}{1-1/s+1/\beta}}, \qquad \rho_n = C_{11}(L_\beta/n)^{\frac{1-1/p}{1-1/s+1/\beta}} = C_{11}\big(L_\beta \alpha_n n^{-1}\big)^\nu.$$

We remark that $N \to \infty$ as $n \to \infty$. It remains to check that σ_l, $l = 1, \dots, d$ are small enough. It follows from (7.34) that if $r_l > 1$, then $\sigma_l \to 0$ as $n \to \infty$ since $A \to 0$. If $r_l = 1$, then

$$\sigma_l = C_{12} C_9^{1/\beta_l} L_l^{-1/\beta_l} \leq C_{12}\big(C_9/L_0\big)^{1/\beta_l}.$$

Choosing C_9 small enough we guarantee that $\sigma_l \leq (20\varkappa)^{-1}$, for all $l = 1, \dots, d$. This condition is required in the construction of the family G_m, $m \in \mathcal{M}$. Thus, Lemma 4 can be applied with $\rho_n = C_{11}\big(L_\beta \alpha_n n^{-1}\big)^\nu$, and the result follows.

7.3.2. Dense zone: $\frac{2+1/\beta}{1+1/s} \leq p \leq s(2+\frac{1}{\beta})$

Here, as in the previous case, we let $m = |\mathcal{M}|$. The relationships (7.34) (7.35) and (7.36) remain to be true, but our choice of N will be different.

Let $N = C_{12}$ from some constant C_{12}. This yields in view of (7.35) and (7.36)

$$A = C_{13}(L_\beta/n)^{\frac{\beta}{2\beta+1}}, \qquad \rho_n = C_{14}(L_\beta/n)^{\frac{\beta}{2\beta+1}} = C_{14}\big(L_\beta \alpha_n n^{-1}\big)^\nu.$$

The requirement (7.23) is obviously fulfilled since $A \to 0,\ n \to \infty$. Moreover, we obtain from (7.34) that $\sigma_l \to 0$ as $n \to \infty$ and, therefore, $\sigma_l \leq (20\varkappa)^{-1}$, $l = 1, \dots, d$ for n large enough. Thus, Lemma 4 can be applied with $\rho_n = C_{14}\big(L_\beta \alpha_n n^{-1}\big)^\nu$ and the result follows.

7.3.3. Sparse zone: $s(2+\frac{1}{\beta}) < p < \infty$, $s < 1$

Let $A = \tilde{C}$ and $N = C_{17}$ and suppose that $\tilde{C} \leq C_{17}^{-1} \varkappa^d$; then (7.23) is satisfied. Moreover (7.31) and (7.32) are reduced to

$$n \prod_{j=1}^d \sigma_j \leq \tilde{C}^{-2} C_{18}, \qquad \sigma_l^{-\beta_l} \left(m \prod_{j=1}^d \sigma_j \right)^{1/r_l} \leq \widetilde{C}^{-1} C_{19} L_l, \quad \forall l = 1, \dots, d. \tag{7.37}$$

Let $\tilde{c_1}$, $\tilde{c}_2$ be constants satisfying $\tilde{c_1} \leq \tilde{C}^{-1} C_{18}$, and $\tilde{c}_2 \leq \tilde{C}^{-1} C_{19}$. It is straightforward to check that if we choose

$$m = \tilde{c}_1^{-1+s} \tilde{c}_2^{s/\beta} L_\beta^s n^{1-s}, \qquad \sigma_l = (\tilde{c}_2 L_l)^{-1/\beta_l} \left(\tilde{c}_1 \tilde{c}_2^{1/\beta} L_\beta n^{-1} \right)^{s/(\beta_l r_l)}, \quad l = 1, \dots, d, \tag{7.38}$$

then inequalities (7.37) are fulfilled. With this choice (7.26) is reduced to

$$\rho_n = \widetilde{C}C_{17}\Big(m\prod_{j=1}^{d}\sigma_j\Big)^{1/p} = \widetilde{C}C_{20}(L_\beta n^{-1})^{s/p} = \widetilde{C}C_{20}\big(L_\beta\alpha_n n^{-1}\big)^{\nu}. \tag{7.39}$$

It remains to verify that σ_l are small enough, and that $m \geq 8$, $|\mathcal{M}|/\mathfrak{m} \geq 1$. Note that $m \to \infty$ as $n \to \infty$ because of $s < 1$. Remind also that

$$|\mathcal{M}| = \prod_{l=1}^{d} M_l = (20\varkappa)^{-d}N^d\prod_{l=1}^{d}\sigma_l^{-1} = (20\varkappa)^{-d}C_{17}^d\tilde{c}_1 n\,;$$

hence $|\mathcal{M}|/m \geq (20\varkappa)^{-d}C_{17}^d(\tilde{c}_1\tilde{c}_2^{1/\beta})^{-s}L_0^{-s/\beta}n^s$. Thus $|\mathcal{M}|/m \geq 1$ for large enough n. We note also that $\sigma_l \leq (\tilde{c}_2 L_0)^{-1/\beta_l}$ for all n large enough. Therefore, if we choose $\tilde{C}$ large enough and put $\tilde{c}_2 = \tilde{C}^{-1}C_{19}$ we can ensure that $\sigma_l \leq (20\varkappa)^{-1}$ for all $l = 1, \dots, d$. Thus, Lemma 4 can be applied with $\rho_n = \widetilde{C}C_{20}\big(L_\beta\alpha_n n^{-1}\big)^{\nu}$ and the result follows.

7.3.4. Sparse zone: $s(2+\frac{1}{\beta}) < p < \infty$, $s \geq 1$

Here we consider another choice of the set W. Let $W = \{e_1, e_2, \dots, e_{|\mathcal{M}|}\}$, where e_j, $j = 1, \dots, |\mathcal{M}|$ is the canonical basis in $\mathbb{R}^{|\mathcal{M}|}$. With this choice

$$S_W = 1, \qquad |W| = N^d\prod_{j=1}^{d}\sigma_j^{-1},$$

and (7.24) holds with $B = 1$. Let $N = C_{14}$; then (7.18) and (7.28) take the form

$$A\sigma_l^{-\beta_l}\Big(\prod_{j=1}^{d}\sigma_j\Big)^{1/r_l} \leq (2C_1)^{-1}L_l, \quad \forall l = 1, \dots, d; \tag{7.40}$$

$$A^2 n\prod_{j=1}^{d}\sigma_j \leq C_{15}\ln\Big(\prod_{j=1}^{d}\sigma_j^{-1}\Big). \tag{7.41}$$

Moreover, we get from (7.26)

$$\rho_n = C_{16}A\Big(\prod_{j=1}^{d}\sigma_j\Big)^{1/p}. \tag{7.42}$$

Put $\varepsilon = \sqrt{\ln n/n}$ and

$$A = c_1 L_\beta^{\frac{1}{2-2/s+1/\beta}}\varepsilon^{\frac{1-1/s}{1-1/s+1/(2\beta)}}, \qquad \sigma_l = c_2 L_\beta^{\frac{1-2/r_l}{\beta_l(2-2/s+1/\beta)}}\varepsilon^{\frac{1-1/s+1/(\beta r_l)}{\beta_l(1-1/s+1/(2\beta))}}L_l^{-1/\beta_l}. \tag{7.43}$$

We have

$$\prod_{l=1}^{d}\sigma_l = c_2^d L_\beta^{-\frac{2}{2-2/s+1/\beta}}\varepsilon^{\frac{1/\beta}{1-1/s+1/(2\beta)}},$$

and it is evident that $\prod_{l=1}^d \sigma_l \le \varepsilon^{1/(\beta+1/2)}$ for all n large enough; hence $\ln(\prod_{l=1}^d \sigma_l^{-1}) \ge \ln n/(2\beta+1)$. Then is is easily checked that our choice (7.43) satisfies (7.40) and (7.41) provided that

$$c_1 \le (2C_1)^{-1}, \qquad c_2 \le 1 \qquad c_1^2 c_2^d \le C_{15}/(2\beta+1). \tag{7.44}$$

Here we have also used that $d - 1/s \ge 0$. Note also that if $s > 1$ then

$$A \to 0, \qquad \max_{l=1,\dots,d} \sigma_l \to 0, \quad n \to \infty,$$

which ensures (7.23) and $\sigma_l \le (20\varkappa)^{-1}$, $l = 1, \dots, d$ for all n large enough.

On the other hand, if $s = 1$ then we should add to (7.44) the conditions

$$c_1 L_\beta^{\frac{1}{2-2/s+1/\beta}} \le C_{14}\varkappa^d, \qquad c_2 \max_{l=1,\dots,d}\left[L_\beta^{\frac{1/\beta_l - 2/(\beta_l r_l)}{2-2/s+1/\beta}} L_0^{-1/\beta_l}\right] \le (20\varkappa)^{-1}.$$

Obviously, both restrictions hold if we choose c_1 and c_2 small enough, but now these constants may depend on $\vec{L}$. Note, however, that if $\max_{l=1,\dots,d} L_l \le L_\infty$ then c_1 and c_2 can be chosen depending on L_0 and L_∞ only.

Using (7.42) and (7.43) we conclude that Lemma 4 is applicable with

$$\rho_n = C_{16} L_\beta^{\frac{1/2-1/p}{1-1/s+1/(2\beta)}} \left(\frac{\ln n}{n}\right)^{\frac{1-1/s+1/(p\beta)}{2(1-1/s+1/(2\beta))}} = C_{16} L_\beta^{\frac{1/2-1/p}{1-1/s+1/(2\beta)} - \nu} \left(\frac{L_\beta \alpha_n}{n}\right)^{\nu}. \tag{7.45}$$

that completes the proof of statement (i) of the theorem.

7.3.5. Proof of statement (ii): sparse zone, $p = \infty$, $s \le 1$

The proof in this case coincides with the one for the sparse zone with $s < 1$. Thus, we keep (7.37), (7.38), and, in view of (7.27), (7.39) is replaced by $\rho_n = \widetilde{C} C_{17}$. Since ρ_n does not tend to 0 as $n \to \infty$, a consistent estimator does not exist. All other details of the proof remain unchanged. This completes the proof of Theorem 3. ∎

7.4. Proof of statement (ii) of Theorem 4

The proof goes along the lines of the proof of Theorem 3 with modifications indicated below.

We start with the following simple observation: for any $M > 0$ and $y > 0$ one has

$$\|g\|_\infty \le M, \ \ \mathrm{supp}\{g\} \subseteq [-y, y]^d \ \ \Rightarrow \ \ g \in \mathbb{G}_\theta\big(M(2y+4)^{d/\theta}\big), \ \ \forall \theta \in (0, 1]. \tag{7.46}$$

This is an immediate consequence of the fact that conditions $\|g\|_\infty \le M$, $\mathrm{supp}\{g\} \subseteq [-y, y]^d$ imply that $\|g^*\|_\infty \le M$ and $\mathrm{supp}\{g\} \subseteq [-y-2, y+2]^d$.

Next, we note that the lower bounds of Theorem 3 in the dense and sparse zones are proved over the set of compactly supported densities. Hence they are valid also on $\mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$, provided that R is large enough. Hence, if $p \ge \frac{2+1/\beta}{1/\theta+1/s}$ the assertion of the theorem follows.

Let $p < \frac{2+1/\beta}{1/\theta+1/s}$. The proof of the lower bound here differs from the proof of Theorem 3 only in construction of the function $f^{(0)}$.

Let $f^{(0)}$ be the function constructed exactly as in the proof of Theorem 3 with $N = N_0$ fixed throughout the asymptotics $n \to \infty$, and such that $f^{(0)} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, 4^{-1}\vec{L}_0, 4^{-1}M)$. Since N_0 is fixed, $f^{(0)}$ is compactly supported and, by (7.46) we have that $f^{(0)} \in \mathbb{G}_\theta(R_1)$ for some large enough $R_1 > 0$. Define

$$\bar{f}^{(\theta)}(x) = \prod_{l=1}^{d} \left[N^{-1/\theta} \int_{\mathbb{R}} \Lambda(y - x_l) \mathbf{1}_{[-\frac{N}{2}, \frac{N}{2}]}(y) \mathrm{d}y \right], \qquad x = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where $N = N(n) \to \infty$ will be specified later. Let $\tilde{f}^{(\theta)}(x) = \varsigma^d \bar{f}^{(\theta)}(\varsigma x)$, where $\varsigma > 0$ is chosen to guarantee $\tilde{f}^{(\theta)} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, 4^{-1}\vec{L}_0, 4^{-1}M)$. We note however that, in contrast to the case $\theta = 1$, $\tilde{f}^{(\theta)}$ is not a probability density. In particular, $\int \tilde{f}^{(\theta)} \to 0$ as $N \to \infty$, because $\theta < 1$. Define

$$f^{(\theta)} = (1 - p_N) f^{(0)} + \tilde{f}^{(\theta)},$$

where $p_N := \int \tilde{f}^{(\theta)}$ ensures $\int f^{(\theta)} = 1$.

Note also that $f^{(1)} = \tilde{f}^{(1)}$ since $\tilde{f}^{(1)}$ is a probability density and, therefore $p_N = 1$. Thus, we can assert that

$$f^{(\theta)} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, 2^{-1}\vec{L}_0, 2^{-1}M), \quad \int f^{(\theta)} = 1, \quad f^{(\theta)} \geq 0.$$

Note that, by construction, $\tilde{f}^{(\theta)}$ is supported on the cube $[(-N/2 - 1)/\varsigma, (N/2 + 1)/\varsigma]^d$ and bounded by $N^{-d/\theta}\varsigma^d$. Therefore, in view of (7.46), $\tilde{f}^{(\theta)} \in \mathbb{G}_\theta(R_2)$ for some large enough R_2.

Let W be the parameter set as defined in the proof of Theorem 3. For any $w \in W$ and any $\theta < 1$ we let

$$f_w^{(\theta)}(x) = f^{(\theta)}(x) + F_w(x), \qquad x \in \mathbb{R}^d,$$

where functions F_w are constructed as in the proof of Theorem 3. If instead of (7.23) we require

$$A \leq \left[\varkappa^d + \varsigma^d\right] N^{-d/\theta}, \tag{7.47}$$

then we obtain in view of (7.11) and (7.47) that $\{F_w,\ w \in W\} \subset \mathbb{G}_\theta(R_3)$ for some large enough R_3. All said above one allows to conclude that $\{f^{(\theta)},\ f_w^{(\theta)},\ w \in W\}$ is a finite set of probability densities from $\mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$ for some large enough $R > 0$, and Lemma 4 is applicable with $\mathcal{J}_n = W$ and $\mathbb{F} = \mathbb{G}_\theta(R) \cap \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, M)$.

Note also that if $\theta = 1$ we come to the construction used in the proof of Theorem 3 and, therefore, the statement of the theorem in the case $\theta = 1$ follows.

Suppose now that $\theta < 1$. We will follow construction of the set W for the tail zone which is given in Subsection 7.3.1. Choose $m = |\mathcal{M}|$ and note that (7.33), (7.34), (7.36) remain unchanged, while (7.35) should be replaced by

$$A^{2+1/\beta} N^{d(1/\theta + 1/s)} \leq C_7 L_\beta n^{-1}. \tag{7.48}$$

Now we choose $N^d = cA^{-\theta}$ with $c \leq \varkappa^d + \varsigma^d$; then (7.47) is valid. We obtain from (7.48) that

$$A = C_8 (L_\beta / n)^{\frac{1}{1 - \theta/s + 1/\beta}}, \qquad \rho_n = C_9 (L_\beta / n)^{\frac{1 - \theta/p}{1 - \theta/s + 1/\beta}}.$$

Finally, because (7.34) remains intact, $\sigma_l \to 0$ as $n \to \infty$ for any $l = 1, \dots, d$; this follows from $A \to 0$ and $\theta < 1$. This completes the proof. ∎

7.5. Proof of the lower bound in (4.4)

The required result will follow from the lower bound of Theorem 3 in the tail zone (see Section 7.3.1) if we will show that for any given $R > 0$ and $\theta \in (0,1)$

$$f^{(0)} \notin \mathbb{G}_\theta(R), \quad f_w \notin \mathbb{G}_\theta(R), \quad \forall w \in W. \tag{7.49}$$

First we note that $f^{(0)} = N^{-d}$ for $x \in [-(N-2)/(2\varkappa), (N-2)/(2\varkappa)]^d$; therefore, $\|[f^{(0)}]^*\|_\theta \to \infty$ as $N \to \infty$, because $\theta < 1$.

Next, in view of (7.21), $f_w(x) = f^{(0)}(x)$ for any $x \notin [-(N-4)/(4\varkappa),\ (N-4)/(4\varkappa)]^d$, which also implies

$$\inf_{w \in W} \|f_w^*\|_\theta \to \infty, \quad N \to \infty.$$

It remains to note that in the tail zone the parameter N is chosen so that $N = N(n) \to \infty$ as $n \to \infty$. This completes the proof of (7.49). ∎

Appendix A: Proofs of auxiliary results of Section 5

A.1. Measurability

Write $f(x, X^{(n)}) := \hat{f}_{\hat{h}(x)}(x)$, and note that the map $f : \mathbb{R}^d \times \mathbb{R}^{dn} \to \mathbb{R}$ is completely determined by the kernel K and the set $\mathcal{H}$. We need to show that f is a Borel function.

Let $R_h(x, X^{(n)}) := \hat{R}_h(x)$, and note that for every $h \in \mathcal{H}$, the map $R_h : \mathbb{R}^d \times \mathbb{R}^{dn} \to \mathbb{R}$ is a continuous function. This follows from the continuity of the kernel K and from the fact that $\mathcal{H}$ is a finite set. The continuity of K also implies that the map $f_h : \mathbb{R}^d \times \mathbb{R}^{dn} \to \mathbb{R}$ is a continuous function for any $h \in \mathcal{H}$, where $f_h(x, X^{(n)}) := \hat{f}_h(x)$. Next, denote by $\mathfrak{B}$ the Borel σ-algebra on $\mathbb{R}^d \times \mathbb{R}^{dn}$, and let $b : \mathbb{R}^d \times \mathbb{R}^{dn} \to \mathcal{H}$ be the function $b(x, X^{(n)}) := \hat{h}(x)$. We obviously have for any given $h \in \mathcal{H}$

$$\Big\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^{dn} : b(x,y) = h\Big\} = \bigcup_{\eta \in \mathcal{H}} \Big\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^{dn} : R_h(x,y) - R_\eta(x,y) \le 0\Big\} \in \mathfrak{B},$$

where the last inclusion follows from the continuity of R_η, $\eta \in \mathcal{H}$. Here we have also used that $\mathcal{H}$ is a finite set. It remains to note that

$$\hat{f}_{\hat{h}(x)}(x) = \sum_{h \in \mathcal{H}} f_h(x, X^{(n)}) \mathbf{1}\big\{b\big(x, X^{(n)}\big) = h\big\},$$

and the required statement follows.

A.2. Proof of Lemma 1

1^0. Note that $\check{M}_h(g,x) = 4^{-1}\hat{M}_h(g,x)$ and let $\mathcal{H}_0 = \{h \in \mathcal{H} : A_h(g,x) \ge 4\varkappa \ln n/(nV_h)\}$.

For any $h \in \mathcal{H}_0$ we have

$$\sqrt{\frac{\varkappa A_h(g,x)\ln n}{nV_h}} \ge \frac{2\varkappa \ln n}{nV_h} \quad \Rightarrow \quad M_h(g,x) \le \frac{3}{4} A_h(g,x).$$

Therefore,

$$|\hat{A}_h(g,x) - A_h(g,x)| \le \chi_h(g,x) + M_h(g,x) \le \chi_h(g,x) + (3/4)A_h(g,x).$$

We have for any $h \in \mathcal{H}_0$

$$\begin{aligned} \left|\check{M}_h(g,x) - M_h(g,x)\right| &= \left|\sqrt{\frac{\varkappa \ln n}{nV_h}} \frac{\hat{A}_h(g,x) - A_h(g,x)}{\hat{A}_h^{1/2}(g,x) + A_h^{1/2}(g,x)}\right| \\ &\le \sqrt{\frac{\varkappa \ln n}{nV_h}} \left(\frac{\chi_h(g,x) + (3/4)A_h(g,x)}{A_h^{1/2}(g,x)}\right) \le \frac{1}{2}\chi_h(g,x) + \frac{3}{4}M_h(g,x). \end{aligned}$$

It yields for any $h \in \mathcal{H}_0$

$$\left[\check{M}_h(g,x) - \frac{7}{4}M_h(g,x)\right]_+ \le \frac{1}{2}\chi_h(g,x), \quad \left[M_h(g,x) - 4\check{M}_h(g,x)\right]_+ \le 2\chi_h(g,x). \quad \text{(A.1)}$$

(b). Now consider the set $\mathcal{H}_1 := \mathcal{H} \setminus \mathcal{H}_0$. Here $A_h(g,x) \le 4\varkappa \ln n/(nV_h)$, and, by definition of M_h we have

$$\frac{1}{4}A_h(g,x) \le \frac{\varkappa \ln n}{nV_h} \le M_h(g,x) \le \frac{3\varkappa \ln n}{nV_h}, \quad \forall h \in \mathcal{H}_1. \quad \text{(A.2)}$$

Note that we have $\hat{M}_h(g,x) \ge \varkappa \ln n/(nV_h)$ for all h. This together with (A.2) shows that

$$[M_h(g,x) - 3\tilde{M}_h(g,x)]_+ = 0, \quad \forall h \in \mathcal{H}_1. \quad \text{(A.3)}$$

Furthermore, for any $h \in \mathcal{H}_1$

$$\hat{A}_h(g,x) \le A_h(g,x) + \chi_h(g,x) + M_h(g,x) \le \chi_h(g,x) + \frac{7\varkappa \ln n}{nV_h}.$$

Therefore

$$\begin{aligned} \check{M}_h(g,x) &= \sqrt{\frac{\varkappa \hat{A}_h(g,x) \ln n}{nV_h}} + \frac{\varkappa \ln n}{nV_h} \le \sqrt{\frac{\varkappa \chi_h(g,x) \ln n}{nV_h}} + (\sqrt{7}+1)\frac{\varkappa \ln n}{nV_h} \\ &\le \frac{1}{2}\chi_h(g,x) + \left(\sqrt{7} + \frac{3}{2}\right)\frac{\varkappa \ln n}{nV_h} \le \frac{1}{2}\chi_h(g,x) + \left(\sqrt{7} + \frac{3}{2}\right)M_h(g,x). \end{aligned}$$

To get the penultimate inequality we have used that $\sqrt{|ab|} \le 2^{-1}(|a| + |b|)$. Thus, it is shown that

$$\left[\check{M}_h(g,x) - \left(\sqrt{7} + \frac{3}{2}\right)M_h(g,x)\right]_+ \le \frac{1}{2}\chi_h(g,x), \quad \forall h \in \mathcal{H}_1. \quad \text{(A.4)}$$

Relations (A.4), (A.3) and (A.1) imply statement of the lemma. ∎

A.3. Proof of Lemma 2

1^0. Let $g : \mathbb{R}^d \to \mathbb{R}^1$ be a fixed bounded function, and let

$$\xi_h(g,x) := \frac{1}{n}\sum_{i=1}^{n} g_h(X_i - x) - \int g_h(t-x)f(t)\mathrm{d}t, \quad h \in \mathcal{H}.$$

With this notation $\xi_h(x) = \xi_h(K,x)$ and $\hat{A}_h(g,x) - A_h(g,x) = \xi_h(|g|,x)$. Therefore moment bounds on $\zeta_1(x)$, $\zeta_3(x)$ and $\zeta_4(x)$ will follow from those on $\xi_h(g,x)$ with substitution $g \in \{K, Q, |K|, |Q|\}$. Since $M_h(g,x)$ depends on g only via $|g|$ and $\|g\|_\infty$ [see (2.6)–(2.7)], $M_h(g,x) = M_h(|g|,x)$, and moment bounds on $\zeta_1(x)$ and $\zeta_3(x)$ are identical. The bound on $\zeta_4(x)$ will follow from bounds on $\zeta_1(x)$ and $\zeta_3(x)$ with only one modification: kernel K should be replaced by Q. As for $\zeta_2(x)$, $\xi_{h,\eta}(x)$ cannot be represented in terms of $\xi_h(g,x)$ with function g independent of h and η; see (2.3). However, the bounds on $\zeta_2(x)$ will be obtained similarly with minor modifications. Thus it suffices to bound $\mathbb{E}_f[\zeta_1(x)]^q$ and $\mathbb{E}_f[\zeta_2(x)]^q$.

2^0. We start with bounding $\mathbb{E}_f[\zeta_1(x)]^q$. For any $z > 0$, $h \in \mathcal{H}$ and $q \geq 1$ one has

$$\mathbb{E}_f\left[|\xi_h(x)| - \sqrt{\frac{2\mathrm{k}_\infty A_h(K,x)z}{nV_h}} - \frac{2\mathrm{k}_\infty z}{3nV_h}\right]_+^q \leq 2\Gamma(q+1)\left[\sqrt{\frac{2\mathrm{k}_\infty A_h(K,x)}{nV_h}} + \frac{2\mathrm{k}_\infty}{3nV_h}\right]^q e^{-z}. \qquad \text{(A.5)}$$

This inequality follows by integration of the Bernstein inequality and the following bound on the second moment of $\xi_h(x)$:

$$\mathbb{E}_f|\xi_h(x)|^2 \;\leq\; \frac{\mathrm{k}_\infty}{nV_h}\int |K_h(t-x)|f(t)\mathrm{d}t = \frac{\mathrm{k}_\infty A_h(K,x)}{nV_h}.$$

We will show that $\mathbb{E}_f[\zeta_1(x)]^q$ is bounded by the expression appearing on the right hand side of (5.2). In fact, we will prove a stronger inequality. Let for some $l > 0$

$$\lambda_h := (1+q)\ln\big(1/V_h\big) + \ln\big(F^{-1}(x) \wedge n^l\big).$$

In suffices to show that (5.2) holds when in the definition of $\zeta_1(x)$ the quantity $M_h(K,x)$ replaced by $\tilde{M}_h(K,x)$, where

$$\tilde{M}_h(K,x) = \sqrt{\frac{2\mathrm{k}_\infty A_h(K,x)\lambda_h}{nV_h}} + \frac{2\mathrm{k}_\infty\lambda_h}{3nV_h}.$$

Indeed, since and $n^{-d} \leq V_h \leq 1$ for any $h \in \mathcal{H}$, we have that

$$\lambda_h \leq (q+1)d\ln n + l\ln n.$$

Therefore $\tilde{M}_h(K,x) \leq M_h(K,x)$ for all $x \in \mathbb{R}^d$ and $h \in \mathcal{H}$ provided that

$$\varkappa \geq \mathrm{k}_\infty\big[d(2q+4) + 2l\big].$$

Thus if we establish (5.2) with $M_h(K,x)$ replaced by $\tilde{M}_h(K,x)$, the required bound for $\mathbb{E}_f[\zeta_1(x)]^q$ will be proved.

We have for any $h \in \mathcal{H}$

$$\exp\{-\lambda_h\} = \big(V_h\big)^{q+1}\big\{F(x) \vee n^{-l}\big\}.$$

Furthermore, taking into account that $A_h(g,x) \le V_h^{-1}\|g\|_\infty$ for any g, we obtain

$$\sqrt{\frac{2\mathrm{k}_\infty A_h(K,x)}{nV_h}} + \frac{2\mathrm{k}_\infty}{3nV_h} \le \frac{2\mathrm{k}_\infty}{\sqrt{n}V_h}.$$

Here we have used that $n \ge 3$. If we set $z = \lambda_h$ then (A.5) together with two previous display formulas yields

$$\begin{aligned}
\mathbb{E}_f[\zeta_1(x)]^q &= \mathbb{E}_f \sup_{h\in\mathcal{H}} \Big[|\xi_h(x)| - M_h(K,x)\Big]_+^q \le \sum_{h\in\mathcal{H}} \mathbb{E}_f\Big[|\xi_h(x)| - \tilde{M}_h(K,x)\Big]_+^q \\
&\le 2\Gamma(q+1)\big(2\mathrm{k}_\infty\big)^q n^{-q/2}\big\{F(x)\vee n^{-l}\big\}\sum_{h\in\mathcal{H}} V_h \\
&\le 2^{d+1}\Gamma(q+1)\big(2\mathrm{k}_\infty\big)^q n^{-q/2}\big\{F(x)\vee n^{-l}\big\}. \qquad \text{(A.6)}
\end{aligned}$$

As it was mentioned above, under the same conditions inequality (A.6) holds for $\mathbb{E}_f[\zeta_3(x)]^q$.

As for the moment bound for $\zeta_4(x)$, in all formulas above K should be replaced by Q and k_∞ by k_∞^2 since $\|Q\|_\infty \le \mathrm{k}_\infty^2$. Specifically, if $\varkappa \ge \mathrm{k}_\infty^2[d(2q+4)+2l]$ then

$$\mathbb{E}_f[\zeta_4(x)]^q \le 2^{d+1}\Gamma(q+1)\big(2\mathrm{k}_\infty^2\big)^q n^{-q/2}\big\{F(x)\vee n^{-l}\big\}.$$

3^0. Now we turn to bounding $\mathbb{E}_f[\zeta_2(x)]^q$. We have similarly to (A.5)

$$\begin{aligned}
&\mathbb{E}_f\left[|\xi_{h,\eta}(x)| - \sqrt{\frac{2\mathrm{k}_\infty^2 A_{h\vee\eta}(Q,x)z}{nV_{h\vee\eta}}} - \frac{2\mathrm{k}_\infty^2 z}{3nV_{h\vee\eta}}\right]_+^q \\
&\qquad\le 2\Gamma(q+1)\left[\sqrt{\frac{2\mathrm{k}_\infty^2 A_{h\vee\eta}(Q,x)}{nV_{h\vee\eta}}} + \frac{2\mathrm{k}_\infty^2}{3nV_{h\vee\eta}}\right]^q e^{-z}.
\end{aligned}$$

Here we have used the following bound on the second moment of $\xi_{h,\eta}(x)$:

$$\begin{aligned}
\mathbb{E}_f|\xi_{h,\eta}(x)|^2 &\le \frac{\|Q_{h,\eta}\|_\infty}{nV_{h\vee\eta}}\int\Big|\frac{1}{V_{h\vee\eta}}Q_{h,\eta}\Big(\frac{t-x}{h\vee\eta}\Big)\Big|f(t)\mathrm{d}t \\
&\le \frac{\|Q\|_\infty}{nV_{h\vee\eta}}\int|Q_{h\vee\eta}(t-x)|f(t)\mathrm{d}t = \frac{\mathrm{k}_\infty^2 A_{h\vee\eta}(Q,x)}{nV_{h\vee\eta}}\ .
\end{aligned}$$

The further proof goes along the same lines as the above proof with the following minor modifications: in all formulas k_∞ should be replaced with k_∞^2, $V_{h\vee\eta}$ should be written instead of V_h, and $\varkappa$ should satisfy $\varkappa \ge \mathrm{k}_\infty^2[d(2q+4)+2l]$. The statement of the lemma holds with constant $C_0 = 2^{d^2+1}\Gamma(q+1)(2\mathrm{k}_\infty^2)^q$. Combining the above bounds we complete the proof. ∎

Appendix B: Proofs of auxiliary results of Section 6

B.1. Proof of Lemma 3

We have

$$B_h(f,x) = \int K(u)\big[f(x+uh)-f(x)\big]\mathrm{d}u = \int\prod_{j=1}^d w_\ell(u_j)\big[f(x+uh)-f(x)\big]\mathrm{d}u.$$

First, we note that $f(x+uh)-f(x)$ can be represented by the telescopic sum

$$f(x+uh)-f(x)=\sum_{j=1}^{d}\Delta_{u_jh_j,j}f(x_1,\dots,x_j,x_{j+1}+u_{j+1}h_{j+1},\dots,x_d+u_dh_d), \tag{B.1}$$

where we put formally $h_{d+1}u_{d+1}=0$.

Next, for any function $g:\mathbb{R}^d\to\mathbb{R}^1$ and $j=1,\dots,d$ we have

$$\begin{aligned}&\int w_\ell(u_j)\Delta_{u_jh_j,j}g(x)\mathrm{d}u_j=\int\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+1}\frac{1}{i}w\Big(\frac{u_j}{i}\Big)\Delta_{u_jh_j,j}g(x)\mathrm{d}u_j\\&=(-1)^{\ell-1}\int w(z)\sum_{i=1}^{\ell}\binom{\ell}{i}(-1)^{i+\ell}\Delta_{izh_j,j}g(x)\mathrm{d}z=(-1)^{\ell-1}\int w(z)\Delta^{\ell}_{zh_j,j}\,g(x)\mathrm{d}z.\end{aligned} \tag{B.2}$$

The last equality follows from the definition of ℓ-th order difference operator (3.1). Thus (B.2) and (B.1) imply that $B_h(f,x)=\sum_{j=1}^{d}B_{h,j}(f,x)$, where

$$\begin{aligned}&(-1)^{\ell-1}B_{h,j}(f,x):=\\&\int\int w(z)\Delta^{\ell}_{zh_j,j}\,f(x_1,\dots,x_j,x_{j+1}+u_{j+1}h_{j+1},\dots,x_d+u_dh_d)\mathrm{d}z\prod_{m=j+1}^{d}w_\ell(u_m)\mathrm{d}u_m.\end{aligned}$$

Therefore, by the Minkowski inequality for integrals [see, e.g., (Folland 1999, Section 6.3)]

$$\begin{aligned}&\big\|B_{h,j}(f,\cdot)\big\|_{r_j}\\&\le\int\int|w(z)|\Big\|\Delta^{\ell}_{zh_j,j}\,f(\cdot,\dots,\cdot,\cdot+u_{j+1}h_{j+1},\dots,\cdot+u_dh_d)\Big\|_{r_j}\mathrm{d}z\prod_{m=j+1}^{d}|w_\ell(u_m)|\mathrm{d}u_m\\&=\int\int|w(z)|\big\|\Delta^{\ell}_{zh_j,j}\,f\big\|_{r_j}\mathrm{d}z\prod_{m=j+1}^{d}|w_\ell(u_m)|\mathrm{d}u_m.\end{aligned}$$

Since $f\in\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ one has

$$\big\|B_{h,j}(f,\cdot)\big\|_{r_j}\le\ L_jh_j^{\beta_j}\int\int|w(z)|\,|z|^{\beta_j}\mathrm{d}z\prod_{m=j+1}^{d}|w_\ell(u_m)|\mathrm{d}u_m\ \le\ C_1L_jh_j^{\beta_j}.$$

This proves (6.6). To get (6.7) we first note that the condition $s\ge1$ implies $\tau(p)>0$ and $\tau_j>0$, $j=1,\dots,d$. Then the inequality in (6.7) follows by the same reasoning with r_j replaced by q_j, β_j replaced by γ_j and with the use of embedding (6.2). ∎

B.2. Proof of Proposition 1

By definition of J_m and $\mathcal{X}_m$,

$$J_m\le2^{p(m+1)}\varphi^p|\mathcal{X}_m|, \tag{B.3}$$

and now we bound from above $|\mathcal{X}_m|$. By definition of $\mathcal{X}_m$ we have for any $h \in \mathcal{H}$

$$|\mathcal{X}_m| \le \Big|\{x : \sup_{\eta \ge h} M_\eta(x) > 2^{m-1}\varphi\}\Big| + \sum_{j=1}^{d} \Big|\{x : B^*_{h,j}(f,x) > 2^{m-1}\varphi\}\Big| \\ =: J_{m,1}(h) + J_{m,2}(h). \tag{B.4}$$

Recall that with the introduced notation

$$M_\eta(x) = \sqrt{\varkappa A_\eta(x)\delta V_\eta^{-1}} + \varkappa\delta V_\eta^{-1}, \quad \eta \in \mathcal{H}.$$

For any $h \in \mathcal{H}$ we have

$$J_{m,2}(h) \;=\; \sum_{j \in I\setminus I_\infty} \Big|\{x : B^*_{h,j}(f,x) > 2^{m-1}\varphi\}\Big| \;+\; \sum_{j \in I_\infty} \Big|\{x : B^*_{h,j}(f,x) > 2^{m-1}\varphi\}\Big| \\ =:\; J^{(1)}_{m,2}(h) + J^{(2)}_{m,2}(h). \tag{B.5}$$

By the Chebyshev inequality and (6.12) for any h

$$J^{(1)}_{m,2}(h) \le \sum_{j \in I\setminus I_\infty} \big[2^{(m-1)}\varphi\big]^{-r_j} \|B^*_{h,j}(f,\cdot)\|^{r_j}_{r_j} \;\le\; c_1 \sum_{j \in I\setminus I_\infty} 2^{-r_j m}\varphi^{-r_j} L_j^{r_j} h_j^{\beta_j r_j}. \tag{B.6}$$

In addition, if $s \ge 1$ then the Chebyshev inequality and (6.13) yield

$$J^{(1)}_{m,2}(h) \le \sum_{j \in j \in I\setminus I_\infty} \big[2^{(m-1)}\varphi\big]^{-q_j} \|B^*_{h,j}(f,\cdot)\|^{q_j}_{q_j} \le \tilde{c}_1 \sum_{j \in I\setminus I_\infty} 2^{-q_j m}\varphi^{-q_j} L_j^{q_j} h_j^{\gamma_j q_j}. \tag{B.7}$$

In order to prove statements (i)–(iv) of the proposition we bound quantities $J_{m,1}(h)$ and $J_{m,2}(h)$ with bandwidth $h = h[m]$ specified in an appropriate way.

B.2.1. Proof of statements (i) *and* (ii)

Note that the result of the statement of (i) coincides formally with those of the statement (ii) when $\theta = 1$. This implies obviously that the statement (ii) in the case $\theta = 1$ follows from the statement (i). So, with slight abuse of the notation we will identify the case $\theta = 1$ with the assumption that f is a probability density.

1^0. We start with bounding the term $J_{m,1}(h)$ on the right hand side of (B.4). Assume that $h \in \mathcal{H}$ is such that

$$\varkappa\delta V_h^{-1} < 2^{m-2}\varphi; \tag{B.8}$$

then by the Chebyshev inequality

$$\begin{aligned} J_{m,1}(h) \;&\le\; \Big|\big\{x : \sup_{\eta \ge h} \sqrt{\varkappa A_\eta(x)\delta V_\eta^{-1}} > 2^{m-2}\varphi\big\}\Big| \\ &\le\; \sum_{\substack{\eta \ge h \\ \eta \in \mathcal{H}}} \Big|\{x : A_\eta(x) > 2^{2m-4}\varphi^2\varkappa^{-1}\delta^{-1}V_\eta\}\Big| \\ &\le\; \big(2^{-2m+4}\varphi^{-2}\delta\varkappa\big)^\theta \sum_{\substack{\eta \ge h \\ \eta \in \mathcal{H}}} \|A_\eta\|^\theta_\theta V_\eta^{-\theta} \;\le\; c_2\big(2^{-2m}\varphi^{-2}\delta\big)^\theta \sum_{\substack{\eta \ge h \\ \eta \in \mathcal{H}}} V_\eta^{-\theta}, \end{aligned}$$

where we have taken into account that, for any η, $\|A_\eta\|_\theta \le R$ if $f \in \mathbb{G}_\theta(R)$ with $\theta < 1$, and $\|A_\eta\|_1 \le \mathrm{k}_\infty^2$. By definition of $\mathcal{H}$, for any $\eta \ge h$, $\eta \in \mathcal{H}$, we have $V_\eta = V_h 2^{k_1+\cdots+k_d}$ for some $k_1, \dots, k_d \ge 1$, which implies that $\sum_{\eta \ge h} V_\eta^{-\theta} \le (1-2^{-\theta})^{-d} V_h^{-\theta}$. Thus, we conclude that for any h satisfying (B.8) one has

$$J_{m,1}(h) \le c_3 \left(2^{-2m}\varphi^{-2}\delta V_h^{-1}\right)^\theta. \tag{B.9}$$

2^0. Let $\tilde{h} = (\tilde{h}_1, \dots, \tilde{h}_d) \in (0,\infty]^d$ be given by

$$\tilde{h}_j = \left(c_4 L_j^{-1}\varphi\right)^{1/\beta_j} 2^{\frac{m}{\beta_j}\left(1-\frac{\theta(2+1/\beta)}{r_j(1+\theta/s)}\right)}, \quad j = 1, \dots, d, \tag{B.10}$$

where constant c_4 will be specified later. Let us prove that $\tilde{h} \in \left[n^{-1}, 1\right]^d$ for large enough n.

Denote

$$a = \frac{1-\theta/s+1/\beta}{1+\theta/s}, \quad b_j = 1 - \frac{\theta(2+1/\beta)}{r_j(1+1/s)},$$

and remark that $a > 0$. We note also that

$$a^{-1}b_j = \frac{(2+1/\beta)(1-\theta/r_j)}{1-\theta/s+1/\beta} - 1.$$

If $b_j \le 0$ then, because $m \le 0$,

$$\tilde{h}_j \ge \left(c_4 L_j^{-1}\varphi\right)^{1/\beta_j} = (c_4 L_j^{-1})^{1/\beta_j} (L_\beta \delta)^{\frac{1}{\beta_j(2+1/\beta)}} > n^{-1}$$

for all large enough n. On the other hand, since $0 \ge m \ge m_0(\theta)$ and $2^{m_0(\theta)a} \le 2^a \hat{c}_1 \varkappa\varphi$ by definition of $m_0(\theta)$,

$$\begin{aligned}\tilde{h}_j &\le (c_4 L_j^{-1}\varphi)^{1/\beta_j} 2^{\frac{m_0(\theta) b_j}{\beta_j}} = (c_4 L_j^{-1}\varphi)^{1/\beta_j} \left(2^{m_0(\theta)a}\right)^{\frac{b_j}{a\beta_j}} \\ &\le (c_4 L_j^{-1})^{1/\beta_j} (2^{-a}\hat{c}_1\varkappa)^{\frac{b_j}{a\beta_j}} \varphi^{\frac{1+b_j/a}{\beta_j}} \le c_5 (c_4 L_0^{-1})^{1/\beta_j},\end{aligned}$$

where we took into account that $1 + b_j a^{-1} > 0$ and $\min_{j=1,\dots,d} L_j \ge L_0 > 0$. Then choosing constant c_4 small enough we have $\tilde{h}_j \le 1$. Thus we showed that $\tilde{h}_j \in [n^{-1}, 1]$ for j such that $b_j \le 0$.

Now consider the case $b_j > 0$. Here

$$\begin{aligned}\tilde{h}_j \ge (c_4 L_j^{-1}\varphi)^{1/\beta_j} \left(2^{m_0(\theta)a}\right)^{b_j a^{-1}/\beta_j} &\ge (c_4 L_j^{-1}\varphi)^{1/\beta_j} \left(\hat{c}_1 \varkappa\varphi\right)^{b_j a^{-1}/\beta_j} \\ &\ge c_6 \left(c_4 L_j^{-1}\right)^{1/\beta_j} \varphi^{\frac{(2+1/\beta)(1-\theta/r_j)}{\beta_j(1-\theta/s+1/\beta)}}.\end{aligned}$$

It remains to note that

$$\frac{1-\theta/r_j}{\beta_j(1-\theta/s+1/\beta)} = \frac{1/\beta_j - \theta/(\beta_j r_j)}{1-\theta/s+1/\beta} < 1, \quad \forall j = 1, \dots, d,$$

in view of the obvious inequality $1/\beta - \theta/s \ge 1/\beta_j - \theta/(\beta_j r_j)$, which, in its turn, follows from the fact that $\theta \in (0,1]$. Thus, we have that $\tilde{h}_j > n^{-1}$ for all large enough n. Furthermore, if $b_j > 0$ then since $m \le 0$

$$\tilde{h}_j \le (c_4 L_j^{-1})^{1/\beta_j} \varphi^{1/\beta_j} \le 1$$

for all large enough n. Thus we have shown that $\tilde{h} \in [n^{-1}, 1]^d$.

3^0. Now we proceed with bounding $J_{m,2}(h)$ for a specific choice of $h = h[m]$, which is defined as follows. Let $h[m] \in \mathcal{H}$ such that $h[m] < \tilde{h} \le 2h[m]$. Let constant c_4 in (B.10) be chosen so that $c_4 < (2\bar{c}_1)^{-1}$, where $\bar{c}_1$ appears on the right hand side of (6.12). With this choice of c_4 by (6.12)

$$\|B^*_{h[m],j}(f,\cdot)\|_\infty \le \bar{c}_1 L_j (h_j[m])^{\beta_j} \le \bar{c}_1 L_j \tilde{h}_j^{\beta_j} \le 2^{m-1}\varphi.$$

Therefore, $J^{(2)}_{m,2}\big(h[m]\big) = 0$, where $J^{(2)}_{m,2}(\cdot)$ is defined in (B.5). Moreover, we obtain from (B.6) and $h[m] \le \tilde{h}_j$ that

$$J^{(1)}_{m,2}\big(h[m]\big) \le c_1 \sum_{I \setminus I_\infty} 2^{-r_j m} \varphi^{-r_j} L_j^{r_j} \tilde{h}_j^{\beta_j r_j} \le c_1 \sum_{j \in I \setminus I_\infty} c_4^{r_j} 2^{-m\left(\frac{2+1/\beta}{1/\theta+1/s}\right)} \le c_7 2^{-m\left(\frac{2+1/\beta}{1/\theta+1/s}\right)}. \quad \text{(B.11)}$$

Note that

$$V_{h[m]} \ge 2^{-d} V_{\tilde{h}} = 2^{-d} c_4^{1/\beta} L_\beta^{-1} \varphi^{1/\beta} 2^{m\left(\frac{1}{\beta} - \frac{2+1/\beta}{1+s/\theta}\right)}. \quad \text{(B.12)}$$

This together with (B.9) yields

$$J_{m,1}\big(h[m]\big) \le c_8 2^{-m\left(\frac{2+1/\beta}{1/\theta+1/s}\right)}. \quad \text{(B.13)}$$

Then it follows from (B.11) and (B.13) that

$$J_{m,1}\big(h[m]\big) + J_{m,2}\big(h[m]\big) \le c_9 2^{-m\left(\frac{2+1/\beta}{1/\theta+1/s}\right)},$$

which combined with (B.3) results in

$$J_m \le c_{10} 2^{m\left(p - \frac{2+1/\beta}{1/\theta+1/s}\right)} \varphi^p.$$

Inequality (B.3) is valid only if (B.8) is fulfilled for $h[m]$, i.e., $\varkappa\delta V^{-1}_{h[m]} < 2^{m-2}\varphi$; now we verify this condition. It is sufficient to check that $\varkappa\delta 2^d V^{-1}_{\tilde{h}} < 2^{m-2}\varphi$. In view of (B.12) this inequality will follow if

$$c_4^{1/\beta} \varphi^{1+1/\beta} 2^{m\left(1+\frac{1}{\beta} - \frac{2+1/\beta}{1+s/\theta}\right)} > 2^{d+2} \varkappa (L_\beta \delta).$$

Taking into account that $L_\beta \delta = \varphi^{2+1/\beta}$ we conclude that (B.8) is fulfilled for $h[m]$ if

$$2^{m\left(\frac{1-\theta/s+1/\beta}{1+\theta/s}\right)} > \hat{c}_1 \varkappa \varphi,$$

which is ensured by the condition $m \ge m_0(\theta)$. This completes the proof of (6.24).

B.2.2. Proof of statement (iii)

1^0. Let $\hat{c}_4$ be a constant to be specified later, and let c_4 be the constant given in (B.10). Let $C_j = c_4$ if $j \in I_\infty$ and $C_j = \hat{c}_4$ if $j \in I \setminus I_\infty$. Define $\tilde{h} = (\tilde{h}_1, \dots, \tilde{h}_d) \in (0, \infty]^d$ by the formula

$$\tilde{h}_j = \big(C_j L_j^{-1} \varphi\big)^{1/\beta_j} 2^{m\left(\frac{1}{\beta_j} - \frac{s(2+1/\beta)}{\beta_j r_j}\right)}, \quad j = 1, \dots, d. \quad \text{(B.14)}$$

Note that if $j \in I_\infty$ the corresponding coordinates of $\tilde{h}$ given by (B.10) and (B.14) are the same.

Let us show that $\tilde{h} \in \big[n^{-1}, 1\big]^d$ for large enough n. First consider the coordinates $\tilde{h}_j$ such that $1 - \frac{s}{r_j}(2 + 1/\beta) \geq 0$. Because $m \geq 0$ we have for all n large enough

$$\tilde{h}_j \geq \big(C_j L_j^{-1}\varphi\big)^{1/\beta_j} \geq \big(C_j L_j^{-1}\big)^{1/\beta_j} (L_\beta \delta)^{\frac{1}{2\beta_j + \beta_j/\beta}} > \delta > n^{-1},$$

where we have used the obvious inequality $\beta_j/\beta > 1$ for any $j = 1, \dots, d$. On the other hand, because $2^m \leq \hat{c}_2 \varphi^{-1}$ we obtain

$$\tilde{h}_j \leq c_{11}\big(\hat{c}_4 L_j^{-1}\big)^{1/\beta_j} \varphi^{\frac{s(2+1/\beta)}{\beta_j r_j}} \to 0, \quad n \to \infty, \quad \forall j \in I \setminus I_\infty;$$

$$\tilde{h}_j \leq c_{11}\big(C_j L_j^{-1}\big)^{1/\beta_j} \varphi^{\frac{s(2+1/\beta)}{\beta_j r_j}} \leq c_{11}\big(c_4 L_0^{-1}\big)^{1/\beta_j}, \quad \forall j \in I_\infty.$$

Thus $\tilde{h}_j \leq 1$ for large enough n if $j \in I \setminus I_\infty$, and $\tilde{h}_j \leq 1$ by choice of constant c_4 if $j \in I_\infty$.

Now consider the case $1 - \frac{s}{r_j}(2 + 1/\beta) < 0$. Since $2^m \leq \hat{c}_2 \varphi^{-1}$

$$\tilde{h}_j \geq c_{12}\big(C_j L_j^{-1}\big)^{1/\beta_j} \varphi^{\frac{s(2+1/\beta)}{\beta_j r_j}} = c_{12}\big(C_j L_j^{-1}\big)^{1/\beta_j} (L_\beta \delta)^{\frac{s}{\beta_j r_j}} > \delta > n^{-1},$$

for all n large enough. Here we have used the obvious inequality $1/s > 1/\beta_j r_j \ \forall j = 1, \dots, d$. On the other hand, since $m \geq 0$, $\tilde{h}_j \leq \big(C_j L_j^{-1}\varphi\big)^{1/\beta_j} \leq 1$ for large enough n. Thus we have proved that $\tilde{h} \in [n^{-1}, 1]^d$ for all large enough n.

2^0. Let $h[m] \in \mathcal{H}$ such that $h[m] < \tilde{h} \leq 2h[m]$ and choose constant c_4 satisfies $c_4 < (2\bar{c}_1)^{-1}$ [see (6.12)]. Recall that formulas (B.10) and (B.14) coincide for $j \in I_\infty$. Therefore, as before, with the indicated choice of c_4 we have

$$J^{(2)}_{m,2}\big(h[m]\big) = 0. \tag{B.15}$$

Let $\beta_\pm$ and β_∞ be defined by expressions $1/\beta_\pm := \sum_{j \in I_+ \cup I_-} 1/\beta_j$ and $1/\beta_\infty := \sum_{j \in I_\infty} 1/\beta_j$. We have

$$V_{h[m]} \geq 2^{-d} V_{\tilde{h}} = 2^{-d} c_4^{1/\beta_\infty} \hat{c}_4^{1/\beta_\pm} L_\beta^{-1} \varphi^{1/\beta} 2^{-2m}. \tag{B.16}$$

This together with $2^m \leq \hat{c}_2 \varphi^{-1}$ shows that $V_{h[m]} \geq c_{13} \hat{c}_4^{1/\beta_\pm} L_\beta^{-1} \varphi^{2+1/\beta} = c_{13} \hat{c}_4^{1/\beta_\pm} \delta$ and, therefore,

$$\varkappa \delta V^{-1}_{h[m]} \leq c_{13}^{-1} \hat{c}_4^{-1/\beta_\pm} \varkappa. \tag{B.17}$$

Remark that $A_\eta(x) \leq 2^d M \mathrm{k}_\infty^2$ for all $x \in \mathbb{R}^d$ and $\eta \in \mathcal{H}$. Hence, in view of (B.17)

$$\begin{aligned} \sup_{\eta \geq h[m]} M_\eta(x) &\leq \mathrm{k}_\infty \sqrt{2^d M} \sqrt{\varkappa \delta V^{-1}_{h[m]}} + \varkappa \delta V^{-1}_{h[m]} \\ &\leq \left(\mathrm{k}_\infty \sqrt{2^d M} + \sqrt{c_{13}^{-1} \hat{c}_4^{-1/\beta_\pm}}\right) \sqrt{\delta V^{-1}_{h[m]}} = c_{14} \sqrt{\delta V^{-1}_{h[m]}}. \end{aligned}$$

It yields together with (B.16)

$$\sup_{\eta \geq \tilde{h}} M_\eta(x) \leq c_{15}\, \hat{c}_4^{-1/(2\beta_\pm)} 2^m \sqrt{(L_\beta \delta)\varphi^{-1/\beta}} = c_{15}\, \hat{c}_4^{-1/(2\beta_\pm)} 2^m \varphi.$$

Setting $\hat{c}_4$ so that $c_{15}\hat{c}_4^{-1/(2\beta_\pm)} < 2^{-1}$, we obtain $\sup_{\eta\geq\tilde{h}} M_\eta(x) \leq 2^{m-1}\varphi$. This implies that

$$J_{m,1}\big(h[m]\big) = 0. \tag{B.18}$$

Moreover, it follows from (B.6) and from inequality $h[m] \leq \tilde{h}$ that

$$J^{(1)}_{m,2}\big(h[m]\big) \leq J^{(1)}_{m,2}\big(\tilde{h}\big) \leq \left[c_1 \sum_{j\in I\setminus I_\infty} \hat{c}_4^{r_j}\right] 2^{-ms(2+1/\beta)}. \tag{B.19}$$

Then (6.25) is a consequence of (B.3), (B.15), (B.18) and (B.19). The statement (ii) is proved.

B.2.3. Proof of statement (iv)

1^0. Let C_j, $j = 1, \ldots, d$ be the same constants in the proof of statement (ii) in the previous section. Define $\tilde{h} = (\tilde{h}_1, \ldots, \tilde{h}_d) \in (0, \infty]^d$ by the following formula

$$\tilde{h}_j \;=\; \big(C_j L_j^{-1}\varphi\big)^{1/\gamma_j} 2^{m\left(\frac{1}{\gamma_j} - \frac{\upsilon(2+1/\gamma)}{\gamma_j q_j}\right)} \left[\frac{L_\gamma \varphi^{1/\beta}}{L_\beta \varphi^{1/\gamma}}\right]^{\frac{\upsilon}{\gamma_j q_j}}, \quad j = 1, \ldots, d, \tag{B.20}$$

where γ_j, q_j are defined in (6.1) and γ, υ and L_γ are given in (6.15).

Let us show that $\tilde{h} \in \big[n^{-1}, 1\big]^d$ for large n. Let $b_j = 1 - \frac{\upsilon}{q_j}(2 + 1/\gamma)$.

First, assume that $b_j < 0$. Since $m > 0$ and $2^m \leq \hat{c}_2\varphi^{-1}$,

$$\begin{aligned}\tilde{h}_j \geq c_{16}\big(C_j L_j^{-1}\big)^{1/\gamma_j} \varphi^{\frac{\upsilon(2+1/\gamma)}{\gamma_j q_j}} \left[\frac{L_\gamma \varphi^{1/\beta}}{L_\beta \varphi^{1/\gamma}}\right]^{\frac{\upsilon}{\gamma_j q_j}} &= c_{16}\big(C_j L_j^{-1}\big)^{1/\gamma_j} \left[L_\gamma / L_\beta\right]^{\frac{\upsilon}{\gamma_j q_j}} \varphi^{\frac{\upsilon(2+1/\beta)}{\gamma_j q_j}} \\ &= c_{16}\big(C_j L_j^{-1}\big)^{1/\gamma_j} \left[L_\gamma\right]^{\frac{\upsilon}{\gamma_j q_j}} \delta^{\frac{\upsilon}{\gamma_j q_j}} > \delta > n^{-1},\end{aligned}$$

where we have used the obvious inequality $1/\upsilon > 1/(\gamma_j q_j)$ for any $j = 1, \ldots, d$. On the other hand, in view of $m \geq m_1$ and by (6.20)

$$\begin{aligned}\tilde{h}_j \;&\leq\; \big(C_j L_j^{-1}\varphi\big)^{1/\gamma_j} 2^{\frac{m_1 b_j}{\gamma_j}} \left[(L_\gamma/L_\beta)\varphi^{1/\beta - 1/\gamma}\right]^{\frac{\upsilon}{\gamma_j q_j}} \\ &\leq\; (C_j L_j^{-1}\varphi)^{1/\gamma_j} 2^{\frac{m_1 b_j}{\gamma_j}} \left[2^{m_1[\upsilon(2+1/\gamma) - s(2+1/\beta)]}\right]^{\frac{1}{\gamma_j q_j}} = \big(C_j L_j^{-1}\varphi\big)^{1/\gamma_j} 2^{m_1\left(\frac{1}{\gamma_j} - \frac{s(2+1/\beta)}{\gamma_j q_j}\right)}.\end{aligned}$$

Then by (6.21)

$$\tilde{h}_j \leq c_{17} C_1(\vec{L}) C_j^{1/\gamma_j} \varphi^{\frac{1}{\gamma_j}\left(1 + [1 - \frac{s}{q_j}(2+1/\beta)]\frac{\upsilon(1/\beta - 1/\gamma)}{\upsilon[2+1/\gamma] - s[2+1/\beta]}\right)}, \tag{B.21}$$

where the expression for constant $C_1(\vec{L})$ is easily found. It remains to note that

$$\upsilon(2 + 1/\gamma) - s(2 + 1/\beta) = s\upsilon\left[(2 + 1/\beta)(1/s - 1/\upsilon) + (1/\gamma - 1/\beta)s^{-1}\right],$$

and in view of (6.16) and (6.17)

$$1 + \frac{[q_j - s(2 + 1/\beta)]\upsilon(1/\beta - 1/\gamma)}{q_j[\upsilon(2 + 1/\gamma) - s(2 + 1/\beta)]} = \frac{2 + 1/\beta}{q_j}\left[\frac{(1/s - 1/\upsilon)q_j + (1/\gamma - 1/\beta)}{(1/s - 1/\upsilon)(2 + 1/\gamma) + (1/\gamma - 1/\beta)\upsilon^{-1}}\right] > 0.$$

This shows that $\tilde{h}_j \leq 1$ for large n.

Now assume that $b_j \geq 0$. Then, similarly to the reasoning that resulted in (B.21) we have

$$\begin{aligned}\tilde{h}_j &\geq \bigl(C_j L_j^{-1}\varphi\bigr)^{1/\gamma_j} 2^{\frac{m_1 b_j}{\gamma_j}} \bigl[(L_\gamma/L_\beta)\varphi^{1/\beta-1/\gamma}\bigr]^{\frac{\upsilon}{\gamma_j q_j}} \\ &\geq C_1(\vec{L}) C_j^{1/\gamma_j} \varphi^{\frac{1}{\gamma_j}\left(1+\left[1-\frac{s}{q_j}(2+1/\beta)\right]\frac{\upsilon(1/\beta-1/\gamma)}{\upsilon[2+1/\gamma]-s[2+1/\beta]}\right)}.\end{aligned}$$

Since $\varphi^{2+1/\beta} = L_\beta \delta$,

$$\tilde{h}_j \geq c_{18} C_1(\vec{L}) C_j^{1/\gamma_j} \delta^{\frac{(1/s-1/\upsilon)(1/\gamma_j)+(1/\gamma-1/\beta)(1/\gamma_j r_j)}{(1/s-1/\upsilon)(2+1/\gamma)+(1/\gamma-1/\beta)\upsilon^{-1}}} > \delta > n^{-1}$$

for all n large enough. Here we have used (6.16), (6.17), and obvious inequalities: $2+1/\gamma > 1/\gamma_j$ and $1/\upsilon > 1/\gamma_j r_j$ for all $j = 1, \dots, d$. On the other hand, since $2^m \leq \hat{c}_2 \varphi^{-1}$

$$\begin{aligned}\tilde{h}_j &\leq c_{19}\bigl(C_j L_j^{-1}\bigr)^{1/\gamma_j} \varphi^{\frac{\upsilon(2+1/\gamma)}{\gamma_j q_j}} \left[\frac{L_\gamma \varphi^{1/\beta}}{L_\beta \varphi^{1/\gamma}}\right]^{\frac{\upsilon}{\gamma_j q_j}} = c_{19}\bigl(C_j L_j^{-1}\bigr)^{1/\gamma_j} \bigl[L_\gamma/L_\beta\bigr]^{\frac{\upsilon}{\gamma_j q_j}} \varphi^{\frac{\upsilon(2+1/\beta)}{\gamma_j q_j}} \\ &= c_{19}\bigl(C_j L_j^{-1}\bigr)^{1/\gamma_j} \bigl[L_\gamma\bigr]^{\frac{\upsilon}{\gamma_j q_j}} \delta^{\frac{\upsilon}{\gamma_j q_j}}.\end{aligned}$$

Therefore, $\tilde{h}_j \to 0,\ n \to \infty$, $\forall j \in I \setminus I_\infty$, and $\tilde{h}_j \leq c_{19}\bigl(c_4 L_0^{-1}\bigr)^{1/\gamma_j}$, $\forall j \in I_\infty$. Choosing c_4 small enough we come to required assertion.

3^0. Let $h[m] \in \mathcal{H}$ be such that $h[m] < \tilde{h} \leq 2h[m]$, and let constant c_4 satisfy $c_4 < (2\bar{c}_1)^{-1}$, where $\bar{c}_1$ is given in (6.12). With this choice of c_4, if $j \in I_\infty$ then the corresponding coordinates of $\tilde{h}$ given by (B.14) and (B.20) coincide. Hence we have as before

$$J_{m,2}^{(2)}\bigl(h[m]\bigr) = 0. \tag{B.22}$$

Let $\frac{1}{\gamma_\pm} := \sum_{j \in I_+ \cup I_-} \frac{1}{\gamma_j}$; then

$$V_{h[m]} \geq 2^{-d} V_{\tilde{h}} = 2^{-d}\bigl(c_4\bigr)^{1/\beta_\infty}\bigl(\hat{c}_4\bigr)^{1/\gamma_\pm} L_\beta^{-1} \varphi^{1/\beta} 2^{-2m}. \tag{B.23}$$

We remark that (B.23) and (B.16) coincide up to the change in notation $\beta_\pm \leftrightarrow \gamma_\pm$. Hence all the computations preceding (B.18) remain valid, and we have as before

$$J_{m,1}\bigl(h[m]\bigr) = 0. \tag{B.24}$$

Moreover, we obtain from (B.7)

$$J_{m,2}^{(1)}\bigl(h[m]\bigr) \leq \left[\tilde{c}_1 \sum_{j \in I \setminus I_\infty} \hat{c}_4^{r_j}\right] \left[\frac{L_\gamma \varphi^{1/\beta}}{L_\beta \varphi^{1/\gamma}}\right]^{\upsilon} 2^{-m\upsilon(2+1/\gamma)}. \tag{B.25}$$

The bound given in (6.26) follows now from (B.3), (B.22), (B.24) and (B.25). ∎

B.3. Proof of Proposition 2

In view of (2.14) and (6.11)

$$|\hat{f}(x) - f(x)| \leq c_0[\bar{U}_f(x) + \omega(x)] \leq c_1[U_f(x) + \omega(x)], \tag{B.26}$$

where c_0 and c_1 are appropriate constants, $\bar{U}_f(x)$ and $U_f(x)$ are given by (6.8) and (6.10) respectively, and $\omega(x) := \zeta(x) + \chi(x)$ with $\zeta(x)$ and $\chi(x)$ defined in (2.15) and (2.16).

B.3.1. Proof of statement (i)

Here for brevity we will write $m_0 = m_0(1)$. By (B.26)

$$\int_{\mathcal{X}^-_{m_0}} |\hat{f}(x) - f(x)|^p \mathrm{d}x \;\le\; c_1^{p-1} \int_{\mathcal{X}^-_{m_0}} \big[U_f(x) + \omega(x)\big]^{p-1} |\hat{f}(x) - f(x)| \mathrm{d}x$$
$$\le\; c_2 \bigg[(2^{m_0}\varphi)^{p-1} \int_{\mathbb{R}^d} |\hat{f}(x) - f(x)| \mathrm{d}x + \int_{\mathbb{R}^d} \omega^{p-1}(x) [2^{m_0}\varphi + \omega(x)] \mathrm{d}x \bigg].$$

Noting that $\|\hat{f}\|_1 \le \ln^d(n) \|K\|_1 \le \ln^d(n) \mathrm{k}_\infty$, we have $\|\hat{f} - f\|_1 \le \ln^d(n) \mathrm{k}_\infty + 1$ and, therefore,

$$(2^{m_0}\varphi)^{p-1} \int_{\mathbb{R}^d} |\hat{f}(x) - f(x)| \mathrm{d}x \;\le\; \big(\ln^d(n) \mathrm{k}_\infty + 1 \big) (2^{m_0}\varphi)^{p-1}.$$

Moreover, since $\varkappa = \mathrm{k}_\infty^2 [(4d+2)p + 4(d+1)]$, the second statement of Theorem 1 implies

$$\mathbb{E}_f \int_{\mathbb{R}^d} \omega^{p-1}(x) [2^{m_0}\varphi + \omega(x)] \mathrm{d}x \le c_3 (2^{m_0}\varphi) n^{-(p-1)/2} + c_4 n^{-p/2}.$$

Combining these inequalities and taking into account that $2^{m_0}\varphi \le 1$ we obtain

$$J^-_{m_0} = \mathbb{E}_f \int_{\mathcal{X}^-_{m_0}} |\hat{f}(x) - f(x)|^p \mathrm{d}x \le c_5 \big[\ln^d(n) (2^{m_0}\varphi)^{p-1} + 2^{m_0}\varphi n^{-(p-1)/2} + n^{-p/2} \big]$$
$$\le 2c_5 \big[\ln^d(n) (2^{m_0}\varphi)^{p-1} + n^{-p/2} \big].$$

By definition of $m_0 = m_0(1)$, $2^{m_0}\varphi \le c_6 \big(L_\beta \delta\big)^{1/(1+1/\beta-1/s)}$; therefore

$$J^-_{m_0} \;\le\; c_7 \ln^d(n) \big(L_\beta \delta\big)^{\frac{p-1}{1-1/s+1/\beta}} + c_7 n^{-p/2}.$$

It remains to note that for large n that

$$\big(L_\beta \delta\big)^{\frac{p-1}{1+1/\beta-1/s}} \le \big(L_\beta \delta\big)^{\nu p}, \qquad n^{-p/2} < \big(L_\beta \delta\big)^{\nu p},$$

and (6.27) follows.

B.3.2. Proof of statement (ii)

Let f^* be the maximal operator of f defined in (4.1). It follows from the definition of $M_\eta(x)$ that for any $h \in \mathcal{H}$

$$\sup_{\eta \ge h} M_\eta(x) \le c_8 \sqrt{\frac{\varkappa f^*(x) \ln n}{n V_h}} + \frac{\varkappa \ln n}{n V_h}.$$

Moreover, by definition of $\bar{B}_h(f,x)$, $\bar{B}_h(f,x) \le c_9 [f^*(x) + f(x)] \le 2c_9 f^*(x)$ almost everywhere, where the last inequality follows from the Lebesgue differentiation theorem. Using these two inequalities and setting $h = (1, \ldots, 1)$ in (6.8) we come to the following upper bound on $\bar{U}_f(x)$

$$\bar{U}_f(x) \le c_{10} \big[f^*(x) + \sqrt{f^*(x)\delta} + \delta \big]. \tag{B.27}$$

In view of (6.11) we have that $\mathcal{X}^-_{m_0(\theta)} \subseteq \mathcal{X}^- := \{x \in \mathbb{R}^d : \bar{U}_f(x) \le \mathrm{k}_\infty 2^{m_0(\theta)}\varphi\}$; therefore if we put

$$D_1 := \mathcal{X}^- \cap \{x \in \mathbb{R}^d : f^*(x) \le \delta\}, \;\; D_2 := \mathcal{X}^- \cap \{x \in \mathbb{R}^d : f^*(x) > \delta\}$$

then

$$J^-_{m_0(\theta)} \le \mathbb{E}_f \int_{D_1} |\hat{f}(x) - f(x)|^p \mathrm{d}x + \mathbb{E}_f \int_{D_2} |\hat{f}(x) - f(x)|^p \mathrm{d}x =: \mathbb{E}_f S_1 + \mathbb{E}_f S_2. \tag{B.28}$$

We bound from above the two terms on the right hand side of the above inequality.

First consider $\mathbb{E}_f S_1$. By (B.26) for any $\theta \in (0, 1]$ we have

$$\begin{aligned} S_1 &= \int_{D_1} |\hat{f}(x) - f(x)|^p \mathrm{d}x \;\le\; c_0^{p-\theta} \int_{D_1} \big[\bar{U}_f(x) + \omega(x)\big]^{p-\theta} |\hat{f}(x) - f(x)|^\theta \mathrm{d}x \\ &\le c_{11}\bigg\{\delta^{p-\theta} \int_{\mathbb{R}^d} |\hat{f}(x) - f(x)|^\theta \mathrm{d}x + \int_{\mathbb{R}^d} \omega^{p-\theta}(x)[\delta + \omega(x)]^\theta \mathrm{d}x\bigg\}. \end{aligned}$$

Here we have used that, by (B.27), $\bar{U}_f(x) \le 2c_{10}\delta$ for all $x \in D_1$. Remind that $\hat{f}(x) = \hat{f}_{\hat{h}(x)}(x)$; therefore, for any $\theta \in (0, 1]$

$$\mathbb{E}_f |\hat{f}(x)|^\theta \le \big(\mathbb{E}_f |\hat{f}(x)|\big)^\theta \;\le\; \Big(\sum_{h \in \mathcal{H}} \mathbb{E}_f \big|\hat{f}_h(x)\big|\Big)^\theta \le c_{12}\big[(\ln n)^d f^*(x)\big]^\theta.$$

Thus, for any $f \in \mathbb{G}_\theta(R)$,

$$\delta^{p-\theta}\mathbb{E}_f \int_{\mathbb{R}^d} |\hat{f}(x) - f(x)|^\theta \mathrm{d}x \;\le\; \delta^{p-\theta}\Big\{\|f\|_\theta^\theta + c_{12}(\ln n)^{d\theta}\|f^*\|_\theta^\theta\Big\} \;\le\; c_{13}\delta^{p-\theta}R^\theta(\ln n)^{d\theta}.$$

Furthermore, because $\varkappa = \mathrm{k}_\infty^2[(4d+2)p + 4(d+1)]$, by the second statement of Theorem 1

$$\mathbb{E}_f \int_{\mathbb{R}^d} \omega^{p-\theta}(x)[\delta + \omega(x)]^\theta \mathrm{d}x \le c_{14}\delta^\theta n^{-(p-\theta)/2} + c_{15}n^{-p/2} \le c_{15}n^{-p/2}.$$

Combining the last two inequalities we obtain

$$\mathbb{E}_f S_1 = \mathbb{E}_f \int_{D_1} |\hat{f}(x) - f(x)|^p \mathrm{d}x \le c_{16}\big[\delta^{p-\theta}R^\theta(\ln n)^{d\theta} + n^{-p/2}\big]. \tag{B.29}$$

Now we proceed with bounding $\mathbb{E}_f S_2$. We have

$$\begin{aligned} \mathbb{E}_f S_2 &= \mathbb{E}_f \int_{D_2} |\hat{f}(x) - f(x)|^p \mathrm{d}x \;\le\; c_0^p\, \mathbb{E}_f \int_{D_2} \big[\bar{U}_f(x) + \omega(x)\big]^p \mathrm{d}x \\ &\overset{\text{(a)}}{\le} c_{17}\Big[(2^{m_0(\theta)}\varphi)^{p-\theta} \int_{D_2} |\bar{U}_f(x)|^\theta \mathrm{d}x + n^{-p/2}\Big] \\ &\overset{\text{(b)}}{\le} c_{17}\big[(2^{m_0(\theta)}\varphi)^{p-\theta}R^\theta + n^{-p/2}\big] \;\le\; c_{18}\big[R^\theta(L_\beta\delta)^{\frac{p-\theta}{1-\theta/s+1/\beta}} + n^{-p/2}\big]. \end{aligned} \tag{B.30}$$

Here (a) follows from the second statement of Theorem 1 and $\bar{U}_f(x) \le 2^{m_0(\theta)}\varphi$ for $x \in D_2$, and (b) is valid because $\bar{U}_f(x) \le 3c_{10}f^*(x)$ for all $x \in D_2$, see (B.27).

Note that $\delta^{1-\frac{1}{1-\theta/s+1/\beta}}(\ln n)^{\frac{d\theta}{p-\theta}} \to 0$ as $n \to \infty$ since $\theta \le 1$ and $1/\beta > 1/s$, where the latter inequality follows from $r \in (1, \infty]^d$.

Thus, combining (B.29) and (B.30) with (B.28), we obtain

$$J^{-}_{m_0(\theta)} \;\le\; c_{19}\Big[\big(L_\beta\delta\big)^{\frac{p-\theta}{1-\theta/s+1/\beta}} + n^{-p/2}\Big] \le c_{20}\big(L_\beta\delta\big)^{p\nu(\theta)},$$

as claimed. ∎

References

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

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

Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47**, 119–137.

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3**, 215228.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation. The* $\mathbb{L}_1$ *View.* Wiley & Sons, New York.

Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimation. *Ann. Statist.* **24**, 2499–2512.

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

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation.* Springer, New York.

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. *Ann. Statist.* **36**, 1127–1155.

Efroimovich, S. Yu. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Theory Probab. Appl.* **30**, 557–568.

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

Garling, D. J. H. (2007). *Inequalities: A Journey Into Linear Analysis.* Cambridge University Press, Cambridge.

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

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

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

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.

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

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

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

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

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

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

Kerkyacharian, G., Lepski, O. and Picard, D. (2008). Nonlinear estimation in anisotropic multiindex denoising. Sparse case. *Theory Probab. Appl.* **52**, 58–77.

Lepski, 0. (2012). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Manuscript.*

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

Massart, P. (2007). *Concentration Inequalities and Model Selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin.

Nemirovski, A. S. (1985). Nonparametric estimation of smooth regression functions. *Soviet J. Comput. Systems Sci.* **23** (1985), no. 6, 111 (1986); translated from *Izv. Akad. Nauk SSSR Tekhn. Kibernet.* 1985, no. 3, 50–60, (Russian)

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.

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, Ph. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12**, 351–370.

Rigollet, Ph. 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.

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